

ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมิน  
ระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์: การประยุกต์ใช้โมเดลหลายองค์ประกอบของ  
ราล์ฟ และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด



บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)  
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)  
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต  
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา  
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2560  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

ALIGNMENT BETWEEN NATIONAL AND CLASSROOM SCIENCE TEST ITEMS: AN APPLICATION OF MANY-FACET RASCH MEASUREMENT AND GENERALIZABILITY THEORY

Miss Budsayarat Janprasert



A Dissertation Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy Program in Educational Measurement and  
Evaluation

Department of Educational Research and Psychology

Faculty of Education

Chulalongkorn University

Academic Year 2017

Copyright of Chulalongkorn University



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**



หัวข้อวิทยานิพนธ์	ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราส์ซ และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด
โดย	นางสาวบุษยรัตน์ จันทร์ประเสริฐ
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ ังดกระโทก

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศึกษาศาสตรบัณฑิต

..... คณบดีคณะครุศาสตร์

(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร.โชติกา ภาษีผล)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(ผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ ังดกระโทก)

..... กรรมการ

(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

..... กรรมการ

(รองศาสตราจารย์ ดร.กมลวรรณ ตังชนกานนท์)

..... กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร.สุวิมล กฤษศยาสา)

บุษยารัตน์ จันทร์ประเสริฐ : ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราสช์ และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (ALIGNMENT BETWEEN NATIONAL AND CLASSROOM SCIENCE TEST ITEMS: AN APPLICATION OF MANY-FACET RASCH MEASUREMENT AND GENERALIZABILITY THEORY) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร.ณัฐภรณ์ หลาวทอง, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ผศ. ดร.สังวรณ์ จักกระโทก, หน้า.

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) ตรวจสอบอิทธิพลของผู้ประเมิน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น 2) เปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน 3) ศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน และ 4) ประเมินค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน ตัวอย่างในการศึกษา ประกอบด้วย ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น ของโรงเรียนมัธยมศึกษาในกรุงเทพมหานคร จำนวน 1,089 ข้อ และผู้เชี่ยวชาญในการประเมินความสอดคล้องในแนวเดียวกันจำนวน 20 คน เครื่องมือที่ใช้ในการวิจัยคือ แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน การวิเคราะห์ข้อมูลโดยประยุกต์ใช้ MFRM การทดสอบสถิติที่ (paired-samples t-test) การวิเคราะห์ค่าดัชนีความสอดคล้อง (alignment index) และการใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ผลการวิจัยสรุปได้ดังนี้

1. เกิดอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน โดยผู้ประเมินมีแนวโน้มที่จะมีลักษณะของการกดคะแนนมากกว่าปล่อยคะแนน (คะแนนโลจิทอยู่ระหว่าง -3.24 ถึง 1.83) ผู้ประเมินส่วนใหญ่แสดงรูปแบบของความแม่นยำในการให้คะแนน จำนวน 16 คน คิดเป็นร้อยละ 80.00 ไม่มีผู้ประเมินที่แสดงอิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน และมีผู้ประเมิน จำนวน 4 คนที่แสดงรูปแบบอื่นๆ คิดเป็นร้อยละ 20.00

2. ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ระหว่างก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ( $t = 17.044$ ,  $p = .00$ ) และเมื่อมีการควบคุมอิทธิพลการกดหรือปล่อยคะแนน มีข้อสอบจำนวน 21 ข้อ คิดเป็นร้อยละ 1.93 ที่มีการเปลี่ยนแปลงผลการประเมิน และมีข้อสอบที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัด และสอดคล้องกับโมเดล จำนวน 902 ข้อ คิดเป็นร้อยละ 82.33 (Fair-M Average อยู่ระหว่าง 3.06 ถึง 3.97, infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.50 ถึง 1.50)

3. โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน มีค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียนที่ใกล้เคียงกัน โดยมีดัชนีความสอดคล้อง อยู่ระหว่าง 0.436 ถึง 0.588

4. ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับตัดสินใจเชิงสัมบูรณ์ ของคะแนนการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ประเมินเพิ่มขึ้น ในทุกรูปแบบการออกแบบการประเมิน โดยในการประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ ด้วยผู้ประเมินจำนวน 2 คน และ 3 คน ตามลำดับ ทำให้ได้สัมประสิทธิ์การสรุปอ้างอิงสูงตามเกณฑ์ที่ยอมรับได้

ภาควิชา	วิจัยและจิตวิทยาการศึกษา	ลายมือชื่อนิสิต .....
สาขาวิชา	การวัดและประเมินผลการศึกษา	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
ปีการศึกษา	2560	ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

# # 5784209827 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION

KEYWORDS: ALIGNMENT / MANY-FACET RASCH MEASUREMENT / GENERALIZABILITY THEORY

BUDSAYARAT JANPRASERT: ALIGNMENT BETWEEN NATIONAL AND CLASSROOM SCIENCE TEST ITEMS: AN APPLICATION OF MANY-FACET RASCH MEASUREMENT AND GENERALIZABILITY THEORY. ADVISOR: ASST. PROF. NUTTAPORN LAWTHONG, Ph.D., CO-ADVISOR: ASST. PROF. SUNGWORN NGUDGRATOKE, Ph.D., pp.

The objectives of this research were to 1) examine rater effects in the evaluation of the science standards and indicators-classroom test items alignment in the junior secondary school education, 2) compare the extent to which standards & indicators and test items was aligned after and before controlling for severity/leniency effects, 3) investigate and compare the degree of alignment between national and classroom science item tests assessed by Porter's alignment index among schools with different science achievements, and 4) estimate and compare the generalizability coefficients of the evaluation results of the standards and indicators-classroom test items alignment when the different number of raters and evaluation designs vary. Research subjects were 1,089 science classroom test items used in junior secondary school under the Office of the Basic Education Commission in Bangkok, and 20 expert panelists who evaluate alignment. Research instrument were the evaluation of the science standards and indicators-items alignment scale. MFRM, Paired-samples t-test, alignment index analysis, and G-theory were employed to analyze the data. Research results were as follows:

1. There were severity/leniency effects in the evaluation of the science standards and indicators-classroom test items alignment. The raters tend to be severe rather than a lenient (rater logit ranged from -3.24 to 1.83). The majority of 16 raters (80.00%) fit the profile of accurate raters. No raters exhibited central tendency effect, restriction of range effect, randomness effect. The 4 raters (20.00%) exhibited other profile.

2. There was a statistically significant difference at .01 level between before and after controlling for severity/leniency effects in the evaluation of the standards and indicators-classroom test items alignment ( $t = 17.044$ ,  $p = .00$ ). When the severity/leniency effects were controlled, there were 21 item tests (1.93%) changes in the alignment evaluation results, and there were 901 item tests (82.74%) that alignment with standards and indicators and fit to model (Fair-M Average ranged from 3.06 to 3.97, infit MNSQ and outfit MNSQ ranged from 0.50 to 1.50).

3. The schools with different science achievements have similar alignment indices between national and classroom science item tests. The alignment indices were between 0.436 and 0.588.

4. The generalizability coefficient for an absolute decision of the evaluation scores of the standards and indicators-classroom test items alignment increased when the number of raters increased in all evaluation designs. In the cognitive demand evaluation and the alignment between items and indicators level evaluation with 5-point rating scale the number of raters at least equal 2 and 3 raters respectively yield an acceptably high generalizability coefficient.

Department: Educational Research and Psychology Student's Signature .....

Field of Study: Educational Measurement and Evaluation Advisor's Signature .....

Evaluation Co-Advisor's Signature .....

Academic Year: 2017

## กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จได้ก็ด้วยความกรุณาอย่างสูงยิ่งจากอาจารย์ที่ปรึกษา วิทยานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง และผู้ช่วยศาสตราจารย์ ดร.สังวรณัฏฐ์ ภัทรโกท ที่ได้ให้ความรู้ทั้งในหลักวิชาและคำแนะนำอันเป็นประโยชน์ในการทำวิทยานิพนธ์ ทั้งยังได้ดูแลเอาใจใส่ คอยติดตามความก้าวหน้าของผู้วิจัยเป็นอย่างดีตลอดมา ผู้วิจัยรู้สึกซาบซึ้งและขอกราบขอบพระคุณอย่างยิ่ง

ขอกราบขอบพระคุณประธานและคณะกรรมการสอบทุกท่าน ที่ได้กรุณาสละเวลาอันมีค่ามาร่วมสอบวิทยานิพนธ์ ตลอดจนให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการทำวิทยานิพนธ์ฉบับนี้

ขอกราบขอบพระคุณคณาจารย์ภาควิชาวิจัยและจิตวิทยาทุกท่าน ที่เป็นผู้ประสิทธิ์ประสาทวิชาความรู้ทั้งปวงให้แก่ผู้วิจัย อันเป็นประโยชน์อย่างสูงในการทำวิทยานิพนธ์

ขอขอบคุณผู้เชี่ยวชาญในการประเมินข้อสอบ ผู้เชี่ยวชาญในการประเมินคู่มือฯ และผู้เชี่ยวชาญในการตรวจเครื่องมือวิจัยทุกท่าน ที่สละเวลาอันมีค่าในการทำหน้าที่ของผู้เชี่ยวชาญอย่างตั้งใจ และเต็มความรู้ความสามารถ

ขอขอบคุณโรงเรียนกลุ่มตัวอย่าง ที่ให้ความอนุเคราะห์ข้อสอบและข้อมูลการทำวิทยานิพนธ์

ขอขอบคุณทุน “90 ปี จุฬาลงกรณ์มหาวิทยาลัย” จากกองทุนรัชดาภิเษกสมโภช จุฬาลงกรณ์มหาวิทยาลัย ที่ให้การสนับสนุนงบประมาณในการทำวิทยานิพนธ์

ขอขอบคุณมหาวิทยาลัยเทคโนโลยีมหานคร ที่ให้โอกาสและสนับสนุนทุนการศึกษาตลอดหลักสูตร

ขอขอบคุณอาจารย์สำนักประกันคุณภาพการศึกษา มหาวิทยาลัยเทคโนโลยีมหานครทุกท่าน ที่คอยรับฟังปัญหา ให้กำลังใจ และปลอบประโลมในยามที่ผู้วิจัยเผชิญกับอุปสรรคต่างๆ

ขอกราบขอบพระคุณ คุณพ่อสุทัศน์ คุณแม่ศตพร จันท์ประเสริฐ ที่ให้กำเนิด อบรมเลี้ยงดู ให้โอกาสในการศึกษาและเป็นกำลังใจสำคัญที่ทำให้ผู้วิจัยผ่านพ้นอุปสรรคและประสบความสำเร็จในทุกวันนี้ และขอขอบคุณพี่สุรเชษฐ์ พี่ศิรินุช น้องมุสซา ตลอดจนญาติพี่น้องและเพื่อนๆ อีกมากมายที่ไม่ได้เอ่ยนามในที่นี้ ที่ความช่วยเหลือ เป็นกำลังใจและอยู่เคียงข้างผู้วิจัยเสมอมา

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ญ
สารบัญภาพ .....	ฐ
บทที่ 1 บทนำ.....	1
ความเป็นมาและความสำคัญของการศึกษา .....	1
คำถามการวิจัย .....	11
วัตถุประสงค์ของการวิจัย.....	12
สมมติฐานการวิจัย .....	13
ขอบเขตการวิจัย.....	15
คำจำกัดความที่ใช้ในการวิจัย.....	17
ประโยชน์ที่ได้รับ.....	22
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	24
ตอนที่ 1 มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลาง การศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ของระดับชั้นมัธยมศึกษาตอนต้น.....	24
ตอนที่ 2 การทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน.....	33
ตอนที่ 3 การศึกษาความสอดคล้องในแนวเดียวกัน (alignment study) .....	39
ตอนที่ 4 พฤติกรรมการเรียนรู้ด้านพุทธิพิสัย ตามแนวคิดของ Bloom ที่ได้รับการปรับปรุง ใหม่ .....	75
ตอนที่ 5 อิทธิพลของผู้ประเมิน.....	83
ตอนที่ 6 โมเดลหลายองค์ประกอบของราล์ฟ.....	87

ตอนที่ 7 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด.....	116
ตอนที่ 8 มาตรฐานของการประเมิน .....	124
ตอนที่ 9 กรอบแนวคิดการวิจัย.....	126
บทที่ 3 วิธีดำเนินการวิจัย .....	128
1. ประชากรและตัวอย่างในการวิจัย .....	128
2. เครื่องมือในการวิจัย .....	130
3. การเก็บรวบรวมข้อมูล.....	142
4. การวิเคราะห์ข้อมูล.....	145
5. การจัดทำคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบ ในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน .....	155
บทที่ 4 ผลการวิเคราะห์ข้อมูล .....	159
ตอนที่ 1 ข้อมูลพื้นฐานของตัวอย่างในการวิจัย .....	160
ตอนที่ 2 ผลการตรวจสอบอิทธิพลของผู้ประเมินในการประเมินความสอดคล้องในแนว เดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน .....	172
ตอนที่ 3 ผลการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐาน และตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพล การกดหรือปล่อยคะแนน.....	177
ตอนที่ 4 ผลการศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตาม แนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมิน ระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน.....	187
ตอนที่ 5 ผลการประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของผลการ ประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการ ประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน	208
ตอนที่ 6 ผลการตรวจสอบคุณภาพคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนว เดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน	.214

บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ .....	219
สรุปผลการวิจัย.....	220
อภิปรายผลการวิจัย.....	224
ข้อเสนอแนะ .....	235
รายการอ้างอิง .....	239
ภาคผนวก.....	252
ภาคผนวก ก รายนามผู้เชี่ยวชาญในการวิจัย.....	253
ภาคผนวก ข เครื่องมือในการวิจัย.....	256
ภาคผนวก ค ผลการวิเคราะห์ด้วยโปรแกรม FACETS .....	281
ภาคผนวก ง คู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่าง ข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน .....	304
ประวัติผู้เขียนวิทยานิพนธ์ .....	335

## สารบัญตาราง

ตาราง 1	สรุปสาระ มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษา ตอนต้น ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.....	26
ตาราง 2	จำนวนข้อสอบ O-NET ชั้นมัธยมศึกษาปีที่ 3 จำแนกตามรูปแบบของข้อสอบ ประจำปีการศึกษา 2559 .....	35
ตาราง 3	จำนวนข้อสอบ O-NET วิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำแนกตามรูปแบบของข้อสอบ.....	36
ตาราง 4	งานวิจัยที่เกี่ยวข้องกับการศึกษาความสอดคล้องในแนวเดียวกัน.....	69
ตาราง 5	มิติความรู้ (knowledge dimension) และมิติกระบวนการทางปัญญา (cognitive process dimension).....	77
ตาราง 6	ลำดับชั้นของกระบวนการทางปัญญา revised Bloom’s taxonomy .....	77
ตาราง 7	คำสำคัญและตัวอย่างพฤติกรรมในกระบวนการทางปัญญา ตามแนวคิด revised Bloom’s taxonomy .....	78
ตาราง 8	งานวิจัยที่เกี่ยวข้องกับการใช้โมเดลหลายองค์ประกอบของราส์ในการตรวจสอบ อิทธิพลของผู้ตรวจ.....	110
ตาราง 9	สรุปจำนวนสาระ มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จำแนกตามระดับชั้น .....	132
ตาราง 10	สรุปตัวชี้วัดตามรูปแบบของข้อสอบ O-NET วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 ของ สทศ.....	133
ตาราง 11	ตัวอย่างแบบบันทึกระดับความซับซ้อนทางปัญญาของข้อสอบในการประเมิน ระดับชาติ .....	136
ตาราง 12	ตัวอย่างแผนผังการจัดข้อสอบที่ใช้ในการศึกษาความสอดคล้องในแนวเดียวกัน .....	145
ตาราง 13	ค่าสถิติในการตรวจสอบอิทธิพลของผู้ประเมินและความแม่นยำในการให้คะแนน ด้วย MFRM .....	148



ตาราง 14	คุณลักษณะจำเพาะของข้อสอบ (item-test specification) ในการทดสอบ O-NET วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 2559 ประจำปีการศึกษา 3.....	160
ตาราง 15	ข้อมูลเกี่ยวกับการออกข้อสอบในการประเมินระดับชั้นเรียน จำแนกตามโรงเรียน ...	163
ตาราง 16	ข้อมูลเกี่ยวกับกระบวนการในการออกข้อสอบของโรงเรียน .....	165
ตาราง 17	จำนวนข้อสอบในการศึกษาความสอดคล้องในแนวเดียวกันจำแนกตามโรงเรียน ระดับชั้น มาตรฐานและตัวชี้วัด.....	168
ตาราง 18	ข้อมูลพื้นฐานของผู้เชี่ยวชาญ .....	170
ตาราง 19	ค่าสถิติแคปปาของฟลีของการประเมินความสอดคล้องในแนวเดียวกันระหว่าง ข้อสอบกับตัวชี้วัด .....	171
ตาราง 20	สถิติสหสัมพันธ์ภายในชั้นของการประเมินระดับความสอดคล้องในแนวเดียวกัน ระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ.....	172
ตาราง 21	ผลการวิเคราะห์ฟาเซทผู้ประเมินด้วยโปรแกรม FACETS.....	174
ตาราง 22	ลักษณะของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกันระหว่าง มาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน .....	176
ตาราง 23	จำนวนและร้อยละของข้อสอบจากการวิเคราะห์ด้วย MFRM จำแนกตามค่าเฉลี่ย คะแนนความสอดคล้องฯ ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน.....	178
ตาราง 24	จำนวนและร้อยละของข้อสอบจากการวิเคราะห์ด้วย MFRM จำแนกตามค่าดัชนี ความสอดคล้องระหว่างข้อสอบเชิงประจักษ์กับโมเดลและค่าเฉลี่ยคะแนนความ สอดคล้องฯ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน.....	179
ตาราง 25	ข้อสอบที่มีค่าเฉลี่ยคะแนนความสอดคล้องในแนวเดียวกันฯ หลังควบคุมอิทธิพล การกดหรือปล่อยคะแนน (Fair-M Avge) ต่ำกว่า 3.00 คะแนน (จำนวน 76 ข้อ)....	180
ตาราง 26	ตัวอย่างข้อสอบที่มีค่าเฉลี่ยคะแนนความสอดคล้องในแนวเดียวกันฯ หลังควบคุม อิทธิพลการกดหรือปล่อยคะแนน ตั้งแต่ 3.00 คะแนนขึ้นไป และมีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.50 – 1.50 (ตัวอย่างจำนวน 30 ข้อ จาก 902 ข้อ) .....	183

ตาราง 27	ตัวอย่างข้อสอบที่มีค่าเฉลี่ยคะแนนความสอดคล้องในแนวเดียวกันฯ หลังควบคุมอิทธิพล การกดหรือปล่อยคะแนน ตั้งแต่ 3.00 คะแนนขึ้นไป แต่มีข้อมูลเชิงประจักษ์ไม่สอดคล้องกับโมเดล (ตัวอย่างจำนวน 15 ข้อ จาก 111 ข้อ).....	184
ตาราง 28	ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนการประเมินความสอดคล้องในแนวเดียวกัน ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน.....	185
ตาราง 29	ข้อสอบที่มีการเปลี่ยนแปลงผลการประเมินความสอดคล้องในแนวเดียวกันฯ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน (ข้อ 21 จำนวน).....	186
ตาราง 30	สรุปข้อมูลความถี่ของข้อสอบ O-NET วิทยาศาสตร์ ในแต่ละระดับความซับซ้อนทางปัญญา จำแนกตามสาระและระดับความซับซ้อนทางปัญญา.....	187
ตาราง 31	ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน.....	206
ตาราง 32	สรุปร้อยละของสัดส่วนระดับความซับซ้อนทางปัญญา ของข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน.....	207
ตาราง 33	G-Study สำหรับประมาณค่าองค์ประกอบความแปรปรวนของการวัดในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบฯ.....	210
ตาราง 34	D-Study ของการประเมินระดับความซับซ้อนทางปัญญา.....	212
ตาราง 35	D-Study ของการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ.....	214
ตาราง 36	ผลการตรวจสอบคุณภาพด้านความตรงตามเนื้อหาของคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน.....	215
ตาราง 37	ผลการประเมินคุณภาพของคู่มือฯ ตามมาตรฐานการประเมิน.....	216

## สารบัญภาพ

ภาพ 1	ตัวอย่างเมทริกซ์มาตรฐาน และเมทริกซ์เครื่องมือประเมิน (Y) ของวิชาคณิตศาสตร์ ในระดับประถมศึกษา .....	52
ภาพ 2	กรอบแนวคิดการวิจัย.....	127
ภาพ 3	ตัวอย่างข้อรายการของแบบประเมินความสอดคล้องในแนวเดียวกันระหว่าง มาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนฯ.....	139
ภาพ 4	ตัวอย่างเมทริกซ์ข้อสอบในการประเมินตัวอย่างเมทริกซ์ข้อสอบในการประเมินระดับ ชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์Y).....	151
ภาพ 5	Venn diagram แสดงแหล่งความแปรปรวนของรูปแบบการประเมิน $i \times r$ design และ $r : i$ design .....	154
ภาพ 6	แผนผังแสดงขั้นตอนการดำเนินการวิจัย.....	158
ภาพ 7	แผนที่ตัวแปร (variables map) จากการวิเคราะห์ด้วยโปรแกรม FACETS .....	175
ภาพ 8	เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชาติ (เมทริกซ์ผลรวม X).....	189
ภาพ 9	เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชาติ (เมทริกซ์สัดส่วน X).....	190
ภาพ 10	เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียนโดยภาพรวม (เมทริกซ์ ผลรวม $Y_{\text{ภาพรวม}}$ ).....	191
ภาพ 11	เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียนโดยภาพรวม (เมทริกซ์ สัดส่วน $Y_{\text{ภาพรวม}}$ ).....	191
ภาพ 12	เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน <sub>ONETสูง 1</sub> (เมทริกซ์ ผลรวม $Y_{\text{โรงเรียนONETสูง 1}}$ ).....	192
ภาพ 13	เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน <sub>ONETสูง 1</sub> (เมทริกซ์ สัดส่วน $Y_{\text{โรงเรียนONETสูง 1}}$ ) .....	192
ภาพ 14	เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน <sub>ONETสูง 2</sub> (เมทริกซ์ ผลรวม $Y_{\text{โรงเรียนONETสูง 2}}$ ).....	193

ภาพ 15	เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน <sub>ONETสูง2</sub> (เมทริกซ์สัดส่วน Y โรงเรียน <sub>ONETสูง2</sub> ).....	193
ภาพ 16	เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน <sub>ONETต่ำ 1</sub> (เมทริกซ์ ผลรวม Y โรงเรียน <sub>ONETต่ำ 1</sub> ).....	194
ภาพ 17	เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน <sub>ONETต่ำ 1</sub> (เมทริกซ์ สัดส่วน Y โรงเรียน <sub>ONETต่ำ 1</sub> ).....	194
ภาพ 18	เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน <sub>ONETต่ำ 2</sub> (เมทริกซ์ ผลรวม Y โรงเรียน <sub>ONETต่ำ 2</sub> ).....	195
ภาพ 19	เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน <sub>ONETต่ำ 2</sub> (เมทริกซ์ สัดส่วน Y โรงเรียน <sub>ONETต่ำ 2</sub> ).....	195
ภาพ 20	เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบใน การประเมินระดับชั้นเรียน ของโรงเรียนโดยภาพรวม.....	196
ภาพ 21	แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับ ชั้นเรียน ของโรงเรียนโดยภาพรวม.....	197
ภาพ 22	เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบใน การประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 1 .....	198
ภาพ 23	แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับ ชั้นเรียนของโรงเรียน <sub>ONETสูง1</sub> .....	199
ภาพ 24	เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบใน การประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 2 .....	200
ภาพ 25	แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับ ชั้นเรียน ของโรงเรียน <sub>ONETสูง2</sub> .....	201
ภาพ 26	เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบใน การประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 1 .....	202
ภาพ 27	แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับ ชั้นเรียน ของโรงเรียน <sub>ONETต่ำ 1</sub> .....	203

ภาพ 28	เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 2.....	204
ภาพ 29	แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน ONETต่ำ 2.....	205
ภาพ 30	สรุปร้อยละของสัดส่วนระดับความซับซ้อนทางปัญญา ของข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน.....	208



## บทที่ 1

### บทนำ

#### ความเป็นมาและความสำคัญของการศึกษา

ระบบการศึกษาแบบอิงมาตรฐาน (standard-based educational system) ประกอบด้วย 3 องค์ประกอบหลักที่สำคัญ ได้แก่ มาตรฐานการเรียนรู้ การเรียนการสอน และการประเมินผล ความสอดคล้องไปในทิศทางเดียวกันของทั้ง 3 องค์ประกอบดังกล่าวจึงเป็นเงื่อนไขพื้นฐานที่สำคัญของระบบการจัดการศึกษาแบบอิงมาตรฐาน (Fuhrman, 2001; Smith & O'Day, 1990) เนื่องจากการช่วยในการส่งเสริมการเรียนรู้ของผู้เรียน (Anderson, 2002; Biggs, 2003; La Marca, Redfield, & Winter, 2000) ช่วยให้ระบบการศึกษาเกิดประสิทธิผล (Webb, 1997) มีความสำคัญต่อการประเมินการปฏิรูปทางการศึกษา และยังให้สารสนเทศที่มีประโยชน์ต่อทั้งกับผู้เรียน ผู้ปกครอง สาธารณชน และผู้วางนโยบาย (Herman, Webb, & Stephen, 2007) ความสอดคล้องกันดังกล่าว ใช้คำว่า “alignment” แปลว่า ความสอดคล้องในแนวเดียวกัน ซึ่งในทางการศึกษา หมายถึง ระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างองค์ประกอบต่างๆ ในระบบการศึกษา อันได้แก่ ผลการเรียนรู้ที่คาดหวัง (เช่น มาตรฐานการเรียนรู้ วัตถุประสงค์หรือตัวชี้วัด) การเรียนการสอน และการประเมินผล (เช่น เครื่องมือในการประเมิน) (Ananda, 2003; Bholá, Impara, & Buckendahl, 2003; Biggs, 2003; Case, Jorgensen, & Zucker, 2004; La Marca, 2001; Webb, 1997)

การศึกษาเรื่องความสอดคล้องในแนวเดียวกันของระบบการศึกษาในยุคการศึกษาอิงมาตรฐานได้รับความสนใจอย่างมากในช่วงปี 1990 เป็นต้นมา โดยความสนใจนี้เกิดขึ้นในหมู่นักวิชาการ นักการศึกษา และผู้กำหนดนโยบายการศึกษาของประเทศสหรัฐอเมริกาเป็นส่วนใหญ่ สาเหตุสำคัญของการจุดประกายให้มีความสนใจในเรื่องนี้ คือ การปฏิรูปการศึกษาที่ต้องการเห็นระบบการศึกษาต่างๆ มีความสอดคล้องและประสานกันทุกระบบเพื่อให้เกิดความมั่นใจว่าการดำเนินงานภายใต้การปฏิรูปการศึกษาจะประสบความสำเร็จ ความสำคัญเช่นนี้ นำสู่การออกกฎหมายทางการศึกษาหลายฉบับที่ต้องการให้ผู้ปฏิบัติจัดระบบการศึกษาให้สอดคล้องกัน เช่น ในปี ค.ศ. 1995 กฎหมายการปฏิรูปโรงเรียนของสหรัฐอเมริการะบุว่ารัฐต่างๆ ต้องทำการประเมินให้สอดคล้องหรือสอดคล้องกับหลักสูตร รวมทั้งการปฏิรูปการศึกษาล่าสุดของอเมริกา คือ No Child Left Behind หรือ NCLB ที่เริ่มใช้ในปี ค.ศ. 2001 ดังนั้น การศึกษาเรื่องความสอดคล้องไม่ใช่เรื่องใหม่สำหรับนักวิชาการในแวดวงการศึกษาและประเมินทางการศึกษา เพราะการศึกษาความสอดคล้องในบริบทของการวัดและประเมินผลอิงมาตรฐานเป็นประเด็นเกี่ยวกับการศึกษาเรื่อง “ความตรง” (validity) นักวิชาการหลายท่าน เช่น Porter (2002) และ Webb (1997) กล่าวว่าความสอดคล้อง

เป็นเงื่อนไขที่จำเป็นเงื่อนไขหนึ่งที่จะทำให้การใช้ผลการวัดความรู้ของผู้เรียนมีความถูกต้องและเหมาะสม การประเมินความสอดคล้องยิ่งทวีความสำคัญมากขึ้นในบริบทที่มีการจัดการศึกษาที่อิงมาตรฐาน เพราะเป็นเรื่องของการตรวจสอบความเหมาะสมของการวัดและการประเมินทางการศึกษา หากต้องการประเมินคุณภาพผู้เรียนตามหลักสูตรอิงมาตรฐาน แต่พบว่ามาตรฐานที่กำหนดในหลักสูตรกับเครื่องมือประเมินไม่สอดคล้องกันก็จะทำให้การใช้ผลการประเมินคุณภาพผู้เรียนนั้นไม่ถูกต้อง และการประเมินผลความสำเร็จของการจัดการศึกษาก็จะไม่ถูกต้องด้วยเช่นกัน (สังวรณัฎฐ์ จัตตะระโทก, 2555)

สำหรับในประเทศไทยที่เริ่มใช้หลักสูตรอิงมาตรฐานตั้งแต่ที่กระทรวงศึกษาได้ประกาศใช้หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 และต่อมาได้มีการปรับปรุงไปสู่หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ซึ่งได้กำหนดมาตรฐานและตัวชี้วัด เป็นสิ่งที่สะท้อนคุณภาพผู้เรียนในแต่ละกลุ่มสาระการเรียนรู้ตามมาตรฐานและตัวชี้วัดที่กำหนดไว้ในแต่ละกลุ่มสาระการเรียนรู้ โดยถือว่าครูผู้สอนเป็นผู้ที่มีบทบาทสำคัญที่สุดในการจัดการเรียนการสอนที่เน้นผู้เรียนให้มีคุณภาพตามมาตรฐานและตัวชี้วัด และยึดมาตรฐานและตัวชี้วัดเป็นกรอบในการจัดการเรียนรู้ในแต่ละหน่วยการเรียนรู้ รวมถึงการวัดและประเมินผลคุณภาพผู้เรียน เพื่อตรวจสอบหรือยืนยันว่าผู้เรียนมีความรู้ความสามารถตามที่หลักสูตรต้องการ ตลอดจนปรับปรุงและพัฒนาผู้เรียนที่ไม่ผ่านในมาตรฐานหรือตัวชี้วัดตามที่หลักสูตรกำหนด เพื่อให้พร้อมเข้ารับการศึกษาในระดับชาติขั้นพื้นฐาน (O-NET) ในระดับชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ซึ่งใช้มาตรฐานและตัวชี้วัดของหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 เป็นกรอบในการทดสอบคุณภาพผู้เรียนเช่นเดียวกัน ดังนั้น ถ้าโรงเรียนสามารถออกข้อสอบที่ใช้ในการวัดและประเมินผลในชั้นเรียนได้ตรงตามมาตรฐานและตัวชี้วัดตามหลักสูตรก็จะสามารถสะท้อนคุณภาพของผู้เรียนตามหลักสูตรได้อย่างชัดเจน นำไปสู่การยกระดับคุณภาพของผู้เรียนในการทดสอบระดับชาติอีกด้วย (สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2559)

อย่างไรก็ตาม สำหรับประเทศไทย มีการศึกษาเรื่องความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานกับการประเมินในระดับชั้นเรียน เช่น กฤตวรรณ ทองมี (2557) นราศักดิ์ ไชยเรือง (2557) และพดุม ไบระหมาน, สังวรณัฎฐ์ จัตตะระโทก, และ นลินี ณ นคร (2559) แต่ยังไม่ปรากฏถึงการศึกษาวิธีการวัดความสอดคล้องในแนวเดียวกันของการประเมินระดับใหญ่ ซึ่งมีผลกระทบสูง ซึ่งอาจเป็นเพราะประเด็นเกี่ยวกับการวัดความสอดคล้องต่างๆ ยังไม่เป็นที่รู้จักแพร่หลายในหมู่นักวิชาการมากนัก ถึงแม้ว่าประเด็นเรื่องเกี่ยวกับการวัดความสอดคล้องในแนวเดียวกันจะเป็นสิ่งจำเป็นมากในการวัดและประเมินผลที่อิงมาตรฐาน

การพัฒนาแบบส่งเสริมสนับสนุนคุณภาพการจัดการศึกษา ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ได้ให้ความสำคัญกับระบบการวัดและประเมินผลอย่างยิ่ง ซึ่งตามหลักสูตร

แกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กำหนดให้มีการวัดผลการเรียนรู้ของผู้เรียนไว้ 3 ระดับ คือ ระดับชั้นเรียน ระดับสถานศึกษา และระดับชาติ (สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2559) โดยการวัดและประเมินผลระดับชั้นเรียนเป็นหน้าที่ของครูผู้สอนที่จะต้องหาคำตอบว่าการจัดกิจกรรมการเรียนการสอนส่งผลให้ผู้เรียนมีความก้าวหน้าด้านความรู้ ทักษะกระบวนการ ด้านคุณธรรม จริยธรรม และค่านิยมอันพึงประสงค์มากน้อยเพียงใด การวัดและประเมินผลในระดับสถานศึกษา เป็นการประเมินเพื่อตรวจสอบความก้าวหน้าด้านการเรียนรู้ของผู้เรียนทุกคนที่เรียนในชั้นปีสุดท้ายของโรงเรียน ได้แก่ ชั้นประถมศึกษาปีที่ 3 และ 6 ชั้นมัธยมศึกษาปีที่ 3 และ 6 ส่วนการวัดและประเมินผลระดับชาติ เป็นการทดสอบความรู้ทางการศึกษาในระดับชาติขั้นพื้นฐานของผู้เรียนทุกคน ทุกสังกัดที่กำลังศึกษาในชั้นปีสุดท้าย ได้แก่ ชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ซึ่งประเมินตามมาตรฐานการเรียนรู้ในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จำนวน 51 มาตรฐานการเรียนรู้ ครอบคลุม 5 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ศาสนา และวัฒนธรรม และภาษาต่างประเทศ (ภาษาอังกฤษ) โดยมีวัตถุประสงค์เพื่อทดสอบความรู้และความคิดของนักเรียนชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 และนำผลการทดสอบไปใช้เพื่อเป็นองค์ประกอบหนึ่งในการจบการศึกษา ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 เพื่อปรับปรุงคุณภาพการเรียนการสอนของโรงเรียน ตลอดจนเพื่อนำผลการทดสอบไปใช้ในการประเมินผลการเรียนรู้ของนักเรียนระดับชาติ และเพื่อนำผลการทดสอบไปใช้ในวัตถุประสงค์อื่น (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2559: เว็บไซต์) ดังนั้น การวัดและประเมินผลของแต่ละระดับ ทั้งในระดับชั้นเรียน ระดับสถานศึกษา และระดับชาติควรมีความสอดคล้องกัน ตลอดจนมีสอดคล้องกับมาตรฐานและตัวชี้วัดของหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ก็จะสามารถสร้างความมั่นใจด้านคุณภาพของผู้เรียนให้กับผู้ที่เกี่ยวข้องทั้งภายในและภายนอกสถานศึกษาได้มากยิ่งขึ้น

อย่างไรก็ตาม การทดสอบความรู้ทางการศึกษาในระดับชาติขั้นพื้นฐานในแต่ละปีการศึกษาที่ผ่านมาสะท้อนให้เห็นถึงความรู้ความสามารถของผู้เรียนที่อยู่ในระดับต่ำ จากงานวิจัยของ เอื้อมพร หลินเจริญ, สิริศักดิ์ อาจวิชัย, และภริภา จันทรอินทร์ (2552) สรุปสาเหตุที่ทำให้คะแนนการทดสอบ O-NET ต่ำมาจากหลายปัจจัย เช่น ปัจจัยด้านผู้เรียน ปัจจัยด้านนโยบายในการพัฒนาคุณภาพวิชาการของโรงเรียนหรือสภาพแวดล้อมของโรงเรียน ปัจจัยด้านครูผู้สอน ซึ่งในการเพิ่มผลสัมฤทธิ์ของการสอบ O-NET นั้น ปัจจัยด้านครูผู้สอนนั้น ถือว่ามีความสำคัญและต้องได้รับพัฒนาเป็นลำดับต้นๆ เนื่องจากครูผู้สอนเป็นผู้ที่มีบทบาทสำคัญที่สุดในการจัดการเรียนการสอนที่เน้นผู้เรียนให้มีคุณภาพตามมาตรฐานและตัวชี้วัด ตลอดจนทำการวัดและประเมินผลระดับชั้นเรียนเพื่อหาคำตอบว่าการจัดกิจกรรมการเรียนการสอนส่งผลให้ผู้เรียนมีความก้าวหน้าด้านความรู้ ทักษะกระบวนการ ด้าน



คุณธรรม จริยธรรม และค่านิยมอันพึงประสงค์ตามมาตรฐานและตัวชี้วัดอย่างน้อยเพียงใด ซึ่งบริษัท อักษรเจริญทัศน์ จำกัด (2559) ได้จัดทำสารสนเทศเกี่ยวกับการเพิ่มผลสัมฤทธิ์ของการสอบ O-NET โดยได้ให้ข้อเสนอแนะว่าครูยังต้องมีบทบาทในเรื่องต่างๆ โดยสรุปคือ การสอนโดยยึดตามหลักสูตร และสอนให้ครอบคลุมหลักสูตร จัดกิจกรรมการเรียนการสอนที่ส่งเสริมการคิดวิเคราะห์ (critical thinking) เพื่อให้สอดคล้องกับข้อสอบ O-NET ของ สทศ. ที่เน้นการคิดวิเคราะห์ ศึกษาแนวข้อสอบ O-NET ย้อนหลัง และควรทำ Test Blueprint ของข้อสอบ O-NET จัดให้ผู้เรียนได้ฝึกทำแบบทดสอบเสมือนจริง โดยจำลองบรรยากาศการสอบ O-NET ทำการวัดและประเมินผลโดยใช้ข้อสอบให้สอดคล้องหรือคล้ายคลึงกับข้อสอบ O-NET โดยเฉพาะข้อสอบแบบปรนัยที่มีรูปแบบข้อสอบที่แตกต่างจากที่โรงเรียนใช้ ทำโครงการต่างๆ เพื่อเพิ่มผลสัมฤทธิ์ของการสอบ O-NET ตลอดจนสร้างความตระหนักให้ผู้เรียนเห็นความสำคัญของการสอบ O-NET ดังนั้น การตรวจสอบว่าครูทำการวัดและประเมินผลโดยใช้ข้อสอบให้สอดคล้องหรือคล้ายคลึงกับข้อสอบ O-NET รวมถึงออกข้อสอบให้เน้นระดับความซับซ้อนทางปัญญา เช่น การคิดวิเคราะห์ ให้สอดคล้องกับข้อสอบ O-NET ที่กล่าวข้างต้นนั้นจึงเป็นเรื่องที่มีความสำคัญ ซึ่งสามารถตรวจสอบได้จากการประเมินความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ (ข้อสอบ O-NET) กับข้อสอบประเมินระดับชั้นเรียน เพื่อให้ได้สารสนเทศว่าโรงเรียนสามารถวัดและประเมินผลในระดับชั้นเรียนได้สอดคล้องกับการประเมินระดับชาติอย่างน้อยเพียงใด อันจะนำไปสู่ การปรับปรุงการเรียนการสอน และพัฒนาการวัดและประเมินในชั้นเรียน เพื่อสร้างความมั่นใจด้านคุณภาพของผู้เรียน และสนับสนุนเพิ่มผลสัมฤทธิ์ของการทดสอบระดับชาติ ซึ่งอาจนำไปสู่การเพิ่มความรู้ความสามารถของผู้เรียน นอกจากนี้ การประเมินดังกล่าวสามารถตอบคำถามในเชิงประจักษ์เพื่อยืนยันได้อีกว่าการวัดและประเมินผลของครูผู้สอนในระดับชั้นเรียน เป็นปัจจัยสำคัญที่มีความสัมพันธ์กับคะแนนการสอบ O-NET หรือไม่ หรือเป็นผลจากปัจจัยหลักด้านอื่นๆ

กระทรวงศึกษาธิการ ได้เริ่มเล็งเห็นถึงความสำคัญของความสอดคล้องกันระหว่างการออกข้อสอบระดับในการประเมินระดับชาติกับข้อสอบในการประเมินระดับชั้นเรียน ตลอดจนความสำคัญของการศึกษาความสอดคล้องในแนวเดียวกันระหว่างตัวชี้วัดตามมาตรฐานการเรียนรู้ กับข้อสอบในการประเมินระดับโรงเรียน โดยได้มีการประชุมหารือกับ สทศ. เมื่อวันอังคารที่ 16 กุมภาพันธ์ พ.ศ. 2559 และได้มอบหมายให้สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) ส่ง test blueprint ข้อสอบแต่ละระดับชั้นให้ สทศ. นำไปวิเคราะห์และบูรณาการกับการออกข้อสอบของ สทศ. ต่อไป ซึ่งจะช่วยให้การออกข้อสอบเป็นไปในทิศทางเดียวกัน และช่วยสร้างความตระหนักให้กับโรงเรียนในการออกข้อสอบให้ตรงตัวชี้วัด ตลอดจนช่วยให้ สทศ. สามารถออกข้อสอบได้ตรงกับที่นักเรียนเรียนมากขึ้น นอกจากนี้ ยังมอบหมายให้ สพฐ. ทำการพิจารณาการออกข้อสอบของครูในโรงเรียนว่าตรงกับตัวชี้วัดหรือไม่ ทั้งข้อสอบกลางภาคและปลายภาค จึงเป็นที่มาของการศึกษาวิจัยเพื่อติดตามและ

ตรวจสอบคุณภาพของข้อสอบที่ใช้ในการวัดและประเมินผลในชั้นเรียนของสถานศึกษา สังกัด สพฐ. ในปี พ.ศ. 2559 โดยทำการศึกษาข้อสอบในการประเมินระดับชั้นเรียน ระดับชั้นประถมศึกษาปีที่ 6 และมัธยมศึกษาปีที่ 3 ของโรงเรียนในสังกัด สพฐ. ผลการศึกษาพบว่า ทั้งระดับชั้นประถมศึกษาปีที่ 6 และมัธยมศึกษาปีที่ 3 มีข้อสอบที่มีความสอดคล้องกับตัวชี้วัดมากกว่าร้อยละ 94.00 ในทุกกลุ่มสาระการเรียนรู้ (สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2559) อย่างไรก็ตาม การศึกษาความสอดคล้องในแนวเดียวกันดังกล่าว ยังไม่ได้อยู่บนพื้นฐานของวิธีการศึกษาความสอดคล้องในแนวเดียวกันที่เป็นสากล ซึ่งอาจให้สารสนเทศที่สำคัญของการศึกษาความสอดคล้องในแนวเดียวกันที่ยังไม่ครบถ้วน ตลอดจนขาดการควบคุมแหล่งความคลาดเคลื่อนที่อาจส่งผลต่อคุณภาพการประเมินความสอดคล้องได้

ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ได้มีการกำหนดมาตรฐานการเรียนรู้และตัวชี้วัดสาระการเรียนรู้แกนกลาง เพื่อระบุถึงคุณภาพที่ต้องการให้เกิดขึ้นในตัวผู้เรียน เป็นสิ่งที่ผู้เรียนพึงรู้และปฏิบัติได้เมื่อจบการศึกษาขั้นพื้นฐาน แบ่งตามกลุ่มสาระการเรียนรู้เป็น 8 กลุ่ม ได้แก่ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคม ศาสนาและวัฒนธรรม สุขศึกษาและพลศึกษา ศิลปะ การงานอาชีพและเทคโนโลยี และภาษาต่างประเทศ กลุ่มสาระวิทยาศาสตร์เป็นหนึ่งในกลุ่มสาระหลักที่มีความสำคัญในการพัฒนาผู้เรียน เนื่องจากวิทยาศาสตร์มีบทบาทสำคัญยิ่งในสังคมโลกปัจจุบันและอนาคต เพราะวิทยาศาสตร์เกี่ยวข้องกับทุกคนทั้งในชีวิตประจำวันและการงานอาชีพต่างๆ ตลอดจนเทคโนโลยี เครื่องมือเครื่องใช้และผลผลิตต่างๆ ที่มนุษย์ได้ใช้เพื่ออำนวยความสะดวกในชีวิตและการทำงาน ดังนั้น การศึกษาความสอดคล้องในแนวเดียวกันในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ในโรงเรียนจึงมีความสำคัญ เนื่องจากช่วยในการตรวจสอบว่าผู้เรียนทุกคนมีโอกาสในการเรียนรู้ในสิ่งที่ผู้เรียนพึงรู้และปฏิบัติได้ในด้านวิทยาศาสตร์ตามที่มาตรฐานกำหนดอย่างเท่าเทียมกัน

ปัจจุบันวิธีการศึกษาความสอดคล้องในแนวเดียวกันที่พบในงานวิจัยส่วนใหญ่มี 3 วิธี ได้แก่ (1) วิธีการของ Webb (Webb, 1997, 1999, 2002, 2007) (2) วิธีการของ Porter หรือเรียกว่าวิธีการสำรวจหลักสูตรที่นำไปใช้จริง (survey of enacted curriculum: SEC) (Porter & Smithson, 2001, 2002) และ (3) วิธีการของบริษัท Achieve (Rothman, Slattery, & Vranek (2002)) โดยวิธีการที่ (1) และ (3) เป็นวิธีการที่ใช้ในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างเนื้อหาของมาตรฐานกับเนื้อหาของการประเมินที่มีความครอบคลุม ในขณะที่วิธีการที่ (2) เป็นวิธีการที่สามารถใช้ในการศึกษาระดับความสอดคล้องเชิงปริมาณได้ทั้ง 3 องค์ประกอบของระบบการศึกษา ได้แก่ มาตรฐาน การประเมิน ตำราเรียน และแม้แต่การสอนของครูในห้องเรียน (Porter, Smithson, Blank, & Zeidner, 2007) ดังนั้น วิธีการของ Porter จึงเป็นวิธีการเดียวที่มีความยืดหยุ่นและเหมาะสมที่สุดในการนำมาประยุกต์ใช้ในงานวิจัยนี้ ซึ่งเป็นการศึกษาความสอดคล้องในแนวเดียวกัน

ระหว่าง 2 การประเมิน ได้แก่ การประเมินระดับชาติ (ข้อสอบ O-NET) และการประเมินระดับชั้นเรียน (ข้อสอบในการประเมินผลการเรียนรู้ที่ครูผู้สอนสร้างขึ้น)

วิธีการศึกษาความสอดคล้องในแนวเดียวกันของ Porter เป็นวิธีการที่พัฒนาขึ้นมาจากแนวคิดการประเมินหลักสูตร โดยให้ความสนใจกับคำว่า เนื้อหา (content) ครอบคลุมการวัดความสอดคล้องเริ่มจากการแบ่งประเภทของเนื้อหาออกเป็นเรื่องๆ (topic) ในแต่ละเรื่องจะทำการประเมินความลึกของความรู้ โดยความลึกของความรู้เป็นทักษะทางปัญญาที่ต้องการให้ผู้เรียนบรรลุในระหว่างการเรียนการสอน เช่น การจำ ความเข้าใจ การนำไปใช้ เป็นต้น การวัดความสอดคล้องด้วยวิธีการ ของ Porter เป็นการนำผลการประเมินของผู้เชี่ยวชาญมาสร้างเมทริกซ์ 2 เมทริกซ์ เช่น เมทริกซ์ข้อสอบการประเมินระดับชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) แล้วคำนวณค่าดัชนีความสอดคล้อง (alignment index) ที่เป็นดัชนีสำหรับใช้ประเมินว่า 2 องค์ประกอบทางการศึกษาที่นำมาประเมินมีความสอดคล้องกันมากน้อยเพียงใด ซึ่งจุดเด่นของวิธีการนี้ คือ มีความยืดหยุ่นและสามารถให้ค่าดัชนีความสอดคล้อง 1 ค่า ซึ่งทำให้ง่ายต่อการเข้าใจและเปรียบเทียบ

การศึกษาความสอดคล้องในแนวเดียวกัน ส่วนใหญ่เป็นการศึกษาความสอดคล้องใน 2 มิติ คือ มิติด้านเนื้อหาและมิติด้านความซับซ้อนทางปัญญาหรือระดับความลึกของความรู้ ซึ่งการจำแนกความซับซ้อนทางปัญญาสามารถจำแนกได้หลายรูปแบบ ขึ้นอยู่กับแนวคิดที่ใช้ในการจำแนก เช่น แนวคิดการจำแนกลำดับชั้นของกระบวนการทางปัญญาของ Bloom ที่ได้รับการปรับปรุงใหม่ (revised Bloom's taxonomy) (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths, & Wittrock, 2001) จำแนกความซับซ้อนทางปัญญาได้เป็น 6 ชั้น ได้แก่ จำ เข้าใจ ประยุกต์ใช้ วิเคราะห์ ประเมินค่า และสร้างสรรค์ แนวคิดของ Webb จำแนกได้เป็น 4 ระดับ ได้แก่ ระลึกได้ ทักษะ/ความคิดรวบยอด การคิดเชิงกลยุทธ์ และการขยายความคิด หรือแนวคิดของ Porter จำแนกได้เป็น 5 ระดับ ได้แก่ จำ ปฏิบัติตามขั้นตอน สื่อสารความเข้าใจ แก้ปัญหาในสถานการณ์ใหม่ และคาดคะเน/สรุปอ้างอิง/พิสูจน์ แนวคิดของ Marzano (Marzano & Kendall, 2001) จำแนกได้เป็น 4 ระดับ ได้แก่ ดึงความรู้เดิมออกมาใช้ สร้างความคิดรวบยอด วิเคราะห์ความรู้ ใช้ประโยชน์จากความรู้ อภิปัญญา และจัดระบบแห่งตน อย่างไรก็ตาม Näsström และ Henriksson (2008) ได้ทำวิเคราะห์กรอบแนวคิดหรือลำดับชั้นในการจำแนกความซับซ้อนทางปัญญาจากแนวคิด 9 แนวคิด และสรุปว่า revised Bloom's taxonomy มีความครอบคลุมและจำเพาะ ตลอดจนมีค่าความเที่ยงระหว่างผู้ตรวจสูงกว่าเกณฑ์การจำแนกตามแนวคิดอื่นๆ นอกจากนี้ แนวคิด revised Bloom's taxonomy ยังเป็นการจำแนกพฤติกรรมการเรียนรู้ทางปัญญาที่ง่ายต่อการสื่อความให้เป็นที่เข้าใจตรงกันในหมู่ครูและนักวิชาการของประเทศไทย ด้วยเหตุผลที่กล่าวข้างต้น งานวิจัยนี้จึงเลือกใช้แนวคิด revised Bloom's taxonomy มาใช้ในการศึกษาระดับความซับซ้อนทางปัญญาของข้อสอบ

อย่างไรก็ตาม การศึกษาความสอดคล้องในแนวเดียวกัน ต้องอาศัยผู้เชี่ยวชาญในการพิจารณาตัดสินหรือให้คะแนนความสอดคล้องขององค์ประกอบต่างๆ โดยผู้เชี่ยวชาญจะต้องมีความรู้และคุ้นเคยกับเนื้อหาวิชา ลักษณะของผู้สอบ หลักสูตร ตลอดจนมาตรฐานการเรียนรู้ที่ต้องการประเมิน (Davis-Becker and Buckendahl 2013; La Marca, Redfield, & Winter, 2000) ซึ่งการใช้ดุลพินิจในการตัดสินหรือให้คะแนนจากผู้เชี่ยวชาญในการพิจารณาให้คะแนน อาจเกิดความคลาดเคลื่อนในการวัดอันเนื่องมาจากอัตวิสัยในการพิจารณาตัดสินของผู้เชี่ยวชาญหรือผู้ตรวจประเมิน (subjectivity of rater's judgment) (Song & Wolfe, 2015: 2) การลดความคลาดเคลื่อนนี้ โดยทั่วไปทำได้โดยการคัดเลือกผู้เชี่ยวชาญที่มีคุณสมบัติเหมาะสม การฝึกอบรมผู้เชี่ยวชาญให้มีความรู้ความเข้าใจในการประเมิน ตลอดจนการกำกับติดตามการดำเนินงานของผู้เชี่ยวชาญ (Wolfe, 2004: 36) นอกจากนี้ การใช้วิธีการทางสถิติในการควบคุมอิทธิพลของผู้ประเมิน (rater effect) ที่อาจส่งต่อผลการพิจารณาของผู้ประเมินหรือผู้เชี่ยวชาญ ยังเป็นอีกวิธีการหนึ่งที่จะช่วยลดความคลาดเคลื่อนอันเกิดจากผู้ประเมินได้

โมเดลหลายองค์ประกอบของราส์ช (many-facets Rasch measurement: MFRM) เป็นโมเดลทางสถิติที่นิยมนำมาใช้ในการตรวจสอบและควบคุมความคลาดเคลื่อนอันเกิดจากอิทธิพลของผู้ประเมิน (rater effect) ซึ่งเป็นกลุ่มของผลกระทบของความแปรปรวนอย่างเป็นระบบในการให้คะแนนที่เกี่ยวข้องอย่างใดอย่างหนึ่งกับผู้ประเมิน และเป็นสิ่งที่ไม่ได้เกี่ยวข้องกับความสามารถที่แท้จริงของผู้ถูกประเมิน (Scullen, Mount, & Goff, 2000) อิทธิพลของผู้ประเมินเป็นรูปแบบของการประเมินที่มีความคลาดเคลื่อนในการวัด (Engelhard, 1994; Myford & Wolfe, 2003, 2004; Wolfe, 2004; Wolfe & McVay, 2012) อิทธิพลของผู้ประเมินพบบ่อย ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน (severity/leniency effect) อิทธิพลแนวโน้มการให้คะแนนค่ากลาง (central tendency effect) อิทธิพลฮาโล (halo effect) อิทธิพลการจำกัดช่วง (restriction of range effect) และอิทธิพลของความไม่มีแบบแผน (randomness effect)

Linacre (1994) ได้ปรับขยาย MFRM มาจากโมเดลของราส์ช สำหรับการตอบข้อสอบแบบลำดับชั้น (order categories) ข้อสอบที่มีการตรวจให้คะแนนแบบบางส่วน (partial credit test items) และมีการปรับขยายโมเดลไปสู่สถานการณ์ที่มากกว่า 2 องค์ประกอบ (facets) โดยเพิ่มองค์ประกอบที่สนใจเข้าไปในโมเดล ทำให้สามารถศึกษาได้มากกว่า 2 องค์ประกอบ เช่น เพิ่มองค์ประกอบผู้ประเมินเข้าไปในโมเดล จึงทำให้สามารถประมาณค่าพารามิเตอร์ของผู้ประเมิน (rater parameter) Linacre (1994) มีแนวคิดว่าการได้คะแนนในระดับใดระดับหนึ่งของผู้สอบได้รับอิทธิพลจาก 4 องค์ประกอบ คือ ความสามารถของผู้สอบ (ability) ความยากของข้อสอบ (difficulty) การกดคะแนนของผู้ประเมิน (rater severity) และโครงสร้างของมาตราประมาณค่า (rating scale) ซึ่งฟังก์ชันความสัมพันธ์ขององค์ประกอบทั้งหมดจะแสดงด้วยสมการทางคณิตศาสตร์ MFRM สามารถบ่งชี้พฤติกรรมการตรวจให้คะแนนของผู้ประเมินว่ามีการกดหรือปล่อยคะแนนมากเพียงใด

การวิเคราะห์ข้อมูลโดยใช้ MFRM จึงเป็นการศึกษาอิทธิพลของแหล่งความคลาดเคลื่อนต่างๆ ที่มีต่อคุณภาพของการวัดประเมิน ผู้วิจัยสามารถใช้โครงสร้างทางสถิติในการวิเคราะห์ผลการประเมินเพื่อการสรุปรูปแบบ (pattern) ของผลหลักสำหรับผู้ประเมิน (rater) ผู้สอบ (ratee) และคุณลักษณะ (trait) หรือข้อสอบ (item) และตัวแปรอื่นๆ ที่สนใจ ทั้งนี้ การวิเคราะห์ด้วย MFRM สามารถแยกคำนวณในแต่ละองค์ประกอบ (facet) อย่างเป็นอิสระจากองค์ประกอบอื่น (Myford & Wolfe, 2003) นอกจากนี้ ยังสามารถประเมินในระดับบุคคล (individual – level effects) ได้แก่ ผู้ประเมินแต่ละคน ผู้สอบแต่ละคน คุณลักษณะหรือข้อสอบแต่ละข้อ ซึ่งเป็นการศึกษาในระดับสมาชิกภายในแต่ละองค์ประกอบว่ามีลักษณะใด

Anderson, Irvin, Alonzo, และ Tindal (2015) ได้ประยุกต์ใช้ MFRM เพื่อควบคุมอิทธิพลของผู้ประเมิน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับมาตรฐานผ่านระบบออนไลน์ ผู้ประเมินความสอดคล้องเป็นครูคณิตศาสตร์ จำนวน 25 คน ทำการประเมินความสอดคล้องโดยใช้มาตรประมาณค่าแบบเรียงอันดับ 4 อันดับ การวิเคราะห์ข้อมูลโดยประยุกต์ใช้ MFRM ในการควบคุมอิทธิพลการกดหรือปล่อยคะแนนผู้ประเมิน และให้ค่าสถิติการวินิจฉัยในระดับข้อสอบ (item-level diagnostic statistics) การประยุกต์ใช้ MFRM มีการปรับแก้เล็กน้อยเพื่อให้เหมาะสมกับข้อมูลการศึกษาความสอดคล้องในแนวเดียวกัน โดยที่ให้ข้อสอบ (items) เป็นสิ่งที่ถูกประเมิน แทนที่จะเป็นผู้ถูกประเมิน (persons) และผู้ประเมิน (raters) แทนข้อสอบ (items) ทั้งนี้ ผู้ประเมินจะเป็นผู้ตัดสินคุณลักษณะแฝง (latent trait) ของข้อสอบ ข้อตกลงเบื้องต้นคือ ข้อสอบอยู่บนเงื่อนไขของการกดหรือปล่อยคะแนนของผู้ประเมินแต่ละคน ผลการวิจัยพบว่า มีความแตกต่างอย่างมีนัยสำคัญในการศึกษาความสอดคล้องระหว่างข้อสอบกับมาตรฐาน เมื่อมีการใช้สถิติในการควบคุมอิทธิพลของผู้ประเมิน

อย่างไรก็ตาม งานวิจัยของ Anderson, Irvin, Alonzo, และ Tindal (2015) เป็นการประยุกต์ใช้ MFRM เพื่อตรวจสอบและควบคุมอิทธิพลของผู้ประเมินเฉพาะอิทธิพลด้านการกดหรือปล่อยคะแนนเท่านั้น แต่ความคลาดเคลื่อนที่เกิดจากผู้ประเมินที่สามารถส่งผลกระทบต่อความถูกต้องในการประเมินอาจมีได้หลายประเภท ดังนั้น เพื่อให้เกิดความครอบคลุมในการศึกษาและสอดคล้องกับบริบทของการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบ งานวิจัยนี้จึงได้เพิ่มการตรวจสอบอิทธิพลของผู้ประเมินประเภทอื่น นอกเหนือจากอิทธิพลด้านการกดหรือปล่อยคะแนน ได้แก่ อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน นอกจากนี้ งานวิจัยของ Anderson, Irvin, Alonzo, และ Tindal (2015) ได้ศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับมาตรฐานในงานวิจัยเฉพาะมิติด้านเนื้อหาเท่านั้น ซึ่งการศึกษาความสอดคล้องในแนวเดียวกันที่ทำให้ได้สารสนเทศอย่างครอบคลุม ควรศึกษาความสอดคล้องอย่างน้อยใน 2 มิติ ได้แก่ มิติเนื้อหาและมิติความซับซ้อนทางปัญญา (Näsström &

Henriksson, 2008) ดังนั้น งานวิจัยนี้ จึงได้ศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียน ครอบคลุม 2 มิติ ได้แก่ มิติด้านเนื้อหา และมิติความซับซ้อนทางปัญญา โดยประยุกต์วิธีการศึกษาความสอดคล้องตามแนวคิดของ Porter และแนวคิด revised Bloom's taxonomy (Anderson et al., 2001) เพื่อให้ได้ข้อมูลสำหรับวิเคราะห์โดยประยุกต์ใช้ MFRM ในการตรวจสอบอิทธิพลของผู้ประเมิน และเพื่อใช้ในการตัดสินใจว่าข้อสอบแต่ละสอดคล้องหรือวัดตัวชี้วัดที่ครูระบุจริงหรือไม่ ก่อนคัดเลือกข้อสอบไปใช้ในการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (Y) และคำนวณค่าดัชนีความสอดคล้อง (alignment index) ระหว่างข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter ต่อไป

จำนวนผู้เชี่ยวชาญในการประเมินความสอดคล้องในเดียวกันมีผลต่อความเที่ยงของผลการประเมิน แต่ยังไม่มีการกำหนดถึงจำนวนผู้เชี่ยวชาญที่เหมาะสมในการประเมินความสอดคล้องในแนวเดียวกันว่าควรเป็นเท่าไร (Anderson, Irvin, Alonso, & Tindal, 2015) ซึ่งงานวิจัยเกี่ยวกับการศึกษาความสอดคล้องในแนวเดียวกันส่วนใหญ่ มีจำนวนผู้เชี่ยวชาญอยู่ระหว่าง 2 - 10 คน (Blank & Smith, 2010; Blank, Porter, & Smithson, 2001; Fonthal, 2004; Näsström & Henriksson, 2008; Polikoff & Porter, 2014; Polikoff, 2015; Shelton & Brown, 2008; Webb, 1999, 2002) ทั้งนี้ Webb (2007) กล่าวว่า การเพิ่มจำนวนผู้เชี่ยวชาญจะส่งผลให้ค่าความเที่ยงเพิ่มสูงขึ้น และเสนอแนะว่าในการศึกษา ความสอดคล้องในแนวเดียวกัน ควรผู้เชี่ยวชาญจำนวน 5 - 7 คน Herman, Webb, และ Zuniga (2007) และ Lombardi, Sebuen, Conley, และ Snow (2010) กล่าวว่า ผู้เชี่ยวชาญจำนวน 6 คน จะทำให้ความเที่ยงมีค่าอยู่ในเกณฑ์ที่ยอมรับได้ หรือทำให้ผลการประเมินมีความน่าเชื่อถือ อย่างไรก็ตามจำนวนผู้เชี่ยวชาญที่เหมาะสมนั้นยังไม่มีข้อสรุปที่ชัดเจน และหากเป็นงานวิจัยที่ต้องมีการประเมินความสอดคล้องของข้อสอบจำนวนมาก จำเป็นต้องใช้ผู้เชี่ยวชาญที่มากขึ้น ส่งผลต่อการบริหารจัดการ งบประมาณ เวลาในการศึกษา ดังนั้นเพื่อเป็นการลดปัญหาดังกล่าว จึงจำเป็นต้องมีการศึกษาการออกแบบการประเมินความสอดคล้องในแนวเดียวกันที่ผู้เชี่ยวชาญแต่ละคนอาจไม่จำเป็นต้องประเมินข้อสอบทุกข้อ แต่จะมีการวิเคราะห์ความน่าเชื่อถือของผลการประเมินสำหรับสถานการณ์ของการประเมินลักษณะต่างๆ ให้สามารถสรุปความเที่ยงในสถานการณ์หรือเงื่อนไขการวัดในลักษณะอื่นๆ ได้ และให้ข้อสรุปถึงสถานการณ์ที่ให้ความเที่ยงที่เหมาะสมในระดับที่สามารถยอมรับได้ จึงต้องอาศัยหลักการแนวคิดของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (generalizability theory: G-theory) เพื่อให้ได้คำตอบดังกล่าว

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด เป็นทฤษฎีที่ได้ขยายแนวคิดของความเที่ยงในทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory) ให้สามารถสรุปอ้างอิงค่าความเที่ยงไปยังสถานการณ์หรือเงื่อนไขการวัดในลักษณะต่างๆ ได้ เนื่องจากทฤษฎีการทดสอบแบบ

ดั้งเดิมไม่ได้ให้ความสนใจต่อสถานการณ์หรือเงื่อนไขการวัดและปฏิสัมพันธ์ระหว่างผู้สอบกับเงื่อนไขของการวัดที่สามารถส่งผลกระทบต่อความน่าเชื่อถือของผลการวัดและเสนอว่ามีความคลาดเคลื่อนจากการวัดแหล่งเดียว แต่ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดได้เสนอว่ามีความคลาดเคลื่อนมากกว่าหนึ่งแหล่ง (multiple error source) โดยแยกส่วนความคลาดเคลื่อนของการวัดเป็นสองแหล่งประกอบด้วย ความคลาดเคลื่อนอย่างมีระบบ (systematic source) และความคลาดเคลื่อนอย่างสุ่ม (random source) (ศิริชัย กาญจนวาสี, 2555) ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดเป็นทฤษฎีที่ให้สารสนเทศที่สามารถนำไปสรุปอ้างอิงไปยังเงื่อนไขการวัดอื่นๆ ได้ โดยทำการประมาณค่าความแปรปรวนจากแหล่งต่างๆ ด้วยการใช้การวิเคราะห์ความแปรปรวน (analysis of variance: ANOVA) รวมทั้งสามารถจำแนกแหล่งความแปรปรวนที่เป็นไปได้ทั้งหมด เพื่อให้ได้ค่าความแปรปรวนของความคลาดเคลื่อนที่ถูกต้อง ซึ่งจะช่วยให้ประมาณค่าความเที่ยงได้ถูกต้องแม่นยำ สามารถใช้ออกแบบรูปแบบการวัด เพื่อให้ได้ค่าความเที่ยงที่สูงขึ้นได้ในหลายสถานการณ์วัด โดยการเพิ่มหรือลดจำนวนองค์ประกอบความแปรปรวนหรือพาเซตต่างๆ เช่น จำนวนข้อสอบ จำนวนครั้งในการวัด จำนวนผู้ตรวจ หรือจำนวนผู้ประเมิน เป็นต้น ซึ่งแนวคิดของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดสามารถนำมาประยุกต์ใช้ในการศึกษาความน่าเชื่อถือของการประเมินความสอดคล้องในแนวเดียวกันได้ (Lombardi, Sebuen, Conley, & Snow, 2010) ดังนั้น ในงานวิจัยนี้จึงเลือกใช้ทฤษฎีการสรุปอ้างอิง ในการศึกษาถึงจำนวนผู้เชี่ยวชาญและรูปแบบการออกแบบการประเมินที่เหมาะสมต่อการศึกษาคือความสอดคล้องในแนวเดียวกัน โดยกำหนดให้ข้อสอบ (item) และจำนวนผู้เชี่ยวชาญ (rater) เป็นองค์ประกอบความแปรปรวนหรือพาเซต (facet) ในโมเดลของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ดังที่ได้กล่าวมานี้ ผู้วิจัยจึงสนใจศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติซึ่งเป็นข้อสอบในการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (ordinary national educational test: O-NET) กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น โดยประยุกต์ใช้วิธีการศึกษาความสอดคล้องในแนวเดียวกันของ Porter (2001) และจำแนกระดับความซับซ้อนทางปัญญาตามแนวคิด revised Bloom's taxonomy (Anderson et al., 2001) และมีการประยุกต์ใช้ MFRM ในการตรวจสอบอิทธิพลของผู้ประเมิน ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ตัดสินผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน โดยพิจารณา ค่าเฉลี่ยคะแนนหลังการควบคุมอิทธิพลการกดหรือปล่อยคะแนน เพื่อคัดเลือกข้อสอบดังกล่าวไปใช้ในการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) และนำไปคำนวณค่าดัชนีความสอดคล้อง (alignment index) ต่อไป ซึ่งจะช่วยให้ผลการประเมินความสอดคล้องในแนวเดียวกัน

ที่มีความน่าเชื่อถือมากยิ่งขึ้น เนื่องจากมีการใช้โมเดลทางสถิติในการตรวจสอบและควบคุมความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้เชี่ยวชาญที่ทำหน้าที่ประเมินความสอดคล้อง นอกจากนี้ในงานวิจัยนี้ยังมีการศึกษาถึงจำนวนผู้เชี่ยวชาญและรูปแบบการออกแบบการประเมินที่จะทำให้ผลการประเมินความสอดคล้องในแนวเดียวกันมีความน่าเชื่อถือในระดับที่ยอมรับได้ โดยการใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดในประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของผลการประเมินความสอดคล้องในแนวเดียวกัน เมื่อกำหนดเงื่อนไขในการประเมินในเรื่องจำนวนผู้ประเมินและรูปแบบการออกแบบประเมินที่แตกต่างกัน การวิจัยนี้แสดงให้เห็นถึงการปรับปรุงขั้นตอนการศึกษาความสอดคล้องในแนวเดียวกัน ที่อยู่บนพื้นฐานของวิธีการศึกษาความสอดคล้องในแนวเดียวกันของ Porter แต่มีการเพิ่มขึ้นขั้นตอนในการคัดเลือกข้อสอบเพื่อใช้ในการสร้างเมทริกซ์ ที่มีการลดความคลาดเคลื่อนอันเกิดจากอิทธิพลของผู้ประเมินโดยใช้โมเดลทางสถิติ ผลการวิจัยนี้จะเป็นแนวทางสำหรับครูผู้สอน และ/หรือผู้บริหารสถานศึกษา ในการประยุกต์หรือเลือกใช้แนวคิดในการศึกษาความสอดคล้องในแนวเดียวกัน เพื่อออกแบบการประเมินความสอดคล้ององค์ประกอบทางการศึกษา อันได้แก่ ระหว่างมาตรฐาน การเรียนการสอน และการประเมินที่ยึดตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ของประเทศไทย ตลอดจนให้สารสนเทศเพื่อเป็นข้อมูลป้อนกลับไปยังผู้มีส่วนได้ส่วนเสียทางการศึกษา (เช่น ผู้วางนโยบายหรือผู้กำหนด/เขียนหลักสูตร ครู และนักการศึกษา) เกี่ยวกับความสอดคล้องของมาตรฐานและการประเมินวิทยาศาสตร์ในโรงเรียนมัธยมศึกษา เพื่อใช้เป็นแนวทางในการพัฒนาการออกแบบการประเมินในแต่ละระดับให้มีความสอดคล้องกัน และสอดคล้องกับมาตรฐานที่กำหนด ให้มุ่งไปสู่เป้าหมายเดียวกัน อันเป็นประโยชน์ต่อการสร้างโอกาสในการเรียนรู้อย่างเท่าเทียมกันให้กับผู้เรียนทุกคน

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

### คำถามการวิจัย

1. เกิดอิทธิพลของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียนกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น หรือไม่ อย่างไร
2. ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ก่อนและหลังควบคุมอิทธิพลของผู้ประเมิน เหมือนหรือแตกต่างกัน อย่างไร



3. โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน จะมีผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น เหมือนหรือแตกต่างกัน อย่างไร

4. จำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินแบบใด ที่จะทำให้ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น มีความน่าเชื่อถือ

### วัตถุประสงค์ของการวิจัย

1. เพื่อตรวจสอบอิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น

2. เพื่อเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

3. เพื่อศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน

4. เพื่อประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน

## สมมติฐานการวิจัย

การวิจัยครั้งนี้ผู้วิจัยจึงตั้งสมมติฐานการวิจัย 3 ประการ ดังนี้

1. จากการศึกษาเกี่ยวกับการตรวจสอบและควบคุมอิทธิพลการกดหรือปล่อยคะแนนด้วย MFRM พบว่า MFRM เป็นเครื่องมือที่ดีเยี่ยมในการศึกษาอิทธิพลของผู้ประเมิน (Iramaneerat & Yudkowsky, 2007: 270) ทำให้การตรวจให้คะแนนมีความเป็นปรนัยมากขึ้น (Linacre, 1994) โดยสามารถใช้ในการควบคุมอิทธิพลการกดหรือปล่อยคะแนน ประเมินความสอดคล้องระหว่างผู้ประเมิน (intrarater consistency) และให้ค่าสถิติระดับข้อสอบได้ ทั้งนี้ จากงานวิจัยของ Anderson, Irvin, Alonzo, และ Tindal (2015) พบว่า มีความแตกต่างอย่างมีนัยสำคัญในการศึกษาความสอดคล้องระหว่างข้อสอบกับมาตรฐาน เมื่อมีการใช้สถิติในการควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยมีการเปลี่ยนแปลงของผลประเมินระดับความสอดคล้องในแนวเดียวกันๆ ระหว่างก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ทั้งจาก “สอดคล้อง” เป็น “ไม่สอดคล้อง” และจาก “ไม่สอดคล้อง” เป็น “สอดคล้อง” ดังนั้น ผู้วิจัยจึงตั้งสมมติฐานของการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ดังนี้

“หากเกิดอิทธิพลการกดหรือปล่อยคะแนน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้นแล้ว น่าจะมีการเปลี่ยนแปลงของค่าเฉลี่ยคะแนน ระหว่างคะแนนผลการประเมินความสอดคล้องในแนวเดียวกันๆ ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน ซึ่งเป็นค่าเฉลี่ยคะแนนดิบของข้อสอบ (item raw mean) กับคะแนนผลการประเมินความสอดคล้องในแนวเดียวกันๆ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ซึ่งเป็นค่าเฉลี่ยคะแนนที่ได้จากโมเดล ที่มีการปรับแก้อิทธิพลการกดหรือปล่อยคะแนนแล้ว (item adjusted mean)”

2. จากการศึกษาเกี่ยวกับผลการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (Ordinary National Educational Test: O-NET) ที่ผ่านมา ซึ่งสะท้อนให้เห็นถึงความรู้ความสามารถของผู้เรียนที่อยู่ในระดับต่ำ โดยสาเหตุที่ทำให้คะแนนการทดสอบ O-NET ต่ำมาจากหลายปัจจัย ซึ่งปัจจัยด้านครูผู้สอน ถือว่ามีความสำคัญและต้องได้รับพัฒนาเป็นลำดับต้นๆ (เอี่ยมพร หลินเจริญ, สิริศักดิ์ อาจิวชัย, และภิรภา จันทร์อินทร์, 2552) และการที่ครูผู้สอนมีการวัดและประเมินผลโดยใช้ข้อสอบให้สอดคล้องหรือคล้ายคลึงกับข้อสอบ O-NET เป็นปัจจัยหนึ่งที่สามารถส่งเสริมการเพิ่มผลสัมฤทธิ์ของ

การสอบ O-NET ของผู้เรียนได้ (บริษัทอักษรเจริญทัศน์ จำกัด, 2559) ดังนั้น ผู้วิจัยจึงตั้งสมมติฐานของการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน ดังนี้

“โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกันน่าจะมีค่าดัชนีความสอดคล้อง (alignment index) ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น แตกต่างกัน โดยโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง น่าจะมีค่าดัชนีความสอดคล้องสูงกว่าโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ”

3. จากการศึกษาเกี่ยวกับความเที่ยงในการประเมินความสอดคล้องในแนวเดียวกันของผู้เชี่ยวชาญ พบว่า จำนวนผู้เชี่ยวชาญในการประเมินความสอดคล้องในแนวเดียวกันมีผลต่อความเที่ยงของผลการประเมิน แต่ยังไม่มีการกำหนดชัดเจนถึงจำนวนผู้เชี่ยวชาญที่เหมาะสมในการประเมินความสอดคล้องในแนวเดียวกัน (Anderson, Irvin, Alonso, & Tindal, 2015) ซึ่งงานวิจัยเกี่ยวกับการศึกษาความสอดคล้องในแนวเดียวกันส่วนใหญ่ มีจำนวนผู้เชี่ยวชาญอยู่ระหว่าง 2 - 10 คน (Blank & Smith, 2010; Blank, Porter, & Smithson, 2001; Contio, 2013; Fonthal, 2004; Liu & Fulmer, 2007; Lombardi, Sebuén, Conley, & Snow, 2010; Näsström & Henriksson, 2008; Polikoff & Porter, 2014; Polikoff, 2015; Shelton & Brown, 2008; Webb, 1999; 2002) อย่างไรก็ตาม การเพิ่มจำนวนผู้เชี่ยวชาญจะส่งผลให้ค่าความเที่ยงเพิ่มสูงขึ้น (Herman, Webb, & Zuniga, 2005; Lombardi, Sebuén, Conley, & Snow, 2010) ดังนั้น ผู้วิจัยจึงตั้งสมมติฐานของการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น เมื่อจำนวนผู้เชี่ยวชาญและรูปแบบการออกแบบการประเมินต่างกัน ดังนี้

“ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจของคะแนนผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ที่มีจำนวนผู้เชี่ยวชาญที่มากกว่า น่าจะมีค่าสูงกว่าจำนวนผู้ประเมินที่น้อยกว่า ในทุกรูปแบบการออกแบบการประเมิน”

## ขอบเขตการวิจัย

### 1. ประชากรในการวิจัยครั้งนี้ แบ่งเป็น 2 ส่วน ดังนี้

1.1 ข้อสอบ แบ่งเป็น 2 ส่วนคือ ข้อสอบในการประเมินระดับชั้นเรียน และข้อสอบในการประเมินระดับชาติ

1.1.1 ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 – 3) ของโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร ที่ถูกจัดอยู่ในกลุ่มโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ สูงสุดและต่ำสุด ซึ่งประกอบด้วยข้อสอบที่ครูสร้างขึ้น (teacher-made test) เพื่อประเมินผลการเรียนรู้ของผู้เรียน ทั้งข้อสอบระหว่างภาคและข้อสอบประจำภาค ในปีการศึกษา 2559

1.1.2 ข้อสอบในการประเมินระดับชาติ ได้แก่ ข้อสอบในการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (ordinary national educational test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำนวน 45 ข้อ

1.2 ผู้เชี่ยวชาญในการประเมินความสอดคล้องในแนวเดียวกัน ได้แก่ ครูสอนวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น และ/หรือนักวิชาการ หรืออาจารย์ระดับอุดมศึกษาที่สอนในรายวิชาวิทยาศาสตร์ และ/หรือสาขาที่เกี่ยวข้องกับการวัดและประเมินผลการศึกษา

### 2. ตัวแปรที่ศึกษา

#### 2.1 ตัวแปรตาม มีดังนี้

2.1.1 ความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ในงานวิจัยนี้ เป็นการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กับการประเมิน ได้แก่ ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ของโรงเรียนมัธยมศึกษา ซึ่งวัดความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนโดยให้ผู้เชี่ยวชาญเป็นผู้ประเมิน ด้วยแบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น ที่ผู้วิจัยสร้างขึ้น ที่มีลักษณะเป็นมาตรฐานประมาณค่า

2.1.2 ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ในงานวิจัยนี้เป็นการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (ordinary national educational

test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 กับข้อสอบในการประเมินระดับ ชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 – 3) ของโรงเรียนมัธยมศึกษา ในปีการศึกษา 2559 ซึ่งวัดความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน โดยให้ผู้เชี่ยวชาญเป็นผู้ประเมิน แล้วนำคะแนนจากผู้เชี่ยวชาญมาหาค่าดัชนีความสอดคล้อง (alignment index) ตามแนวคิดของ Porter

2.1.3 ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจ เป็นค่าความเชื่อมั่นของคะแนนที่ได้จากแบบประเมินความสอดคล้องในแนวเดียวกันในเงื่อนไขหรือสถานการณ์ที่กำหนด

## 2.2 ตัวแปรอิสระ มีดังนี้

2.2.1 จำนวนผู้ประเมิน ในงานวิจัยนี้ เป็นจำนวนผู้เชี่ยวชาญในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กับข้อสอบในการประเมินผลการเรียนรู้ กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 โดยมีจำนวนผู้เชี่ยวชาญตั้งแต่ 1 – 20 คน

2.2.2 รูปแบบการออกแบบการประเมิน ในงานวิจัยนี้ เป็นลักษณะของออกแบบการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 แบ่งรูปแบบการออกแบบการประเมินเป็น 2 รูปแบบ คือ 1) ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ (i x r design) และ 2) ผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อ (r : i design)

2.2.3 ผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียน ในงานวิจัยนี้ พิจารณาจากค่าเฉลี่ยคะแนนการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 ของโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร จำแนกเป็น 2 กลุ่ม คือ (1) โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูงสุด เป็นโรงเรียนที่มีคะแนน O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 สูงสุด 10 อันดับแรก และ (2) โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำสุด เป็นโรงเรียนที่มีคะแนน O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 ต่ำสุด 10 อันดับสุดท้าย

2.2.4 อิทธิพลของผู้ประเมิน ในงานวิจัยนี้ เป็นอิทธิพลของผู้ประเมิน 4 ชนิด ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และ อิทธิพลของความไม่มีแบบแผน

### คำจำกัดความที่ใช้ในการวิจัย

**สาระ** หมายถึง กลุ่มย่อยภายในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ซึ่งในงานวิจัยนี้กำหนดสาระในการศึกษาทั้งหมด 7 สาระ ได้แก่ สาระที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต สาระที่ 2 ชีวิตกับสิ่งแวดล้อม สาระที่ 3 สารและสมบัติของสาร สาระที่ 4 แรงและการเคลื่อนที่ สาระที่ 5 พลังงานและสาระที่ 7 ดาราศาสตร์และอวกาศ

**มาตรฐาน** หมายถึง คุณภาพที่ต้องการให้เกิดขึ้นในตัวผู้เรียน เป็นสิ่งที่ผู้เรียนพึงรู้และปฏิบัติได้ในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ซึ่งในงานวิจัยนี้กำหนดมาตรฐานในการศึกษาทั้งหมด 12 มาตรฐาน ภายใต้ 7 สาระ (สาระที่ 1 - 7)

**ตัวชี้วัด** หมายถึง สิ่งที่ผู้เรียนพึงรู้และปฏิบัติได้ในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ซึ่งสะท้อนถึงมาตรฐานการเรียนรู้ มีลักษณะเฉพาะเจาะจงและมีความเป็นรูปธรรมในการนำไปใช้ในการกำหนดเนื้อหา การเรียนการสอน และเป็นเกณฑ์สำคัญสำหรับการวัดและประเมินผลเพื่อตรวจสอบคุณภาพผู้เรียน ซึ่งในงานวิจัยนี้กำหนดตัวชี้วัดในการศึกษาทั้งหมด 41 ตัวชี้วัด ภายใต้ 12 มาตรฐาน จาก 7 สาระ (สาระที่ 1 - 7) ที่ สทศ. ใช้ในการออกข้อสอบ O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559

**ข้อสอบในการประเมินระดับชั้นเรียน** หมายถึง ข้อคำถามที่ใช้ในการประเมินผลการเรียนรู้ ผู้เรียนระดับชั้นเรียน ที่ครูสร้างขึ้น (teacher-made test) ในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 - 3) ของโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร ทั้งข้อสอบระหว่างภาคและข้อสอบประจำภาค ในปีการศึกษา 2559

**ข้อสอบในการประเมินระดับชาติ** หมายถึง ข้อสอบในการทดสอบทางการศึกษาระดับชาตินี้ขั้นพื้นฐาน (ordinary national educational test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559

**ความสอดคล้องในแนวเดียวกัน** หมายถึง ระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างองค์ประกอบต่างๆ ในระบบการศึกษา ได้แก่ ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน และความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน

**ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน** หมายถึง ระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างข้อสอบ

ในการประเมินผลระดับชาติ ได้แก่ ข้อสอบในการทดสอบทางการศึกษาระดับชาติด้านขั้นพื้นฐาน (ordinary national educational test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 กับการประเมินผลระดับชั้นเรียน ได้แก่ ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ในปีการศึกษา 2559 ซึ่งวัดความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน โดยการคำนวณค่าดัชนีความสอดคล้อง (alignment index) ตามแนวคิดของ Porter (2001)

**ความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน** หมายถึง ระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างผลการเรียนรู้ที่คาดหวัง ได้แก่ มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กับการประเมิน ได้แก่ ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ของโรงเรียนมัธยมศึกษา การวัดความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ทำโดยให้ผู้เชี่ยวชาญเป็นผู้ประเมิน ด้วยแบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น ที่ผู้วิจัยสร้างขึ้น

**วิธีการศึกษาความสอดคล้องในแนวเดียวกันของ Porter** หรือวิธีการสำรวจหลักสูตรที่นำไปใช้จริง (Survey of enacted curriculum (SEC) methodology) หมายถึง การวัดระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างเนื้อหาของ 2 องค์ประกอบในการศึกษาที่นำมาประเมิน แล้วสร้างเมทริกซ์ เมทริกซ์ ได้แก่ เมทริกซ์ข้อสอบในการประเมินระดับชาติ 2(เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) ก่อนนำมาคำนวณค่าดัชนีความสอดคล้อง (alignment index)

**ความซับซ้อนทางปัญญา** หมายถึง ลำดับขั้นของกระบวนการทางสมองหรือความลึกของความรู้ของข้อสอบ วัดได้จากการพิจารณาของผู้เชี่ยวชาญ ซึ่งความซับซ้อนทางปัญญาจำแนกตามแนวคิดการจำแนกพฤติกรรมการเรียนรู้ด้านพุทธิพิสัยของ Bloom ที่ได้รับการปรับปรุงใหม่ (revised Bloom's taxonomy) (Anderson et al., 2001) โดยแบ่งลำดับขั้นของกระบวนการทางปัญญา (cognitive process dimension) ได้เป็น 6 ขั้น เรียงลำดับจากความรู้ระดับต่ำไปยังความรู้ระดับสูง ดังนี้

1. จำ (remember) เป็นความสามารถในการระลึกหรือจำความรู้หรือสารสนเทศที่เก็บไว้ในสมอง ซึ่งเป็นความจำระยะยาว

2. เข้าใจ (understand) เป็นความสามารถในการสร้างความหมายหรือความรู้จากสื่อหรือเครื่องมือทางการศึกษาด้วยตนเอง เช่น จากการอ่าน การอธิบายของครู ทักษะย่อยของความสามารถในขั้นนี้ ได้แก่ การแปลความหมาย (interpreting) การให้ตัวอย่าง (exemplifying) การจัดจำแนก (classifying) การสรุป (summarizing) การสรุปอ้างอิง (inferring) การเปรียบเทียบ (comparing) และการอธิบาย (explaining)

3. ประยุกต์ใช้ (apply) เป็นความสามารถในการใช้กระบวนการที่ได้เรียนรู้มาในสถานการณ์ใหม่หรือสถานการณ์ที่คล้ายคลึงกัน

4. วิเคราะห์ (analyze) เป็นความสามารถในการแยกความรู้ออกเป็นส่วนๆ โดยสามารถให้เหตุผลว่า ความรู้ส่วนย่อยที่แยกแต่ละส่วนมีความเกี่ยวข้องกับโครงสร้างของความรู้ทั้งหมดอย่างไร สามารถจำแนกความแตกต่าง จัดระบบความรู้ และบอกที่มาของความรู้หรือองค์ประกอบแต่ละส่วนได้

5. ประเมินค่า (evaluate) เป็นความสามารถในการตรวจสอบ วิพากษ์ และตัดสิน

6. สร้างสรรค์ (create) เป็นความสามารถการสร้างสิ่งใหม่จากสิ่งที่เคยเรียนรู้ หรือสิ่งที่พบเห็นในบริบทต่างๆ สามารถสร้างสรรค์งาน แผนงาน หรือผลิตภัณฑ์ หรือชิ้นงานที่แปลกใหม่

**ผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียน** หมายถึง ค่าเฉลี่ยคะแนนการทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน (ordinary national educational test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 ของโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร ที่ดำเนินการสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) ซึ่งการจำแนกระดับผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียน เป็น 2 กลุ่ม คือ (1) โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูงสุด เป็นโรงเรียนที่มีคะแนน O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 สูงสุด 10 อันดับแรก ในกลุ่มโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร และ (2) โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำสุด เป็นโรงเรียนที่มีคะแนน O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 ต่ำสุด 10 อันดับสุดท้าย ในกลุ่มโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร

**โมเดลหลายองค์ประกอบของรารซ์ หรือ MFRM** หมายถึง รูปแบบสมการเชิงเส้นตรงทางคณิตศาสตร์ที่ใช้ในการแปลงค่าคะแนนที่ได้จากการสังเกตในทุกองค์ประกอบให้มีหน่วยเดียวกัน เพื่อใช้ในการเปรียบเทียบระหว่างองค์ประกอบ ตลอดจนสามารถคำนวณความน่าจะเป็น ของคะแนนความสอดคล้อง และวิเคราะห์ความคลาดเคลื่อนและความแม่นยำในการให้คะแนนของผู้ประเมิน ซึ่งมาจากการประมาณค่าพารามิเตอร์ในโมเดล

**อิทธิพลของผู้ประเมิน** หมายถึง กลุ่มของผลกระทบของความแปรปรวนอย่างเป็นระบบในการให้คะแนนที่เกี่ยวข้องอย่างใดอย่างหนึ่งกับผู้ประเมิน ทำให้เกิดความคลาดเคลื่อนในการวัด



อิทธิพลของผู้ประเมิน ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ซึ่งได้จากการวิเคราะห์โดยใช้ MFRM จากโปรแกรม FACETS ดังนี้

1. อิทธิพลการกดหรือปล่อยคะแนน (severity/leniency effect) หมายถึง ความคลาดเคลื่อนที่เกิดจากการที่ผู้ประเมินหรือผู้เชี่ยวชาญในการประเมินมีแนวโน้มจะให้คะแนนผู้ถูกตรวจประเมิน หรือสิ่งที่ถูกประเมินสูงหรือต่ำกว่าความเป็นจริง โดยผู้ประเมินมีลักษณะกดคะแนน (severity) จะมีแนวโน้มให้คะแนนโดยเฉลี่ยต่ำกว่าผู้ประเมินคนอื่น ซึ่งเป็นการให้คะแนนต่ำกว่าความเป็นจริงตลอดช่วงความสามารถ ส่วนผู้ประเมินมีลักษณะปล่อยคะแนน (leniency) จะมีแนวโน้มให้คะแนนโดยเฉลี่ยสูงกว่าผู้ประเมินคนอื่นตลอดช่วงความสามารถ ซึ่งตรวจสอบอิทธิพลการกดหรือปล่อยคะแนนโดยใช้ MFRM โดยคะแนนในหน่วยโลจิทของฟาเซทของผู้ประเมิน (rater logit) หากมีคะแนนต่ำ หมายถึง ผู้ประเมินมีลักษณะปล่อยคะแนน แต่หากมีคะแนนสูง หมายถึง ผู้ประเมินมีลักษณะกดคะแนน

2. อิทธิพลแนวโน้มการให้คะแนนค่ากลาง (central tendency effect) หมายถึง ความคลาดเคลื่อนที่เกิดจากการให้คะแนนส่วนใหญ่มีแนวโน้มอยู่ตรงกลางของมาตรฐานประเมิน โดยผู้ประเมินหรือผู้เชี่ยวชาญในการประเมินจะแสดงแนวโน้มในการใช้ระดับตรงกลางของมาตรฐานมากเกินไป ซึ่งตรวจสอบอิทธิพลแนวโน้มการให้คะแนนค่ากลางทำการวิเคราะห์ด้วย MFRM เพื่อการประมาณค่าผู้ประเมิน (rater estimate) และพิจารณาค่า infit MNSQ และ outfit MNSQ ของผู้ประเมินที่มีค่าน้อยกว่า 0.70 จากฟาเซทผู้ประเมิน จะแสดงว่าเกิดอิทธิพลในการให้คะแนนค่ากลาง

3. อิทธิพลการจำกัดช่วง (restriction of range effect) หมายถึง ความคลาดเคลื่อนที่เกิดจากการให้คะแนนที่กระจุกอยู่จุดใดจุดหนึ่งของช่วงที่ต่อเนื่อง (continuum) บนมาตรฐานประมาณค่า ไม่ว่าจะเป็นจุดที่ให้คะแนนสูง (ปล่อยคะแนน) จุดที่ให้คะแนนต่ำ (กดคะแนน) หรือจุดกลางๆ (แนวโน้มให้คะแนนค่ากลาง) ซึ่งตรวจสอบอิทธิพลการจำกัดช่วงทำการวิเคราะห์ด้วย MFRM เพื่อประมาณค่าผู้ประเมิน (rater estimate) และพิจารณาค่า infit MNSQ และ outfit MNSQ ของผู้ประเมินที่มีค่าน้อยกว่า 0.60 จากฟาเซทผู้ประเมิน จะแสดงว่าเกิดอิทธิพลการจำกัดช่วง

4. อิทธิพลของความไม่มีแบบแผน (randomness effect) หรือความไม่สอดคล้องในการให้คะแนนของผู้ประเมิน (inconsistency) หมายถึง แนวโน้มที่ผู้ประเมินหรือผู้เชี่ยวชาญในการประเมินใดๆ ใช้มาตรฐานประเมินในลักษณะที่ไม่สอดคล้องกับผู้ประเมินหรือผู้เชี่ยวชาญในการประเมินคนอื่นๆ ใช้ ซึ่งการตรวจสอบอิทธิพลของความไม่มีแบบแผนทำการวิเคราะห์ด้วย MFRM เพื่อประมาณค่าผู้ประเมิน (rater estimate) และพิจารณาค่า infit MNSQ และ outfit MNSQ ที่มีค่าสูง จากฟาเซทผู้ประเมิน คือ มีค่า infit MNSQ และ outfit MNSQ มากกว่า 1.30 จะแสดงว่าเกิดอิทธิพลของความไม่มีแบบแผน หรือความไม่สอดคล้องในการประเมิน

**ความแม่นยำในการให้คะแนน (accurate)** หมายถึง แนวโน้มที่ผู้ตรวจประเมินสามารถพิจารณาตัดสินหรือให้คะแนนได้อย่างถูกต้องแม่นยำ การตรวจสอบความแม่นยำในการให้คะแนนทำโดยการวิเคราะห์ด้วย MFRM เพื่อประมาณค่าผู้ประเมิน (rater estimates) และพิจารณา infit MNSQ และ outfit MNSQ ที่มีค่าตั้งแต่ 0.70 – 1.30 จากฟิเชอร์ผู้ประเมิน จะแสดงถึงความแม่นยำในการให้คะแนน

**การตรวจสอบอิทธิพลของผู้ประเมิน** หมายถึง ผลการวิเคราะห์เพื่อตรวจสอบความคลาดเคลื่อนของผู้เชี่ยวชาญในการพิจารณาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินผลการเรียนรู้ โดยใช้ MFRM จากโปรแกรม FACETS ในการตรวจสอบอิทธิพลของผู้ประเมิน 4 ชนิด ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน

**การควบคุมอิทธิพลของผู้ประเมิน** หมายถึง ผลการประมาณค่าของผู้ประเมินในค่า Measure ที่เป็นคะแนนในหน่วยโลจิท (logits) ที่แสดงระดับการกดคะแนนของผู้ประเมินแต่ละคน จากการวิเคราะห์โดยประยุกต์ใช้ MFRM ด้วยโปรแกรม FACETS

**โมเดลการสรุปอ้างอิง** หมายถึง โมเดลของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (generalizability theory: G-Theory) ที่จำแนกแหล่งความคลาดเคลื่อนได้หลายแหล่ง ซึ่งในงานวิจัยนี้ศึกษา 2 แหล่ง คือ ข้อสอบ (item: i) และจำนวนผู้ประเมิน (rater: r) การประมาณค่าขนาดความแปรปรวน และค่าสัมประสิทธิ์การสรุปอ้างอิง โดยการวิเคราะห์ความแปรปรวนแบบสองทาง (two-way ANOVA)

**ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจ** หมายถึง ค่าความเชื่อมั่นของคะแนนที่ได้จากแบบประเมินความสอดคล้องในแนวเดียวกันในเงื่อนไขหรือสถานการณ์ที่กำหนด ซึ่งแสดงถึงระดับแม่นยำและความน่าเชื่อถือของการสรุปอ้างอิงคะแนนที่ได้จากแบบประเมินไปยังคะแนนเอกภพของข้อสอบ ประมาณค่าจากอัตราส่วนของความแปรปรวนของคะแนนเอกภพกับความแปรปรวนของคะแนนสังเกต

**จำนวนผู้ประเมิน** หมายถึง จำนวนผู้เชี่ยวชาญในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 โดยมีจำนวนผู้เชี่ยวชาญตั้งแต่ 1 – 20 คน

**รูปแบบการออกแบบการประเมิน** หมายถึง ลักษณะของออกแบบการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กับข้อสอบในการประเมินผลการเรียนรู้ กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 แบ่งรูปแบบการออกแบบการประเมิน

เป็น 2 รูปแบบ คือ 1) ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ (i x r design) และ 2) ผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อ (r : i design)

การตรวจสอบคุณภาพของคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน หมายถึง การตรวจสอบคุณภาพของคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน โดยผู้ทรงคุณวุฒิในด้านความตรงตามเนื้อหา และด้านความเหมาะสมตามมาตรฐานการประเมิน 4 ด้าน ของคณะกรรมการมาตรฐานสำหรับการประเมินทางการศึกษา (The Joint Committee on Standards for Educational Evaluation, 1994) ได้แก่ ด้านอรรถประโยชน์ (utility) ด้านความเป็นไปได้ (feasibility) ด้านความเหมาะสม (property) และด้านความถูกต้อง (accuracy)

## ประโยชน์ที่ได้รับ

### 1. ประโยชน์ด้านวิชาการ

1.1 ได้ผลตรวจสอบความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน ที่มีความเที่ยงตรงและยุติธรรม ทำให้ครูและผู้บริหารสถานศึกษาที่ออกข้อสอบสามารถคัดเลือกข้อสอบที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัดไปใช้ในการพัฒนาฐานข้อมูลหรือคลังข้อสอบต่อไป นอกจากนี้ ยังได้สารสนเทศในการพัฒนาหรือปรับปรุงการออกข้อสอบของสถานศึกษาให้ตรงกับมาตรฐานและตัวชี้วัดตามตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน

1.2 ได้รูปแบบการดำเนินงานศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter ที่เพิ่มขึ้นตอนในการตรวจสอบและควบคุมความคลาดเคลื่อนในการให้คะแนนอันเกิดจากผู้ประเมินด้วยวิธีการทางสถิติ เพื่อเป็นแนวทางสำหรับนักวัดผลการศึกษาหรือผู้ที่สนใจนำไปใช้เป็นแนวทางในการออกแบบการศึกษาความสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบทางการศึกษา ซึ่งทำให้ผลการศึกษาน่าเชื่อถือมากยิ่งขึ้น

1.3 ได้สารสนเทศเพื่อเป็นข้อมูลป้อนกลับไปยังผู้มีส่วนได้ส่วนเสียทางการศึกษา (เช่น ผู้วางนโยบายหรือผู้กำหนด/เขียนหลักสูตร ครู และนักการศึกษา) เกี่ยวกับความสอดคล้องระหว่างการประเมินระดับชั้นเรียนกับการประเมินระดับชาติ เพื่อใช้เป็นแนวทางในการพัฒนาการออกแบบการประเมิน รวมถึงการเรียนการสอนให้สอดคล้องกับมาตรฐานที่กำหนด ให้มุ่งไปสู่เป้าหมายเดียวกัน อันเป็นประโยชน์ต่อการสร้างโอกาสในการเรียนรู้อย่างเท่าเทียมกันให้กับผู้เรียนทุกคน

## 2. ประโยชน์ด้านการปฏิบัติ

2.1 ได้คู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน โดยใช้แนวคิดของ Porter เพื่อให้ครูผู้สอน ผู้บริหารสถานศึกษา และ/หรือผู้บริหารเขตพื้นที่การศึกษาสามารถนำไปใช้ในการศึกษาความสอดคล้องฯ และพัฒนา/ปรับปรุงคุณภาพการออกข้อสอบได้ในทุกกลุ่มสาระการเรียนรู้และทุกระดับชั้น

2.2 ได้แนวทางสำหรับออกแบบสถานการณ์หรือเงื่อนไขของการประเมินความสอดคล้องในแนวเดียวกัน ตามหลักของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ที่มีแหล่งความแปรปรวนต่างๆ เช่น จำนวนผู้เชี่ยวชาญ ลักษณะข้อรายการประเมิน รูปแบบการออกแบบการประเมิน และนำไปสู่การตัดสินใจเลือกสถานการณ์การประเมินที่มีความน่าเชื่อถือ บนพื้นฐานของความเป็นไปได้ในทางปฏิบัติ

2.3 ได้เครื่องมือสำหรับประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ทั้งที่อยู่ในรูปแบบเอกสารและออนไลน์

## บทที่ 2

### เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยเรื่อง ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับ ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราล์ซ และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง โดยนำเสนอเป็น 9 ตอน ได้แก่

ตอนที่ 1 มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลาง การศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ของระดับชั้นมัธยมศึกษาตอนต้น

ตอนที่ 2 การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน

ตอนที่ 3 แนวคิดเกี่ยวกับการศึกษาความสอดคล้องในแนวเดียวกัน

ตอนที่ 4 พฤติกรรมการเรียนรู้ด้านพุทธิพิสัย ตามแนวคิดของ Bloom ที่ได้รับการปรับปรุงใหม่

ตอนที่ 5 อิทธิพลของผู้ประเมิน

ตอนที่ 6 แนวคิดทฤษฎีเกี่ยวกับโมเดลหลายองค์ประกอบของราล์ซ

ตอนที่ 7 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ตอนที่ 8 มาตรฐานของการประเมิน

ตอนที่ 9 กรอบแนวคิดการวิจัย

รายละเอียดของแต่ละตอน มีดังต่อไปนี้

ตอนที่ 1 มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลาง การศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ของระดับชั้นมัธยมศึกษาตอนต้น

ผู้วิจัยได้ทบทวนเอกสารที่เกี่ยวข้องกับมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้ วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ของระดับชั้น มัธยมศึกษาตอนต้น โดยแบ่งการนำเสนอออกเป็น 2 ประเด็นหลัก ได้แก่ 1) สาระการเรียนรู้ วิทยาศาสตร์ และ 2) มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษา ตอนต้น รายละเอียดของแต่ละประเด็นมีดังนี้

## 1.1 สารการเรียนรู้วิทยาศาสตร์

กลุ่มสาระการเรียนรู้วิทยาศาสตร์มุ่งหวังให้ผู้เรียนได้เรียนรู้วิทยาศาสตร์ที่เน้นการเชื่อมโยงความรู้กับกระบวนการ มีทักษะสำคัญในการค้นคว้าและสร้างองค์ความรู้ โดยใช้กระบวนการในการสืบเสาะหาความรู้ และการแก้ปัญหาที่หลากหลาย ให้ผู้เรียนมีส่วนร่วมในการเรียนรู้ทุกขั้นตอน มีการทำกิจกรรมด้วยการลงมือปฏิบัติจริงอย่างหลากหลาย เหมาะสมกับระดับชั้น โดยได้กำหนดสาระสำคัญไว้ (กระทรวงศึกษาธิการ, 2552) ดังนี้

**สาระที่ 1: สิ่งมีชีวิตกับกระบวนการดำรงชีวิต** ว่าด้วยเรื่องสิ่งมีชีวิต หน่วยพื้นฐานของสิ่งมีชีวิต โครงสร้างและหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิต และกระบวนการดำรงชีวิต ความหลากหลายทางชีวภาพ การถ่ายทอดทางพันธุกรรม การทำงานของระบบต่างๆ ของสิ่งมีชีวิต วิวัฒนาการและความหลากหลายของสิ่งมีชีวิต และเทคโนโลยีชีวภาพ

**สาระที่ 2: ชีวิตกับสิ่งแวดล้อม** ว่าด้วยเรื่องสิ่งมีชีวิตที่หลากหลายรอบตัว ความสัมพันธ์ระหว่างสิ่งมีชีวิตกับสิ่งแวดล้อม ความสัมพันธ์ของสิ่งมีชีวิตต่างๆ ในระบบนิเวศ ความสำคัญของทรัพยากรธรรมชาติการใช้และจัดการทรัพยากรธรรมชาติ ในระดับท้องถิ่น ประเทศ และโลก ปัจจัยที่มีผลต่อการอยู่รอดของสิ่งมีชีวิตในสภาพแวดล้อมต่างๆ

**สาระที่ 3: สารและสมบัติของสาร** ว่าด้วยเรื่องสมบัติของวัสดุและสาร แรงแยัดเหนียวระหว่างอนุภาค การเปลี่ยนสถานะ การเกิดสารละลายและการเกิดปฏิกิริยาเคมีของสาร สมการเคมี และการแยกสาร

**สาระที่ 4: แรงแยกและการเคลื่อนที่** ว่าด้วยเรื่องธรรมชาติของแรงแม่เหล็กไฟฟ้า แรงโน้มถ่วง แรงนิวเคลียร์ การออกแรงกระทำต่อวัตถุ การเคลื่อนที่ของวัตถุ แรงเสียดทาน โมเมนต์การเคลื่อนที่แบบต่างๆ ในชีวิตประจำวัน

**สาระที่ 5: พลังงาน** ว่าด้วยเรื่องพลังงานกับการดำรงชีวิต การเปลี่ยนรูปพลังงาน สมบัติและปรากฏการณ์ของแสง เสียง และวงจรไฟฟ้า คลื่นแม่เหล็กไฟฟ้า กัมมันตภาพรังสีและปฏิกิริยานิวเคลียร์ ปฏิสัมพันธ์ระหว่างสารและพลังงานการอนุรักษ์พลังงาน ผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม

**สาระที่ 6: กระบวนการเปลี่ยนแปลงของโลก** ว่าด้วยเรื่องโครงสร้างและองค์ประกอบของโลก ทรัพยากรทางธรณี สมบัติทางกายภาพของดิน หิน น้ำ อากาศ สมบัติของผิวโลก และบรรยากาศ กระบวนการเปลี่ยนแปลงของเปลือกโลก ปรากฏการณ์ทางธรณี ปัจจัยที่มีผลต่อการเปลี่ยนแปลงของบรรยากาศ

**สาระที่ 7: ดาราศาสตร์และอวกาศ** ว่าด้วยเรื่องวิวัฒนาการของระบบสุริยะ กาแล็กซี เอกภพ ปฏิสัมพันธ์และผลต่อสิ่งมีชีวิตบนโลก ความสัมพันธ์ของดวงอาทิตย์ ดวงจันทร์ และโลก ความสำคัญของเทคโนโลยีอวกาศ

**สาระที่ 8: ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี** ว่าด้วยเรื่องกระบวนการทางวิทยาศาสตร์ การสืบเสาะหาความรู้ การแก้ปัญหา และจิตวิทยาศาสตร์

### 1.2 มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น

จากการศึกษาเอกสารมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ของระดับชั้นมัธยมศึกษาตอนต้น โดยมีการกำหนดสาระสำคัญไว้ จำนวน 8 สาระ 13 มาตรฐาน ตัวชี้วัดจำแนกตามระดับชั้นมัธยมศึกษาปีที่ 1 2 และ 3 มีจำนวนเท่ากับ 42 38 และ 40 ตัวชี้วัด ตามลำดับ (กระทรวงศึกษาธิการ, 2552) รายละเอียดของสาระ มาตรฐานและตัวชี้วัดสรุปได้ตาราง 1

**ตาราง 1** สรุปสาระ มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
สาระที่ 1 สิ่งมีชีวิตกับ กระบวนการ ดำรงชีวิต	มาตรฐาน ว 1.1 เข้าใจ หน่วยพื้นฐานของ สิ่งมีชีวิต ความสัมพันธ์ ของโครงสร้าง และ หน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงาน สัมพันธ์กัน มี กระบวนการสืบเสาะหา ความรู้ สื่อสารสิ่งที่ เรียนรู้และนำความรู้ไป ใช้ในการดำรงชีวิตของ ตนเองและดูแลสิ่งมีชีวิต	1. สังเกตและอธิบายรูปร่าง ลักษณะของเซลล์ของสิ่งมีชีวิต เซลล์เดียวและเซลล์ของสิ่งมีชีวิต หลายเซลล์ 2. สังเกตและเปรียบเทียบ ส่วนประกอบสำคัญของเซลล์พืช และเซลล์สัตว์ 3. ทดลองและอธิบายหน้าที่ของ ส่วนประกอบที่สำคัญของเซลล์ พืชและเซลล์สัตว์ 4. ทดลองและอธิบาย กระบวนการสารผ่านเซลล์โดย การแพร่และออสโมซิส 5. ทดลองหาปัจจัยบางประการที่ จำเป็นต่อการสังเคราะห์ด้วยแสง ของพืชและอธิบายว่าแสง คลอโรฟิลล์ แก๊สคาร์บอนได- ออกไซด์ น้ำ เป็นปัจจัย ที่จำเป็น ต้องใช้ในการสังเคราะห์ด้วยแสง 6. ทดลองและอธิบายผลที่ได้จาก การสังเคราะห์ด้วยแสงของพืช 7. อธิบายความสำคัญของ กระบวนการสังเคราะห์ด้วยแสง ของพืชต่อสิ่งมีชีวิตและสิ่งแวดล้อม 8. ทดลองและอธิบายกลุ่มเซลล์ที่ เกี่ยวข้องกับการลำเลียงน้ำของพืช	1. อธิบายโครงสร้างและการ ทำงานของระบบย่อยอาหาร ระบบหมุนเวียนเลือด ระบบ หายใจ ระบบขับถ่าย ระบบ สืบพันธุ์ ของมนุษย์และสัตว์ รวมทั้งระบบประสาทของมนุษย์ 2. อธิบายความสัมพันธ์ของระบบ ต่างๆ ของ มนุษย์และนำความรู้ ไปใช้ประโยชน์ 3. สังเกตและอธิบายพฤติกรรม ของมนุษย์และสัตว์ที่ตอบสนอง ต่อสิ่งเร้าภายนอกและภายใน 4. อธิบายหลักการและผลของ การใช้เทคโนโลยีชีวภาพในการ ขยายพันธุ์ ปรับปรุงพันธุ์ และ เพิ่มผลผลิตของสัตว์และนำ ความรู้ไปใช้ประโยชน์ 5. ทดลอง วิเคราะห์ และอธิบาย สารอาหารในอาหารมีปริมาณ พลังงานและสัดส่วนที่เหมาะสม กับเพศและวัย 6. อภิปรายผลของสารเสพติดต่อ ระบบต่างๆ ของร่างกาย และ แนวทางในการป้องกันตนเองจาก สารเสพติด	-

ตาราง 1 (ต่อ)

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
		<p>9. สังเกตและอธิบายโครงสร้างที่เกี่ยวข้องกับระบบลำเลียงน้ำและอาหารของพืช</p> <p>10. ทดลองและอธิบายโครงสร้างของดอกที่เกี่ยวข้องกับการสืบพันธุ์ของพืช</p> <p>11. อธิบายกระบวนการสืบพันธุ์แบบอาศัยเพศของพืชดอกและการสืบพันธุ์แบบไม่อาศัยเพศของพืช โดยใช้ส่วนต่างๆ ของพืชเพื่อช่วยในการขยายพันธุ์</p> <p>12. ทดลองและอธิบายการตอบสนองของพืชต่อแสง น้ำ และการสัมผัส</p> <p>13. อธิบายหลักการและผลของการใช้เทคโนโลยี ชีวภาพในการขยายพันธุ์ ปรับปรุงพันธุ์ เพิ่มผลผลิตของพืชและนำความรู้ไปใช้ประโยชน์</p>		
	<p>มาตรฐาน ว 1.2 เข้าใจกระบวนการและความสำคัญของการถ่ายทอดลักษณะทางพันธุกรรม วิวัฒนาการของสิ่งมีชีวิต ความหลากหลายทางชีวภาพ การใช้เทคโนโลยีชีวภาพที่มีผลกระทบต่อมนุษย์ และสิ่งแวดล้อม มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสาร สิ่งที่เรียนรู้ และนำความรู้ไปใช้ประโยชน์</p>			<p>1. สังเกตและอธิบายลักษณะของโครโมโซมที่มีหน่วยพันธุกรรมหรือยีนในนิวเคลียส</p> <p>2. อธิบายความสำคัญของสารพันธุกรรมหรือดีเอ็นเอ และกระบวนการถ่ายทอดลักษณะทางพันธุกรรม</p> <p>3. อภิปรายโรคทางพันธุกรรมที่เกิดจากความผิดปกติของยีนและโครโมโซมและนำความรู้ไปใช้ประโยชน์</p> <p>4. สืบเสาะหาและอธิบายความหลากหลายทางชีวภาพในท้องถิ่นที่ทำให้สิ่งมีชีวิตดำรงชีวิตอยู่ได้อย่างสมดุล</p> <p>5. อธิบายผลของความหลากหลายทางชีวภาพที่มีต่อมนุษย์ สัตว์ พืช และสิ่งแวดล้อม</p> <p>6. อภิปรายผลของเทคโนโลยีชีวภาพต่อการดำรงชีวิตของมนุษย์และสิ่งแวดล้อม</p>



ตาราง 1 (ต่อ)

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
สาระที่ 2 ชีวิตกับ สิ่งแวดล้อม	มาตรฐาน ว 2.1 เข้าใจ สิ่งแวดล้อมในท้องถิ่น ความสัมพันธ์ระหว่าง สิ่งแวดล้อมกับสิ่งมีชีวิต ความสัมพันธ์ระหว่าง สิ่งมีชีวิตต่างๆ ในระบบ นิเวศ มีกระบวนการ สืบเสาะหาความรู้และ จิตวิทยาศาสตร์สื่อสาร สิ่งที่เรียนรู้และนำความรู้ ไปใช้ประโยชน์	-	-	1. สำรวจระบบนิเวศต่างๆใน ท้องถิ่นและอธิบาย ความสัมพันธ์ ขององค์ประกอบภายในระบบนิเวศ 2. วิเคราะห์และอธิบาย ความสัมพันธ์ของการถ่ายทอด พลังงานของสิ่งมีชีวิตในรูปของโซ่ อาหารและสายใยอาหาร 3. อธิบายวัฏจักรน้ำ วัฏจักร คาร์บอน และความสำคัญที่มีต่อ ระบบนิเวศ 4. อธิบายปัจจัยที่มีผลต่อการ เปลี่ยนแปลงขนาดของประชากร ในระบบนิเวศ
	มาตรฐาน ว 2.2 เข้าใจ ความสำคัญของ ทรัพยากรธรรมชาติ การ ใช้ทรัพยากรธรรมชาติใน ระดับท้องถิ่น ประเทศ และโลกนำความรู้ไปใช้ ในการจัดการ ทรัพยากรธรรมชาติและ สิ่งแวดล้อมในท้องถิ่น อย่างยั่งยืน	-	-	1. วิเคราะห์สภาพปัญหา สิ่งแวดล้อม ทรัพยากรธรรมชาติ ในท้องถิ่น และเสนอแนวทางใน การแก้ปัญหา 2. อธิบายแนวทางการรักษา สมดุลของระบบนิเวศ 3. อภิปรายการใช้ทรัพยากร ธรรมชาติอย่างยั่งยืน 4. วิเคราะห์และอธิบายการใช้ ทรัพยากรธรรมชาติ ตามปรัชญา เศรษฐกิจพอเพียง 5. อภิปรายปัญหาสิ่งแวดล้อมและ เสนอแนะแนวทางการแก้ปัญหา 6. อภิปรายและมีส่วนร่วมใน การดูแลและอนุรักษ์สิ่งแวดล้อม ในท้องถิ่นอย่างยั่งยืน
สาระที่ 3 สารและ สมบัติของ สาร	มาตรฐาน ว 3. 1 เข้าใจ สมบัติของสาร ความสัมพันธ์ระหว่าง สมบัติของสารกับ โครงสร้างและแรงยึด เหนี่ยวระหว่างอนุภาค มีกระบวนการสืบเสาะ หาความรู้และจิตวิทยา ศาสตร์สื่อสารสิ่งที่เรียนรู้ นำความรู้ไปใช้ประโยชน์	1. ทดลองและจำแนกสารเป็น กลุ่มโดยใช้เนื้อสารหรือขนาด อนุภาคเป็นเกณฑ์ และอธิบาย สมบัติของสารในแต่ละกลุ่ม 2. อธิบายสมบัติและการเปลี่ยน สถานะของสาร โดยใช้ แบบจำลองการจัดเรียงอนุภาค ของสาร 3. ทดลองและอธิบายสมบัติความเป็นกรด เบส ของสารละลาย 4. ตรวจสอบค่า pH ของสารละลาย และนำความรู้ไปใช้ประโยชน์	1. สำรวจและอธิบาย องค์ประกอบ สมบัติของธาตุและ สารประกอบ 2. สืบค้นข้อมูลและเปรียบเทียบ สมบัติของธาตุโลหะ ธาตุอโลหะ ธาตุกึ่งโลหะและธาตุกัมมันตรังสี และนำความรู้ไปใช้ประโยชน์ 3. ทดลองและอธิบายหลักการ แยกสารด้วยวิธีการกรอง การตก ผลึก การสกัด การกลั่น และโคร มาโทกราฟี และนำความรู้ไปใช้ ประโยชน์	-

ตาราง 1 (ต่อ)

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
	มาตรฐาน ว เข้าใจ 3.2 หลักการและธรรมชาติของการเปลี่ยนแปลงสถานะของสาร การเกิดสารละลาย การเกิดปฏิกิริยา มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ และนำความรู้ไปใช้ประโยชน์	<p>1. ทดลองและอธิบายวิธีเตรียมสารละลายที่มีความเข้มข้นเป็นร้อยละ และอภิปรายการนำความรู้เกี่ยวกับสารละลายไปใช้ประโยชน์</p> <p>2. ทดลองและอธิบายการเปลี่ยนแปลงสมบัติ มวลและพลังงานของสาร เมื่อสารเปลี่ยนสถานะและเกิดการละลาย</p> <p>3. ทดลองและอธิบายปัจจัยที่มีผลต่อการเปลี่ยนสถานะ และการละลายของสาร</p>	<p>1. ทดลองและอธิบายการเปลี่ยนแปลงสมบัติ มวล และพลังงานเมื่อสารเกิดปฏิกิริยาเคมี รวมทั้งอธิบายปัจจัยที่มีผลต่อการเกิดปฏิกิริยาเคมี</p> <p>2. ทดลอง อธิบายและเขียนสมการเคมีของปฏิกิริยาของสารต่างๆ และนำความรู้ไปใช้ประโยชน์</p> <p>3. สืบค้นข้อมูลและอภิปรายผลของสารเคมี ปฏิกิริยาเคมีต่อสิ่งมีชีวิตและสิ่งแวดล้อม</p> <p>4. สืบค้นข้อมูลและอธิบายการใช้สารเคมีอย่างถูกต้อง ปลอดภัย วิธีป้องกันและแก้ไขอันตรายที่เกิดขึ้นจากการใช้สารเคมี</p>	-
สาระที่ 4 แรงและการเคลื่อนที่	มาตรฐาน ว 4. 1 เข้าใจธรรมชาติของแรง แม่เหล็ก ไฟฟ้า แรงโน้มถ่วง และแรงนิวเคลียร์ มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์อย่างถูกต้อง และมีคุณธรรม	<p>1. สืบค้นข้อมูล และอธิบายปริมาณ สเกลาร์ ปริมาณเวกเตอร์</p> <p>2. ทดลองและอธิบายระยะทางการกระจัด อัตราเร็วและความเร็ว ในการเคลื่อนที่ของวัตถุ</p>	<p>1. ทดลองและอธิบายการหาแรงลัพธ์ของแรงหลายแรงในระนาบเดียวกันที่กระทำต่อวัตถุ</p> <p>2. อธิบายแรงลัพธ์ที่กระทำต่อวัตถุที่หยุดนิ่งหรือวัตถุเคลื่อนที่ด้วยความเร็วคงตัว</p>	<p>1. อธิบายความเร่งและผลของแรงลัพธ์ที่กระทำต่อวัตถุ</p> <p>2. ทดลองและอธิบายแรงกิริยาและแรงปฏิกิริยาระหว่างวัตถุ และนำความรู้ไปใช้ประโยชน์</p> <p>3. ทดลองและอธิบายแรงพวงของของเหลวที่กระทำต่อวัตถุ</p>
	มาตรฐาน ว 4.2 เข้าใจลักษณะการเคลื่อนที่แบบต่างๆ ของวัตถุในธรรมชาติมีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์	-	-	<p>1. ทดลองและอธิบายความแตกต่างระหว่างแรงเสียดทานสถิตกับแรงเสียดทานจลน์ และนำความรู้ไปใช้ประโยชน์</p> <p>2. ทดลองและวิเคราะห์โมเมนต์ของแรง และนำความรู้ไปใช้ประโยชน์</p> <p>3. สังเกตและอธิบายการเคลื่อนที่ของวัตถุที่เป็นแนวตรง และแนวโค้ง</p>
สาระที่ 5 พลังงาน	มาตรฐาน ว 5. 1 เข้าใจความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การเปลี่ยนรูปพลังงาน ปฏิสัมพันธ์ระหว่างสารและพลังงาน ผลของการใช้พลังงานต่อ	<p>1. ทดลองและอธิบายอุณหภูมิและการวัดอุณหภูมิ</p> <p>2. สังเกตและอธิบายการถ่ายโอนความร้อน และนำความรู้ไปใช้ประโยชน์</p> <p>3. อธิบายการดูดกลืน การคายความร้อน โดยการแผ่รังสี และ</p>	<p>1. ทดลองและอธิบายการสะท้อนของแสง การหักเหของแสง และนำความรู้ไปใช้ประโยชน์</p> <p>2. อธิบายผลของความสว่างที่มีต่อมนุษย์และสิ่งมีชีวิตอื่นๆ</p> <p>3. ทดลองและอธิบายการดูดกลืนแสงสี การมองเห็นสีของวัตถุ</p>	<p>1. อธิบายงาน พลังงานจลน์ พลังงานศักย์โน้มถ่วง กฎการอนุรักษ์พลังงาน และความสัมพันธ์ระหว่างปริมาณเหล่านี้ รวมทั้งนำความรู้ไปใช้ประโยชน์</p> <p>2. ทดลองและอธิบาย</p>

ตาราง 1 (ต่อ)

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
	ชีวิตและสิ่งแวดล้อม มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์	นำความรู้ไปใช้ประโยชน์ 4. อธิบายสมมูลความร้อนและผลของความร้อนต่อการขยายตัวของสาร และนำความรู้ไปใช้ในชีวิตประจำวัน	และนำความรู้ไปใช้ประโยชน์	ความสัมพันธ์ระหว่างความต่างศักย์ กระแสไฟฟ้า ความต้านทาน และนำความรู้ไปใช้ประโยชน์ 3.คำนวณพลังงานไฟฟ้าของเครื่องใช้ไฟฟ้า และนำความรู้ไปใช้ประโยชน์ 4.สังเกตและอภิปรายการต่อวงจรไฟฟ้าในบ้านอย่างถูกต้องปลอดภัย และประหยัด 5.อธิบายตัวต้านทาน ไดโอด ทรานซิสเตอร์ และทดลองต่อวงจรอิเล็กทรอนิกส์เบื้องต้นที่มีทรานซิสเตอร์
สาระที่ 6 กระบวนการเปลี่ยนแปลงของโลก	มาตรฐาน ว 6.1 เข้าใจกระบวนการต่างๆ ที่เกิดขึ้นบนผิวโลกและภายในโลก ความสัมพันธ์ของกระบวนการต่างๆ ที่มีผลต่อการเปลี่ยนแปลงภูมิอากาศ ภูมิประเทศ และลักษณะของโลก มีกระบวนการสืบเสาะหาความรู้และจิตวิทยา ศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์	1. สืบค้นและอธิบายองค์ประกอบและการแบ่งชั้นบรรยากาศที่ปกคลุมผิวโลก 2. ทดลองและอธิบายความสัมพันธ์ระหว่าง อุณหภูมิ ความชื้นและความกดอากาศที่มีผลต่อปรากฏการณ์ทางลมฟ้าอากาศ 3. สังเกต วิเคราะห์และ อภิปรายการเกิดปรากฏการณ์ทางลมฟ้าอากาศที่มีผลต่อมนุษย์ 4. สืบค้น วิเคราะห์ และแปลความหมายข้อมูลจากการพยากรณ์อากาศ 5. สืบค้น วิเคราะห์ และอธิบายผลของลมฟ้าอากาศต่อการดำรงชีวิตของสิ่งมีชีวิต และสิ่งแวดล้อม 6. สืบค้น วิเคราะห์ และอธิบายปัจจัยทางธรรมชาติและการกระทำของมนุษย์ที่มีผลต่อการเปลี่ยนแปลงอุณหภูมิของโลก รุโหวโอโซน และฝนกรด 7. สืบค้น วิเคราะห์และอธิบายผลของภาวะโลกร้อน รุโหวโอโซน และฝนกรด ที่มีต่อสิ่งมีชีวิตและสิ่งแวดล้อม	1. สสำรวจ ทดลองและอธิบายลักษณะของชั้นหน้าตัดดิน สมบัติของดิน และกระบวนการเกิดดิน 2. สสำรวจ วิเคราะห์และอธิบายการใช้ประโยชน์และการปรับปรุงคุณภาพของดิน 3. ทดลองเลียนแบบเพื่ออธิบายกระบวนการเกิด และลักษณะองค์ประกอบของหิน 4. ทดสอบ และสังเกตองค์ประกอบและสมบัติของหินเพื่อจำแนกประเภทของหิน และนำความรู้ไปใช้ประโยชน์ 5. ตรวจสอบและอธิบายลักษณะทางกายภาพของแร่ และการนำไปใช้ประโยชน์ 6. สืบค้นและอธิบายกระบวนการเกิดลักษณะและสมบัติของปิโตรเลียม ถ่านหิน หินน้ำมัน และการนำไปใช้ประโยชน์ 7. สสำรวจและอธิบายลักษณะแหล่งน้ำธรรมชาติ การใช้ประโยชน์และการอนุรักษ์แหล่งน้ำในท้องถิ่น 8. ทดลองเลียนแบบ และอธิบายการเกิดแหล่งน้ำบนดิน แหล่งน้ำใต้ดิน 9. ทดลองเลียนแบบและอธิบายกระบวนการพุพุ่งอยู่กับที่	-

ตาราง 1 (ต่อ)

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
			การกร่อน การพัดพา การทับถม การตกผลึกและผลของกระบวนการดังกล่าว 10. สืบค้น สร้างแบบจำลองและอธิบายโครงสร้างและองค์ประกอบของโลก	
สาระที่ 7 ดาราศาสตร์ และอวกาศ	มาตรฐาน ว 7.1 เข้าใจวิวัฒนาการของระบบสุริยะ กาแล็กซีและเอกภพการปฏิสัมพันธ์ภายในระบบสุริยะและผลต่อสิ่งมีชีวิตบนโลก มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ การสื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์	-	-	1. สืบค้นและอธิบายความสัมพันธ์ระหว่างดวงอาทิตย์ โลก ดวงจันทร์และดาวเคราะห์อื่นๆ และผลที่เกิดขึ้นต่อสิ่งแวดล้อมและสิ่งมีชีวิตบนโลก 2. สืบค้นและอธิบายองค์ประกอบ ของเอกภพ กาแล็กซี และระบบสุริยะ 3. ระบุตำแหน่งของกลุ่มดาวและนำความรู้ไปใช้ประโยชน์
	มาตรฐาน ว 7.2 เข้าใจความสำคัญของเทคโนโลยีอวกาศที่นำมาใช้ในการสำรวจอวกาศและทรัพยากร ธรรมชาติ ด้านการเกษตรและการสื่อสาร มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์อย่างมีคุณธรรมต่อชีวิตและสิ่งแวดล้อม	-	-	1. สืบค้นและอภิปรายความก้าวหน้าของเทคโนโลยีอวกาศที่ใช้สำรวจอวกาศ วัตถุท้องฟ้า สภาวะอากาศ ทรัพยากรธรรมชาติการเกษตรและการสื่อสาร
สาระที่ 8 ธรรมชาติ ของ วิทยาศาสตร์ และ เทคโนโลยี	มาตรฐาน ว 8.1 ใช้กระบวนการทางวิทยาศาสตร์และจิตวิทยาศาสตร์ในการสืบเสาะหาความรู้ การแก้ปัญหา รู้ว่าปรากฏการณ์ทางธรรมชาติที่เกิดขึ้นส่วนใหญ่มีรูปแบบที่แน่นอน สามารถอธิบายและตรวจสอบได้ ภายได้	1. ตั้งคำถามที่กำหนดประเด็นหรือตัวแปรที่สำคัญในการสำรวจตรวจสอบ หรือศึกษาค้นคว้าเรื่องที่สนใจได้อย่างครอบคลุมและเชื่อถือได้ 2. สร้างสมมติฐานที่สามารถตรวจสอบได้และวางแผนการสำรวจตรวจสอบหลายๆ วิธี 3. เลือกเทคนิควิธีการสำรวจตรวจสอบทั้งเชิงปริมาณและเชิงคุณภาพที่ได้ผลเที่ยงตรงและ	1. ตั้งคำถามที่กำหนดประเด็นหรือตัวแปรที่สำคัญในการสำรวจตรวจสอบ หรือศึกษาค้นคว้าเรื่องที่สนใจได้อย่างครอบคลุมและเชื่อถือได้ 2. สร้างสมมติฐานที่สามารถตรวจสอบได้และวางแผนการสำรวจตรวจสอบหลายๆ วิธี 3. เลือกเทคนิควิธีการสำรวจตรวจสอบทั้งเชิงปริมาณและเชิงคุณภาพที่ได้ผลเที่ยงตรงและ	1. ตั้งคำถามที่กำหนดประเด็นหรือตัวแปรที่สำคัญในการสำรวจตรวจสอบ หรือศึกษาค้นคว้าเรื่องที่สนใจได้อย่างครอบคลุมและเชื่อถือได้ 2. สร้างสมมติฐานที่สามารถตรวจสอบได้และวางแผนการสำรวจตรวจสอบหลายๆ วิธี 3. เลือกเทคนิควิธีการสำรวจตรวจสอบทั้งเชิงปริมาณและเชิงคุณภาพที่ได้ผลเที่ยงตรงและ

ตาราง 1 (ต่อ)

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
	ข้อมูลและเครื่องมือที่มีอยู่ในเวลานั้นๆ เข้าใจว่า วิทยาศาสตร์ เทคโนโลยี สังคม และสิ่งแวดล้อม มีความเกี่ยวข้องสัมพันธ์กัน	ปลอดภัย โดยใช้วัสดุและเครื่องมือที่เหมาะสม 4. รวบรวมข้อมูล จัดกระทำ ข้อมูลเชิงปริมาณและคุณภาพ 5. วิเคราะห์และประเมินความ สอดคล้องของประจักษ์พยานกับ ข้อสรุป ทั้งที่สนับสนุนหรือขัดแย้ง กับสมมติฐาน และความผิดปกติ ของข้อมูลจากการสำรวจตรวจสอบ 6. สร้างแบบจำลอง หรือรูปแบบ ที่อธิบายผลหรือแสดงผลของ การสำรวจตรวจสอบ 7. สร้างคำถามที่นำไปสู่การสำรวจ ตรวจสอบ ในเรื่องที่เกี่ยวข้อง และ นำความรู้ที่ได้ไปใช้ในสถานการณ์ ใหม่หรืออธิบายเกี่ยวกับแนวคิด กระบวนการ และผลของ โครงการหรือชิ้นงานให้ผู้อื่นเข้าใจ	ปลอดภัย โดยใช้วัสดุและ เครื่องมือที่เหมาะสม 4. รวบรวมข้อมูล จัดกระทำ ข้อมูลเชิงปริมาณและคุณภาพ 5. วิเคราะห์และประเมินความ สอดคล้องของประจักษ์พยานกับ ข้อสรุป ทั้งที่สนับสนุนหรือขัดแย้ง กับสมมติฐาน และความผิดปกติ ของข้อมูลจากการสำรวจตรวจสอบ 6. สร้างแบบจำลอง หรือรูปแบบ ที่อธิบายผลหรือแสดงผลของ การสำรวจตรวจสอบ 7. สร้างคำถามที่นำไปสู่การสำรวจ ตรวจสอบ ในเรื่องที่เกี่ยวข้อง และ นำความรู้ที่ได้ไปใช้ในสถานการณ์ ใหม่หรืออธิบายเกี่ยวกับแนวคิด กระบวนการ และผลของ โครงการหรือชิ้นงานให้ผู้อื่นเข้าใจ	ปลอดภัย โดยใช้วัสดุและ เครื่องมือที่เหมาะสม 4. รวบรวมข้อมูล จัดกระทำ ข้อมูลเชิงปริมาณและคุณภาพ 5. วิเคราะห์และประเมินความ สอดคล้องของประจักษ์พยานกับ ข้อสรุป ทั้งที่สนับสนุนหรือขัดแย้ง กับสมมติฐาน และความผิดปกติ ของข้อมูลจากการสำรวจตรวจสอบ 6. สร้างแบบจำลอง หรือรูปแบบ ที่อธิบายผลหรือแสดงผลของ การสำรวจตรวจสอบ 7. สร้างคำถามที่นำไปสู่การสำรวจ ตรวจสอบ ในเรื่องที่เกี่ยวข้อง และ นำความรู้ที่ได้ไปใช้ในสถานการณ์ ใหม่หรืออธิบายเกี่ยวกับแนวคิด กระบวนการ และผลของ โครงการหรือชิ้นงานให้ผู้อื่นเข้าใจ
8	13	42	38	40

จากมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลาง การศึกษาขั้นพื้นฐาน พุทธศักราช 2551 โดยมีการกำหนดสาระสำคัญไว้ จำนวน 8 สาระ ประกอบด้วย 13 มาตรฐาน รวมจำนวนตัวชี้วัดทั้งหมด 119 ตัวชี้วัด ทั้งนี้ เมื่อพิจารณารายละเอียด ของแต่ละมาตรฐานและตัวชี้วัดแล้ว จะเห็นว่าตัวชี้วัดของมาตรฐานต่างๆ ส่วนใหญ่สามารถ ประเมินผลการเรียนรู้โดยใช้ข้อสอบ ยกเว้นตัวชี้วัดและมาตรฐานในสาระที่ 8 ซึ่งเป็นเรื่องที่มีนัย การใช้กระบวนการทางวิทยาศาสตร์และจิตวิทยาศาสตร์ในการสืบเสาะหาความรู้ การแก้ปัญหา รู้ว่า ปรากฏการณ์ทางธรรมชาติที่เกิดขึ้นส่วนใหญ่มีรูปแบบที่แน่นอน สามารถอธิบายและตรวจสอบได้

ภายใต้ข้อมูลและเครื่องมือที่มีอยู่ในช่วงเวลานั้นๆ เข้าใจว่า วิทยาศาสตร์ เทคโนโลยี สังคม และสิ่งแวดล้อมมีความเกี่ยวข้องสัมพันธ์กัน โดยครูผู้สอนจำเป็นต้องใช้การประเมินทักษะการปฏิบัติต่างๆ ในการประเมินผลผู้เรียน ซึ่งจะเหมาะสมกว่าการประเมินด้วยข้อสอบ ดังนั้น ในงานวิจัยนี้ที่ต้องการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดๆ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น จึงได้มีการคัดเลือกเฉพาะมาตรฐานและตัวชี้วัดที่เหมาะสมสำหรับการประเมินผลด้วยข้อสอบเท่านั้น โดยในงานวิจัยนี้ จะศึกษาเฉพาะตัวชี้วัดที่อยู่ในสาระที่ 1 - 7 เท่านั้น ซึ่งสอดคล้องกับรูปแบบของข้อสอบ O-NET ที่ สทศ. มีการออกข้อสอบโดยใช้ตัวชี้วัดที่อยู่ในสาระที่ 1 - 7 โดยสำหรับสาระที่ 8 นั้น สทศ. ให้เหตุผลว่ามีการแทรกอยู่ในสาระที่ 1 - 7 แล้ว ดังนั้น ในงานวิจัยนี้ มีจำนวนตัวชี้วัดที่ใช้ในการศึกษาทั้งหมด 92 ตัวชี้วัด (จากทั้งหมด 119 ตัวชี้วัด) โดยไม่รวมตัวชี้วัดของสาระที่ 8 มาตรฐาน ว 8.1 ที่ประกอบด้วยตัวชี้วัดของระดับชั้นมัธยมศึกษาปีที่ 1 2 และ 3 จำนวนระดับชั้นละ 9 ตัวชี้วัด

## ตอนที่ 2 การทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน

การทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (Ordinary National Educational Test: O-NET) เป็นการทดสอบเพื่อวัดความรู้และความคิดของนักเรียนชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ประเมินตามมาตรฐานการเรียนรู้ในหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จำนวน 51 มาตรฐานการเรียนรู้ ครอบคลุม 5 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษา ศาสนา และวัฒนธรรม และภาษาอังกฤษ

การทดสอบ O-NET ดำเนินการจัดสอบโดยสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ซึ่งเป็นหน่วยงานระดับชาติที่จัดตั้งขึ้นโดยมีวัตถุประสงค์เพื่อบริหารจัดการและดำเนินการเกี่ยวกับการศึกษา วิจัย พัฒนา และให้บริการทางการประเมินผลทางการศึกษาและทดสอบทางการศึกษาอย่างต่อเนื่อง รวมทั้งเป็นศูนย์กลางความร่วมมือด้านการทดสอบทางการศึกษาในระดับชาติและระดับนานาชาติ สทศ. มีหน้าที่ในการดำเนินการเกี่ยวกับการจัดทำระบบ วิธีการทดสอบและพัฒนาเครื่องมือวัดและประเมินผลตามมาตรฐานการศึกษา รวมทั้งดำเนินการประเมินผลการจัดการศึกษาและการทดสอบทางการศึกษาระดับชาติ ตลอดจนให้ความร่วมมือและสนับสนุนการทดสอบทั้งระดับเขตพื้นที่การศึกษา และสถานศึกษา

### 2.1 วัตถุประสงค์ของการทดสอบ O-NET

สถาบันทดสอบทางการศึกษาแห่งชาติ (2559: เว็บไซต์) กล่าวถึงวัตถุประสงค์ของการทดสอบ O-NET ดังนี้

1. เพื่อทดสอบความรู้และความคิดของนักเรียนชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6
2. เพื่อนำผลการทดสอบไปใช้เป็นองค์ประกอบหนึ่งในการจบการศึกษา ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551
3. เพื่อนำผลการทดสอบไปใช้ในการปรับปรุงคุณภาพการเรียนการสอนของโรงเรียน
4. เพื่อนำผลการทดสอบไปใช้ในการประเมินผลการเรียนรู้ของนักเรียนระดับชาติ
5. เพื่อนำผลการทดสอบไปใช้ในวัตถุประสงค์อื่น

## 2.2 ความสำคัญของการทดสอบ O-NET

สถาบันทดสอบทางการศึกษาแห่งชาติ (2559: เว็บไซต์) กล่าวถึงความสำคัญของการทดสอบ O-NET ที่มีต่อการศึกษา และต่อบุคคลที่เกี่ยวข้อง ดังนี้

1. การปรับการศึกษาให้มีมาตรฐานเดียวกัน คณะแผนกเฉลี่ยวการสอบ O-NET ในแต่ละปีจะบ่งบอกได้ว่า โรงเรียนแต่ละแห่งพัฒนาด้านคุณภาพการศึกษาไปในแนวทางใด นักเรียนมีคุณภาพการศึกษาอย่างไร เพื่อเป็นแนวทางในการปรับการศึกษาให้มีมาตรฐานเดียวกัน
2. การจัดกลุ่มโรงเรียน คณะแผนกเฉลี่ยว O-NET ใช้จัดกลุ่มโรงเรียนแต่ละแห่ง ซึ่งอาจแบ่งจาก ปัจจัยของโรงเรียนโดยเปรียบเทียบว่า คณะแผนกเฉลี่ยวของโรงเรียนอยู่ในระดับใด จะมีการพิจารณาตัวแปรหรือปัจจัยในการแบ่งกลุ่มว่าจะมีการกำหนดอย่างไรให้เกิดความยุติธรรมกับโรงเรียน การจัดอันดับจะจัดกลุ่มโรงเรียนที่มีความสามารถใกล้เคียงกัน เพื่อให้แต่ละโรงเรียนได้ทราบถึงระดับคุณภาพของโรงเรียน และจะได้มีการพัฒนาคุณภาพโรงเรียนของตนเองให้ดียิ่งขึ้นต่อไป
3. โรงเรียนและครูผู้สอน สามารถนำข้อมูลที่ได้จากการสอบไปปรับปรุงและพัฒนาการเรียนการสอนของโรงเรียนต่อไป รวมทั้งนำผลการประเมินของนักเรียนแต่ละคนไปใช้เป็นข้อมูลในการจัดการเรียนและกิจกรรมต่างๆ ให้สอดคล้องต่อความสนใจ ความต้องการเรียนรู้ เพื่อพัฒนาความรู้ความสามารถของนักเรียน
4. ผู้ปกครอง สามารถใช้คะแนนสอบของนักเรียนเป็นแนวทางแก้ไขและสนับสนุนบุตรหลานให้กระตือรือร้นในการเรียนรู้มากขึ้น เพื่อให้เกิดการพัฒนาไปในแนวทางที่ดี
5. นักเรียน สามารถใช้ประโยชน์จากผลคะแนนที่นักเรียนได้รับ เป็นหลักฐานสำคัญในการศึกษาต่อระดับสูงขึ้นไป ซึ่งจะนำไปสู่ความสำเร็จในการเรียนรู้ และวางแผนอนาคต

## 2.3 จำนวนข้อสอบ O-NET ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำแนกตามรูปแบบของข้อสอบ

จำนวนข้อสอบ O-NET ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำแนกตามรูปแบบของข้อสอบ ซึ่งแบ่งเป็น 6 รูปแบบ ได้แก่ (1) ปรนัย (2) ปรนัยหลายตัวเลือก 1 คำตอบ (3) ปรนัยหลายตัวเลือก มากกว่า 1 คำตอบ (4) เลือกคำตอบจากแต่ละหมวดที่สัมพันธ์กัน (5) ระบายคำตอบที่เป็นค่า/ตัวเลข และ (6) อัตนัย ของทั้ง 5 สาระการเรียนรู้ แสดงดังตาราง 2

ตาราง 2 จำนวนข้อสอบ O-NET ชั้นมัธยมศึกษาปีที่ 3 จำแนกตามรูปแบบของข้อสอบ ประจำปีการศึกษา 2559

ที่	รูปแบบ	สาระการเรียนรู้									
		ภาษาไทย		สังคมศึกษา		ภาษาอังกฤษ		คณิตศาสตร์		วิทยาศาสตร์	
		จำนวนข้อ	คะแนน	จำนวนข้อ	คะแนน	จำนวนข้อ	คะแนน	จำนวนข้อ	คะแนน	จำนวนข้อ	คะแนน
1	ปรนัย										
	1.1) 4 ตัวเลือก 1 คำตอบ	40	80	45	90	50	100	25	80	40	80
	1.2) 4 ตัวเลือก 2 คำตอบ	-	-	-	-	-	-	-	-	-	-
	1.3) 5 ตัวเลือก 1 คำตอบ	-	-	-	-	-	-	-	-	-	-
	1.4) 5 ตัวเลือก 2 คำตอบ	10	20	-	-	-	-	-	-	-	-
2	ปรนัย หลายตัวเลือก 1 คำตอบ	-	-	5	10	-	-	-	-	-	-
3	ปรนัย หลายตัวเลือก มากกว่า 1 คำตอบ	-	-	-	-	-	-	-	-	5	20
4	เลือกคำตอบจากแต่ละหมวดที่สัมพันธ์กัน	-	-	-	-	-	-	5	20	-	-
5	ระบายคำตอบที่เป็นค่าตัวเลข/	-	-	-	-	-	-	-	-	-	-
6	อัตนัย	-	-	-	-	-	-	-	-	-	-
รวม		50	100	50	100	50	100	30	100	45	100
จำนวนเวลาที่ใช้สอบ (นาที)		90 นาที		90 นาที		90 นาที		90 นาที		90 นาที	

## 2.4 จำนวนข้อสอบ O-NET วิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำแนกตามรูปแบบของข้อสอบ

ข้อสอบ O-NET วิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 มีจำนวนทั้งหมด 45 ข้อ ประกอบด้วยรูปแบบข้อสอบ 2 รูปแบบ ได้แก่ (1) รูปแบบปรนัย 4 ตัวเลือก 1 คำตอบ จำนวน 40 ข้อ ขอละ 2 คะแนน และ (2) รูปแบบปรนัยหลายตัวเลือก มากกว่า 1 คำตอบ จำนวน 5 ข้อ ขอละ 4 คะแนน (สถาบันทดสอบทางการศึกษาแห่งชาติ, 2559: เว็บไซต์) รายละเอียดแสดงดังตาราง 3



ตาราง 3 จำนวนข้อสอบ O-NET วิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำแนกตามรูปแบบของข้อสอบ

สาระ	มาตรฐาน	คุณภาพผู้เรียน	ตัวชี้วัด	สรุปจำนวน ตัวชี้วัด	รูปแบบข้อสอบ (จำนวนข้อ)	คะแนน	
1	ว 1.1	เข้าใจลักษณะและองค์ประกอบที่สำคัญของเซลล์สิ่งมีชีวิตความสัมพันธ์ของการทำงานของระบบต่างๆ การถ่ายทอดลักษณะทางพันธุกรรมเทคโนโลยีชีวภาพ ความหลากหลายของสิ่งมีชีวิต พฤติกรรมและการตอบสนองต่อสิ่งเร้าของสิ่งมีชีวิต ความสัมพันธ์ระหว่างสิ่งมีชีวิตในสิ่งแวดล้อม	ม.1/4 ทดลองและอธิบายกระบวนการสารผ่านเซลล์โดยการแพร่และออสโมซิส	6	4 ตัวเลือก 1 คำตอบ (6 ข้อ)	12	
			ม.1/5 ทดลองหาปัจจัยบางประการที่จำเป็นต่อการสังเคราะห์ด้วยแสงของพืชและอธิบายว่าแสง คลอโรฟิลล์ แก๊สคาร์บอนไดออกไซด์ น้ำ เป็นปัจจัย ที่จำเป็นต้องใช้ในการสังเคราะห์ด้วยแสง				
			ม.2/2 อธิบายความสัมพันธ์ของระบบต่างๆ ของมนุษย์และนำความรู้ไปใช้ประโยชน์				
			ม.2/3 สังเกตและอธิบายพฤติกรรมของมนุษย์และสัตว์ที่ตอบสนองต่อสิ่งเร้าภายนอกและภายใน				
			ม.2/4 อธิบายหลักการและผลของการใช้เทคโนโลยีชีวภาพในการขยายพันธุ์ ปรับปรุงพันธุ์ และเพิ่มผลผลิตของสัตว์และนำความรู้ไปใช้ประโยชน์				
	ม.2/5 ทดลอง วิเคราะห์ และอธิบายสารอาหารในอาหารมีปริมาณพลังงานและสัดส่วนที่เหมาะสมกับเพศและวัย						
	ว 1.2		ม.3/3 อภิปรายโรคทางพันธุกรรมที่เกิดจากความผิดปกติของยีนและโครโมโซมและนำความรู้ไปใช้ประโยชน์	2	4 ตัวเลือก 1 คำตอบ (1 ข้อ) 6 ตัวเลือก 2 คำตอบ (1 ข้อ)	6	
			ม.3/6 อภิปรายผลของเทคโนโลยี ชีวภาพต่อการดำรงชีวิตของมนุษย์และสิ่งแวดล้อม				
	2	ว 2.1		ม.3/1 สืบหากระบวนการในท้องถิ่นและอธิบายความสัมพันธ์ขององค์ประกอบภายในระบบนิเวศ	2	4 ตัวเลือก 1 คำตอบ (1 ข้อ) 6 ตัวเลือก 2 คำตอบ (1 ข้อ)	8
				ม.3/3 อธิบายวัฏจักรน้ำ วัฏจักรคาร์บอน และความสัมพันธ์ที่มีต่อระบบนิเวศ			
ว 2.2			ม.3/2 อธิบายแนวทางการรักษาสมดุลของระบบนิเวศ	2	4 ตัวเลือก 1 คำตอบ (2 ข้อ)	4	
			ม.3/4 วิเคราะห์และอธิบายการใช้ทรัพยากรธรรมชาติ ตามปรัชญาเศรษฐกิจพอเพียง				

ตาราง 3 (ต่อ)

สาระ	มาตรฐาน	คุณภาพผู้เรียน	ตัวชี้วัด	สรุปจำนวน ตัวชี้วัด	รูปแบบข้อสอบ (จำนวนข้อ)	คะแนน	
3	ว 3.1	เข้าใจองค์ประกอบและสมบัติของสารละลาย สารบริสุทธิ์ การเปลี่ยนแปลงของสาร ในรูปแบบของการเปลี่ยนสถานะ การเกิดสารละลาย และการเกิดปฏิกิริยาเคมี	ม.1/1 ทดลองและจำแนกสารเป็นกลุ่มโดยใช้เนื้อสารหรือขนาดอนุภาคเป็นเกณฑ์ และอธิบายสมบัติของสารในแต่ละกลุ่ม	5	4 ตัวเลือก 1 คำตอบ (4 ข้อ)	8	
			ม.1/2 อธิบายสมบัติและการเปลี่ยนสถานะของสาร โดยใช้แบบจำลองการจัดเรียงอนุภาคของสาร				
			ม.1/4 ตรวจสอบค่า pH ของสารละลายและนำความรู้ไปใช้ประโยชน์				
			ม.2/2 สืบค้นข้อมูลและเปรียบเทียบสมบัติของธาตุโลหะ ธาตุอโลหะ ธาตุกึ่งโลหะและธาตุกัมมันตรังสีและนำความรู้ไปใช้ประโยชน์				
			ม.2/3 ทดลองและอธิบายหลักการแยกสารด้วยวิธีการกรอง การตกผลึก การสกัด การกลั่น และโครมาโทกราฟี และนำความรู้ไปใช้ประโยชน์				
	ว 3.2		ม.1/1 ทดลองและอธิบายวิธีเตรียมสารละลายที่มีความเข้มข้นเป็นร้อยละ และอภิปรายการนำความรู้เกี่ยวกับสารละลายไปใช้ประโยชน์	3	4 ตัวเลือก 1 คำตอบ (2 ข้อ)	4	
			ม.1/2 ทดลองและอธิบายการเปลี่ยนแปลงสมบัติมวลและพลังงานของสาร เมื่อสารเปลี่ยนสถานะและเกิดการละลาย				
			ม.2/1 ทดลองและอธิบายการเปลี่ยนแปลงสมบัติมวล และพลังงานเมื่อสารเกิดปฏิกิริยาเคมี รวมทั้งอธิบายปัจจัยที่มีผลต่อการเกิดปฏิกิริยาเคมี				
	4	ว 4.1	เข้าใจแรงเสียดทาน โมเมนตัมของแรง การเคลื่อนที่แบบต่างๆ ในชีวิตประจำวัน กฎการอนุรักษ์พลังงาน การถ่ายโอนพลังงาน สมดุลความร้อน การสะท้อน การหักเห และความเข้มของแสง	ม.1/2 ทดลองและอธิบายระยะทาง การกระจัด อัตราเร็วและความเร็ว ในการเคลื่อนที่ของวัตถุ	4	4 ตัวเลือก 1 คำตอบ (7 ข้อ) 6 ตัวเลือก 2 คำตอบ (1 ข้อ)	18
				ม.2/2 อธิบายแรงลัพธ์ที่กระทำต่อวัตถุที่หยุดนิ่งหรือวัตถุเคลื่อนที่ด้วยความเร็วคงตัว			
ม.3/1 อธิบายความเร่งและผลของแรงลัพธ์ที่กระทำต่อวัตถุ							
ม.3/2 ทดลองและอธิบายแรงกิริยาและแรงปฏิกิริยาระหว่างวัตถุ และนำความรู้ไปใช้ประโยชน์							
ว 4.2			ม.3/1 ทดลองและอธิบายความแตกต่างระหว่างแรงเสียดทานสถิตกับแรงเสียดทานจลน์ และนำความรู้ไปใช้ประโยชน์	3	4 ตัวเลือก 1 คำตอบ (4 ข้อ)	12	
			ม.3/2 ทดลองและวิเคราะห์โมเมนต์ของแรง และนำความรู้ไปใช้ประโยชน์				
			ม.3/3 สังเกตและอธิบายการเคลื่อนที่ของวัตถุที่เป็นแนวตรง และแนวโค้ง				

ตาราง 3 (ต่อ)

สาระ	มาตรฐาน	คุณภาพผู้เรียน	ตัวชี้วัด	สรุปจำนวน ตัวชี้วัด	รูปแบบข้อสอบ (จำนวนข้อ)	คะแนน	
5	ว 5.1	เข้าใจความสัมพันธ์ระหว่างปริมาณทางไฟฟ้าหลักการต่อวงจรไฟฟ้าในบ้านพลังงานไฟฟ้าและหลักการเบื้องต้น ของวงจรอิเล็กทรอนิกส์	ม.1/2	สังเกตและอธิบายการถ่ายโอนความร้อนและนำความรู้ไปใช้ประโยชน์	8	4 ตัวเลือก 1 คำตอบ (6 ข้อ) 6 ตัวเลือก 2 คำตอบ (1 ข้อ)	16
			ม.2/1	ทดลองและอธิบายการสะท้อนของแสง การหักเหของแสง และนำความรู้ไปใช้ประโยชน์			
			ม.2/2	อธิบายผลของความสว่างที่มีต่อมนุษย์และสิ่งมีชีวิตอื่นๆ			
			ม.2/3	ทดลองและอธิบายการดูดกลืนแสงสี การมองเห็นสีของวัตถุ และนำความรู้ไปใช้ประโยชน์			
			ม.3/1	อธิบายงาน พลังงานจลน์ พลังงานศักย์โน้มถ่วง กฎการอนุรักษ์พลังงาน และความสัมพันธ์ระหว่างปริมาณเหล่านี้ รวมทั้งนำความรู้ไปใช้ประโยชน์			
			ม.3/3	คำนวณพลังงานไฟฟ้าของเครื่องใช้ไฟฟ้าและนำความรู้ไปใช้ประโยชน์			
			ม.3/4	สังเกตและอภิปรายการต่อวงจรไฟฟ้าในบ้านอย่างถูกต้องปลอดภัย และประหยัด			
6	ว 6.1	เข้าใจกระบวนการเปลี่ยนแปลงของเปลือกโลก แพลงทรพยากรธรณีปฏิกิริยาที่มีผลต่อการเปลี่ยนแปลงของบรรยากาศ ปฏิสัมพันธ์ภายในระบบสุริยะ และผลที่มีต่อสิ่งต่างๆ บนโลก ความสำคัญของเทคโนโลยี	ม.1/6	สืบค้น วิเคราะห์ และอธิบายปัจจัยทางธรรมชาติและการทำงานของมนุษย์ที่มีผลต่อการเปลี่ยนแปลงอุณหภูมิของโลกหรือโอโซนและฝนกรด	3	4 ตัวเลือก 1 คำตอบ (1 ข้อ)	2
			ม.2/2	สำรวจ วิเคราะห์และอธิบายการใช้ประโยชน์และการปรับปรุงคุณภาพของดิน			
			ม.2/4	ทดสอบ และสังเกตองค์ประกอบและสมบัติของหิน เพื่อจำแนกประเภทของหิน และนำความรู้ไปใช้ประโยชน์			
7	ว 7.1	อวกาศ	ม.3/1	สืบค้นและอธิบายความสัมพันธ์ระหว่างดวงอาทิตย์ โลก ดวงจันทร์และดาวเคราะห์อื่นๆ และผลที่เกิดขึ้นต่อสิ่งแวดล้อมและสิ่งมีชีวิตบนโลก	2	4 ตัวเลือก 1 คำตอบ (5 ข้อ)	10
			ม.3/3	ระบุตำแหน่งของกลุ่มดาว และนำความรู้ไปใช้ประโยชน์			
	ว 7.2		ม.3/1	สืบค้นและอภิปรายความก้าวหน้าของเทคโนโลยีอวกาศที่ใช้สำรวจอวกาศ วัตถุท้องฟ้า สภาวะอากาศ ทรัพยากรธรรมชาติ การเกษตร และการสื่อสาร	1		
<b>รวม</b>				<b>40</b>	<b>45</b>	<b>100</b>	
<b>จำนวนเวลาที่ใช้สอบ</b>				<b>90 นาที</b>			

### ตอนที่ 3 การศึกษาความสอดคล้องในแนวเดียวกัน (alignment study)

ผู้วิจัยได้ทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาความสอดคล้องในแนวเดียวกัน (alignment study) โดยแบ่งการนำเสนอออกเป็น 6 ประเด็นหลัก ได้แก่

- 3.1 ความเป็นมาของการศึกษาความสอดคล้องในแนวเดียวกัน
- 3.2 ความหมายของความสอดคล้องในแนวเดียวกัน
- 3.3 ความสำคัญของการศึกษาความสอดคล้องในแนวเดียวกัน
- 3.4 วิธีการในการศึกษาความสอดคล้องในแนวเดียวกัน

รายละเอียดของแต่ละประเด็นมีดังนี้

#### 3.1 ความเป็นมาของการศึกษาความสอดคล้องในแนวเดียวกัน

การศึกษาเรื่องความสอดคล้องในแนวเดียวกันของระบบการศึกษาในยุคการศึกษาอิงมาตรฐานเริ่มได้รับความสนใจอย่างมากในช่วงปี ค.ศ. 1990 เป็นต้นมา โดยความสนใจนี้เกิดขึ้นในหมู่นักวิชาการ นักการศึกษา และผู้กำหนดนโยบายการศึกษาของประเทศสหรัฐอเมริกาเป็นส่วนใหญ่ เหตุผลหลักของการให้ความสนใจในเรื่องนี้มากจากการปฏิรูปการศึกษาที่ต้องการเห็นระบบการศึกษาต่างๆ มีความสอดคล้องและประสานกันทุกระบบเพื่อให้เกิดความมั่นใจว่าการดำเนินงานภายใต้การปฏิรูปการศึกษาจะประสบความสำเร็จ ทั้งนี้ สหรัฐอเมริกาได้มีการออกกฎหมายทางการศึกษาหลายฉบับที่ต้องการให้ผู้ปฏิบัติจัดระบบการศึกษาให้สอดคล้องกัน เช่น ในปี ค.ศ. 1995 กฎหมายการปฏิรูปโรงเรียน ที่ระบุว่ารัฐต่างๆ ต้องทำการประเมินให้สอดคล้องหรือสอดคล้องกับหลักสูตร รวมทั้งการปฏิรูปการศึกษาล่าสุดของอเมริกา คือ No Child Left Behind หรือ NCLB ที่เริ่มใช้ในปี ค.ศ. 2001 ก็ยังคงมาตรการเดิมนี้ไว้ และระบุให้ชัดเจนขึ้นมากกว่าเดิม โดยระบุว่า รัฐต่างๆ จะต้องทำการศึกษาหาข้อมูลเพื่อยืนยันว่าข้อสอบวิชาการอ่านภาษาอังกฤษและคณิตศาสตร์ของผู้เรียนระดับเกรด 3 ถึง เกรด 8 (ประถมศึกษาปีที่ 3 ถึง มัธยมศึกษาปีที่ 2) ที่รัฐพัฒนาขึ้นเองสำหรับใช้ประเมินผู้เรียนในรัฐของตนนั้นสอดคล้องกับมาตรฐานด้านเนื้อหาที่กำหนดขึ้น ดังนั้น วิธีการวัดและประเมินความสอดคล้องจึงเป็นประเด็นสำคัญมากในยุคนั้น

นักวิชาการด้านการศึกษาที่เล็งเห็นความสำคัญของการประเมินความสอดคล้องระหว่างมาตรฐานกับเครื่องมือที่ใช้ประเมินในยุคแรกๆ ได้แก่ Norman Webb ซึ่ง ในปี ค.ศ. 1997 Webb ได้พัฒนาวิธีการและเกณฑ์สำหรับประเมินความสอดคล้องขึ้นมา และเขียนเป็นรายงานนำเสนอต่อสถาบันการสอนวิทยาศาสตร์แห่งชาติอเมริกัน (National Institute for Science Education) โดยเกณฑ์การประเมินความสอดคล้องที่ Webb เสนอขึ้นมา มี 5 ประเด็น คือ การมุ่งเน้นเนื้อหา (content focus) การสอน (pedagogical implications) ความเสมอภาค (equity) ความเชื่อมโยง

กันระหว่างระดับชั้น (articulation across grades) และ ความสามารถในการนำไปใช้อย่างเป็นระบบ (system applicability)

ในช่วงระหว่างปี ค.ศ. 1995 สมาคมผู้บริหารโรงเรียนระดับสูงของรัฐ (The Council of Chief State School Officers: CCSSO) ได้ดำเนินการประเมินมาตรฐานการศึกษาของรัฐต่างๆ และมีความสนใจในการประเมินความสอดคล้องของมาตรฐานการศึกษาของรัฐกับเครื่องมือประเมิน และ Webb ได้มีบทบาทในการพัฒนากระบวนการวิเคราะห์เนื้อหา (content analysis) สำหรับตัดสินความสอดคล้องของมาตรฐานและเครื่องมือการประเมิน โดยการวิเคราะห์เนื้อหาที่มี 6 เกณฑ์ ได้แก่ ความสอดคล้องของเนื้อหา (categorical concurrence) ความสอดคล้องด้านความลึกของความรู้ (depth of knowledge consistency) ความสอดคล้องด้านขอบเขตของความรู้ (range of knowledge correspondence) และความสมดุลของเนื้อหา (balance of representation) โครงสร้างของเนื้อหา (structure of knowledge) และนิสัยของความสอดคล้อง (dispositional consonance) อย่างไรก็ตาม มีเพียง 4 เกณฑ์แรกเท่านั้นที่ถูกนำไปใช้ในการศึกษาความสอดคล้องในแนวเดียวกัน (Webb, 1999; 2007)

ปี ค.ศ. 1998 เป็นปีแรกที่มีการนำวิธีการประเมินความสอดคล้องของ Webb ไปใช้ประเมินใน 4 รัฐ มีผู้เชี่ยวชาญร่วมกันประมาณ 4-5 คน หลังจากได้มีการนำวิธีการประเมินที่พัฒนาขึ้นไปใช้แล้ว Webb ได้ปรับปรุงวิธีการประเมินความสอดคล้องให้มีความเหมาะสมยิ่งขึ้น ก่อนนำไปใช้ประเมินในรัฐอื่นๆ ตลอดจนจัดทำและเผยแพร่คู่มืออธิบายวิธีการและเกณฑ์สำหรับประเมินความสอดคล้องอีกครั้งในปี ค.ศ. 2002 ต่อมาในปี ค.ศ. 2003 รัฐโอกาโฮมาได้ริเริ่มการพัฒนาระบบคอมพิวเตอร์เพื่อใช้สำหรับการประเมินความสอดคล้อง โดยได้รับการสนับสนุนจากสมาคมผู้บริหารโรงเรียนระดับสูงของรัฐ (CCSSO) โดยได้รับทุนจากกองทุนด้านวิทยาศาสตร์แห่งชาติอเมริกัน (National Science Foundation: NSF) และกระทรวงศึกษาธิการของสหรัฐ หลังจากนั้นในปีเดียวกัน Webb ได้พัฒนาเครื่องมือสำหรับประเมินความสอดคล้องขึ้นทางเว็บไซต์ (Web Alignment Tool) หรือเรียกกันว่า WAT ซึ่งเป็นเครื่องมือที่ช่วยให้ผู้ใช้สามารถทำการวิเคราะห์และประเมินความสอดคล้องอย่างสะดวกมากขึ้น และได้เผยแพร่เครื่องมือนี้ในปี ค.ศ. 2004 และมีการประชุมสัมมนาและสาธิตการใช้เครื่องมือนี้ในปี ค.ศ. 2005

นอกจากนี้ ยังมีนักวิชาการอื่นๆ ที่พัฒนาวิธีการประเมินความสอดคล้องขึ้นมาด้วยเช่นกัน เช่น Porter (2002) ได้พัฒนาวิธีการที่เรียกว่า “การสำรวจหลักสูตรที่นำไปใช้จริง” (survey of the enacted curriculum: SEC) หรือที่เรียกว่า วิธีการของ Porter โดยมีจุดเน้นคือการวิเคราะห์เมทริกซ์ความสอดคล้องของเนื้อหาตามหลักสูตรกับระดับของความซับซ้อนทางปัญญาหรือทางการคิด (content-by-cognitive level matrix) ผู้ประเมินจะทำการวิเคราะห์มาตรฐาน การประเมินหลักสูตรหรือการสอน แล้วนำเสนอในรูปแบบของเมทริกซ์ การประเมินความสอดคล้องพิจารณาได้จาก

ระดับของความซ้อนทับกัน (overlap) ระหว่างสิ่งที่ต้องการประเมินที่แสดงในรูปเมทริกซ์ เช่น ความสอดคล้องระหว่างการประเมินและมาตรฐาน หรือระหว่างหลักสูตรกับมาตรฐาน เป็นต้น

จากพัฒนาการของการศึกษาเรื่องความสอดคล้องที่กล่าวมา จะสามารถกล่าวได้ว่าการศึกษาเรื่องความสอดคล้องไม่ใช่เรื่องใหม่สำหรับนักวิชาการในแวดวงการวัดและประเมินทางการศึกษา เพราะการศึกษาความสอดคล้องในบริบทของการวัดและประเมินผลอิงมาตรฐานเป็นประเด็นเดียวกับการศึกษาเรื่อง “ความตรง” (validity) นักวิชาการหลายท่าน เช่น Porter (2002) และ Webb (1997) กล่าวว่าความสอดคล้องเป็นเงื่อนไขที่จำเป็นเงื่อนไขหนึ่งที่จะทำให้การใช้ผลการวัดความรู้ของผู้เรียนมีความถูกต้องและเหมาะสมมากขึ้น ดังนั้น การศึกษาเรื่องความสอดคล้องจึงเป็นประเด็นหนึ่งของการศึกษาเรื่องความตรง มาตรฐานการทดสอบด้านการจิตวิทยา และการทดสอบทางการศึกษา เน้นย้ำว่า เมื่อใดก็ตามที่มีการประเมินอิงมาตรฐาน ควรมีการประเมินความสอดคล้องด้วย ผลการวิเคราะห์ความสอดคล้องจะเป็น “หลักฐานของการประเมินความตรงของการแปลและใช้ผลการประเมิน” (evidence of assessment’s validity) ให้มีความเหมาะสม มีความหมาย และถูกต้อง ซึ่งความตรงในที่นี้มีความหมายกว้างรวมไปถึงระดับความเหมาะสมของการใช้และการแปลความหมายผลการประเมินด้วย (APA, AERA, & NCME, 1999; Webb, 1977)

การประเมินความสอดคล้องยิ่งทวีความสำคัญมากขึ้นในบริบทที่มีการจัดการศึกษาที่อิงมาตรฐาน เพราะเป็นเรื่องของการตรวจสอบความเหมาะสมของการวัดและการประเมินทางการศึกษา หากต้องการประเมินคุณภาพผู้เรียนตามหลักสูตรอิงมาตรฐาน แต่พบว่ามาตรฐานที่กำหนดในหลักสูตรกับเครื่องมือประเมินไม่สอดคล้องกันก็จะทำให้การใช้ผลการประเมินคุณภาพผู้เรียนนั้นไม่ถูกต้อง และการประเมินผลความสำเร็จของการจัดการศึกษาก็จะไม่ถูกต้องด้วยเช่นกัน

ปัจจุบันนี้แนวโน้มของการใช้ผลการประเมินทางการศึกษาเพื่อประเมินโรงเรียน ครู และบุคลากรทางการศึกษาสูงมากขึ้นตามหลักการและแนวคิดของการประกันคุณภาพการศึกษา (educational accountability) การประเมินความสอดคล้องระหว่างเครื่องมือประเมินและมาตรฐาน ก็ยังมีความจำเป็นมากขึ้น ตัวอย่างเช่น การดำเนินนโยบายทางการศึกษาที่เรียกว่า No Child Left Behind หรือ NCLB ของประเทศสหรัฐอเมริกาที่รัฐบาลมอบอำนาจให้รัฐต่างๆ กำหนดมาตรฐานการศึกษาของวิชาหลักด้วยตนเอง และกำหนดให้รัฐแต่ละรัฐประเมินความรู้และความสามารถของผู้เรียนด้วยตนเองเพื่อต้องการนำผลการประเมินมาเป็นส่วนหนึ่งของการประกันคุณภาพการศึกษา รัฐแต่ละรัฐนอกจากจะรายงานผลการประเมินผู้เรียนแล้ว ต้องทำการประเมินความสอดคล้องเพื่อเป็นหลักประกันหรือเป็นหลักฐานว่าการประเมินและการใช้คะแนนจากการประเมินผู้เรียนโดยบุคคลและหน่วยงานที่เกี่ยวข้อง เช่น ครู และโรงเรียน สามารถนำไปใช้ประเมินความรู้ความสามารถของผู้เรียนได้อย่างเหมาะสม

ดังนั้น จะเห็นว่าการวัดและการศึกษาเรื่องความสอดคล้องของมาตรฐานกับองค์ประกอบของการจัดการศึกษามีรากฐานแนวคิดมาจากการประเมินความตรงในบริบทของการจัดการศึกษาอิงมาตรฐาน การประเมินความสอดคล้องของมาตรฐานกับองค์ประกอบของการจัดการศึกษาจะมีความสำคัญมากขึ้น เพราะผลการประเมินความสอดคล้องมีความสำคัญต่อการตัดสินใจเกี่ยวกับประเด็นต่างๆ อันเกี่ยวข้องกับคะแนนหรือผลการประเมินตามมาตรฐาน

สำหรับประวัติความเป็นมาของการศึกษาความสอดคล้องในประเทศไทยยังไม่ปรากฏว่ามี การพัฒนาวิธีการวัดความสอดคล้อง ทั้งที่ในปัจจุบันประเทศไทยใช้หลักสูตรอิงมาตรฐาน อาจเป็น เพราะประเด็นเกี่ยวกับการวัดความสอดคล้องต่างๆ ยังไม่เป็นที่รู้จักแพร่หลายในหมู่นักวิชาการมากนัก ถึงแม้ว่าประเด็นเรื่องกับการวัดความสอดคล้องเป็นสิ่งจำเป็นมากในการวัดและประเมินผลอิงมาตรฐาน

### 3.2 ความหมายของความสอดคล้องในแนวเดียวกัน

โดยทั่วไประบบการศึกษาแบบอิงมาตรฐานประกอบด้วย 3 องค์ประกอบหลัก ได้แก่ มาตรฐาน (standard) การเรียนการสอน (instruction) และการประเมิน (assessment) ซึ่ง มาตรฐานเป็นเอกสารเชิงนโยบายที่แสดงถึงเป้าหมายที่คาดหวังให้ผู้เรียนมีความรู้หรือความสามารถ ในเรื่องใด และใช้เป็นแนวทางในการดำเนินการให้ผู้เรียนบรรลุถึงความรู้และทักษะที่ต้องการนั้น (Popham, 2003) ในขณะที่การเรียนการสอนเป็นสิ่งผู้สอนจัดให้กับผู้เรียน เพื่อให้ผู้เรียนมีโอกาส บรรลุตามความรู้และทักษะที่คาดหวัง (Fuhrman, 2001) และการประเมินเป็นการให้สารสนเทศว่า ผู้เรียนสามารถบรรลุความรู้และทักษะที่คาดหวังได้มากน้อยเพียงไร (Biggs, 2003) ดังนั้น ความ สอดคล้องในแนวเดียวกัน (alignment) ระหว่าง 3 องค์ประกอบจึงเป็นเงื่อนไขของระบบการศึกษา แบบอิงมาตรฐาน (standard-based education) (Smith & O'Day, 1990; Fuhrman, 2001)

ความหมายของคำว่า “ความสอดคล้องในแนวเดียวกัน” มาจากคำภาษาอังกฤษ “alignment” ซึ่งตามพจนานุกรม Oxford ให้ความหมายของ alignment ไว้ว่า 2 ความหมาย คือ (1) การจัดวางให้อยู่ในแนวเส้นตรง หรือปรับให้อยู่ในตำแหน่งที่ถูกต้อง (arrangement in a straight line or in correct relative positions) และ (2) ตำแหน่งของความเห็นพ้องหรือเกี่ยวข้องสัมพันธ์ (a position of agreement or alliance)

ในทางการศึกษานักวิชาการได้ให้ความหมายของคำว่า “ความสอดคล้องในแนวเดียวกัน” ดังนี้ Webb (1997: 3) ให้นิยามว่าเป็น “ระดับความเห็นพ้องและเชื่อมโยงซึ่งกันและกันของ ผลการเรียนรู้ที่คาดหวัง (expectations) และการประเมิน (assessments) เพื่อให้แนวทางของระบบ ที่เกี่ยวข้องกับช่องการเรียนรู้ของผู้เรียน ในสิ่งที่คาดหวังว่าผู้เรียนควรรู้และสามารถทำได้” Bhola, Impara, และ Buckendahl (2003: 21) นิยามว่าเป็น “ระดับความสอดคล้องระหว่างมาตรฐาน การเรียนรู้ของรัฐในเนื้อหารายวิชากับเนื้อหาในการประเมินที่ใช้วัดผลสัมฤทธิ์ของผู้เรียนตาม

มาตรฐาน” ซึ่งคล้ายคลึงกับ La Marca (2001: 3) ที่ให้นิยามว่าเป็น “ระดับของความสอดคล้องระหว่างเนื้อหาในการทดสอบกับมวลเนื้อหาสาระที่กำหนดตามมาตรฐานการเรียนรู้ของรัฐ” และ Case, Jorgensen, และ Zucker (2004: 2) ที่ให้นิยามว่าเป็น “ระดับของการทำงานประสานกันขององค์ประกอบในระบบการศึกษา เช่น มาตรฐาน หลักสูตร การประเมิน และการเรียนการสอน เพื่อบรรลุเป้าหมายที่ต้องการ”

จากนิยามข้างต้นสรุปได้ว่า ความสอดคล้องในแนวเดียวกัน หมายถึง ระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างองค์ประกอบต่างๆ ในระบบการศึกษา อันได้แก่ ผลการเรียนรู้ที่คาดหวัง (เช่น มาตรฐาน วัตถุประสงค์การเรียนรู้) การเรียนการสอน และการประเมิน (เช่น เครื่องมือในการประเมิน) ทั้งนี้ การวัดความสอดคล้องในแนวเดียวกันจึงหมายถึง การวัดระดับความเชื่อมโยงและประสานความสัมพันธ์ หรือความสอดคล้องระหว่างองค์ประกอบต่างๆ ในระบบการศึกษา งานวิจัยความสอดคล้องในแนวเดียวกันที่พบเป็นการศึกษาความสอดคล้องขององค์ประกอบต่างๆ ดังนี้ (1) ความสอดคล้องระหว่างมาตรฐานกับการเรียนการสอน (2) ความสอดคล้องระหว่างมาตรฐานกับการประเมิน และ (3) ความสอดคล้องระหว่างการเรียนการสอนกับการประเมิน

### 3.3 ความสำคัญของการศึกษาความสอดคล้องในแนวเดียวกัน

การจัดการศึกษาอิงมาตรฐานให้ประสบความสำเร็จ สิ่งที่มีความจำเป็นและมีความสำคัญมากคือ การทำให้องค์ประกอบของระบบการศึกษาที่อิงมาตรฐานมีความสอดคล้องและเป็นไปในทิศทางเดียวกัน การที่องค์ประกอบต่างๆ ในระบบการศึกษามีความสอดคล้องกันจะสามารถช่วยในการปรับปรุงการเรียนรู้ของผู้เรียน (La Marca, Redfield, Winter, Bailey & Hansche, 2000; Anderson, 2002; Farenga, Joyce & Ness, 2002; Biggs, 2003) ช่วยให้ระบบการศึกษาเกิดประสิทธิผล (Webb, 1997) นอกจากนี้ ความสอดคล้องระหว่างองค์ประกอบต่างๆ ยังมีความสำคัญต่อการประเมินการปฏิรูปทางการศึกษา และยังให้สารสนเทศที่มีประโยชน์ต่อทั้งกับผู้เรียน ผู้ปกครอง สาธารณชน และผู้วางนโยบาย (Herman, Webb, & Stephen, 2007) ซึ่งหากองค์ประกอบของระบบการศึกษาไม่สอดคล้องกันก็จะทำให้ระบบการศึกษามีความบกพร่อง ส่งสัญญาณเกี่ยวกับการจัดการศึกษาที่ผิดพลาด จนทำให้การตัดสินใจดำเนินงานจัดการศึกษาไม่มีประสิทธิภาพ กล่าวโดยสรุป การวัดความสอดคล้องมีความสำคัญดังต่อไปนี้ (สังวรณ์ รัตกระโทก, 2555)

1. ทำให้ระบบการศึกษาทั้งหลายมุ่งไปสู่เป้าหมายเดียวกัน ผู้เกี่ยวข้องในแต่ละระบบมีความเข้าใจตรงกันสามารถขับเคลื่อนการจัดการศึกษาให้สอดคล้องกันจนบรรลุถึงเป้าหมายเดียวกัน
2. ช่วยลดการประเมินอื่นๆ ที่ซ้ำซ้อน
3. ช่วยในการวางแผนและปรับปรุงที่เกี่ยวกับการจัดการศึกษาให้มีความสอดคล้องกัน เช่น การพัฒนาครู การเลือกสื่อ และตำราเรียน



### 3.4 วิธีการศึกษาความสอดคล้องในแนวเดียวกัน

ปัจจุบันวิธีการศึกษาเพื่อวัดความสอดคล้องในแนวเดียวกันที่นิยมใช้มี 3 วิธี คือ (1) วิธีการของ Webb (Webb methodology) (2) วิธีการของ Porter หรือวิธีการสำรวจหลักสูตรที่นำไปใช้จริง (Survey of enacted curriculum (SEC) methodology) และ (3) วิธีการของบริษัท Achieve (Achieve methodology)

#### 3.4.1 วิธีการศึกษาความสอดคล้องในแนวเดียวกันของ Webb

Webb ได้พัฒนาวิธีการประเมินความสอดคล้องขึ้นในปี ค.ศ. 1997 และได้นำเสนอในที่ประชุมต่างๆ เช่น การประชุมประจำปีของสมาคมการวิจัยทางการศึกษาของสหรัฐอเมริกา รวมทั้งการจัดการอบรมให้ผู้ที่สนใจด้วย ดังนั้น วิธีการของ Webb จึงเป็นวิธีการที่นิยมใช้กันอย่างแพร่หลายในแวดวงของการวัดและประเมินผลทางการศึกษา Webb (1997, 1999) ได้พัฒนาวิธีการที่มีความครอบคลุมและซับซ้อนในการศึกษาระดับของความสอดคล้องในแนวเดียวกัน ระหว่างการประเมินกับมาตรฐาน โดยอาศัยผู้เชี่ยวชาญพิจารณาและลงรหัส (coding) การจับคู่ความสอดคล้องกัน (match) ระหว่างเนื้อหาในมาตรฐานกับเนื้อหาในเครื่องมือการประเมิน การตัดสินระดับความสอดคล้องในแนวเดียวกันจากเกณฑ์การประเมินความสอดคล้อง รายละเอียดของวิธีการศึกษาความสอดคล้องของ Webb จะแบ่งการนำเสนอใน 3 ประเด็นได้แก่ (1) เกณฑ์ประเมินความสอดคล้องในแนวเดียวกัน (2) ขั้นตอนการศึกษาความสอดคล้องในแนวเดียวกันของ Webb รายละเอียดของแต่ละประเด็นมีดังนี้

##### 3.4.1.1 เกณฑ์ประเมินความสอดคล้องในแนวเดียวกัน

มิติการมุ่งเน้นเนื้อหาของ Webb ประกอบด้วยการวิเคราะห์ความสอดคล้องใน 4 เกณฑ์ ได้แก่ ความสอดคล้องของเนื้อหา (categorical concurrence) ความสอดคล้องด้านความลึกของความรู้ (depth of knowledge consistency) ความสอดคล้องด้านขอบเขตของความรู้ (range of knowledge correspondence) และความสมดุลของเนื้อหา (balance of representation) (Webb, 1999, 2007) ความหมาย หลักการ และการพิจารณาตัดสินของแต่ละเกณฑ์ประเมินความสอดคล้องในแนวเดียวกัน (Webb, 1997, 1999, 2002, 2007) มีรายละเอียดต่อไปนี้

##### 1. ความสอดคล้องของเนื้อหา (Categorical Concurrence)

ความสอดคล้องของเนื้อหาเป็นการเปรียบเทียบความเหมือนกันระหว่างเนื้อหาของผลการเรียนรู้ที่คาดหวังที่แสดงในมาตรฐานกับเนื้อหาในการประเมิน ซึ่งถือเป็นเกณฑ์พื้นฐานและมีความสำคัญในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานกับการประเมิน เกณฑ์นี้มีความคล้ายคลึงกับความตรงตามเนื้อหา (content validity) เป็นอย่างมาก คล้ายกับการตรวจสอบความตรงโดยการเปรียบเทียบแผนผังข้อสอบ (test blueprint) แต่ความสอดคล้องของเนื้อหาเป็นการพิจารณาเนื้อหากว้างๆ เช่น ความรู้สึกเชิงตัวเลขและเรขาคณิต

การดำเนินการศึกษาความสอดคล้องของเนื้อหาเป็นการพิจารณาว่าข้อคำถาม/กิจกรรมในการประเมินสามารถวัดเนื้อหาจากแต่ละมาตรฐานหรือไม่ ทำโดยให้ผู้เชี่ยวชาญพิจารณาและระบุความสอดคล้องตรงกัน (map) ของเนื้อหาระหว่างข้อคำถาม/กิจกรรมในการประเมินกับจุดประสงค์ โดยข้อคำถาม/กิจกรรมที่ความสอดคล้องตรงกันกับ 1 จุดประสงค์ ถือเป็น 1 ฮิต ทั้งนี้ ใน 1 ข้อคำถาม/กิจกรรม อาจมีได้มากกว่า 1 ฮิต (เรียกว่า multiple hits) เนื่องจากมีความสอดคล้องกับจุดประสงค์มากกว่า 1 จุดประสงค์ ซึ่งพบว่าข้อคำถาม 1 ข้อ อาจมีจำนวนฮิต ได้ถึง 3 ฮิต โดยแต่ละฮิต อยู่บนจุดประสงค์ที่แตกต่างกัน

การตัดสินความสอดคล้องของเกณฑ์ประเมินความสอดคล้องของเนื้อหาพิจารณาได้จากค่าฮิตของแต่ละมาตรฐาน ทำได้โดยจากการคำนวณค่าเฉลี่ยของฮิตที่ได้จากการลงรหัส (code) เพื่อระบุความสอดคล้องตรงกันจากผู้เชี่ยวชาญแต่ละคน ในทุกจุดประสงค์ที่อยู่ภายใต้มาตรฐานนั้น ทั้งนี้ กำหนดว่า **“แต่ละมาตรฐานจะต้องมีจำนวนฮิต อย่างน้อย 6 ฮิต” กล่าวคือ “เครื่องมือในการประเมินแต่ละมาตรฐานควรมีข้อคำถามอย่างน้อย 6 ข้อ”** จึงจะถือว่าเป็นระดับที่ยอมรับได้ (Webb, 2007: 10)

## 2. ความสอดคล้องด้านความลึกของความรู้ (depth of knowledge consistency: DOK consistency)

Webb (1997, 1999, 2002, 2007) กล่าวว่า “มาตรฐานและการประเมินมีความสอดคล้องในแนวเดียวกันก็ต่อเมื่อผลที่ได้จากการประเมินผู้เรียนอยู่ในระดับความซับซ้อนทางปัญญา (cognitive demand) ระดับเดียวกับสิ่งที่คาดหวังว่าผู้เรียนจะต้องรู้และสามารถทำได้”

ระดับความลึกของความรู้ (depth of knowledge) คำนวณจากความถี่ของความสอดคล้องตรงกัน (hit) กับข้อคำถาม/กิจกรรม ของทุกจุดประสงค์ในแต่ละมาตรฐาน และพิจารณาตัดสินระดับความลึกของความรู้ด้วยร้อยละของข้อคำถาม/กิจกรรมที่มีระดับความลึก ที่อยู่ในระดับต่ำกว่า หรือระดับเดียวกัน หรือระดับที่สูงกว่าระดับความลึกในจุดประสงค์นั้น การตัดสินความสอดคล้องของเกณฑ์นี้ พิจารณาได้จาก **“มีข้อคำถาม/กิจกรรมอย่างน้อยร้อยละ 50 ของข้อคำถาม/กิจกรรมทั้งหมด ที่มีระดับความลึกอยู่ในระดับเดียวกันหรือสูงกว่าระดับความลึกในจุดประสงค์นั้น”** จึงจะถือว่าอยู่ในระดับที่สามารถยอมรับได้

การตีความและกำหนดระดับความลึกของความรู้ให้กับจุดประสงค์ภายใต้มาตรฐาน และข้อคำถามในการประเมินเป็นสิ่งสำคัญต่อการวิเคราะห์ความสอดคล้องในแนวเดียวกัน Webb (1999, 2007) ซึ่งแต่ละระดับมีรายละเอียดดังนี้

ระดับที่ 1 ระลึกได้ (recall) เป็นการระลึกสารสนเทศได้ เช่น ข้อเท็จจริง การนิยาม คำศัพท์ หรือขั้นตอนง่ายๆ และหรือการแสดงกระบวนการง่ายๆ หรือการประยุกต์ใช้สูตรง่ายๆ

ระดับที่ 2 ทักษะ/ความคิดรวบยอด (skill/concept) เป็นระดับการทำงานของกระบวนการประมวลผลทางสมอง (mental processing) ที่มากกว่าการตอบสนองตามปกติวิธีสัชข้อคำถามในการประเมินในระดับนี้ต้องการให้ผู้เรียนทำการตัดสินใจแก้ปัญหาตามวิธีการของตนเอง ในขณะที่ระดับที่ 1 เป็นการให้ผู้เรียนตอบโดยใช้การท่องจำ หรือทำตามขั้นตอนที่มีอยู่แล้ว

ระดับที่ 3 การคิดเชิงกลยุทธ์ (strategic thinking) เป็นการให้เหตุผล การวางแผน การใช้หลักฐาน และใช้ความคิดระดับสูงกว่าระดับที่ 2 การคิดในระดับที่ 3 นี้ ต้องการให้ผู้เรียนอธิบายความคิดของตนเองและต้องการหาข้อสรุปจากการอนุมาน ความซับซ้อนทางปัญญาในระดับที่ 3 มีความซับซ้อนและเป็นนามธรรม ความซับซ้อนไม่ได้มีผลมาจากข้อเท็จจริงที่มีหลายคำตอบที่มีความเป็นไปได้สำหรับทั้งระดับ 1 และ 2 เป็นงาน (task) ที่ต้องการมากกว่าการให้เหตุผล กิจกรรมที่มีมากกว่าหนึ่งคำตอบที่เป็นไปได้ และต้องการให้ผู้เรียนแสดงให้เห็นว่าคำตอบที่ให้มีความเป็นไปได้มากที่สุดในระดับ 3 กิจกรรมอื่นๆ

ระดับที่ 4 การขยายความคิด (extended thinking) เป็นการให้เหตุผลที่ซับซ้อน การวางแผน การพัฒนา และการคิดส่วนใหญ่เป็นลักษณะการคิดที่ต้องใช้เวลา แต่การคิดที่ต้องใช้เวลาไม่ใช่เงื่อนไขที่สำคัญถ้าการทำงานเป็นเพียงการทำงานซ้ำๆ และไม่เป็นการประยุกต์ใช้ความเข้าใจเชิงความคิดรวบยอด และใช้การคิดในระดับสูง

### 3. ความสอดคล้องด้านขอบเขตของความรู้ (range of knowledge correspondence)

หลักเกณฑ์ประเมินความสอดคล้องด้านขอบเขตของความรู้ (range of knowledge correspondence) ใช้ในการตัดสินว่าขอบเขตที่ครอบคลุมของความรู้ (span of knowledge) ที่คาดหวังของผู้เรียนตามมาตรฐาน มีความเหมือนหรือความสอดคล้องกับขอบเขตของความรู้ที่ผู้เรียนจำเป็นต้องใช้ในการตอบคำถามหรือทำกิจกรรมได้อย่างถูกต้องหรือไม่

เกณฑ์ของความสอดคล้องระหว่างขอบเขตของเนื้อหาระหว่างมาตรฐานกับการประเมิน พิจารณาได้จากจำนวนจุดประสงค์ภายในมาตรฐาน ที่มีความสัมพันธ์กับข้อคำถาม/กิจกรรมอย่างน้อย 1 ข้อ/กิจกรรม จำนวนจุดประสงค์ภายใต้มาตรฐานที่ผู้เชี่ยวชาญกำหนดว่ามี อย่างน้อย 1 อธิจะถูกหาค่าเฉลี่ยภายในกลุ่มผู้เชี่ยวชาญเช่นเดียวกับเกณฑ์อื่นๆ ซึ่งเป็นไปได้ว่าผู้เชี่ยวชาญอาจกำหนดอธิในจุดประสงค์ที่แตกต่าง

ระดับที่สามารถยอมรับได้การตัดสินความสอดคล้องของเกณฑ์นี้ พิจารณาได้จาก **“ในแต่ละมาตรฐานมีจุดประสงค์อย่างน้อยร้อยละ 50 ที่มีความสอดคล้องกับข้อคำถาม/กิจกรรมอย่างน้อย 1 ข้อ/กิจกรรม”**

#### 4. ความสมดุลของเนื้อหา (balance of representation)

มาตรฐานและการประเมินที่มีความสอดคล้องในแนวเดียวกันจะต้องมีการแจกแจงหรือการกระจายของเนื้อหาอย่างเท่าเทียมกันด้วย การคำนวณค่าดัชนีความสมดุลพิจารณาจากผลต่างระหว่างสัดส่วนของมาตรฐานต่อจำนวนรวมของจุดประสงค์ที่ถูกระบุว่าสอดคล้องกับข้อคำถาม (objectives hit) ในหนึ่งมาตรฐาน กับสัดส่วนของจำนวนข้อคำถามที่ถูกระบุว่าสอดคล้องกับจุดประสงค์ต่อจำนวนรวมของข้อคำถามที่ถูกระบุว่าสอดคล้องที่สอดคล้องกับจุดประสงค์ในมาตรฐานนั้น **กำหนดให้ค่าดัชนีตั้งแต่ 0.7 ขึ้นไป เป็นระดับที่สามารถยอมรับได้ของเกณฑ์นี้** เนื่องจากค่าดัชนีตั้งแต่ 0.7 ขึ้นไป แสดงให้เห็นว่าข้อคำถาม/กิจกรรมมีการแจกแจง ทั่วทุกจุดประสงค์อย่างน้อยในระดับหนึ่ง สูตรในการคำนวณค่าดัชนีความสมดุลเป็นดังนี้

$$\text{ดัชนีความสมดุล (BALANCE INDEX)} = 1 - \frac{\sum_{k=1}^0 \left| \frac{1}{O} - \frac{I(k)}{H} \right|}{2}$$

เมื่อ O = จำนวนรวมของจุดประสงค์ที่ถูกระบุว่าสอดคล้องกับข้อคำถาม (objectives hit) ในหนึ่งมาตรฐาน

$I(k)$  = จำนวนข้อคำถามที่ถูกระบุว่าสอดคล้องกับจุดประสงค์ (k) (items hit)

H = จำนวนรวมของข้อคำถามที่ถูกระบุว่าสอดคล้องกับจุดประสงค์ในหนึ่งมาตรฐาน

##### 3.4.1.2 ขั้นตอนการศึกษาความสอดคล้องในแนวเดียวกันของ Webb

วิธีการของ Webb แบ่งเป็น 2 ขั้นตอนหลัก ขั้นตอนแรกเป็นการให้ผู้เชี่ยวชาญลงรหัส (code) เพื่อระบุระดับความลึกของความรู้ ให้กับมาตรฐานต่างๆ เพื่อใช้เป็นข้อมูลในการเปรียบเทียบความสอดคล้องกับระดับความลึกของความรู้ในข้อคำถาม/กิจกรรมในการประเมินต่อไป ขั้นตอนถัดมา เป็นการให้ผู้เชี่ยวชาญลงรหัสเพื่อระบุระดับความลึกของความรู้ในข้อคำถาม/กิจกรรมในการประเมิน และตรวจสอบความสอดคล้องกับมาตรฐานหลักสูตร (curriculum standard) หรือจุดประสงค์ที่อยู่ภายใต้มาตรฐาน จากนั้นทำการตัดสินผลโดยเปรียบเทียบกับระดับที่สามารถยอมรับได้ที่กำหนดในแต่ละเกณฑ์การพิจารณาความสอดคล้อง 4 เกณฑ์

##### 3.4.2 วิธีการของ Porter หรือ วิธีการสำรวจหลักสูตรที่นำไปใช้จริง (survey of enacted curriculum: SEC)

วิธีการของ Porter พัฒนาขึ้นมาจากแนวคิดของการประเมินหลักสูตร โดยให้ความสนใจกับคำว่า เนื้อหา (content) ของหลักสูตรเป็นพิเศษ เพราะเนื้อหาเป็นสิ่งที่ผู้เรียนได้เรียนรู้จากหลักสูตร เนื้อหาจึงเป็นสิ่งที่ควรประเมินเป็นอันดับแรกในการประเมินหลักสูตร ซึ่ง Porter และ Smithson ได้นิยามความหมายของคำว่า เนื้อหา ว่าหมายถึง เรื่อง (topic) ที่ครูใช้สอน หรือเรื่องที่ใช้ทดสอบ หรือเรื่องที่มีอยู่ในมาตรฐาน ซึ่งหลักสูตรที่ปฏิบัติได้จริง (enacted curriculum) หมายถึง

เนื้อหาตามหลักสูตรที่ผู้เรียนได้เรียน Porter และ Smithson (2001: 2) กล่าวว่า การเรียนรู้ของผู้เรียนเกิดขึ้นในหลักสูตรประเภทนี้มากที่สุด ดังนั้น หลักสูตรประเภทนี้จึงมีความสำคัญมากที่สุด และการประเมินเนื้อหาที่ผู้เรียนได้เรียนรู้เป็นประเด็นสำคัญของหลักสูตรประเภทนี้ ตัวชี้วัดของการประเมินหลักสูตรนี้ ได้แก่ หัวข้อที่สอน ความซับซ้อนเนื้อหาและงานที่ครูให้ผู้เรียนทำ และรูปแบบการสอนของครู การประเมินนี้มุ่งเสนอผลการประเมินว่า ครูได้สอนเนื้อหาอะไรบ้าง เนื้อหาที่สอนแต่ละเรื่องใช้วิธีสอนแบบใด (เช่น การทดลอง การอธิบาย การให้ผู้เรียนทำใบงาน เป็นต้น) และเนื้อหาที่ครูสอนและงานที่ครูมอบหมายให้ผู้เรียนทำใช้ทักษะทางคิดมากหรือน้อย หรือมีความท้าทายในการคิดระดับใด การประเมินให้ลึกซึ้งแบบนี้จะมีประโยชน์ต่อการนำเสนอแนวทางในการกำหนดยุทธศาสตร์มาก เพราะจะช่วยให้ทราบว่าผู้เรียนที่มีผลสัมฤทธิ์สูงมาจากการสอนของครูแบบใด

Porter และ Smithson (2001) ได้พัฒนาวิธีการวัดความสอดคล้องขึ้นมาอีกวิธีหนึ่ง ถึงแม้ว่าการประเมินความสอดคล้องด้วยวิธีการของ Webb จะใช้กันอย่างแพร่หลาย แต่จะเห็นว่าเป็นวิธีการที่ค่อนข้างซับซ้อนและมีหลายขั้นตอน วิธีศึกษาความสอดคล้องที่พัฒนาโดย Porter และ Smithson (2001) เป็นการศึกษาระดับความสอดคล้องเชิงปริมาณได้ทั้ง 3 องค์ประกอบของระบบการศึกษา ซึ่งสามารถใช้ในการศึกษาความสอดคล้องระหว่างมาตรฐาน การทดสอบ ตำราเรียน และแม้แต่การสอนของครูในห้องเรียน (Porter, Smithson, Blank, & Zeidner, 2007: 27) ดังนั้นวิธีการนี้จึงสามารถช่วยให้ผู้ที่เกี่ยวข้องกับกระบวนการทางการศึกษาได้เห็นความเชื่อมโยงระหว่างสิ่งที่เป็นเป้าหมายกับสิ่งที่สอนในห้องเรียนและสิ่งที่ประเมิน และยังสามารถเปรียบเทียบระดับความสอดคล้องได้ทั้งระดับโรงเรียนและระดับรัฐ วิธีการนี้ต้องอาศัยการประเมินโดยผู้เชี่ยวชาญ การวัดความสอดคล้องจะเริ่มโดยการแบ่งประเภทของเนื้อหาออกเป็นเรื่องๆ (topic) เช่น พืชชนิด เรขาคณิต ฯลฯ ในแต่ละเรื่องจะทำการประเมินความลึกของความรู้ ซึ่งเป็นทักษะทางพุทธิปัญญาที่ต้องการให้ผู้เรียนบรรลุในระหว่างการเรียนการสอน เช่น การจำ ความเข้าใจ การนำไปใช้ เป็นต้น การวัดความสอดคล้องด้วยวิธีการของ Porter นี้จะนำผลการประเมินของผู้เชี่ยวชาญมาสร้างเมตริกซ์ 2 เมตริกซ์ เช่น เมตริกซ์มาตรฐานหรือเนื้อหาตามมาตรฐาน (เมตริกซ์ X) และเมตริกซ์ของเครื่องมือประเมิน (เมตริกซ์ Y) แล้วคำนวณค่าดัชนีความสอดคล้อง (alignment index) ซึ่งเป็นดัชนีสำหรับใช้ประเมินว่ามาตรฐานและเครื่องมือประเมินสอดคล้องกันระดับใด รายละเอียดของวิธีการของ Porter มีดังนี้

#### 3.4.2.1 มิติของการประเมินความสอดคล้อง

มิติของการประเมินความสอดคล้องของวิธีการนี้ มี 3 มิติ ได้แก่ ความสอดคล้องของเนื้อหา (content match) ความสามารถที่คาดหวังของผู้เรียน (expectations for student performance) และเนื้อหาของการสอน (instructional content)

### 1. ความสอดคล้องของเนื้อหา (content match)

วิธีการของ Porter จะสร้างเมทริกซ์ของเนื้อหาของรายวิชาขึ้นมา โดยแยกย่อยเนื้อหาทั้งหมดออกเป็นหลายๆ เรื่อง (topic) ตามระดับชั้นต่างๆ ซึ่งเมทริกซ์สามารถทำได้ 2 รูปแบบ คือ เมทริกซ์แบบละเอียด (fine grain level) จะแสดงหัวข้อเรื่องทั้งหมด และเมทริกซ์แบบหยาบ (coarse grain level) ที่จะแสดงหัวข้อเรื่องกว้าง วิธีการนี้ให้สารสนเทศที่คล้ายคลึงกับวิธีการตรวจสอบความตรงเชิงเนื้อหา แต่สารสนเทศที่ได้จะอยู่ในระดับที่น้อยกว่าและเป็นไปได้ที่จะก่อให้เกิดการเปลี่ยนแปลงทางด้านการสอนและหลักสูตรที่มากขึ้น (Porter & Smithson, 2002)

### 2. ความสามารถที่คาดหวังของผู้เรียน (expectations for student performance)

ข้อสอบ มาตรฐาน และการสอน ยังสามารถลงรหัสบนพื้นฐานของผลการเรียนรู้ที่คาดหวังของผู้เรียน การวัดนี้คล้ายคลึงกับเกณฑ์การวัดระดับความลึกของความรู้ตามแนวคิดของ Webb และการวัดจุดเน้นด้านความสามารถของบริษัท Achieve วิธีการของ Porter ระดับความสามารถที่คาดหวังของผู้เรียน หรือที่เรียกว่าความซับซ้อนทางปัญญา (cognitive demand) โดย Porter และ Smith (2002) ได้จำแนกเป็น 5 ระดับ บนพื้นฐานของการจำแนกพฤติกรรมการเรียนรู้ของ revised Bloom's taxonomy ซึ่งแต่ละระดับไม่ได้เป็นลำดับขั้น (Polikoff, Zhou, & Campbell, 2015: 3) ได้แก่ จำ (memorize) ปฏิบัติตามขั้นตอน (perform procedures) สื่อสารความเข้าใจ (communicate understanding) แก้ปัญหาในสถานการณ์ใหม่ (solve nonroutine problems) และ คาดคะเน/สรุปอ้างอิง/พิสูจน์ (conjecture/ generalize/ prove) (Porter, 2002)

### 3. เนื้อหาของการสอน (instructional content)

Porter และ Smithson (2002) ต้องการให้ความสำคัญขององค์ประกอบด้านเนื้อหาในการสอน เพราะการสอนทำหน้าที่เป็นตัวแปรแทรกแซง (intervening variable) ในการศึกษาผลสัมฤทธิ์ของผู้เรียน การเก็บข้อมูลเกี่ยวกับเนื้อหาที่สอน สามารถทำได้ 2 ประเด็น คือ การประเมินเนื้อหาที่ครูสอน (level of coverage) ซึ่งวัดจากจำนวนเวลาที่ครูใช้สอนแต่ละเนื้อหา ซึ่งอาจให้ครูประเมินการสอนของตนเองว่าได้ใช้เวลาในการสอนแต่ละเรื่องประมาณกี่คาบเรียน ส่วนประเด็นที่สอง คือ การเน้นด้านความลึกของความรู้ ซึ่งเป็นการวัดว่าครูได้เน้นให้ผู้เรียนใช้ความซับซ้อนทางปัญญาหรือทักษะทางการคิดในระดับใด

#### 3.4.2.2 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในวิธีการของ Porter ขึ้นอยู่กับองค์ประกอบที่ต้องการวิเคราะห์ความสอดคล้อง ในการอธิบายเนื้อหาของ **เครื่องมือประเมิน** และเนื้อหาของ **มาตรฐาน** ใช้กระบวนการวิเคราะห์เนื้อหา โดยให้ผู้เชี่ยวชาญด้านเนื้อหาพิจารณาเนื้อหาของเครื่องมือประเมินและเนื้อหามาตรฐาน และใช้กระบวนการที่เป็นระบบในการลงรหัส (code) ลงในตารางเนื้อหา 2 มิติ

หรือเมทริกซ์ ซึ่งหัวข้อเรื่องถูกจัดอยู่ในแนวแถว และระดับความสามารถที่คาดหวังของผู้เรียน หรือ ความซับซ้อนทางปัญญาจัดอยู่ในแนวคอลัมน์ ตัวอย่างดังภาพ 4 สำหรับเครื่องมือในการประเมิน หน่วยในการพิจารณา คือ ข้อคำถามในการประเมิน ส่วนมาตรฐานหรือกรอบหลักสูตร หน่วยในการพิจารณาจะถูกจัดระบบให้เป็นเนื้อหาย่อยที่เล็กที่สุด เช่น หัวข้อเรื่อง (topic) หรือหัวข้อเรื่องย่อย (subtopic) สำหรับเนื้อหาของการสอนของคุณ ใช้การสำรวจหลักสูตรที่ครูนำไปใช้จริง ซึ่งครูจะ รายงานเนื้อหาที่ครูสอนในห้องเรียน ซึ่งโดยทั่วไปครูจะรายงานข้อมูลนี้เมื่อสิ้นปีการศึกษา

ถัดมา ผลจากการเก็บข้อมูลเนื้อหาของหลักสูตรที่ถูกจัดให้อยู่ในรูปแบบเมทริกซ์ แต่ละช่องหรือเซลล์ (cell) ของเมทริกซ์จะถูกคำนวณเป็นสัดส่วนของเนื้อหาในช่องนั้นต่อเนื้อหา ทั้งหมดในเมทริกซ์ ผลจากการเก็บข้อมูลเนื้อหาการสอนของคุณ อธิบายได้ด้วยสัดส่วนของเวลาที่ใช้ ในการสอนเนื้อหาในช่องนั้นๆ ต่อเวลาที่ใช้สอนเนื้อหาทั้งหมด และผลจากการเก็บข้อมูลเนื้อหาของ เครื่องมือ อธิบายได้ด้วยสัดส่วนของข้อคำถามในช่องนั้นๆ กับข้อคำถามทั้งหมดในเมทริกซ์

จากที่ได้กล่าวข้างต้นว่าวิธีการนี้สามารถศึกษาสอดคล้องเชิงปริมาณได้ทั้งกับ มาตรฐาน การประเมิน และการเรียนการสอน ในที่นี้จะกล่าวถึงประเด็นสำคัญ 2 ประเด็น คือ การศึกษาความสอดคล้องของมาตรฐานกับเครื่องมือประเมิน กับการศึกษาความสอดคล้องของ มาตรฐานกับการเรียนการสอน ดังนี้

### 1. การศึกษาความสอดคล้องของมาตรฐานกับเครื่องมือประเมิน

การศึกษาความสอดคล้องของมาตรฐานกับเครื่องมือประเมิน มีขั้นตอนดังนี้

1) สร้างเมทริกซ์เนื้อหา (context matrix) 2 เมทริกซ์ ซึ่งประกอบด้วยเมทริกซ์ เนื้อหาของมาตรฐาน และเมทริกซ์เนื้อหาของเครื่องมือประเมิน ซึ่งการประเมินความสอดคล้องของ มาตรฐานกับเครื่องมือประเมินตามวิธีของ Porter (Porter, 2002) เป็นการเปรียบเทียบเมทริกซ์ มาตรฐาน และเมทริกซ์เครื่องมือประเมิน ซึ่งเรียกว่าข้อมูล X และข้อมูล Y ตามลำดับ ในการทำ เมทริกซ์ X และ Y นั้น มีหลักการว่า ในการประเมินความสอดคล้องแต่ละมาตรฐานจะกำหนดให้แถว ของเมทริกซ์ (แนวนอน) เป็นหัวข้อเรื่อง (topic) และกำหนดให้คอลัมน์ (แนวตั้ง) เป็นระดับ ความซับซ้อนทางปัญญา หรือความลึกของความรู้ ในวิธีการของ Webb นั้นเอง ตัวอย่างเมทริกซ์ มาตรฐานและเมทริกซ์เครื่องมือประเมินแสดงดังภาพ 1 โดยความซับซ้อนทางปัญญาในตัวอย่างนี้ Porter ประเมินความสอดคล้องวิชาคณิตศาสตร์ ใช้ระดับการวัด 5 ระดับ คือ จำ ปฏิบัติตามขั้นตอน สื่อสารความเข้าใจ แก้ปัญหาในสถานการณ์ใหม่ และคาดคะเน/สรุปอ้างอิง/พิสูจน์

ในการทำตาราง X และ Y ต้องทำแยกกัน และต้องใช้ผู้เชี่ยวชาญประมาณ 3-5 คน ผู้เชี่ยวชาญแต่ละคนจะต้องเติมตัวเลขลงไปบนช่องของตาราง X และ Y โดยแต่ละช่องในตาราง X คือความถี่ที่ผู้เชี่ยวชาญระบุว่าหัวข้อเรื่องหรือตัวชี้วัดนั้นวัดได้ระดับใด เช่น ผู้เชี่ยวชาญ คนที่ 1 พิจารณาว่า หัวข้อเรื่อง “ความรู้สึกเชิงตัวเลข/คุณสมบัติ/ความสัมพันธ์” อาจเป็นไปได้ที่จะวัดระดับ

จำ ปฏิบัติตามขั้นตอน และสื่อสารความเข้าใจ ก็จะใส่หมายเลข 1 ในช่องมาตรฐานที่ 1 (แถวที่ 1) ที่ตัดกับระดับจำ ปฏิบัติตามขั้นตอน และสื่อสารความเข้าใจ ทั้ง 3 ช่อง และเติมหมายเลข 0 ลงในช่องที่มาตรฐานนั้นวัดไม่ถึง ส่วนผู้เชี่ยวชาญคนอื่นๆ ก็ทำเช่นเดียวกันแล้วบันทึกข้อมูลความถี่ลงในเอกสารเมทริกซ์ที่แต่ละคนได้รับ

ส่วนการทำเมทริกซ์เครื่องมือประเมิน (Y) นั้น ทำโดยใช้กลุ่มผู้เชี่ยวชาญ โดยผู้เชี่ยวชาญแต่ละคนจะพิจารณาข้อสอบและทำการระบุว่าข้อสอบข้อหนึ่งๆ วัดความรู้ระดับใด ใน 5 ระดับ แล้วจึงเติมเลข 1 ลงในช่องที่คิดว่าข้อสอบตรงกับระดับการคิดนั้น Porter เสนอว่า ผู้เชี่ยวชาญอาจมีความเห็นว่าข้อสอบข้อหนึ่งๆ อาจวัดความรู้หลายระดับ ผู้เชี่ยวชาญจึงสามารถเติมตัวเลขในช่องได้มากที่สุด 3 ระดับ เช่น ตัวอย่างข้อสอบข้อที่ 1 มี 1 คะแนน ผู้เชี่ยวชาญคนที่ 1 คิดว่าข้อนี้มีความซับซ้อนและวัดความรู้ 3 ระดับ เขาจึงต้องแบ่งคะแนนเติมลงในช่องของ 3 ระดับนั้น ช่องละเท่าๆ กัน คือ  $1/3=0.33$  ข้อสอบบางข้ออาจมีคะแนนมากกว่า 1 คะแนน โดยเฉพาะข้อสอบภาคปฏิบัติ เช่น ตัวอย่างข้อสอบข้อที่ 2 มี 5 คะแนน ผู้เชี่ยวชาญอาจคิดว่าข้อสอบข้อนี้วัดความรู้ระดับได้ 3 ระดับ ดังนั้น จึงเติมตัวเลขลงที่เติมในแต่ละช่องของ 3 เป็นค่า  $5/3=1.67$  ผู้เชี่ยวชาญคนอื่นก็ดำเนินการเช่นเดียวกัน ทำการรวมคะแนนของแต่ละหัวข้อเรื่องในแต่ละระดับการคิด

2. นำผลรวมในเมทริกซ์มาตรฐาน และเมทริกซ์เครื่องมือประเมิน (บรรทัดสุดท้ายของทุกสาระที่ทำการประเมิน) ของผู้เชี่ยวชาญแต่ละคนรวมกัน การหาผลรวมนี้เพื่อให้ได้ภาพรวมการประเมินของผู้เชี่ยวชาญแต่ละคน ดังนั้น ผลการหาผลรวมจะได้เมทริกซ์ X และเมทริกซ์ Y ของผู้เชี่ยวชาญแต่ละคน แยกจากกัน

3. ทำค่าในเมทริกซ์มาตรฐานและเมทริกซ์เครื่องมือประเมินให้เป็นเมทริกซ์สัดส่วน การทำให้สมาชิกในเมทริกซ์ทั้งสองให้เป็นค่าสัดส่วนเป็นการทำให้ค่าในเมทริกซ์ X และ Y อยู่มาตรฐานเดียวกัน ซึ่งช่วยทำให้สามารถนำมาเปรียบเทียบกันได้ การทำค่าสมาชิกในเมทริกซ์ให้เป็นสัดส่วนทำได้โดยนำค่าผลรวมไปหารสมาชิกในเมทริกซ์เดียวกัน ผลของการหารสมาชิกทุกตัวของเมทริกซ์ด้วยผลรวม จะได้เมทริกซ์สัดส่วน และทำให้ผลรวมของค่าสัดส่วนในเมทริกซ์จะเท่ากับ 1.00 ส่วนการทำเมทริกซ์สัดส่วนของ Y ก็ทำในทำนองเดียวกัน



เมตริกซ์มาตรฐาน (เมตริกซ์ X)					
เรื่อง	ระดับความซับซ้อนทางปัญญา/ระดับการคิด				
	จำ	ปฏิบัติตามขั้นตอน	สื่อสารความเข้าใจ	แก้ปัญหาในสถานการณ์ใหม่	คาดคะเน/สรุปร่างอิงพิสูจน์/
ความรู้สึกเชิงตัวเลข/ ความสัมพันธ์/คุณสมบัติ					
การวัด					
การวิเคราะห์ข้อมูลสถิติ ,					
พีชคณิต					
เรขาคณิต					

เมตริกซ์เครื่องมือประเมิน (เมตริกซ์ Y)						
	ข้อ	จำ	ปฏิบัติตาม ขั้นตอน	สื่อสารความ เข้าใจ	แก้ปัญหาใน สถานการณ์ใหม่	คาดคะเนสรุป/ พิสูจน์/อ้างอิง
ความรู้สึกเชิงตัวเลข/ ความสัมพันธ์/คุณสมบัติ	1					
	2					
	3					
	รวม					
...	...					
...	...					
...	...					
เรขาคณิต	1					
	2					
	3					
	รวม					

ภาพ 1 ตัวอย่างเมตริกซ์มาตรฐาน และเมตริกซ์เครื่องมือประเมิน (Y) ของวิชาคณิตศาสตร์ในระดับประถมศึกษา

5. นำเมตริกซ์ X และ Y มาเปรียบเทียบกันและคำนวณดัชนีความสอดคล้อง โดยการหาผลต่างของเมตริกซ์ X และ Y หรือ X-Y ถ้ามาตรฐานและเครื่องมือประเมินว่าสอดคล้องกัน ผลลัพธ์ของผลต่างควรมีค่าเป็นศูนย์หรือเข้าใกล้ศูนย์ Porter เสนอสูตรการคำนวณความสอดคล้อง ดังนี้

$$\text{alignment index} = 1 - \frac{\sum |X_i - Y_i|}{2}$$

เมื่อ  $X_i$  คือ ค่าสมาชิกของเมตริกซ์ X ตำแหน่งที่ i

$Y_i$  คือ ค่าสมาชิกของเมตริกซ์ Y ตำแหน่งที่ i

ค่าความสอดคล้องมีค่าตั้งแต่ 0.0 ถึง 1.00 โดย 0.0 หมายถึง ไม่มีความสอดคล้อง และ 1.00 คือ มีความสอดคล้องอย่างสมบูรณ์ อย่างไรก็ตาม ยังไม่มีเกณฑ์ที่แน่ชัดว่า ค่าเท่าใดจึงจะสรุปว่ามีความสอดคล้อง การประเมินความสอดคล้องจึงต้องใช้การตัดสินใจของแต่ละบุคคลตัวอย่าง การคำนวณความสอดคล้องด้วยดัชนีความสอดคล้อง ยกตัวอย่างการหาค่าดัชนีความสอดคล้องจากข้อมูลสมมติในภาพ 2 ดังนี้

### ระดับความซับซ้อนทางปัญญาระดับการคิด/

เรื่อง	เครื่องมือประเมิน			มาตรฐาน		
	0.3	0	0.1	0.2	0	0.1
0	0.1	0	0	0	0.2	0
0	0.2	0.1	0.1	0.1	0.2	0.1
0	0.1	0.1	0	0	0	0.1

ซึ่งได้ผลลัพธ์เมตริกซ์ของผลต่างของ X และ Y ดังนี้

0.10	0.00	0.00
0.00	0.10	0.00
0.10	0.00	0.00
0.00	0.10	0.00

ผลรวมของผลต่างระหว่างเมตริกซ์ X กับ เมตริกซ์ Y เท่ากับ 0.4

เมื่อแทนค่าในสูตร  $1 - \frac{\sum |X_i - Y_i|}{2}$  จะได้  $1 - \frac{0.4}{2} = 0.8$

ค่าความสอดคล้องที่ได้มีค่าเท่ากับ 0.8 แสดงว่ามีความสอดคล้องของมาตรฐานและการประเมินค่อนข้างมาก จะเห็นว่าการหาความสอดคล้องมีหลายขั้นตอน การนำไปใช้ในการทำงาน อาจมีความซับซ้อนในระยะเริ่มแรกของการทำงาน ดังนั้น ก่อนที่มีการประเมินความสอดคล้องควรมีการขั้นตอนของการอบรมและชี้แจงกระบวนการทำงานให้ผู้เชี่ยวชาญทราบ และให้ผู้เชี่ยวชาญได้ทดลองทำพร้อมกับการวิพากษ์วิจารณ์ ให้ข้อมูลย้อนกลับแก่ผู้เชี่ยวชาญ ซึ่งจะทำให้ผู้เชี่ยวชาญคุ้นเคย และสามารถดำเนินงานให้ถูกต้องและมีความเหมาะสมได้ต่อไป

## 2. การประเมินความสอดคล้องของมาตรฐานกับการเรียนการสอน

วิธีการของ Porter มีข้อดีสองประการคือ ประการแรก สามารถรายงานความสอดคล้องออกมาเป็นเชิงปริมาณ เป็นค่าเดียว ที่ช่วยให้เข้าใจง่าย ประการที่สอง คือ วิธีการนี้เป็นวิธีการที่สามารถนำไปใช้ประยุกต์ใช้สำหรับการประเมินความสอดคล้องของสิ่งต่างๆ ได้หลากหลาย หนึ่งในวิธีการประยุกต์ใช้วิธีการ วิธีการของ Porter คือ การประเมินความสอดคล้องของมาตรฐานกับการเรียนการสอน ซึ่งเป็นการประเมินว่า การจัดการเรียนการสอนได้สอดคล้องกับมาตรฐานหรือไม่ ในการประเมินนี้จะเน้นเปรียบเทียบที่เนื้อหาที่ครูสอนกับเนื้อหาที่ปรากฏในมาตรฐาน

Porter (2002) เสนอว่า ในการวัดเนื้อหาที่ครูสอน สามารถทำได้สองประเด็น ประการแรก คือ การประเมินเนื้อหาที่ครูสอน (level of coverage) ซึ่งวัดจากจำนวนเวลาที่ครูใช้สอนแต่ละเนื้อหา ซึ่งอาจให้ครูประเมินการสอนของตนเองว่าได้ใช้เวลาในการสอนแต่ละเรื่องประมาณกี่คาบ ส่วนประเด็นที่สอง คือ การเน้นด้านความลึกของความรู้ ซึ่งเป็นการวัดว่าครูได้เน้นให้ผู้เรียนใช้ความซับซ้อนทางปัญญาหรือทักษะทางการคิดระดับใด เช่น การให้ครูตอบว่าได้เน้นให้ผู้เรียนใช้ทักษะการแก้ปัญหาในสถานการณ์ใหม่ในเรื่องเรขาคณิตมากเพียงใด โดยใช้รูปแบบการตอบแบบมาตรประมาณค่า หลังจากนั้นเราสามารถนำผลการตอบของครูมาสร้างเป็นเมทริกซ์ และนำไปเปรียบเทียบกับเมทริกซ์ X และคำนวณความสอดคล้องของมาตรฐานกับการเรียนการสอนด้วยสูตรดัชนีความสอดคล้องที่กล่าวมาแล้ว

### 3.4.3 วิธีการศึกษาความสอดคล้องในแนวเดียวกันของบริษัท Achieve

การศึกษาความสอดคล้องในแนวเดียวกันตามวิธีการของบริษัท Achieve เป็นการวัดความสอดคล้องระหว่างเนื้อหาของมาตรฐานกับเนื้อหาในการประเมิน ซึ่งให้ทั้งข้อมูลความสอดคล้องทั้งเชิงปริมาณและเชิงคุณภาพ (Rothman, Slattery, & Vranek, 2002) โดยวิธีการนี้ถูกพัฒนาขึ้นเพื่อเปรียบเทียบความสอดคล้องของมาตรฐานการศึกษาของรัฐต่างๆ หรือประเทศต่างๆ ผลของการเปรียบเทียบสามารถนำไปพัฒนาบุคลากรทางการศึกษา การปฏิรูปการศึกษา และการสร้างเครื่องมือประเมิน การนำเสนอการศึกษาความสอดคล้องวิธีการนี้แบ่งเป็น 2 ประเด็น ได้แก่ มิติในการประเมินความสอดคล้อง และขั้นตอนในการประเมินความสอดคล้อง รายละเอียดมีดังนี้

#### 3.4.3.1 มิติในการประเมินความสอดคล้อง

การวัดความสอดคล้องของบริษัท Achieve ได้วิเคราะห์ความสอดคล้องใน 4 มิติ ได้แก่ จุดเน้นของเนื้อหา (content centrality) จุดเน้นด้านความสามารถ (performance centrality) ความท้าทาย (challenge) ความสมดุลและความครอบคลุม (balance and range)

1. **จุดเน้นของเนื้อหา (content centrality)** เป็นเกณฑ์ที่ให้การวิเคราะห์เชิงลึกถึงความสอดคล้อง (match) ระหว่างเนื้อหาของข้อสอบแต่ละข้อกับเนื้อหาของมาตรฐาน โดย

การตรวจสอบระดับหรือคุณภาพของความสอดคล้อง ผู้เชี่ยวชาญจะเป็นผู้กำหนดระดับของความสอดคล้องของข้อสอบแต่ละข้อ โดยใช้แบบมาตรฐานประมาณ 5 ระดับ รายละเอียดจะกล่าวถึงต่อไปนี้ ไม่มีกำหนดระดับที่สามารถยอมรับได้ของเกณฑ์ ซึ่งแตกต่างกับเกณฑ์การประเมินความสอดคล้องของ Webb การให้คะแนนจะถูกทบทวนในภาพรวมด้วยผู้เชี่ยวชาญ และสรุปผลในรายงาน

**2. จุดเน้นด้านความสามารถ (performance centrality)** เป็นเกณฑ์ที่ทำให้ความสำคัญกับระดับของความสอดคล้องระหว่างรูปแบบของความสามารถ (ความซับซ้อนทางปัญญา) ของข้อสอบแต่ละข้อ กับรูปแบบของความสามารถที่กำหนดในมาตรฐาน ข้อสอบแต่ละข้อจะมีการกำหนดมีรูปแบบความซับซ้อนทางปัญญาที่แน่นอน (เช่น ต้องการให้คัดเลือก ระบุ เปรียบเทียบ หรือวิเคราะห์) ซึ่งผู้เชี่ยวชาญจะต้องระบุระดับความซับซ้อนทางปัญญาให้กับข้อสอบแต่ละข้อ โดยประเมินด้วยมาตรฐานค่า 4 ระดับ รายละเอียดจะกล่าวถึงต่อไป และไม่มีกำหนดระดับที่สามารถยอมรับได้ของเกณฑ์นี้เช่นกัน และผลของการลงรหัสของข้อสอบทุกข้อจะถูกทบทวนเพื่อตัดสินจุดเน้นด้านความสามารถในภาพรวม ซึ่งจะแสดงการสรุปผลเชิงปริมาณในรายงาน

**3. ความท้าทาย (challenge)** เป็นเกณฑ์ที่ใช้ในการพิจารณาตัดสินว่าชุดของข้อสอบมีความสามารถวัดความรู้ของผู้เรียนในเนื้อหาที่ต้องการได้ดีหรือไม่ ซึ่งผู้เชี่ยวชาญต้องพิจารณาใน 2 ปัจจัย ได้แก่ แหล่งความท้าทาย (source of challenge) และระดับความท้าทาย (level of challenge)

3.1 แหล่งความท้าทาย (source of challenge) เป็นการวัดความเชื่อมั่นว่าข้อสอบมีการสร้างอย่างความยุติธรรมและได้ออกแบบมาเพื่อลวงผู้เรียน และข้อสอบจะถูกตรวจสอบให้มั่นใจว่าไม่มีข้อผิดพลาดในเชิงเทคนิค (จากการตรวจสอบด้านเนื้อหาและผลจากการวิเคราะห์ข้อสอบ) เช่น ข้อสอบแต่ละข้อที่ยาก เพราะความรู้และทักษะที่ต้องการวัด หรือเป็นเพราะเหตุผลอื่นไม่เกี่ยวข้องกันกับเนื้อหาวิชา เช่น ขึ้นกับความไม่เสมอภาคกันของความรู้ภูมิหลังผู้เรียน ข้อสอบแต่ละข้อจะถูกประเมินและให้คะแนน โดย “1” หมายถึง แหล่งท้าทายมีความเหมาะสม และ “0” หมายถึง แหล่งท้าทายมีความเหมาะสม ถ้าข้อสอบได้คะแนนเป็น 0 แสดงว่าข้อสอบวัดจุดประสงค์นั้นไม่ได้

3.2 ระดับความท้าทาย (level of challenge) เป็นการเปรียบเทียบการให้ความสำคัญของความสามารถที่ต้องการของชุดข้อสอบกับการให้ความสำคัญของความสามารถที่ต้องการที่ระบุไว้ในมาตรฐาน ผู้เชี่ยวชาญจะประเมินความสอดคล้องโดยการพิจารณาตัดสินว่าชุดของข้อสอบมีความขอบเขตของความยากที่เหมาะสมกับผู้เรียนตามมาตรฐาน การประเมิน และสิ่งที่สนับสนุน (supporting material) ในระดับนั้นๆ หรือไม่ ซึ่งผู้เชี่ยวชาญจะเขียนสรุปพอสังเขปถึงระดับความท้าทายของข้อสอบแต่ละชุด

**4. ความสมดุลและความครอบคลุม (balance and range)** เป็นการพิจารณาความครอบคลุมของมาตรฐานด้วยของการให้ความสำคัญที่เท่าเทียมกันในมาตรฐาน การประเมินความสมดุลและความครอบคลุมจะให้สารสนเทศทั้งเชิงปริมาณและเชิงคุณภาพ โดย **ความสมดุล (balance)** เป็นการประเมินภาพรวมในระดับมาตรฐานมากกว่าระดับจุดประสงค์ โดยเป็นการเปรียบเทียบการให้ความสำคัญของเนื้อหาในชุดของข้อสอบ กับการให้ความสำคัญของเนื้อหาที่กำหนดในมาตรฐาน ผู้เชี่ยวชาญต้องพิจารณาเพื่อตัดสินว่าชุดของข้อสอบให้ความสำคัญกับเนื้อหาในระดับชั้นใดมากกว่าหรือไม่ โดยให้ผู้เชี่ยวชาญเขียนสรุปพอสังเขปเกี่ยวกับสมดุลเนื้อหาในแต่ละชุดข้อสอบ สำหรับ **ความครอบคลุม (range)** เป็นการวัดความครอบคลุมหรือหลากหลายของเนื้อหา ซึ่งวัดจากสัดส่วนของจุดประสงค์ในมาตรฐานที่ถูกวัดด้วยข้อสอบอย่างน้อย 1 ข้อ ซึ่งหากสัดส่วนมีค่าระหว่าง 0.50 – 0.66 ถือว่าสามารถยอมรับได้ และหากสูงกว่า 0.67 ถือว่ามีความครอบคลุมหรือหลากหลายที่ดี ซึ่งเกณฑ์นี้มีความคล้ายกับเกณฑ์ขอบเขตของความรู้ตามวิธีการของ Webb แม้ว่าอาจจะแตกต่างกันในเรื่องการกำหนดระดับที่สามารถยอมรับได้ ซึ่ง Webb กำหนดระดับได้ไว้ที่ร้อยละ 50

#### 3.4.3.2 ขั้นตอนในการศึกษาความสอดคล้องในแนวเดียวกันของบริษัท Achieve

Rothman, Slattery, & Vranek (2002) มีการดำเนินการศึกษาความสอดคล้องตามวิธีการของบริษัท Achieve โดยแบ่งเป็น 3 ขั้นตอนใหญ่ๆ ได้แก่ ขั้นตอนที่ 1 ตรวจสอบการจับคู่สอดคล้อง (match) ระหว่างการประเมินกับมาตรฐาน โดยวิเคราะห์เป็นรายข้อ (item-by-item) ขั้นตอนที่ 2 ตรวจสอบความท้าทายของแบบทดสอบ และขั้นตอนที่ 3 ตรวจสอบความสมดุลและความครอบคลุมของแบบทดสอบ

จากวิธีการศึกษาความสอดคล้องในแนวเดียวกันทั้ง 3 วิธี ได้แก่ วิธีการของ Webb วิธีการ ของ Porter และวิธีการของบริษัท Achieve จะพบว่าทุกวิธีต้องอาศัยผู้เชี่ยวชาญในการพิจารณาตัดสินความสอดคล้องขององค์ประกอบทางการศึกษา โดยองค์ประกอบในการประเมินความสอดคล้องตามวิธีการของ Webb และวิธีการของบริษัท Achieve คือ มาตรฐานและเครื่องมือในการประเมิน ซึ่งวิธีการของบริษัท Achieve ก็เป็นการพัฒนามาจากแนวคิดการศึกษาความสอดคล้องตามวิธีการของ Webb แต่วิธีการของบริษัท Achieve จะให้ข้อมูลเชิงคุณภาพที่ลึกและครอบคลุมเพิ่มขึ้น โดยเพิ่มเกณฑ์ที่ไม่มีในวิธีการของ Webb ได้แก่ เกณฑ์การประเมินแหล่งของความท้าทาย และระดับของความท้าทาย สำหรับวิธีการของ Porter สามารถศึกษาองค์ประกอบอื่นเพิ่มอีก ได้แก่ การเรียนการสอน ตลอดจนเนื้อหาของหลักสูตร และตำราต่างๆ วิธีการนี้อาจไม่ได้ศึกษาลงลึกในรายละเอียดเท่ากับอีกสองวิธี แต่ก็ เป็นวิธีการมีความยืดหยุ่นและสามารถคำนวณค่าดัชนีความสอดคล้องได้เป็นตัวเลขค่าเดียว ซึ่งสามารถเข้าใจได้ง่ายและสะดวกต่อการเปรียบเทียบ นอกจากนี้ จะเห็นว่าวิธีการของ Webb เพียงวิธีการเดียวที่มีการกำหนดระดับที่สามารถยอมรับได้

ของแต่ละเกณฑ์ประเมินความสอดคล้อง อย่างไรก็ตาม ทั้ง 3 วิธีการต้องอาศัยผู้เชี่ยวชาญในการให้คะแนนองค์ประกอบต่างๆ โดยคุณสมบัติของผู้เชี่ยวชาญจะต้องมีความคุ้นเคยกับมาตรฐานการประเมิน และการเรียนการสอน และมีความเชี่ยวชาญกับระดับความรู้และทักษะของผู้เรียนวิชา และระดับชั้นที่ศึกษา

จากการวิเคราะห์รายละเอียดวิธีการศึกษาความสอดคล้องในแนวเดียวกันทั้ง 3 วิธี จะเห็นว่า วิธีการของ Webb และวิธีการของบริษัท Achieve เป็นวิธีการที่ใช้ในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างเนื้อหาของมาตรฐานกับเนื้อหาของการประเมิน ในขณะที่วิธีการของ Porter สามารถใช้ในการศึกษาระดับความสอดคล้องเชิงปริมาณได้ทั้ง 3 องค์ประกอบของระบบการศึกษา ได้แก่ มาตรฐาน การประเมิน ตำราเรียน และแม้แต่การสอนของครูในห้องเรียน (Porter, Smithson, Blank, & Zeidner, 2007: 27) จึงเป็นวิธีการที่มีความยืดหยุ่นและเหมาะสมที่สุดในการนำมาประยุกต์ใช้ในงานวิจัยนี้ ที่มีวัตถุประสงค์หลักในการศึกษาความสอดคล้องในแนวเดียวกันระหว่าง 2 การประเมิน ได้แก่ การประเมินระดับชาติ (ข้อสอบ O-NET) และการประเมินระดับชั้นเรียน (ข้อสอบในการประเมินผลการเรียนรู้ระดับชั้นเรียนที่ครูผู้สอนสร้างขึ้น) โดยจะนำผลการประเมินของผู้เชี่ยวชาญมาสร้างเมทริกซ์ 2 เมทริกซ์ คือ เมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับห้องเรียน (Y) แล้วคำนวณค่าดัชนีความสอดคล้อง (alignment index) ซึ่งเป็นดัชนีสำหรับใช้ประเมินว่าการประเมินของ 2 ระดับดังกล่าว มีความสอดคล้องกันมากน้อยเพียงใด โดยการให้ค่าดัชนีความสอดคล้อง 1 ค่า จะทำให้ง่ายต่อการเข้าใจและเปรียบเทียบ

### 3.5 งานวิจัยที่เกี่ยวข้องกับการศึกษาความสอดคล้องในแนวเดียวกัน

Webb (1999) ได้พัฒนาการตรวจสอบความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานการเรียนรู้ของรัฐกับเครื่องมือที่ใช้การประเมินของรัฐ ในวิชาคณิตศาสตร์และวิทยาศาสตร์ โดยทำการศึกษาใน 4 รัฐ โดยพิจารณาความสอดคล้องจากเกณฑ์พื้นฐาน 4 เกณฑ์ของ Webb ผู้เชี่ยวชาญจำนวน 6 คน ทำการเปรียบเทียบความสอดคล้อง (match) ระหว่างข้อคำถามในการประเมินและมาตรฐาน และ/หรือจุดประสงค์ ผลการศึกษาแสดงให้เห็นถึงระดับความสอดคล้องที่หลากหลายตามระดับชั้นและรัฐต่างๆ เกณฑ์ที่พบความสอดคล้องกันมากที่สุดคือเกณฑ์ประเมินความสอดคล้องกัน อย่างไรก็ตาม มี 3 จาก 4 รัฐที่บรรลุตามเกณฑ์นี้ คือ ใน 1 มาตรฐาน จะถูกวัดด้วยข้อคำถามอย่างน้อย 6 ข้อ ส่วนอีกหนึ่งรัฐมีมาตรฐานจำนวนมากว่า 1 ใน 4 ของมาตรฐานทั้งหมดที่ถูกวัดด้วยข้อคำถามที่น้อยกว่า 6 ข้อ สำหรับเกณฑ์ความสมดุลของเนื้อหา พบว่ามาตรฐานที่ถูกประเมินมีข้อคำถามที่กระจายอยู่ตามจุดประสงค์ต่างๆ อย่างเท่าเทียมกัน สำหรับเกณฑ์ที่พบว่าค่อนข้างมีปัญหาของความสอดคล้องคือ เกณฑ์ประเมินความสอดคล้องด้านความลึกของความรู้ และเกณฑ์ประเมินความสอดคล้องด้านขอบเขตของความรู้ แสดงให้เห็นว่าข้อคำถามในการทดสอบ มีจุดมุ่งหมายทั่วไปที่

ต่ำกว่าระดับความรู้ที่กำหนดไว้ในมาตรฐาน ตลอดจนมีขอบเขตเนื้อหาไม่ครอบคลุมตามที่มาตรฐานกำหนด ซึ่งทำให้เห็นว่าการทดสอบมาตรฐาน (standardized testing) ไม่ได้ทดสอบความคิดที่ซับซ้อน และเป็นการทดสอบเพียงบางส่วนของเนื้อหาเท่านั้น ดังนั้น เพื่อแก้ปัญหาดังกล่าว รัฐจึงควรออกแบบการประเมินที่มีการระบุประเด็นของการประเมินที่ชัดเจน ในการศึกษาจะเห็นว่าแต่ละเกณฑ์จะวัดความสอดคล้องในแนวเดียวกันของแ่งมุมที่แตกต่างกัน

Blank, Porter, และ Smithson (2001) ได้ทำการศึกษาระดับความสอดคล้องในแนวเดียวกันระหว่างการสอนและการประเมินของรัฐ 6 รัฐ ด้วยวิธีการของ Porter โดยให้ผู้เชี่ยวชาญจำนวน 4 คน ประชุมกันเพื่อประเมินความสอดคล้องระหว่างกับข้อคำถามในการประเมินกับมาตรฐาน ใช้ระยะเวลา 2 วัน สำหรับข้อมูลเกี่ยวกับเนื้อหาของการสอน ผู้วิจัยได้ทำการเก็บข้อมูลโดยครูผู้สอนจำนวน 600 คน จากโรงเรียนใน 6 รัฐ จำนวน 200 โรงเรียน โดยใช้แบบสำรวจเกี่ยวกับเนื้อหาในการสอนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 (เกรด 8) ผลการศึกษาพบว่า ความสอดคล้องในแนวเดียวกันระหว่างการประเมินกับการสอนภายในรัฐมีความคล้ายคลึงกับความสอดคล้องในแนวเดียวกันระหว่างรัฐ นั่นคือ ดัชนีความสอดคล้องของแบบทดสอบกับมาตรฐานที่ได้จากการเปรียบเทียบระหว่างรัฐมีค่าใกล้เคียงกับค่าดัชนีความสอดคล้องที่ได้จากการเปรียบเทียบกันภายในรัฐ ซึ่งตามหลักการแล้ว ค่าดัชนีความสอดคล้องภายในรัฐควรจะมากกว่าดัชนีความสอดคล้องระหว่างรัฐ (เนื่องจากถือว่ามาตรฐานของรัฐมีความแตกต่างกันอย่างมีนัยสำคัญ) นอกจากนี้ ยังมีการศึกษาความสอดคล้องของการประเมินของรัฐกับการประเมินความก้าวหน้าทางการศึกษาแห่งชาติ (NAEP) ด้านคณิตศาสตร์และการอ่าน ระดับเกรด 8 พบว่า มีความสอดคล้องระหว่างการประเมินของรัฐกับการสอนภายในรัฐเพิ่มขึ้นเล็กน้อยเมื่อเทียบกับระหว่างการสอนภายในรัฐกับการประเมิน NAEP ค่าเฉลี่ยของดัชนีความสอดคล้องระหว่างการสอนภายในรัฐกับการประเมินของรัฐจาก 6 รัฐที่ศึกษา มีค่าอยู่ในช่วง 0.23 (วิทยาศาสตร์ ระดับเกรด 8) ถึง 0.42 (คณิตศาสตร์ ระดับเกรด 4) และค่าเฉลี่ยดัชนีความสอดคล้องระหว่างการสอนภายในรัฐกับการประเมิน NAEP มีค่าอยู่ในช่วง 0.14 (วิทยาศาสตร์ ระดับเกรด 8) ถึง 0.41 (คณิตศาสตร์ ระดับเกรด 4)

Rothman, Slattery, Vranek, และ Resnick (2002) ได้ประยุกต์ใช้วิธีการศึกษาความสอดคล้องของบริษัท Achieve เพื่อตรวจสอบการประเมินใน 5 รัฐ โดยเริ่มจากการฝึกอบรมผู้เชี่ยวชาญ โดยผู้เชี่ยวชาญประกอบด้วย ครูผู้สอน ผู้เชี่ยวชาญด้านหลักสูตร และผู้เชี่ยวชาญด้านเนื้อหาและการประเมิน ในการฝึกอบรมจะมีการคัดสรรข้อสอบอย่างพิถีพิถันมาใช้เพื่อแสดงให้เห็นถึงลักษณะของแต่ละมิติหรือเกณฑ์การประเมินความสอดคล้อง ผลการศึกษาพบว่า รัฐที่เขียนจุดประสงค์โดยใช้ถ้อยคำที่กว้าง จะมีคะแนนต่ำ เพราะยากที่จะพิจารณาตัดสินความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ได้อย่างถูกต้อง โดยภาพรวมพบว่า ข้อสอบมีความสอดคล้องที่ดีกับเนื้อหาและความสามารถที่ระบุในมาตรฐาน รัฐส่วนใหญ่ไม่ได้คำนึงถึงการตรวจสอบในเกณฑ์ที่

เกี่ยวกับแหล่ง ความท้าทาย (source of challenge) และทุกรัฐไม่มีการตรวจสอบเกณฑ์ ความครอบคลุมของเนื้อหา (range) ในมาตรฐานและจุดประสงค์ที่เพียงพอ รวมถึงความท้าทาย ส่วนใหญ่ของมาตรฐานและจุดประสงค์ก็ถูกละเลยในการตรวจสอบ สำหรับเกณฑ์ความสมดุล (balance) พบว่าชุดของข้อสอบมีการให้ความสำคัญกับจุดประสงค์ที่ไม่สำคัญมากเกินไป การศึกษานี้แสดงให้เห็นจุดอ่อนและจุดแข็งของวิธีการศึกษาความสอดคล้องของบริษัท Achieve ซึ่งเป็นวิธีที่ใช้เวลาและค่าใช้จ่ายค่อนข้างมาก แต่สามารถให้ข้อสรุปที่แสดงจุดแข็งและจุดอ่อนของระบบการประเมินใน แต่ละรัฐอย่างละเอียด

Webb (2002) ได้ทำการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและการประเมินวิชาคณิตศาสตร์ของ 3 รัฐ ผู้เชี่ยวชาญระดับชั้นละ 4 คน ทำหน้าที่วิเคราะห์ความสอดคล้องระหว่างมาตรฐานและการประเมิน โดยแบ่งเป็นระดับชั้น ได้แก่ ระดับประถมศึกษา (เกรด 4, 6) ระดับมัธยมศึกษา (เกรด 7, 8, 9, 11) ใช้โมเดลการศึกษาความสอดคล้องของ Webb (1997) ซึ่งพิจารณาความสอดคล้องตาม 4 เกณฑ์ และมีการตรวจสอบความเที่ยงในการประเมินของกลุ่มผู้เชี่ยวชาญด้วยการคำนวณค่าแอลฟา (alpha) หรือสหสัมพันธ์ภายในชั้น (intraclass correlation) ซึ่งพบว่าโดยส่วนใหญ่แต่ละกลุ่มผู้เชี่ยวชาญมีความสอดคล้องในการให้คะแนนอยู่ในระดับปานกลางถึงสูง ( $\alpha > 0.75$ ) ผลการศึกษาโดยภาพรวมพบว่า (1) ข้อสอบในการประเมินมีความสอดคล้องกับมากกว่า 1 จุดประสงค์ (2) เนื้อหาในการประเมินเป็นเพียงส่วนหนึ่งของเนื้อหาในชุดมาตรฐาน (3) ระดับความลึกของความรู้ (DOK) ของชุดข้อสอบในการประเมินมีความสอดคล้องอย่างชัดเจนกับระดับความลึกของความรู้ในเนื้อหาจุดประสงค์ที่มุ่งวัด แต่พบว่าไม่มีความสอดคล้องกันตามเกณฑ์ขอบเขตของความรู้ (ROK) (4) ข้อสอบในการประเมินวัดเพียงแค่บางส่วนของจุดประสงค์ที่อยู่ภายใต้ 1 มาตรฐาน (5) ข้อสอบในการประเมินส่วนใหญ่สอดคล้องกับเพียง 1 หรือ 2 จุดประสงค์ที่อยู่ภายใต้ 1 มาตรฐาน และมีข้อสอบในการประเมินเพียงแค่ 1 หรือ 2 ข้อ ที่สอดคล้องกับจุดประสงค์อื่น ๆ

Fonthal (2004) ได้ทำการพัฒนาระเบียบวิธีเชิงปริมาณของการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานการเรียนรู้กับการประเมินในการศึกษาระดับอุดมศึกษา การศึกษาค้นคว้านี้เลือกใช้เกณฑ์ในการประเมินความสอดคล้อง 3 เกณฑ์ ได้แก่ ความสอดคล้องด้านขอบเขตเนื้อหา (range) ความสอดคล้องด้านความลึกเนื้อหา (depth) และความสอดคล้องด้านความสมดุลเนื้อหา (balance) นอกจากนี้ยังมีการสร้างดัชนีความสอดคล้อง ซึ่งเป็นสมการทางคณิตศาสตร์ที่เป็น การรวมกันเชิงเส้นตรงของ 3 เกณฑ์ความสอดคล้องดังกล่าว โดยโปรแกรมซอฟต์แวร์ถูกพัฒนาเพื่อเป็นต้นแบบดัชนี (สูตร) ที่เป็นไปได้ของความสอดคล้องในแนวเดียวกันเพื่อใช้ในการวัดความสอดคล้องระหว่างชุดของผลความคาดหวังทางการศึกษาระดับอุดมศึกษา (มาตรฐานที่เสนอโดยสมาคมสถาบันอุดมศึกษาในสหรัฐอเมริกา) กับเครื่องมือการประเมินของรัฐที่คัดเลือกจากทั่วประเทศ โดยคัดเลือกแบบทดสอบของรัฐจำนวน 27 ชุด และวิเคราะห์เพื่อตัดสินความสอดคล้องกับผลการเรียนรู้ที่คาดหวัง



ในระดับอุดมศึกษา ในเนื้อหาวิชาคณิตศาสตร์และภาษาอังกฤษของแต่ละการทดสอบของรัฐ โดยทั่วไป ผู้เชี่ยวชาญ (4 - 6 คน) พบว่า ผลการเรียนรู้ที่คาดหวังวิชาภาษาอังกฤษในระดับมหาวิทยาลัยจะมีความซับซ้อนทางปัญญามากกว่าแบบทดสอบของรัฐ ผู้เชี่ยวชาญยังพบว่า แบบทดสอบวิชาคณิตศาสตร์ของรัฐส่วนใหญ่จะมีความซับซ้อนทางปัญญาพอๆ กับผลการเรียนรู้ที่คาดหวังวิชาคณิตศาสตร์ในระดับมหาวิทยาลัย หรือกล่าวอีกอย่างหนึ่งคือ มีความแตกต่างอย่างชัดเจนของดัชนีความสอดคล้องระหว่างทุกแบบทดสอบของรัฐกับผลการเรียนรู้ที่คาดหวังวิชาภาษาอังกฤษของมหาวิทยาลัย และมีความแตกต่างกันเพียงเล็กน้อยของค่าดัชนีความสอดคล้องระหว่างแบบทดสอบคณิตศาสตร์รัฐกับผลการเรียนรู้ที่คาดหวังคณิตศาสตร์ระดับมหาวิทยาลัย อย่างไรก็ตาม แบบทดสอบของรัฐส่วนใหญ่ได้ค่าดัชนีความสอดคล้องโดยรวมที่ต่ำกว่า 0.50 เล็กน้อย ทั้งในวิชาคณิตศาสตร์และภาษาอังกฤษ ซึ่งถือว่าอยู่ในระดับปานกลางถึงต่ำ ระดับของค่าดัชนีความสอดคล้องโดยรวมมีความสัมพันธ์กับคะแนนการประเมินความก้าวหน้าทางการศึกษาระดับชาติ (national assessment of educational progress: NEAP) ของรัฐ

Eckhout, Plake, Smith, และ Larsen (2007) ได้ทำการศึกษาเพื่อตรวจสอบวิธีการในการศึกษาความสอดคล้องในแนวเดียวกันของมาตรฐานการเรียนรู้ทางเลือก (alternative content standards) ในการศึกษาสำหรับเด็กที่มีความต้องการพิเศษ กับมาตรฐานการเรียนรู้ปกติหรือมาตรฐานการเรียนรู้แกนกลาง (regular content standards/ core content standards) ในด้านการอ่านและคณิตศาสตร์ ผู้เชี่ยวชาญเป็นครูด้านการศึกษาพิเศษจากโรงเรียนระดับประถมศึกษา และมัธยมศึกษา จำนวนทั้งหมด 38 คน แบ่งผู้เชี่ยวชาญเป็น 6 กลุ่มย่อย โดยเป็นกลุ่มผู้เชี่ยวชาญด้านการอ่านและกลุ่มผู้เชี่ยวชาญด้านคณิตศาสตร์อย่างละ 3 กลุ่ม แต่ละกลุ่มย่อยประกอบด้วยผู้เชี่ยวชาญจำนวน 6 - 7 คน แต่ละคนประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างเนื้อหาในมาตรฐานการเรียนรู้ทางเลือก กับเนื้อหาในมาตรฐานการเรียนรู้แกนกลาง โดยใช้มาตรฐานค่า 3 ระดับ จากนั้นผู้เชี่ยวชาญจะประชุมกลุ่มเพื่อหาฉันทามติของการประเมิน ผลการศึกษาพบว่า ทั้งในด้านการอ่านและด้านคณิตศาสตร์ ผู้เชี่ยวชาญทั้งหมดเห็นพ้องต้องกันว่ามาตรฐานทางเลือกมีความสอดคล้องกับมาตรฐานการเรียนรู้ และพบว่าผู้เชี่ยวชาญมีความเห็นพ้องว่าทักษะที่ระบุในมาตรฐานโดยส่วนใหญ่มีความสอดคล้องกับระดับความสามารถ (performance level) ที่มุ่งหวัง

Herman, Webb, และ Stephen (2007) ได้ทำการศึกษาเพื่อตรวจสอบผลกระทบของความเห็นพ้องระหว่างผู้ประเมิน (rater agreement) ในการพิจารณาตัดสินเกี่ยวกับความสอดคล้องในแนวเดียวกันระหว่างการทดสอบคณิตศาสตร์โรงเรียนมัธยมศึกษาตอนปลายของรัฐแคลิฟอร์เนีย (golden state examination in high school mathematics) และประกาศของมหาวิทยาลัยแคลิฟอร์เนียเกี่ยวกับความสามารถทางคณิตศาสตร์ที่คาดหวังในการเข้าเป็นนักศึกษาของวิทยาลัย ในการศึกษาครั้งนี้ ให้อาจารย์ของมหาวิทยาลัยแคลิฟอร์เนียและครูคณิตศาสตร์โรงเรียนมัธยมศึกษา

ตอนปลาย จำนวน 20 คน ทำการให้คะแนนข้อสอบคณิตศาสตร์ของการทดสอบของรัฐ (golden state examination) เกี่ยวกับสิ่งที่คาดหวังที่กำหนดในประกาศความสามารถของมหาวิทยาลัย การกำหนดลักษณะข้อสอบที่สัมพันธ์เนื้อหา ผลการศึกษาพบว่า มีความแปรปรวนอย่างมากในการตัดสินของผู้ประเมิน และภาพที่แตกต่างกันของผลความสอดคล้องขึ้นอยู่กับผลคะแนนจากผู้ประเมินแต่ละคน นอกจากนี้ ยังพบความแตกต่างเล็กน้อยระหว่างรูปแบบของผู้ประเมิน

Liu และ Fulmer (2007) ได้ทำการศึกษาความสอดคล้องในแนวเดียวกันระหว่างหลักสูตรแกนกลางของรัฐนิวยอร์ก (New York state core curricula) และการประเมินของรัฐนิวยอร์ก (New York State Regents exams) ในวิชาฟิสิกส์และเคมี ของเดือนมกราคมและมิถุนายน 2006 และศึกษาจากข้อสอบในเดือนมกราคม 2004 เพื่อตรวจสอบการเปลี่ยนแปลงของความสอดคล้องในแนวเดียวกันเมื่อเวลาผ่านไป ในการศึกษาความสอดคล้องในแนวเดียวกันครั้งนี้ใช้โมเดลที่พัฒนาโดย Porter ซึ่งทั้งหลักสูตรและการประเมิน (ข้อสอบ) ของรัฐนิวยอร์กจะถูกนำมาจัดทำตาราง 2 มิติ ที่ประกอบด้วย หัวข้อเรื่อง (topics) และความซับซ้อนทางปัญญา (cognitive demands) ซึ่งในที่นี้ จำแนกความซับซ้อนทางปัญญาตามแนวคิด revised Bloom's taxonomy โดยที่ค่าในเซลล์ (cell) ของตารางเป็นความสามารถหลักในหลักสูตรและประเด็นการสอบของข้อสอบ ผู้เชี่ยวชาญจำนวน 2 คน ทำหน้าที่ในการลงรหัสในแต่ละเซลล์ ผลการศึกษาพบว่า โดยภาพรวม หลักสูตรแกนกลางของรัฐนิวยอร์ก มีความสอดคล้องในแนวเดียวกันสูงกับการข้อสอบในการประเมินของรัฐนิวยอร์ก โดยวิชาฟิสิกส์มีค่าเฉลี่ยดัชนีความสอดคล้อง เท่ากับ 0.809 ในขณะที่วิชาเคมีมีค่าเฉลี่ยดัชนีความสอดคล้องที่ต่ำกว่า คือเท่ากับ 0.809 และเมื่อพิจารณาเสถียรภาพหรือความคงที่ของค่าความสอดคล้องในแนวเดียวกัน โดยภาพรวมพบว่าค่าเฉลี่ยความสอดคล้องในแนวเดียวกันค่อนข้างมีความคงที่ และเมื่อพิจารณาจำแนกตามรายวิชา พบว่า วิชาฟิสิกส์มีค่าเฉลี่ยความสอดคล้องในแนวเดียวกันที่สูงขึ้นอย่างต่อเนื่องมากกว่าวิชาเคมี นอกจากนี้ เมื่อพิจารณาความสอดคล้องของระดับความซับซ้อนทางปัญญา (cognitive levels) โดยภาพรวมพบว่า มีความแตกต่างกันอย่างมากของการให้ความสำคัญในแต่ละระดับความซับซ้อนทางปัญญาระหว่างหลักสูตรแกนกลางกับข้อสอบ โดยหลักสูตรและข้อสอบวิชาฟิสิกส์มีความซับซ้อนทางปัญญาที่แตกต่างกันในระดับเข้าใจ และประยุกต์ใช้ และหลักสูตรและข้อสอบวิชาเคมีมีความซับซ้อนทางปัญญาที่แตกต่างกันในระดับจำ เข้าใจ และประยุกต์ใช้

Porter, Smithson, Blank, และ Zeidner (2007) ได้ทำการศึกษาเชิงทดลองเพื่อเปรียบเทียบแผนที่เนื้อหาและดัชนีความสอดคล้องในการใช้โปรแกรมการพัฒนาความเป็นวิชาชีพ (professional development program) ของครุคณิตศาสตร์และครูวิทยาศาสตร์ระหว่างโรงเรียนที่เข้าร่วมโปรแกรมกับโรงเรียนที่ไม่ได้เข้าร่วม สุ่มตัวอย่างโรงเรียนจำนวน 55 โรงเรียน จาก 5 เขต แบ่งเป็นโรงเรียนกลุ่มทดลอง จำนวน 28 โรงเรียน และโรงเรียนกลุ่มควบคุม 27 โรงเรียน ตัวแทรกแซง (intervention) ในการศึกษาคือ การอบรมต้นแบบของการเป็นผู้ฝึกอบรมในลักษณะของสมาชิกทีม

ผู้นำของโรงเรียนที่ได้รับการพัฒนาวิชาชีพโดยตรง เพื่อครูผู้เข้าอบรมจะนำไปปฏิบัติเมื่อกลับไปทำงานกับครูคนอื่น ๆ ในโรงเรียน ตัวแปรอิสระคือ กลุ่มในการศึกษา (กลุ่มทดลองและกลุ่มควบคุม) ตัวแปรตามหลักในการศึกษา คือ ระดับความสอดคล้องในแนวเดียวกันระหว่างการสอนของครูกับเป้าหมายของการสอน อันได้แก่ การสอบวัดผลสัมฤทธิ์ของผู้เรียน และมาตรฐานการเรียนรู้ของรัฐ วิเคราะห์ข้อมูลโดยใช้โมเดลเชิงเส้นตรงแบบลดหลั่น (hierarchical linear modeling) โดยกลุ่มในการศึกษา (กลุ่มทดลองและกลุ่มควบคุม) และเขตเป็น fixed effects และครู (ระดับที่ 1) และโรงเรียน (ระดับที่ 2) เป็น random effects ผลการศึกษาพบว่า ดัชนีความสอดคล้องสามารถใช้เป็นตัวแปรตามในการศึกษานัยสำคัญของกลุ่มทดลองเมื่อเทียบกับกลุ่มควบคุม นอกจากนี้ ดัชนีความสอดคล้องที่เป็นข้อมูลเชิงปริมาณ มีประโยชน์ในการพิจารณาความสอดคล้องทั้งระหว่างครูรายบุคคล ระหว่างกลุ่มครู ระหว่างมาตรฐาน ระหว่างการทดสอบผลสัมฤทธิ์ของผู้เรียน ระหว่างเนื้อหาหลักสูตร และอื่นๆ ที่สามารถวิเคราะห์เนื้อหาได้ อีกทั้ง แผนที่ยังสามารถกำหนดและแสดงให้เห็นถึงภาพของเนื้อหาที่ครูให้ความสำคัญ

Näsström และ Henriksson (2008) ได้ทำวิเคราะห์กรอบแนวคิดหรือลำดับชั้นในการจำแนกความซับซ้อนทางปัญญาจาก 9 แนวคิด ได้แก่ Bloom's revised taxonomy, De Block, De Corte Guilford, Marzano, Merrill, PISA, Porter และ TIMSS แล้วทำการคัดเลือก 2 แนวคิดเพื่อใช้ในการศึกษาเปรียบเทียบความเหมาะสมของมิติหรือเกณฑ์ที่ใช้ในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างเกณฑ์การจำแนกของ Revised Bloom's taxonomy กับเกณฑ์การจำแนกตามวิธีการของ Porter รายวิชาเคมีของโรงเรียนมัธยมศึกษาตอนปลายในประเทศสวีเดน ใช้การเปรียบเทียบตามหลักการ 5 ข้อ ตามแนวคิดของ Hauenstein (1998) ได้แก่ ความสามารถในการประยุกต์ใช้ ความครอบคลุม ความชัดเจนใน การแบ่งกลุ่ม (categories) หรือระดับของเกณฑ์ การสอดคล้องตามหลักการของการเรียงลำดับ และคำศัพท์ที่ใช้ในการจัดกลุ่มและกลุ่มย่อยเป็นตัวแทนของการใช้ในสาขาวิชานี้ ผลการศึกษาพบว่า เกณฑ์การจำแนกของ revised Bloom's taxonomy มีความครอบคลุมและจำเพาะ ตลอดจนมีค่าความเที่ยงระหว่างผู้ประเมินที่สูงกว่าเกณฑ์ การจำแนกตามวิธีการของ Porter โดยใช้แนวคิด revised Bloom's taxonomy ให้ค่าสถิติแคปปาของฟลีส (Fleiss' kappa:  $K_f$ ) ในการประเมินมาตรฐานและข้อคำถามในการประเมิน เท่ากับ 0.46 และ 0.36 ตามลำดับ ซึ่งมีความมากกว่าที่ได้จากการใช้เกณฑ์การจำแนกตามวิธีการของ Porter ที่ให้ค่า  $K_f$  ในการประเมินมาตรฐานและข้อคำถามในการประเมิน เท่ากับ 0.07 และ 0.30 ตามลำดับ

Shelton และ Brown (2008) ได้ทำการศึกษาความสอดคล้องในแนวเดียวกันระหว่างเนื้อหาในมาตรฐานการสอบเข้าศึกษาต่อวิทยาลัยชุมชนของรัฐแคลิฟอร์เนีย (the California community college placement test) กับเนื้อหาในแบบทดสอบประเมินตามมาตรฐานรัฐแคลิฟอร์เนีย ระดับมัธยมศึกษาตอนปลาย (the high school level California standards tests) ซึ่งเป็นแบบทดสอบ

เมื่อเรียนจบรายวิชาของวิชาคณิตศาสตร์ทั่วไป พีชคณิต 1 และเรขาคณิต ในการประเมินความสอดคล้อง ใช้ผู้เชี่ยวชาญจำนวน 9 คน พิจารณาความสอดคล้องในแนวเดียวกันโดยใช้ 4 เกณฑ์ของ Webb โดยการประเมินด้านความลึกของความรู้ใช้แนวความคิดจำแนกของ Marzano's taxonomy ทั้งนี้ จากการตรวจสอบความเที่ยงหรือความสอดคล้องในการให้คะแนนของผู้เชี่ยวชาญโดยใช้สัมประสิทธิ์เชิงสรุบบ้างอิง พบว่า แบบสอบคณิตศาสตร์ทั่วไปมี ความสอดคล้องระหว่างผู้เชี่ยวชาญสูง ( $\rho^2_{\text{Rel}}=0.80$ ,  $\rho^2_{\text{Abs}}=0.79$ ) เช่นเดียวกับแบบสอบเรขาคณิต  $\rho^2_{\text{Rel}}=0.75$ ,  $\rho^2_{\text{Abs}}=0.74$ ) ในขณะที่แบบสอบพีชคณิต 1 มีความสอดคล้องระหว่างผู้เชี่ยวชาญที่ต่ำกว่า แบบสอบคณิตศาสตร์ทั่วไปและแบบสอบเรขาคณิต ( $\rho^2_{\text{Rel}}=0.60$ ,  $\rho^2_{\text{Abs}}=0.60$ ) ผลการศึกษาพบว่า โดยภาพรวม เนื้อหาของแบบทดสอบประเมินตามมาตรฐานรัฐแคลิฟอร์เนีย ระดับมัธยมศึกษาตอนปลาย ไม่ค่อยสอดคล้องกับเนื้อหาในการสอบเข้าศึกษาต่อในวิทยาลัยชุมชนของรัฐแคลิฟอร์เนีย ใน 4 เรื่อง ได้แก่ จำนวนเต็มและตรรกะฟังก์ชัน ตรีโกณมิติ และกราฟ นั่นคือ มีเฉพาะร้อยละ 75 ของวัตถุประสงค์หลักที่ใช้ในการตัดสินใจคัดเลือกเข้าศึกษาต่อที่มีเนื้อหาสำคัญในหลักสูตรโรงเรียนมัธยมศึกษา

Andrew และ Dawn (2010) ได้ทำการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานการเรียนรู้ระดับอนุบาลของรัฐอินเดียนา (Indiana's kindergarten content standards) กับข้อรายการในการประเมินจากเครื่องมือมาตรฐานของรัฐอินเดียนาสำหรับการรายงานทางเลือก (Indiana standards tool for alternate reporting: ISTAR) เป็นการศึกษาความสอดคล้องในแนวตั้ง (vertical alignment) โดยประยุกต์ใช้วิธีการศึกษาความสอดคล้องของ Webb เพื่อเปรียบเทียบตัวชี้วัดของหลักสูตรจากระดับที่สูงกว่า (ได้แก่ อนุบาล) กับข้อรายการในการประเมินที่เป็นมาตรฐานประมาณค่าของ ISTAR ในระดับพื้นฐานและเบื้องต้นที่ต่ำกว่า ผู้เชี่ยวชาญในการประเมินมีจำนวน 13 คน ผลการศึกษาพบว่า มีความไม่สอดคล้องกันในเกณฑ์ของ Webb ทั้ง 2 เกณฑ์ที่ใช้ในการศึกษานี้ ได้แก่ เกณฑ์ความลึกของความรู้ (DOK) และเกณฑ์ขอบเขตของความรู้ (ROK) และพบว่ามีความสอดคล้องอย่างมากของ ISTAR ระดับ F1 และ F2 ในมาตรฐานการศึกษาปฐมวัย

Blank และ Smith (2010) ได้ทำการศึกษาความสอดคล้องในแนวเดียวกันระหว่างชุดของข้อสอบการประเมินความก้าวหน้าทางการศึกษาระดับชาติ (NAEP) ในระดับเกรด 4 และ 8 วิชาวิทยาศาสตร์ ในปี ค. ศ. 2009 กับแผนผังวิเคราะห์ข้อสอบ (item specifications) ของ NAEP ในการศึกษาครั้งนี้ มีเกณฑ์ 3 รูปแบบที่ใช้ตัดสินความเพียงพอของผลความสอดคล้องในแนวเดียวกัน ได้แก่ (1) ค่าเฉลี่ยดัชนีความสอดคล้องแห่งชาติระหว่างการประเมินของรัฐกับมาตรฐานของรัฐ ซึ่งอิงจากผลที่ได้จากฐานข้อมูล SEC ([www.SEOnline.org](http://www.SEOnline.org)) (2) ดัชนีความสอดคล้องระหว่างเครื่องมือในการประเมินที่ได้รับการวิเคราะห์ก่อนหน้านี้โดยศูนย์สถิติทางการศึกษาแห่งชาติ (National Center for Education Statistics: NCES) และ (3) ระดับความเข้ม (threshold) ความสอดคล้องที่

คาดหวังของรัฐที่สอดคล้องกันในทุกวิชาและระดับชั้น ซึ่งขึ้นอยู่กับข้อมูลและประสบการณ์ก่อนหน้านี้ ในการวิเคราะห์ความสอดคล้องกันระหว่างรัฐ และประกอบเอกสารระดับชาติและระดับนานาชาติ เกณฑ์มาตรฐานที่กำหนดเป็นค่าพื้นฐาน (baseline) ของดัชนีความสอดคล้อง หรือ AI โดยภาพรวมมีค่าเท่ากับ 0.25 และกำหนดค่าของแต่ละเกณฑ์การประเมินย่อย อันได้แก่ ความสมดุลของเนื้อหา ความสอดคล้องของเนื้อหา และความซับซ้อนทางปัญญา มีค่าเท่ากับ 0.50 ผลการศึกษาความสอดคล้อง ระหว่างการประเมิน NAEP กับแผนผังข้อสอบวิชาวิทยาศาสตร์ระดับเกรด 4 และ 8 พบว่า มีความสอดคล้องกันดีระหว่างเครื่องมือการประเมิน NAEP กับแผนผังข้อสอบ (ค่าดัชนีความสอดคล้อง ระดับเกรด 4 มีค่าเท่ากับ 0.37 และระดับเกรด 8 มีค่าเท่ากับ 0.42) และเมื่อเปรียบเทียบค่าของ เกณฑ์ประเมินความสอดคล้องกับ baseline (AI=0.25) พบว่าทุกค่าอยู่ในเกณฑ์ที่ดี (เช่น NAEP คณิตศาสตร์ ระดับเกรด 4 มีค่า AI = 0.40, คณิตศาสตร์ ระดับเกรด 8 มีค่า AI เท่ากับ 0.36)

Lombardi, Sebuen, Conley, และ Snow (2010) ได้ใช้ G-theory ในการศึกษา ความเที่ยง ของผู้ประเมินความสอดคล้องในแนวเดียวกันระหว่างข้อสอบคัดเลือกเข้ามหาวิทยาลัย กับมาตรฐาน การเตรียมความพร้อมเข้ามหาวิทยาลัย ในวิชาคณิตศาสตร์และภาษาอังกฤษ โดยให้ผู้เชี่ยวชาญวิชาละ 6 คน ทำการประเมินความสอดคล้องในแนวเดียวกันผ่านทางระบบออนไลน์ ด้วยเครื่องมือ 2 ลักษณะ ได้แก่ (1) การประเมินระดับความซับซ้อนทางปัญญา และ (2) rigor rating ซึ่งเป็นการประเมินโดยใช้ ระดับที่คาดหวังในการเข้าเรียนมหาวิทยาลัยของผู้เรียนเป็นจุดอ้างอิง (1 = ต่ำกว่าระดับที่คาดหวัง, 2 = เท่ากับระดับที่คาดหวัง, 3 = สูงกว่าระดับที่คาดหวัง) กำหนดให้ข้อสอบ และมาตรฐานเป็นสิ่งที่ ถูกวัด และผู้เชี่ยวชาญ เป็นแหล่งความคลาดเคลื่อน ใช้รูปแบบการออกแบบการประเมินแบบ  $i \times r$  design คือ ผู้เชี่ยวชาญแต่ละคนประเมินข้อสอบทุกข้อ และรูปแบบ  $s \times r$  design คือ ผู้เชี่ยวชาญแต่ละคนประเมินทุกมาตรฐาน ผลการศึกษาพบว่า ข้อสอบและมาตรฐานวิชาคณิตศาสตร์มีผลต่อ ความเที่ยงในการประเมิน ประเภทของมาตรที่ใช้ในการประเมินข้อสอบมีผลต่อ ความเที่ยงใน การประเมิน โดยมาตรประเมินระดับความซับซ้อนทางปัญญาให้ค่าความเที่ยงสูงกว่ามาตรประเมิน ค่าแบบ rigor rating ทั้งวิชาคณิตศาสตร์และภาษาอังกฤษ และทั้งข้อสอบและมาตรฐาน และพบ อิทธิพลของปฏิสัมพันธ์ระหว่าง ผู้เชี่ยวชาญกับสิ่งที่ถูกประเมิน (ข้อสอบ หรือ มาตรฐาน) สำหรับ เครื่องมือที่เป็นมาตรประเมินระดับความซับซ้อนทางปัญญานอกจากนี้ ยังพบว่า ผู้เชี่ยวชาญจำนวน 6 คน เป็นจำนวนที่เพียงพอสำหรับการประเมินระดับความซับซ้อนทางปัญญาในวิชาคณิตศาสตร์ แต่ ไม่เพียงพอสำหรับการประเมินโดยใช้มาตรประเมินค่าแบบ rigor rating

Contio (2013) ได้ทำการวิเคราะห์ความสอดคล้องในแนวเดียวกันระหว่างเนื้อหามาตรฐาน การศึกษาวิทยาศาสตร์ระดับชาติ (national Science Education Standards: NSES) ของรัฐ นิวยอร์ก ในหลักสูตรแกนกลางด้านกายภาพและธรณีวิทยา กับการทดสอบของรัฐในหลักสูตรด้าน กายภาพและธรณีวิทยา (regent exam) มีผู้ประเมินความสอดคล้อง จำนวน 2 คน ซึ่งได้มี

การวิเคราะห์ความเที่ยงระหว่างผู้ประเมิน (inter-rater reliability) ด้วยดัชนีความเที่ยงแคปอาของ โคเฮน (Cohen's Kappa index of reliability) ผลการศึกษาพบว่า NSES มีร้อยละ 49 ที่ซ้อนทับ (overlap) กับหลักสูตรแกนกลาง และร้อยละ 2 ที่ซ้อนทับกับ Regent Exam หลักสูตรแกนกลาง และการสอบของรัฐ แสดงในรูปของเมทริกซ์ที่ประกอบด้วยตัวชี้วัดด้านความสามารถและระดับ ความซับซ้อนทางปัญญา เปรียบเทียบความสอดคล้องโดยใช้ค่าดัชนีความสอดคล้องของ Porter ผลการศึกษาพบว่า มีค่าดัชนีความสอดคล้องเท่ากับ 0.35 ซึ่งเป็นค่าน้อย เนื่องจากทำให้ ความสำคัญที่แตกต่างกันในระดับความซับซ้อนทางปัญญา โดยที่หลักสูตรแกนกลางมุ่งเน้นที่ทักษะ การวัดด้านความเข้าใจและประยุกต์ใช้ ในขณะที่การสอบของรัฐมุ่งเน้นการทักษะประยุกต์ใช้มากกว่า รองลงมาคือ ระดับความเข้าใจ และการจำ ผลการศึกษานี้เสนอให้เห็นว่า NSES ควรมีการปรับปรุง แก้ไข และหลักสูตรแกนกลางควรปรับปรุงให้มีปริมาณที่เน้นความสำคัญในทักษะความเข้าใจเป็นหลัก

Polikoff และ Porter (2014) ได้ทำการวิจัยเพื่อศึกษาบทบาทของความสอดคล้องใน แนวเดียวกันการเรียนการสอนในการทำนายตัวชี้วัดอื่นๆ ของการเรียนการสอนที่มีประสิทธิภาพ ที่กำหนดไว้โดยโมเดลมูลค่าเพิ่ม (value-add model: VAM) หรือการวัดแบบหลายองค์ประกอบ (multiple-measure composite) ตัวอย่างในการวิจัยเป็นครุจำนวน 327 คน จากการศึกษาขนาดใหญ่ของการวัดประสิทธิภาพการสอน (measures of effective teaching study: MET study) ผู้วิจัยได้สำรวจครูถึงความครอบคลุมในหัวข้อเรื่อง (topic) และระดับความซับซ้อนทางปัญญาที่ พึ่งประสงค์โดยใช้ การสำรวจหลักสูตรที่นำไปใช้จริง (surveys of enacted curriculum: SEC) งานวิจัยนี้ได้เปรียบเทียบรายงานการสำรวจด้วยการวิเคราะห์เนื้อหาของมาตรฐานเป้าหมายและ การประเมินเพื่อประเมินความสอดคล้อง และวิเคราะห์เนื้อหาโดยกลุ่มผู้เชี่ยวชาญจำนวน 3 คน จาก ผลการวิเคราะห์ความสัมพันธ์แบบ zero-order correlation และการวิเคราะห์ถดถอย พบว่า ดัชนี ความสอดคล้องมีความสัมพันธ์กับคะแนนโมเดลมูลค่าเพิ่มเพียงเล็กน้อย และเมื่อมีการควบคุม คุณภาพการสอน ดัชนีความสอดคล้องจะไม่มีความสัมพันธ์กับคะแนนโมเดลมูลค่าเพิ่ม และพบ อิทธิพลของปฏิสัมพันธ์ (interaction effects) เพียงเล็กน้อย โดยหนึ่งในความสัมพันธ์อย่างมี นัยสำคัญที่ไม่พบชี้ไปในทิศทางที่คาดว่าเป็นบวก (ความสัมพันธ์ของความสอดคล้องกับการสอนตาม โมเดลมูลค่าเพิ่มมีค่าเป็นบวกมากขึ้นเมื่อการสอนมีคุณภาพสูง) และพบว่า ความสอดคล้องของ การสอนไม่มีความสัมพันธ์กับองค์ประกอบในการวัดประสิทธิภาพของการเรียนการสอน

Anderson, Irvin, Alonzo, และ Tindal (2015) ทำการประเมินความสอดคล้องใน แนวเดียวกันระหว่างข้อสอบกับมาตรฐานผ่านระบบออนไลน์ โดยประยุกต์ใช้ MFRM เพื่อควบคุม อิทธิพลของผู้ประเมิน (rater effects) ข้อสอบที่ใช้เป็นข้อสอบคณิตศาสตร์สำหรับผู้เรียนระดับ 6 – 8 จำนวน 1,345 ข้อ ที่สุ่มมาจากคลังข้อสอบขนาดใหญ่ (ทั้งหมด 2,700 ข้อ) ผู้ประเมินเป็นครู คณิตศาสตร์จำนวน 25 คน ทำการประเมินว่าข้อสอบมีความสอดคล้องกับมาตรฐานที่กำหนดไว้

หรือไม่ ผ่านทางระบบออนไลน์ที่เรียกว่า Distributed Item Review (DIR) ข้อสอบแต่ละข้อจะถูกประเมินด้วยผู้ประเมินจำนวน 3 คน โดยผู้ประเมินแต่ละคนจะประเมินข้อสอบคนละประมาณ 270 ข้อ ด้วยมาตราประมาณค่าแบบเรียงอันดับ 4 อันดับ (4-point ordinal alignment scale) วิเคราะห์ข้อมูลโดยประยุกต์ใช้ MFRM ซึ่งมีการปรับแก้เล็กน้อยเพื่อให้เหมาะสมกับข้อมูลการศึกษา ความสอดคล้องในแนวเดียวกัน โดยที่ให้ข้อสอบ (items) เป็นสิ่งที่ถูกประเมินแทนที่จะเป็นผู้ถูกประเมิน (persons) และผู้ประเมิน (raters) จะเป็นสิ่งที่แสดงถึงคุณลักษณะแฝง (latent trait) ของข้อคำถาม เนื่องจากเงื่อนไขของการวัดเป็นการปรับเทียบให้อยู่บนมาตราเดียว โดยตำแหน่งของข้อคำถามอยู่บนเงื่อนไขของความเข้มงวดของผู้ประเมินให้คะแนนแต่ละคน ความสอดคล้องภายใน ผู้ประเมินถูกประเมินด้วยการทดสอบส่วนเหลือ (residual) ระหว่างค่าที่คาดหวังจากโมเดล (model-based expected value) และค่าที่สังเกตได้ (observed value) ซึ่งค่าที่สังเกตได้ขึ้นอยู่กับ (1) ผู้ประเมินให้คะแนนตัดสินข้อคำถามทั้งหมดที่เหลืออย่างไร (เช่น ความเข้มงวดของผู้ประเมิน) และ (2) ผู้ประเมินให้คะแนนทั้งหมดที่เหลือตัดสินข้อคำถามอย่างไร (เช่น ความยากในการเลือกข้อคำถามที่สอดคล้อง) ถ้าค่าที่คาดหวังมีค่าใกล้เคียงกับค่าที่สังเกตได้ ถือว่าผู้ประเมินสามารถประเมินได้สอดคล้องกัน ผลการวิจัยพบว่า มีความแตกต่างอย่างมีนัยสำคัญในการศึกษาความสอดคล้องระหว่างข้อสอบกับมาตรฐาน เมื่อมีการใช้สถิติในการควบคุมอิทธิพลของผู้ประเมิน

Shankararaman และ Gottipati (2014) ได้ทำการศึกษาความสอดคล้องในแนวเดียวกัน ระหว่างการประเมินกับสมรรถนะ (competencies) ตามหลักสูตรโปรแกรมการบริหารจัดการระบบสารสนเทศในระดับปริญญาตรี ใช้การศึกษาความสอดคล้องแบบอัตโนมัติซึ่งอยู่บนพื้นฐานของเทคนิคการสกัดวลีที่สำคัญ (keyphrase extraction technique) และโมเดลความสอดคล้องของ Porter วิเคราะห์ข้อมูลความสอดคล้องในแนวเดียวกันระหว่างสมรรถนะกับการประเมินด้วยดัชนีความสอดคล้องของ Porter นอกจากนี้ ยังมีการวิเคราะห์ค่าความไวของความสอดคล้อง (sensitivity alignment score) ซึ่งเป็นโมเดลการให้คะแนนความสอดคล้องแบบใหม่ ที่ใช้วัดสัดส่วนของผลบวกที่เกิดขึ้นจริง ซึ่งจะมีการระบุอย่างถูกต้อง เช่น ร้อยละของสมรรถนะทางพุทธิปัญญาที่ได้รับการประเมินโดยการประเมินในทุกหัวข้อเรื่อง ผลการศึกษาพบว่า ดัชนีความสอดคล้องของ Porter เท่ากับ 0.371 แสดงให้เห็นว่าสมรรถนะกับการประเมินมีความสอดคล้องกันต่ำ สำหรับค่าความไวของความสอดคล้อง มีค่าเท่ากับร้อยละ 90.8 แสดงว่าการประเมินมีความสอดคล้องกับสมรรถนะสูง และมีความไม่สอดคล้อง (misalignment) ร้อยละ 9.2

Polikoff (2015) ได้ทำการศึกษาความสอดคล้องของเนื้อหาของตำราเรียนคณิตศาสตร์ ระดับชั้นประถมศึกษาปีที่ 4 กับมาตรฐานแกนกลาง (common core standards) โดยมีวัตถุประสงค์เพื่อพิจารณาระดับของความสอดคล้องอย่างแท้จริงระหว่างมาตรฐานแกนกลางของรัฐ (common core state standard: CCSS) กับเอกสารตำราเรียน ในการศึกษาที่ใช้ข้อมูลจาก

การสำรวจหลักสูตรที่นำไปใช้จริง (SEC) ของวิชาคณิตศาสตร์ ที่ถูกจัดทำขึ้นโดย Porter ในปี 2002 ซึ่งข้อมูลที่ถูกนำมาใช้ประกอบด้วย 2 มิติ คือ มิติหัวข้อเรื่อง 183 หัวข้อ และมิติความซับซ้อนทางปัญหาที่อยู่บนพื้นฐานของแนวความคิดการจำแนกกระบวนการทางปัญหาแบบปรับใหม่ (modified Bloom's taxonomy) ที่ประกอบด้วย 5 ระดับ ดังนั้น ในการสำรวจ SEC ทางคณิตศาสตร์นี้มีทั้งหมด 915 เซลล์ข้อมูล (cell) วิเคราะห์เนื้อหาของทั้งมาตรฐานการเรียนรู้ และตำราเรียนคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 4 ที่ใช้ในโรงเรียน จำนวน 7 เล่ม ผู้เชี่ยวชาญจำนวน 3-4 คนลงรหัสความสอดคล้องในแนวเดียวกัน เพื่อนำมาคำนวณค่าดัชนีความสอดคล้องของ Porter (main alignment index) และดัชนีความสอดคล้องทางเลือก (alternative alignment index) ผลการศึกษาพบว่า เนื้อหาจำนวนมากที่ไม่สอดคล้องกัน (misalignment) ระหว่างตำราเรียนกับมาตรฐานแกนกลาง โดยเฉพาะอย่างยิ่ง ตำราเรียนมักให้ความสำคัญกับด้านกระบวนการและความจำมากกว่าด้านอื่นๆ

นราศักดิ์ ไชยเรือง (2557) ผลกระทบของความไม่สอดคล้องระหว่างมาตรฐานการประเมินกับการประเมินของครูต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 4 ในจังหวัดเชียงใหม่ โดยมีวัตถุประสงค์เพื่อประเมินความไม่สอดคล้องระหว่างการประเมินของครูคณิตศาสตร์กับมาตรฐานการประเมิน วิเคราะห์ผลกระทบของความไม่สอดคล้องระหว่างมาตรฐานการประเมินกับการประเมินของครูต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 4 และวิเคราะห์ปฏิสัมพันธ์ของความไม่สอดคล้องระหว่างมาตรฐานการประเมินกับการประเมินของครูกับปัจจัยของนักเรียนต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 4 กลุ่มตัวอย่าง ได้แก่ ครูคณิตศาสตร์จำนวน 33 คน และนักเรียนจำนวน 957 คน ได้จากการสุ่มแบบสองขั้นตอน เครื่องมือที่ใช้คือ แบบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์แบบสอบถามนักเรียนและแบบสอบถามครู ซึ่งใช้เก็บข้อมูลตัวแปรระดับนักเรียน 3 ตัวแปร ได้แก่ ความรู้เดิมของนักเรียน จำนวนชั่วโมงในการเรียนพิเศษ และเจตคติต่อวิชาคณิตศาสตร์ และตัวแปรระดับครู 4 ตัวแปร ได้แก่ การประเมินที่ครูใช้ วิทยฐานะ ประสบการณ์สอน เจตคติต่อการประเมินความไม่สอดคล้องระหว่างการประเมินของครูกับมาตรฐานการประเมิน คำนวณจากดัชนีความสอดคล้องของ Porter วิเคราะห์ข้อมูลโดยใช้การวิเคราะห์พหุระดับ ผลการศึกษาพบว่า ความไม่สอดคล้องระหว่างมาตรฐานการประเมินกับการประเมินของครูในภาพรวมมีค่าน้อย และความไม่สอดคล้องของการประเมินของครูมีผลกระทบทางลบต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และพบว่า ความไม่สอดคล้องของการประเมินของครูมีปฏิสัมพันธ์กับความรู้เดิมของนักเรียนต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05



กฤตวรรณ ทงมี (2557) ได้ทำการศึกษาการประเมินความสอดคล้องระหว่างมาตรฐานคุณวุฒิระดับอุดมศึกษาแห่งชาติกับการจัดการเรียนการสอนของภาควิชาศัลยกรรม คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร อาจารย์ศัลยแพทย์ทั่วไป จำนวน 16 คน และนิสิตแพทย์ชั้นปีที่ 4 ปีการศึกษา 2556 จำนวน 135 คน โดยเลือกมาแบบเจาะจง ทำหน้าที่ประเมินเมตริกซ์มาตรฐาน ในขณะเดียวกันผู้เชี่ยวชาญ จำนวน 5 คน ก็ทำหน้าที่ประเมินเมตริกซ์มาตรฐานตามกรอบคุณวุฒิมาตรฐาน เครื่องมือวิจัย คือ แบบประเมินความสอดคล้องระหว่างมาตรฐานคุณวุฒิระดับอุดมศึกษาแห่งชาติกับผลการเรียนรู้ของภาควิชาศัลยกรรม คณะแพทยศาสตร์ มหาวิทยาลัยนเรศวร วิเคราะห์ข้อมูลด้วยการคำนวณค่าดัชนีความสอดคล้องของ Porter ผลการศึกษาพบว่า ดัชนีความสอดคล้องระหว่างมาตรฐานคุณวุฒิระดับอุดมศึกษาแห่งชาติกับผลการเรียนรู้ของภาควิชาศัลยกรรม มีค่าเท่ากับ 0.70 แสดงว่ามีความสอดคล้องของมาตรฐานและผลการประเมินค่อนข้างมาก โดยมีความสอดคล้องสูงสุดในด้านคุณธรรม จริยธรรม รองลงมาคือ ด้านทักษะทางปัญญา และด้านทักษะทางคลินิก ตามลำดับ และด้านที่มีความสอดคล้องน้อยที่สุด คือ ด้านความรู้ และพบว่าดัชนีความสอดคล้องระหว่างผลการประเมินการเรียนการสอนของอาจารย์แพทย์และนิสิตแพทย์ ค่าเท่ากับ 0.75 แสดงว่าอาจารย์แพทย์และนิสิตแพทย์มีความเห็นสอดคล้องกันในการจัดการเรียนการสอนตามมาตรฐานคุณวุฒิระดับอุดมศึกษาแห่งชาติ โดยมีความสอดคล้องสูงสุด คือ ด้านทักษะทางปัญญา ด้านทักษะความรู้ ทักษะความสัมพันธ์ระหว่างบุคคลและความรับผิดชอบ และด้านทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร ตามลำดับ และด้านที่มีความสอดคล้องน้อยที่สุด คือ ด้านคุณธรรม จริยธรรมและ ด้านคลินิก ตามลำดับ

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการศึกษาความสอดคล้องแนวโน้มเดียวกันทั้งในด้านองค์ประกอบในการศึกษา ระดับชั้นที่ศึกษา รายวิชา จำนวนผู้เชี่ยวชาญ ตลอดจนเกณฑ์/ มิติ/ ดัชนีที่ใช้ในการประเมิน สามารถสรุปรายละเอียดได้ดังตาราง 4

ตาราง 4 งานวิจัยที่เกี่ยวข้องกับการศึกษาความสอดคล้องในแนวเดียวกัน

ที่	ผู้วิจัย	องค์ประกอบในการศึกษา	ระดับชั้น	รายวิชา	วิธีการศึกษา			จำนวนผู้เชี่ยวชาญ	เกณฑ์ดัชนีความสอดคล้อง / มิติ /					
					Webb	SEC	Achieve		ด้านเนื้อหา	ด้านความซับซ้อนทางปัญญา	ด้านการกระจาย	ด้านคุณภาพข้อสอบ	Porter's Al index	
1	Webb (1999)	มาตรฐาน, การประเมิน	ประถม ม.ต้น ม.ปลาย	คณิต วิทยาศาสตร์	✓			6	✓	✓ (Webb's taxonomy)	✓			
2	Blank, Porter, & Smithson (2001)	มาตรฐาน, การสอนการ, ประเมิน	ประถม ม.ต้น	คณิต วิทยาศาสตร์		✓		≥4	✓	✓ (Porter's taxonomy)				✓
3	Rothman, Slattery, Vranek, & Resnick (2002)	มาตรฐาน, การประเมิน	ประถม ม.ต้น ม.ปลาย	คณิต อังกฤษ			✓	ไม่ระบุ	✓	✓ -performance centrality -level of challenge	✓	✓	✓	
4	Webb (2002)	มาตรฐาน, การประเมิน	ประถม ม.ต้น ม.ปลาย	คณิต	✓			4	✓	✓ (Webb's taxonomy)	✓	✓		
5	Fonthal (2004)	มาตรฐาน, การประเมิน	อุดมศึกษา	คณิต อังกฤษ	✓			4-7		✓ (Marzano's taxonomy)	✓			
6	Eckhout, Plake, Smith, & Larsen (2007)	มาตรฐาน, มาตรฐาน	การศึกษา พิเศษ	การอ่าน คณิต				38	✓	✓ (ปรับจาก Webb's taxonomy)				
7	Herman, Webb, & Stephen (2007)	มาตรฐาน, การประเมิน	ม.ปลาย	คณิต	✓			20	✓	✓ (Webb's taxonomy)				
8	Liu & Fulmer (2008)	มาตรฐาน, การประเมิน	ม.ปลาย	วิทยาศาสตร์ (ฟิสิกส์, เคมี)		✓		2	✓	✓ (revised Bloom's taxonomy)				✓
9	Porter, Smithson, Blank, & Zeidner (2007)	มาตรฐาน, การประเมิน	ประถม ม.ต้น	คณิต วิทยาศาสตร์		✓		ไม่ระบุ	✓	✓ (Porter's taxonomy)				
10	Näsström & Henriksson (2008)	มาตรฐาน, การประเมิน	ม.ปลาย	วิทยาศาสตร์ (เคมี)				2	✓	✓ (revised Bloom's taxonomy)				
11	Shelton & Brown (2008)	มาตรฐาน, การประเมิน	อุดมศึกษา ม.ปลาย	คณิต	✓			9	✓	✓ (Marzano's taxonomy)	✓			
12	Andrew & Dawn (2010)	มาตรฐาน, การประเมิน	ปฐมวัย	-	✓			13	✓	✓ (ปรับจาก Webb's taxonomy)	✓			
13	Blank & Smith (2010)	มาตรฐาน, การประเมิน	ประถม ม.ต้น	วิทยาศาสตร์	✓	✓		3-5	✓	✓ (Porter's taxonomy)	✓			✓
14	Lombardi, L., Sebuen, M., Conley, D., & Snow, E. (2010)	มาตรฐาน, การประเมิน	ม.ปลาย	คณิต, ภาษาอังกฤษ	✓			6	✓	✓ (Marzano's taxonomy)	✓			
16	Contio (2013)	มาตรฐานของรัฐ, หลักสูตรแกนกลาง, การประเมิน	ม.ปลาย	วิทยาศาสตร์		✓		2	✓	✓ (revised Bloom's taxonomy)				✓

ตาราง 4 (ต่อ)

ที่	ผู้วิจัย	องค์ประกอบในการศึกษา	ระดับชั้น	รายวิชา	วิธีการศึกษา			จำนวนผู้ช่วยวิจัย	เกณฑ์ดัชนีความสอดคล้อง / มิติ /					
					Webb	SEC	Achieve		ด้านเนื้อหา	ด้านความซับซ้อนทางปัญญา	ด้านการกระจาย	ด้านคุณภาพข้อสอบ	Porter's Al index	
17	Anderson, Irvin, Alonzo, & Tindal (2015)	มาตรฐาน, การประเมิน	ม.ต้น, ม.ปลาย	คณิต				25 (3 คน/ ข้อสอบ1ข้อ)	✓					
18	Polikoff & Porter (2014)	มาตรฐาน, การประเมิน, การสอน	ม.ปลาย	คณิต อังกฤษ		✓		3	✓	✓ (Porter's taxonomy)			✓	
19	Shankararaman & Gottipati (2014)	สมรรถนะ, การประเมิน	อุดมศึกษา	การจัดการระบบสารสนเทศ		✓		-	✓	✓ (Bloom's taxonomy)			✓	
20	Polikoff (2015)	มาตรฐาน, ตำราเรียน	ประถม	คณิต		✓		3-4	✓	✓ (Porter's taxonomy)			✓	
21	นราศักดิ์ ไชยเรือง (2557)	มาตรฐาน, การประเมิน	ม.ปลาย	คณิต		✓		5	✓	✓ (revised Bloom's taxonomy)			✓	
22	กฤตวรรณ ทองมี (2557)	มาตรฐาน, การสอน	อุดมศึกษา	เกี่ยวกับ ศัลยกรรม		✓		16	✓	✓ (TQF)			✓	

จากการสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการศึกษาความสอดคล้องในแนวเดียวกันตั้งแต่ปี ค.ศ. 1999 จนถึงปัจจุบันสรุปได้ดังนี้

งานวิจัยส่วนใหญ่เป็นการศึกษาความสอดคล้องในแนวเดียวกันระหว่างเนื้อหาขององค์ประกอบด้านมาตรฐานหรือหลักสูตรกับองค์ประกอบด้านการประเมินหรือเครื่องมือในการประเมิน (Anderson, Irvin, Alonzo, & Tindal, 2015; Andrew & Dawn, 2010; Blank & Smith, 2010; Contio, 2013; Fonthal, 2004; Herman, Webb, & Stephen, 2007; Liu & Fulmer, 2007; Lombardi, Sebuon, Conley, & Snow, 2010; Näsström & Henriksson, 2008; Porter, Smithson, Blank, & Zeidner, 2007; Rothman, Slattery, Vranek, & Resnick, 2002; Shankararaman & Gottipati, 2014; Shelton & Brown, 2008; Webb, 1999, 2002; นราศักดิ์ ไชยเรือง, 2557) รองลงมาคือ การศึกษาความสอดคล้องในแนวเดียวกันระหว่างเนื้อหาของทั้ง 3 องค์ประกอบ ได้แก่ องค์ประกอบด้านมาตรฐาน องค์ประกอบด้านการสอน และองค์ประกอบด้านการประเมิน (Blank, Porter, & Smithson, 2001; Polikoff & Porter, 2014) ระหว่างองค์ประกอบด้านมาตรฐานกับองค์ประกอบด้านการสอน (กฤตวรรณ ทองมี, 2557) นอกจากนี้ ยังพบว่ามีการศึกษาความสอดคล้องในแนวเดียวกันระหว่างเนื้อหาของ 2 องค์ประกอบด้านมาตรฐาน (Eckhout, Plake, Smith, & Larsen, 2007) และระหว่างเนื้อหาของมาตรฐานกับตำราเรียน (Polikoff, 2015)

งานวิจัยส่วนใหญ่เป็นการศึกษาความสอดคล้องในแนวเดียวกันขององค์ประกอบทางการศึกษา ในระดับประถมศึกษา (Anderson, Irvin, Alonzo, & Tindal, 2015; Blank, Porter, & Smithson, 2001; Blank & Smith, 2010; Porter, Smithson, Blank, & Zeidner, 2007; Rothman, Slattery, Vranek, & Resnick, 2002) รองลงมาคือ ระดับมัธยมศึกษาตอนต้น (Anderson, Irvin, Alonzo, & Tindal, 2015; Herman, Webb, & Stephen, 2007; Polikoff & Porter, 2014; Rothman, Slattery, Vranek, & Resnick, 2002; Webb, 1999, 2002) ระดับมัธยมศึกษาตอนปลาย (Cimetta, Falco, Smith, VanWinkle, & Powers, 2008; Contio, 2013; Liu & Fulmer, 2007; Lombardi, Sebuen, Conley, & Snow, 2010; Näsström & Henriksson, 2008; Webb, 2002; นราศักดิ์ ไชยเรือง, 2557) ระดับอุดมศึกษา (Fonthal, 2004; Shelton & Brown, 2008; Shankaraman & Gottipati, 2014; กฤตวรรณ ทองมี, 2557) ระดับปฐมวัย (Andrew & Dawn, 2010) นอกจากนี้ ยังพบว่ามีงานวิจัยความสอดคล้องในแนวเดียวกัน ด้านการศึกษาพิเศษ (Eckhout, Plake, Smith, & Larsen, 2007)

งานวิจัยส่วนใหญ่เป็นการศึกษาในวิชาคณิตศาสตร์ (Anderson, Irvin, Alonzo, & Tindal, 2015; Eckhout, Plake, Smith, & Larsen, 2007; Fonthal, 2004; Herman, Webb, & Stephen, 2007; Lombardi, Sebuen, Conley, & Snow, 2010; Porter, Smithson, Blank, & Zeidner, 2007; Polikoff & Porter, 2014; Polikoff, 2015; Rothman, Slattery, Vranek, & Resnick, 2002; Shelton & Brown, 2008; Webb, 1999, 2002) รองลงมาคือ วิชาวิทยาศาสตร์ (Blank, Porter, & Smithson, 2001; Blank & Smith, 2010; Contio, 2013; Liu & Fulmer, 2007; Näsström & Henriksson, 2008; Porter, Smithson, Blank, & Zeidner, 2007; Webb, 1999) วิชาภาษาอังกฤษ (Rothman, Slattery, Vranek, & Resnick, 2002; Fonthal, 2004; Polikoff & Porter, 2014; Lombardi, Sebuen, Conley, & Snow, 2010) การอ่าน (Eckhout, Plake, Smith, & Larsen, 2007) และวิชาเกี่ยวกับศัลยกรรม (กฤตวรรณ ทองมี, 2557)

งานวิจัยส่วนใหญ่ใช้วิธีการของ Porter (Blank, Porter, & Smithson, 2001; Blank & Smith, 2010; Contio, 2013; Liu & Fulmer, 2007; Polikoff, 2015; Polikoff & Porter, 2014; Porter, Smithson, Blank, & Zeidner, 2007; Shankaraman & Gottipati, 2014) รองลงมาคือ วิธีการของ Webb (Andrew & Dawn, 2010; Blank & Smith, 2010; Fonthal, 2004; Herman, Webb, & Stephen, 2007; Lombardi, Sebuen, Conley, & Snow, 2010; Shelton & Brown, 2008; Webb, 1999, 2002) โดยจะสังเกตเห็นว่าในช่วงแรก (ค.ศ.1999-2010) จะพบว่ามีการใช้วิธีการของ Webb มากกว่าวิธีการ SEC แต่ในช่วงหลัง (ค.ศ. 2007 จนถึงปัจจุบัน) มีงานวิจัยที่ใช้วิธีการ SEC จำนวนเพิ่มขึ้นมากกว่าวิธีการของ Webb และพบว่ามีการใช้วิธีการของบริษัท Achieve จำนวนน้อย (Rothman, Slattery, Vranek, & Resnick, 2002)

งานวิจัยการศึกษาความสอดคล้องทั้งสามวิธีต้องใช้ผู้เชี่ยวชาญในการพิจารณา ทั้งนี้ จำนวนผู้เชี่ยวชาญมีจำนวนระหว่าง 2 - 49 คน โดยส่วนใหญ่เป็นงานวิจัยที่มีจำนวนผู้เชี่ยวชาญระหว่าง 2 - 10 คน (Blank & Smith, 2010; Blank, Porter, & Smithson, 2001; Contio, 2013; Fonthal, 2004; Liu & Fulmer, 2007; Lombardi, Sebuen, Conley, & Snow, 2010; Näsström & Henriksson, 2008; Polikoff & Porter, 2014; Polikoff, 2015; Shelton & Brown, 2008; Webb, 1999, 2002) รองลงมาคือ จำนวน 11 - 20 คน (Andrew & Dawn, 2010; Herman, Webb, & Stephen, 2007) จำนวน 21 - 30 คน (Anderson, Irvin, Alonzo, Tindal, & 2015) และจำนวน 31 - 40 คน (Eckhout, Plake, Smith, & Larsen, 2007)

เมื่อพิจารณาเกณฑ์หรือมิติในการประเมินความสอดคล้องในแนวเดียวกัน ซึ่งสรุปได้เป็น 5 เกณฑ์/มิติ ได้แก่ มิติด้านเนื้อหา (content) มิติด้านความซับซ้อนทางปัญญาของเนื้อหา (cognitive complexity) หรือที่เรียกว่าความลึกของความรู้ (depth of knowledge) มิติด้านการกระจาย (distribution) มิติด้านคุณภาพข้อสอบ (item quality) และดัชนีความสอดคล้อง (alignment index) พบว่าแต่ละงานวิจัยมีการศึกษาความสอดคล้องในมิติต่างๆ ที่หลากหลาย ทั้งนี้ขึ้นอยู่กับโมเดลหรือวิธีการที่เลือกใช้ในการศึกษา แต่ทั้งนี้ เกือบทุกงานวิจัยจะเป็นการศึกษาความสอดคล้องในมิติด้านเนื้อหาควบคู่กับมิติด้านความซับซ้อนทางปัญญาหรือความลึกของความรู้ (Andrew & Dawn, 2010; Blank, Porter, & Smithson, 2001; Blank & Smith, 2010; Contio, 2013; Eckhout, Plake, Smith, & Larsen, 2007; Shankararaman & Gottipati, 2014; Herman, Webb, & Stephen, 2007; Liu & Fulmer, 2008; Lombardi, Sebuen, Conley, & Snow, 2010; Näsström & Henriksson, 2008; Porter, Smithson, Blank, & Zeidner, 2007; Polikoff & Porter, 2014; Polikoff, 2015; Rothman, Slattery, Vranek, & Resnick, 2002; Shelton & Brown, 2008; Webb, 1999, 2002; นราศักดิ์ ไชยเรือง, 2557; กฤตวรรณ ทองมี, 2557) ทั้งนี้ มีบางงานวิจัยที่เป็นการศึกษาความสอดคล้องเฉพาะมิติใดมิติหนึ่งในสองมิตินี้ ได้แก่ เฉพาะมิติด้านเนื้อหา (Anderson, Irvin, Alonzo, & Tindal, 2015) หรือเฉพาะมิติด้านความซับซ้อนทางปัญญา (Fonthal, 2004) มิติรองลงมาที่พบว่ามีการใช้ในการศึกษาความสอดคล้องคือ มิติด้านการกระจาย (Andrew & Dawn, 2010; Blank & Smith, 2010; Fonthal, 2004; Rothman, Slattery, Vranek, & Resnick, 2002; Shelton & Brown, 2008; Webb, 1999, 2002) และมีบางงานวิจัยที่มีการศึกษาความสอดคล้องในมิติด้านคุณภาพของข้อสอบ (Rothman, Slattery, Vranek, & Resnick, 2002; Webb, 2002) นอกจากนี้ ยังพบว่าม้งานวิจัยที่มีการคำนวณค่าดัชนีความสอดคล้อง (Porter's alignment index) ซึ่งงานวิจัยเหล่านี้ ทั้งหมดเป็นการศึกษาความสอดคล้องในแนวเดียวกันโดยใช้วิธีการ SEC ของ Porter (Blank, Porter, & Smithson, 2001; Blank & Smith,

2010; Contio, 2013; Liu & Fulmer, 2008; Polikoff, 2015; Polikoff & Porter, 2014; Shankararaman & Gottipati, 2014; นราศักดิ์ ไชยเรือง, 2557; กฤตวรรณ ทองมี, 2557)

เมื่อพิจารณาเฉพาะมิติด้านความซับซ้อนทางปัญญาหรือความลึกของความรู้ พบว่างานวิจัยต่างๆ มีการจำแนกความซับซ้อนทางปัญญาตามแนวคิดที่หลากหลาย ซึ่งแต่ละแนวคิดล้วนแล้วแต่มีพื้นฐานมาจากการจำแนกพฤติกรรมการเรียนรู้ทางพุทธิปัญญาของ Bloom (1956) ทั้งสิ้น โดยงานวิจัยส่วนใหญ่จำแนกความซับซ้อนทางปัญญาตามแนวคิดของ Porter (2002) (Blank, Porter, & Smithson, 2001; Blank & Smith, 2010; Porter, Smithson, Blank, & Zeidner, 2007; Polikoff, 2015; Polikoff & Porter, 2014) ซึ่งจำแนกได้เป็น 5 ระดับ ได้แก่ จำ (memorize) ปฏิบัติตามขั้นตอน (perform procedures) สื่อสารความเข้าใจ (communicate understanding) แก้ปัญหาในสถานการณ์ใหม่ (solve nonroutine problems) และคาดคะเน/สรุปอ้างอิง/พิสูจน์ (conjecture/ generalize/ prove) รองลงมาคือ แนวคิดจำแนกความซับซ้อนทางปัญญาของ Bloom ที่ได้รับการปรับใหม่ (revised Bloom's taxonomy) (Contio, 2013; Näsström & Henriksson, 2008; นราศักดิ์ ไชยเรือง, 2557) ซึ่งจำแนกได้เป็น 6 ระดับ ได้แก่ จำ (remember) เข้าใจ (understand) ประยุกต์ใช้ (apply) วิเคราะห์ (analyze) ประเมินค่า (evaluate) สร้างสรรค์ (create) แนวคิดการจำแนกความซับซ้อนทางปัญญาของ Webb (1999; 2002) (Andrew & Dawn, 2010; Eckhout, Plake, Smith, & Larsen, 2007; Herman, Webb, & Stephen, 2007; Webb, 1999, 2002) ซึ่งจำแนกได้เป็น 4 ระดับ ได้แก่ ระลึกได้ (recall) ทักษะ/ความคิดรวบยอด (skill/concept) การคิดเชิงกลยุทธ์ (strategic thinking) และการขยายความคิด (extended thinking) แนวคิดจำแนกความซับซ้อนทางปัญญาของ Bloom (1956) ซึ่งจำแนกได้เป็น 6 ระดับ ได้แก่ ความรู้ (knowledge) ความเข้าใจ (comprehension) การนำไปใช้ (application) การวิเคราะห์ (analysis) การสังเคราะห์ (synthesis) และการประเมินค่า (Shankararaman & Gottipati, 2014) และแนวคิดจำแนกความซับซ้อนทางปัญญาของ Marzano (Marzano, 2001 อ้างถึงใน Fontha, 2004) (Fonthal, 2004; Lombardi, Sebuen, Conley, & Snow, 2010; Shelton & Brown, 2008) ซึ่งจำแนกได้เป็น 5 ระดับ ได้แก่ กระบวนการค้นคืน (retrieval processes) กระบวนการเข้าใจ (comprehension processes) กระบวนการวิเคราะห์ (analysis processes) กระบวนการใช้ความรู้ (knowledge utilization processes) และกระบวนการอภิปัญญา (meta-cognitive processes) กระบวนการระบบตนเอง (self-system processes) นอกจากนี้ ยังมีการศึกษาความซับซ้อนทางปัญญาโดยประยุกต์ใช้ผลการเรียนรู้ตามกรอบคุณวุฒิระดับอุดมศึกษาแห่งชาติ หลักสูตรแพทยศาสตรบัณฑิตหลักสูตรปรับปรุง พ.ศ. 2555 (กฤตวรรณ ทองมี, 2557) ซึ่งผลการเรียนรู้แบ่งเป็น 6 ด้าน ได้แก่ ด้านคุณธรรม จริยธรรม ด้านความรู้ ด้านทักษะทางปัญญา ด้าน

ทักษะความสัมพันธ์ระหว่างบุคคลและความรับผิดชอบ ด้านทักษะการวิเคราะห์เชิงตัวเลข การสื่อสาร และการใช้เทคโนโลยีสารสนเทศ และด้านทักษะทางคลินิก

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการศึกษาความสอดคล้องในแนวเดียวกัน ซึ่งเกือบทั้งหมดเป็นเอกสารและงานวิจัยของต่างประเทศ สรุปได้ว่า การศึกษาความสอดคล้องในบริบทของการวัดและประเมินผลอิงมาตรฐานเป็นประเด็นเดียวกับการศึกษาเรื่องความตรง (validity) ความสอดคล้องในแนวเดียวกันเป็นเงื่อนไขจำเป็นเงื่อนไขหนึ่งที่จะทำให้การใช้ผลการวัดความรู้ของผู้เรียนมีความถูกต้องและเหมาะสมมาก และมีความสำคัญมากขึ้นในบริบทที่มีการจัดการศึกษาที่อิงมาตรฐาน เพราะเป็นเรื่องของการตรวจสอบความเหมาะสมของการวัดและการประเมินทางการศึกษา หากต้องการประเมินคุณภาพผู้เรียนตามหลักสูตรอิงมาตรฐาน แต่พบว่ามาตรฐานที่กำหนดในหลักสูตรกับเครื่องมือประเมินไม่สอดคล้องกันก็จะทำให้การใช้ผลการประเมินคุณภาพผู้เรียนนั้นไม่ถูกต้อง และการประเมินผลความสำเร็จของการจัดการศึกษาก็จะไม่ถูกต้องด้วยเช่นกัน สำหรับในประเทศไทยที่เริ่มใช้หลักสูตรอิงมาตรฐานตั้งแต่ที่กระทรวงศึกษาได้ประกาศใช้หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 ซึ่งต่อมาได้มีการปรับปรุงไปสู่หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 แต่ยังไม่ชัดเจนว่ามีการศึกษาการวัดความสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบต่างๆ ทางการศึกษา ดังนั้น ในงานวิจัยนี้จึงต้องการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานการเรียนรู้อันประกอบไปด้วยตัวชี้วัดต่างๆ ที่สะท้อนถึงผลการเรียนรู้ที่คาดหวัง ซึ่งเป็นสิ่งที่ต้องการให้ผู้เรียนรู้ตามมาตรฐานที่ระบุในหลักสูตรแกนกลางขั้นพื้นฐาน พุทธศักราช 2551 กับการประเมินผลการเรียนรู้ อันได้แก่ ข้อสอบที่ครูในโรงเรียนสร้างขึ้นเพื่อประเมินผลผู้เรียน หลังจากที่ครูได้ดำเนินการจัดการเรียนรู้เพื่อให้ผู้เรียนบรรลุตามที่มาตรฐานกำหนด โดยจะทำการศึกษาในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 โดยประยุกต์ใช้วิธีการศึกษาความสอดคล้องในแนวเดียวกันตามแนวคิดของ Webb (1997) และแนวคิดของ Porter (2001) ซึ่งเป็นการศึกษาความสอดคล้องทั้งในมิติด้านเนื้อหา มิติด้านความซับซ้อนทางปัญญาหรือความลึกของความรู้ และมิติด้านการกระจายของเนื้อหา ผลที่ได้จากการวิจัยจะเป็นให้สารสนเทศไปยังผู้มีส่วนได้ส่วนเสียทางการศึกษา เช่น ผู้วางนโยบาย ครู หรือนักการศึกษา เกี่ยวกับความสอดคล้องของมาตรฐานและการประเมินวิทยาศาสตร์ในโรงเรียนมัธยมศึกษา ได้พัฒนาแนวทางในการออกแบบการประเมิน รวมถึงการเรียนการสอนให้สอดคล้องกับมาตรฐานที่กำหนด ให้มุ่งไปสู่เป้าหมายเดียวกัน อันเป็นประโยชน์ต่อการสร้างโอกาสในการเรียนรู้อย่างเท่าเทียมกันให้กับผู้เรียนทุกคน ตามหลักการประกันคุณภาพการศึกษาในการจัดการศึกษาแบบอิงมาตรฐานของประเทศไทย

#### ตอนที่ 4 พฤติกรรมการเรียนรู้ด้านพุทธิพิสัย ตามแนวคิดของ Bloom ที่ได้รับการปรับปรุงใหม่

Bloom (1956) ได้จำแนกพฤติกรรมการเรียนรู้ออกเป็น 3 ด้าน ได้แก่ พฤติกรรมการเรียนรู้ด้านพุทธิพิสัยหรือด้านปัญญา (cognitive domain) พฤติกรรมการเรียนรู้ด้านจิตพิสัย (affective domain) และพฤติกรรมการเรียนรู้ด้านทักษะทางกาย (affective domain) โดยทั้ง 3 ด้านมีความเหลื่อมซ้อนกัน มิได้แยกออกจากกันโดยเด็ดขาด พฤติกรรมการเรียนรู้ด้านพุทธิพิสัยหรือด้านปัญญายเป็นด้านที่มีการนำไปใช้มากที่สุด ทั้งในการออกแบบหลักสูตร จัดกิจกรรมการเรียนการสอน ตลอดจนการวัดและประเมินผล ซึ่ง Bloom ได้จำแนกพฤติกรรมการเรียนรู้ด้านพุทธิพิสัยออกเป็น 6 ระดับ เรียงจากระดับพื้นฐานไปยังระดับสูง ได้แก่ ความรู้ (knowledge) ความเข้าใจ (comprehension) การนำไปใช้ (application) การวิเคราะห์ (analysis) การสังเคราะห์ (synthesis) และการประเมินค่า (evaluate) โดยระดับความรู้และความเข้าใจเป็นพฤติกรรมทางปัญญาระดับพื้นฐาน ในขณะที่ระดับการนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมินค่าเป็นพฤติกรรมทางปัญญาระดับสูง

ต่อมาได้มีการปรับปรุงแนวคิดเกี่ยวกับพฤติกรรมการเรียนรู้ของ Bloom โดย Anderson et al. (2001) เพื่อตอบสนองความรู้ใหม่ๆ ที่พัฒนาอย่างมาก ทั้งในเรื่องจิตวิทยา สมองกับการเรียนรู้ ตลอดจนการศึกษาที่อิงมาตรฐาน และการประกันคุณภาพการศึกษาถึงสิ่งที่คุณเรียนพึงรู้ตามมาตรฐาน โดยประเด็นสำคัญที่ทำการปรับปรุงมีดังนี้

1. แนวคิดพฤติกรรมการเรียนรู้ด้านพุทธิพิสัยของ revised Bloom's taxonomy (Anderson et al., 2001) ได้ปรับเปลี่ยนในส่วนของโครงสร้าง ซึ่งจากเดิมมีเพียงหนึ่งมิติ ได้ปรับให้มี 2 มิติ คือ มิติกระบวนการทางปัญญา (cognitive process dimension) และมิติความรู้ (knowledge dimension) และมิติที่มีการกำหนดเพิ่มเข้ามาคือ มิติความรู้ ซึ่งประกอบด้วยความรู้ในข้อเท็จจริง (factual knowledge) ความรู้ในความคิดรวบยอด (conceptual knowledge) ความรู้ในกระบวนการ (procedural knowledge) และความรู้ในอภิปัญญา (meta-cognitive knowledge) มีรายละเอียดดังต่อไปนี้

1) ความรู้ในข้อเท็จจริง (factual knowledge) เป็นความรู้เกี่ยวกับองค์ประกอบพื้นฐานที่ผู้เรียนต้องรู้เพื่อเข้าถึงเนื้อหาของศาสตร์ หรือเพื่อการแก้ปัญหา แบ่งเป็น 2 ประเภทย่อย ดังนี้

1.1) ความรู้เกี่ยวกับนิยามศัพท์ (knowledge of terminology)

1.2) ความรู้ในรายละเอียดและองค์ประกอบ (knowledge of detail and elements)

2) ความรู้ในความคิดรวบยอด (conceptual knowledge) เป็นความรู้เกี่ยวกับความสัมพันธ์กับระหว่างองค์ประกอบต่างๆ ที่อยู่ภายในโครงสร้างหลักขนาดใหญ่ที่ทำให้องค์ประกอบเหล่านั้นทำงานด้วยกันได้ แบ่งเป็น 3 ประเภทย่อย ดังนี้



2.1) การจำแนกประเภทและจัดเข้ากลุ่ม (classifications and categories)

2.2) หลักการและการสรุปเป็นกฎ (principle and generalizations)

2.3) ทฤษฎี รูปแบบและโครงสร้าง (theories, model and structures)

3) ความรู้ในกระบวนการ (procedural knowledge) เป็นความรู้ที่เกี่ยวกับวิธีการทำสิ่งใดสิ่งหนึ่ง วิธีการสืบเสาะหาความรู้ และหลักเกณฑ์ในการใช้ทักษะ ขั้นตอนการแก้ปัญหา เทคนิค และวิธีการต่างๆ แบ่งเป็น 3 ประเภทย่อย ดังนี้

3.1) ทักษะเฉพาะของวิชา (subject specific skills)

3.2) วิธีการเฉพาะวิชา (subject specific techniques)

3.3) ความรู้ว่าจะใช้กระบวนการ/วิธีการที่เหมาะสมเมื่อใด (knowledge of when to use appropriate procedure)

4) ความรู้ในอภิปัญญา (meta-cognitive knowledge) เป็นความรู้เกี่ยวกับกระบวนการทางปัญญาทั่วไป รวมถึงความตระหนักและความรู้เกี่ยวกับกระบวนการทางปัญญาของตนเอง แบ่งเป็น 3 ประเภทย่อย ดังนี้

4.1) ยุทธวิธีหรือการได้มาซึ่งความรู้

4.2) กระบวนการใช้ปัญญา รวมไปถึงการใช้เนื้อหาและเงื่อนไขที่เหมาะสม

4.3) การรู้จักตนเอง

การสร้างตารางแสดงความสัมพันธ์ระหว่างมิติความรู้ (knowledge dimension) และมิติกระบวนการทางปัญญา (cognitive process dimension) สามารถนำไปใช้ในการกำหนดจุดประสงค์การเรียนรู้ โดยกำหนดจุดประสงค์ให้อยู่ในด้านใดด้านหนึ่งใน 4 ด้านของมิติความรู้ และในขณะเดียวกันก็อยู่ภายในระดับใดระดับหนึ่งใน 6 ระดับของมิติกระบวนการทางปัญญา ตารางมิติความรู้ และมิติกระบวนการทางปัญญา แสดงดังตาราง 5

ตาราง 5 มิติความรู้ (knowledge dimension) และมิติกระบวนการทางปัญญา (cognitive process dimension)

มิติความรู้	มิติกระบวนการทางปัญญา					
	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
ความรู้ในข้อเท็จจริง (factual knowledge)						
ความรู้ในความคิดรวบยอด (conceptual knowledge)						
ความรู้ในกระบวนการ (procedural knowledge)						
ความรู้ในอภิปัญญา (meta-cognitive knowledge)						

2. แนวคิดพฤติกรรมการเรียนรู้ด้านพุทธิพิสัยตาม revised Bloom's taxonomy (Anderson et al., 2001) ได้ปรับเปลี่ยนในส่วนของกระบวนการทางปัญญา (cognitive process) โดยจำแนกเป็น 6 ระดับเช่นเดิม แต่มีการเปลี่ยนชื่อที่ใช้เรียกในแต่ละระดับจากการคำนามเป็นคำกริยาเพื่อให้สะท้อนความเป็นกระบวนการของสมองหรือสติปัญญาที่ช่วยให้มนุษย์เกิดความรู้หรือสติปัญญา และเพื่อให้สอดคล้องกับการศึกษาที่อิงมาตรฐานที่ระบุถึงสิ่งที่ผู้เรียนพึงรู้ นอกจากนี้ ยังมีการสลับลำดับขั้นจากเดิมในระดับ “การสังเคราะห์” และ “การประเมินค่า” มาเป็น “ประเมินค่า” และ “สร้างสรรค์” ดังแสดงในตาราง 6

ตาราง 6 ลำดับขั้นของกระบวนการทางปัญญา revised Bloom's taxonomy

Bloom's taxonomy	revised Bloom's taxonomy
1. ความรู้ (knowledge)	1. จำ (remember)
2. ความเข้าใจ (comprehension)	2. เข้าใจ (understand)
3. การนำไปใช้ (application)	3. ประยุกต์ใช้ (apply)
4. การวิเคราะห์ (analysis)	4. วิเคราะห์ (analyze)
5. การสังเคราะห์ (synthesis)	5. ประเมินค่า (evaluate)
6. ประเมินค่า (evaluation)	6. สร้างสรรค์ (create)

กระบวนการทางปัญญาตามการจัดหมวดหมู่ลำดับความรู้ตาม revised Bloom's taxonomy ให้มีความถูกต้องและเหมาะสมกับการจัดการศึกษาในปัจจุบัน ทั้ง 6 ชั้น เรียงลำดับจากความรู้ระดับต่ำไปยังความรู้ระดับสูง มีดังนี้ (Anderson et al., 2001)

1. จำ (remembering) เป็นความสามารถในการระลึกหรือจำความรู้หรือสารสนเทศที่เก็บไว้ในสมอง ซึ่งเป็นความจำระยะยาว

2. เข้าใจ (understanding) เป็นความสามารถในการสร้างความหมายหรือความรู้จากสื่อหรือเครื่องมือทางการศึกษาด้วยตนเอง เช่น จากการอ่าน การอธิบายของครู ทักษะย่อยของความสามารถในขั้นนี้ ได้แก่ การแปลความหมาย (interpreting) การให้ตัวอย่าง (exemplifying) การจัดจำแนก (classifying) การสรุป (summarizing) การสรุปอ้างอิง (inferring) การเปรียบเทียบ (comparing) และการอธิบาย (explaining)

3. ประยุกต์ใช้ (applying) เป็นความสามารถในการใช้กระบวนการที่ได้เรียนรู้มาในสถานการณ์ใหม่หรือสถานการณ์ที่คล้ายคลึงกัน

4. วิเคราะห์ (analyzing) เป็นความสามารถในการแยกความรู้ออกเป็นส่วนๆ โดยสามารถให้เหตุผลว่า ความรู้ส่วนย่อยที่แยกแต่ละส่วนมีความเกี่ยวข้องกับโครงสร้างของความรู้ทั้งหมดอย่างไร สามารถจำแนกความแตกต่าง จัดระบบความรู้ และบอกที่มาของความรู้หรือองค์ประกอบแต่ละส่วนได้

5. ประเมินค่า (evaluating) เป็นความสามารถในการตรวจสอบ วิพากษ์ และตัดสิน

6. สร้างสรรค์ (creating) เป็นความสามารถในการสร้างสิ่งใหม่จากสิ่งที่เคยเรียนรู้ หรือสิ่งที่พบเห็นในบริบทต่างๆ สามารถสร้างสรรค์งาน แผนงาน หรือผลิตภัณฑ์ หรือชิ้นงานที่แปลกใหม่

รายละเอียดของมิติของกระบวนการทางปัญญา (cognitive processes dimensions) คำสำคัญ และตัวอย่างพฤติกรรมตามแนวคิด revised Bloom's taxonomy สรุปได้ดังตาราง 7

**ตาราง 7 คำสำคัญและตัวอย่างพฤติกรรมในกระบวนการทางปัญญา ตามแนวคิด revised Bloom's taxonomy (ปรับจาก Anderson et al., 2001: 66-68)**

กระบวนการทางปัญญา	คำสำคัญ	ความหมายและตัวอย่าง
<b>1. จำ (remember): ระลึกได้จากความจำระยะยาว</b>		
1.1 การระลึกได้ (recognizing)	- การระบุ (identifying)	การหาตำแหน่งองค์ความรู้ในความจำระยะยาวที่สอดคล้องกับเนื้อหาที่นำเสนอ เช่น - บอกวันที่สำคัญของเหตุการณ์ทางประวัติศาสตร์ได้ - บอกได้ว่ารูปหกเหลี่ยมประกอบด้วยด้านทั้งหมดกี่ด้าน - ระบุชนิดต่างๆ ในแผนผังของสัตว์ครึ่งบกครึ่งน้ำ
1.2 การจำได้ (recalling)	- การระลึก (retrieving)	การระลึกถึงความรู้ที่เกี่ยวข้องจากความจำระยะยาว เช่น - จดจำวันที่ของเหตุการณ์ที่สำคัญทางประวัติศาสตร์ได้ - ตอบได้ว่า 7 คูณ 8 เท่ากับเท่าไร - บอกได้ว่าใครเป็นผู้เขียนแต่งเรื่อง Charge of the Light Brigade

ตาราง 7 (ต่อ)

กระบวนการทางปัญญา	คำสำคัญ	ความหมายและตัวอย่าง
<b>2. เข้าใจ (understand): สร้างความหมายจากข้อความการเรียนรู้การสอน ประกอบด้วย พูด เขียน และการสื่อสารกราฟิก</b>		
2.1 การแปลความหมาย (interpreting)	<ul style="list-style-type: none"> <li>- การจัดประเภท (clarifying)</li> <li>- การถอดความ (paraphrasing)</li> <li>- การนำเสนอ (representing)</li> <li>- การแปลความ (translating)</li> </ul>	<p>การแปลงข้อเท็จจริงจากรูปแบบหนึ่ง (เช่น ตัวเลข) ไปยังอีกรูปแบบหนึ่ง (เช่น ภาษา) ได้ เช่น</p> <ul style="list-style-type: none"> <li>- ถอดความที่สำคัญของการกล่าวสุนทรพจน์และเอกสาร</li> <li>- แปลความหมายโจทย์ปัญหาเป็นสมการพีชคณิตได้</li> <li>- วาดรูประบบย่อยอาหารได้</li> <li>- เรียบเรียงคำปราศรัยของประธานาธิบดี Lincoln ได้</li> </ul>
2.2 การให้ตัวอย่าง (exemplifying)	<ul style="list-style-type: none"> <li>- แสดงภาพ (illustrating)</li> <li>- ยกตัวอย่างประกอบ (instantiating)</li> </ul>	<p>การยกตัวอย่างหรือแสดงแนวคิดหรือหลักเกณฑ์ เช่น</p> <ul style="list-style-type: none"> <li>- ยกตัวอย่างของรูปแบบการวาดภาพศิลปะ</li> <li>- วาดรูปที่มีสมบัติคู่ขนานกันได้</li> <li>- บอกชื่อสัตว์เลี้ยงลูกด้วยนมที่อาศัยอยู่ในแถบบ้านเรา</li> </ul>
2.3 การจัดจำแนกหมวดหมู่ (classifying)	<ul style="list-style-type: none"> <li>- จัดประเภท (categorizing)</li> <li>- จัดหมวดหมู่ (instantiating)</li> </ul>	<p>การระบุหมวดหมู่ให้กับสิ่งต่างๆ เช่น</p> <ul style="list-style-type: none"> <li>- จัดจำแนกหมวดหมู่ให้กับผู้มีความผิดปกติทางจิต</li> <li>- ทำรายการประเภทของรัฐบาลที่พบในสหรัฐอเมริกา</li> <li>- จัดกลุ่มสัตว์พื้นเมืองตามสปีชีส์</li> </ul>
2.4 การสรุป (summarizing)	<ul style="list-style-type: none"> <li>- การย่อความ (abstracting)</li> <li>- การสรุปอ้างอิง (generalizing)</li> </ul>	<p>การสรุปประเด็นหลักหรือสาระที่สำคัญ เช่น</p> <ul style="list-style-type: none"> <li>- เขียนสรุปสั้นๆ จากเหตุการณ์ที่เกิดขึ้นในวิดีโอเทป</li> <li>- ตั้งชื่อเรื่องบทความสั้นๆ ได้</li> </ul>
2.5 การสรุปอ้างอิง (inferring)	<ul style="list-style-type: none"> <li>- การสรุป (concluding)</li> <li>- การคาดการณ์ (extrapolating)</li> <li>- การแก้ไขเพิ่มเติม (interpolating)</li> <li>- การทำนาย (predicting)</li> </ul>	<p>การเขียนข้อสรุปอย่างมีเหตุมีผลจากข้อมูลที่นำเสนอ เช่น</p> <ul style="list-style-type: none"> <li>- สรุปหลักไวยากรณ์จากตัวอย่างต่างๆ ในวิชาภาษาต่างประเทศ</li> <li>- อ่านเรื่องราว บทสนทนาระหว่างบุคคลสองบุคลิกพร้อมทั้งสรุปความสัมพันธ์ในอดีตของทั้งสอง</li> <li>- บอกความหมายของคำใหม่จากบริบทของเนื้อเรื่อง</li> <li>- ทำนายจำนวนในอนาคตได้</li> </ul>
2.6 การเปรียบเทียบ (comparing)	<ul style="list-style-type: none"> <li>- การเปรียบเทียบความแตกต่าง (contrasting)</li> <li>- การแสดงแผนผัง (mapping)</li> <li>- การจับคู่ (matching)</li> </ul>	<p>การตรวจสอบความสอดคล้องระหว่าง 2 แนวคิด หรือ 2 สิ่ง เช่น</p> <ul style="list-style-type: none"> <li>- เปรียบเทียบเหตุการณ์ในประวัติศาสตร์กับสถานการณ์ในปัจจุบัน</li> <li>- อธิบายว่าหัวใจมีความเหมือนกับปืมน้ำอย่างไร</li> <li>- ใช้แผนภาพเวนในการอธิบายความเหมือนและความแตกต่างของหนังสือสองเล่มที่แต่งโดย Charles Dickens</li> </ul>
2.7 การอธิบาย (explaining)	<ul style="list-style-type: none"> <li>- การสร้าง (constructing)</li> <li>- การสร้างแบบจำลอง (modeling)</li> </ul>	<p>การสร้างแบบจำลองเชิงสาเหตุและผลของระบบ เช่น</p> <ul style="list-style-type: none"> <li>- อธิบายสาเหตุของเหตุการณ์ที่สำคัญในศตวรรษที่ 18 ของประเทศฝรั่งเศส</li> <li>- วาดแผนผังอธิบายว่าความดันอากาศส่งผลอย่างไรต่ออากาศ</li> </ul>
<b>3. ประยุกต์ใช้ (apply): ดำเนินการตามขั้นตอนหรือใช้ขั้นตอนในสถานการณ์ที่กำหนด</b>		
3.1 การปฏิบัติ (executing)	<ul style="list-style-type: none"> <li>- การดำเนินการให้ สำเร็จ (carrying out)</li> </ul>	<p>การประยุกต์ใช้ขั้นตอนกับงานที่คุ้นเคย เช่น</p> <ul style="list-style-type: none"> <li>- หารจำนวนเต็มจำนวนหนึ่งด้วยจำนวนเต็มอีกจำนวนหนึ่ง โดยที่ทั้ง 2 จำนวนเป็นเลขหลายหลัก</li> </ul>

ตาราง 7 (ต่อ)

กระบวนการทางปัญญา	คำสำคัญ	ความหมายและตัวอย่าง
		<ul style="list-style-type: none"> <li>- บวกลเลขสองหลักในแนวตั้งได้</li> <li>- อ่านออกเสียงเรื่องราวที่เป็นภาษาต่างประเทศได้</li> </ul>
3.2 การนำไปใช้/ การปฏิบัติ (implementing)	- การใช้ (using)	<p>การประยุกต์ใช้ขั้นตอนกับงานที่ไม่คุ้นเคย เช่น</p> <ul style="list-style-type: none"> <li>- ใช้กฎของนิวตันข้อที่ 2 ในสถานการณ์ที่มีความเหมาะสม</li> <li>- ออกแบบการทดลองเกี่ยวกับการเจริญเติบโตของพืชในดินต่างชนิดกัน</li> </ul>
<b>4. วิเคราะห์ (analyze): จำแนกเนื้อหาออกเป็นส่วนย่อยๆ และสามารถบอกได้ว่าแต่ละส่วนมีความสัมพันธ์กับส่วนอื่นๆ และกับโครงสร้างหรือวัตถุประสงค์โดยรวมอย่างไร</b>		
4.1 การหาความเหมือน- ความแตกต่าง (differentiating)	<ul style="list-style-type: none"> <li>- การจำแนก (discriminating)</li> <li>- การจำแนกความแตกต่าง (distinguishing)</li> <li>- การเน้น (focusing)</li> <li>- การเลือก (selecting)</li> </ul>	<p>การจำแนกประเด็นในส่วนที่เกี่ยวข้องกับส่วนที่ไม่เกี่ยวข้อง หรือจำแนกส่วนที่สำคัญจากส่วนที่ไม่สำคัญของเนื้อหาที่นำเสนอ เช่น</p> <ul style="list-style-type: none"> <li>- จำแนกระหว่างตัวเลขที่เกี่ยวข้องกับตัวเลขที่ไม่เกี่ยวข้อง ในโจทย์ปัญหาคณิตศาสตร์</li> <li>- เลือกข้อความหรือคำสำคัญจากโจทย์ปัญหาทางคณิตศาสตร์ และตัดข้อความที่ไม่เกี่ยวข้องออก</li> <li>- วาดแผนภาพแสดงประเด็นหลักและประเด็นรองของบทประพันธ์</li> </ul>
4.2 การจัดระบบ (organizing)	<ul style="list-style-type: none"> <li>- การหาความเชื่อมโยง (finding coherence)</li> <li>- การบูรณาการ (integrating)</li> <li>- การร่างภาพ (outlining)</li> <li>- การแจกแจง (parsing)</li> <li>- การสร้างโครงสร้าง (structuring)</li> </ul>	<p>การระบุส่วนประกอบที่เหมาะสมหรือสัมพันธ์กับโครงสร้าง เช่น</p> <ul style="list-style-type: none"> <li>- จัดวางหนังสือในห้องสมุดให้ตรงตามหมวดหมู่</li> <li>- ทำแผนผังแสดงความสัมพันธ์ระหว่างพืชกับสัตว์</li> </ul>
4.3 การให้เหตุผล หรืออ้างเหตุผล (attributing)	- การสร้างโครงสร้างใหม่ (deconstructing)	<p>การระบุทฤษฎี อะตอม คณิตศาสตร์ หรือเจตนาที่แฝงอยู่ในเนื้อหาที่นำเสนอ เช่น</p> <ul style="list-style-type: none"> <li>- ระบุทฤษฎีที่สำคัญตามมุมมองทางการเมืองของผู้เขียน</li> <li>- ระบุตัวละครหรือเรื่องราวที่เป็นจุดเด่นในบทประพันธ์หรือเรื่องสั้น</li> <li>- ตั้งสมมติฐานเกี่ยวกับมุมมองของนักการเมืองที่มีต่อปัญหาต่างๆ</li> </ul>
<b>5. การประเมิน (evaluating): ตัดสินบนพื้นฐานของเกณฑ์และมาตรฐาน</b>		
5.1 การตรวจสอบ (checking)	<ul style="list-style-type: none"> <li>- การประสาน (coordinating)</li> <li>- การตรวจสอบ (detecting)</li> <li>- การติดตาม (monitoring)</li> <li>- การทดสอบ (testing)</li> </ul>	<p>การตรวจสอบความไม่สอดคล้องหรือข้อผิดพลาดของกระบวนการหรือผลผลิต หรือการตัดสินได้ว่ากระบวนการหรือผลผลิตมีความสอดคล้องภายในหรือไม่ หรือการตรวจสอบประสิทธิผลของขั้นตอนที่ใช้ดำเนินงาน เช่น</p> <ul style="list-style-type: none"> <li>- การตรวจสอบว่าข้อสรุปของนักวิทยาศาสตร์ได้มาจากข้อมูลที่เก็บรวบรวมได้</li> <li>- มีส่วนร่วมในการทำงานกลุ่ม ให้ข้อมูลป้อนกลับในการทำงาน</li> </ul>

## ตาราง 7 (ต่อ)

กระบวนการทางปัญญา	คำสำคัญ	ความหมายและตัวอย่าง
		หรือข้อโต้แย้งอย่างมีเหตุผล - ทบทวนแผนของโครงการว่าได้รวมขั้นตอนที่จำเป็นไว้ใน การดำเนินงานหรือยัง
5.2 การวิพากษ์วิจารณ์ (critiquing)	- การตัดสิน (judging)	การตรวจสอบความไม่สอดคล้องระหว่างผลผลิตกับเกณฑ์ ภายนอก หรือการตัดสินได้ว่าผลผลิตมีความสอดคล้องภายนอก หรือไม่ หรือการตรวจสอบความเหมาะสมของขั้นตอนที่ใช้ สำหรับปัญหาที่กำหนด เช่น - พิจารณาระหว่าง 2 วิธีการ วิธีการใดจะใช้เป็นแนวทางใน การแก้ปัญหาที่ดีที่สุด - ตัดสินว่าควรดำเนินการอย่างไรให้โครงการประสบความสำเร็จ ตามเกณฑ์ - เลือกวิธีการแก้ปัญหาทางคณิตศาสตร์ที่เหมาะสมที่สุด
<b>6. สร้างสรรค์ (creating): รวมองค์ประกอบย่อยๆ ที่มีลักษณะคล้ายๆ กันเข้าด้วยกัน เพื่อสร้างรูปแบบหรือโครงสร้างใหม่</b>		
6.1 การสร้างสิ่งใหม่ (generating)	- การตั้งสมมติฐาน (hypothesizing)	การสร้างสมมติฐานทางเลือกบนพื้นฐานของหลักเกณฑ์ เช่น - สร้างสมมติฐานเพื่ออธิบายปรากฏการณ์ที่เกิดขึ้น - สร้างเกณฑ์หรือทางเลือกสำหรับการพัฒนาความสัมพันธ์ ระหว่างเชื้อชาติในโรงเรียน - เสนอทางเลือกในการลดการพึ่งพาเชื้อเพลิงจากซากพืชซาก สัตว์เพื่อลดปัญหาด้านเศรษฐกิจและสิ่งแวดล้อม
6.2 การวางแผน (planning)	- การออกแบบ (designing)	การออกแบบขั้นตอนการดำเนินงานให้บรรลุความสำเร็จ เช่น - วางแผนการดำเนินการวิจัยตามหัวข้อที่ได้รับ - ทำบทบาทย่นตรในการนำเสนอผลงานเกี่ยวกับแมลง - ออกแบบการทดลองเพื่อตรวจสอบผลของดนตรีประเภทต่างๆ ที่ส่งผลต่อการผลิตไข่ของไก่
6.3 การผลิตผลงาน (producing)	- การสร้าง (constructing)	การสร้างผลิตภัณฑ์ เช่น - สร้างที่อยู่อาศัยตามจุดประสงค์เฉพาะ - แสดงละครเกี่ยวกับเรื่องราวบางตอนในบทประพันธ์ที่ได้จาก การอ่าน

จากการทบทวนเอกสารพบว่าแนวคิดเกี่ยวกับพฤติกรรมการเรียนรู้ตาม revised Bloom's taxonomy สามารถตอบสนองความรู้ใหม่ๆ ที่พัฒนาอย่างมาก ทั้งในเรื่องจิตวิทยา สมองกับการเรียนรู้ ตลอดจนการศึกษาที่อิงมาตรฐาน และการประกันคุณภาพการศึกษาถึงสิ่งที่ผู้เรียนพึงรู้ตามมาตรฐาน เนื่องจากใช้เป็นพื้นฐานในการกำหนดจุดประสงค์การเรียนรู้ ออกแบบการจัดการเรียนรู้ ตลอดจน การวัดและการประเมินผล ทั้งนี้ สามารถนำแนวคิด revised Bloom's taxonomy มาใช้ในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบทางการศึกษาในระบบการศึกษาแบบอิงมาตรฐาน ซึ่งในที่นี้คือความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตาม

หลักสูตรแกนกลางขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน โดยนำมาประยุกต์ใช้กับวิธีการศึกษาความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ในส่วนของการศึกษาระดับความซับซ้อนทางปัญญา หรือความลึกของความรู้ ซึ่งการศึกษาด้านความซับซ้อนทางปัญญาเป็นสิ่งที่สำคัญในการศึกษาความสอดคล้องในแนวเดียวกัน นอกเหนือจากด้านเนื้อหา (Näsström & Henriksson, 2008) จากการทบทวนงานวิจัยเกี่ยวกับการศึกษาความสอดคล้องในแนวเดียวกันพบว่า มีหลายงานวิจัยที่จำแนกความซับซ้อนทางปัญญาโดยใช้แนวคิดพฤติกรรมการเรียนรู้ของ Bloom แบบเก่า (Bloom's taxonomy) เช่น Shankararaman and Gottipati (2014) และแนวคิด revised Bloom's taxonomy เช่น Contio (2013) Fulmer and Polikoff (2014) Näsström and Henriksson (2008) นอกจากนี้ ยังมีงานวิจัยที่จำแนกความซับซ้อนทางปัญญาตามแนวคิดอื่นๆ เช่น แนวคิดของ Porter (เช่น Blank, Porter, & Smithson, 2001; Blank & Smith, 2010; Polikoff & Porter, 2014; Polikoff, 2015) แนวคิดของ Webb (เช่น Webb, 1999, 2002; Herman, Webb, & Stephen, 2007; Andrew & Dawn, 2010) แนวคิดของ Marzano (Fonthal, 2004; Shelton & Brown, 2008) อย่างไรก็ตาม ทุกแนวคิดดังกล่าว ล้วนแล้วแต่มีพื้นฐานมาจากการจำแนกพฤติกรรมการเรียนรู้ของ Bloom (1956) ทั้งสิ้น อีกทั้งจากงานวิจัยของ Näsström and Henriksson (2008) ที่ได้ทำวิเคราะห์กรอบแนวคิดหรือลำดับชั้นในการจำแนกความซับซ้อนทางปัญญาจากหลักเกณฑ์และแนวคิด 9 แนวคิด ได้แก่ แนวคิดลำดับชั้นของกระบวนการทางปัญญาตาม revised Bloom's taxonomy, De Block, De Corte Guilford, Marzano, Merrill, PISA, Porter และ TIMSS แล้วทำการคัดเลือก 2 แนวคิด ได้แก่ แนวคิดลำดับชั้นของกระบวนการทางปัญญาของ Bloom ที่ได้รับการปรับปรุงใหม่ กับแนวคิดการจำแนกตามวิธีการของ Porter เพื่อใช้ในการศึกษาเปรียบเทียบความเหมาะสมของมิติหรือเกณฑ์ที่ใช้ในการศึกษาความสอดคล้องในแนวเดียวกันของด้านเนื้อหาและด้านความซับซ้อนทางปัญญา ผลการวิจัยพบว่า แนวคิด revised Bloom's taxonomy มีความครอบคลุมและจำเพาะ ตลอดจนมีค่าความเที่ยงระหว่างผู้ประเมินที่สูงกว่าเกณฑ์การจำแนกตามวิธีการของ Porter นอกจากนี้ สำหรับประเทศไทย ครูและนักวิชาการส่วนใหญ่มีการใช้แนวคิดเกี่ยวกับพฤติกรรมเรียนรู้ของ Bloom ทั้งที่เป็นแนวคิดแบบเก่าหรือแนวคิดที่ได้รับการปรับปรุงใหม่ ในการกำหนดจุดประสงค์การเรียนรู้ การออกแบบการเรียนการสอน และการกำหนดแนวทางในการวัดและประเมินผลกันอย่างแพร่หลาย ด้วยเหตุนี้ จึงเป็นอีกเหตุผลหนึ่งที่สนับสนุนว่าแนวคิดเกี่ยวกับพฤติกรรมเรียนรู้ของ Bloom เป็นแนวคิดที่ง่ายต่อสื่อความให้เป็นที่เข้าใจตรงกันในหมู่ครูและนักวิชาการของประเทศไทย ดังนั้น จากเหตุผลที่กล่าวมานี้ งานวิจัยนี้จึงเลือกใช้แนวคิด revised Bloom's taxonomy มาประยุกต์ใช้ในการศึกษาความซับซ้อนทางปัญญาของข้อสอบ เพื่อวิเคราะห์ความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ต่อไป

## ตอนที่ 5 อิทธิพลของผู้ประเมิน

อิทธิพลของผู้ประเมิน (rater effect) ซึ่งในบางเอกสารใช้คำว่า ความคลาดเคลื่อนของผู้ประเมิน (rater error) หรือความลำเอียงของผู้ประเมิน (rater bias) หมายถึง กลุ่มของผลกระทบของความแปรปรวนอย่างเป็นระบบในการให้คะแนนความสามารถที่เกี่ยวข้องอย่างใดอย่างหนึ่งกับผู้ประเมินให้คะแนน และเป็นสิ่งที่ไม่ได้เกี่ยวข้องกับความสามารถที่แท้จริงของผู้ถูกให้คะแนน (Scullen, Mount, & Goff, 2000) ซึ่งอิทธิพลของผู้ประเมินเป็นรูปแบบของการให้คะแนนที่มีความคลาดเคลื่อนในการวัด (Engelhard, 1994; Myford & Wolfe, 2000; Myford & Wolfe, 2003, 2004; Wolfe, 2004, 2005; Wolfe & McVay, 2012) ทั้งนี้ อิทธิพลของผู้ประเมินสามารถจำแนกได้เป็นหลายประเภท โดยสามารถตรวจสอบได้ด้วย MFRM จากการศึกษาเอกสารที่เกี่ยวข้องของ Barrett (2005), Myford and Wolfe (2003, 2004, 2009), Wolfe (2000, 2004), Iramaneerat and Yudkowsky (2007), Wolfe and McVay (2012) และ Song and Wolfe (2015) สามารถสรุปอิทธิพลของผู้ประเมินที่พบได้บ่อย และความแม่นยำในการให้คะแนน (accurate) รวมถึงวิธีการตรวจสอบลักษณะผู้ประเมินแต่ละประเภทด้วย MFRM ดังนี้

**5.1 อิทธิพลการกดหรือปล่อยคะแนน (severity/leniency effect)** เป็นแนวโน้มของผู้ประเมินที่จะให้คะแนนผู้ถูกประเมินสูงหรือต่ำกว่าความเป็นจริง โดยผู้ประเมินมีลักษณะกดคะแนน (severity) จะมีแนวโน้มให้คะแนนโดยเฉลี่ยต่ำกว่าผู้ประเมินคนอื่น ซึ่งเป็นการให้คะแนนต่ำกว่าความเป็นจริงตลอดช่วงความสามารถ (continuum) ทำให้มีความผิดพลาดในการกำหนดความสามารถผู้สอบ ส่วนผู้ประเมินมีลักษณะปล่อยคะแนน (leniency) จะมีแนวโน้มให้คะแนนโดยเฉลี่ยสูงกว่าผู้ประเมินคนอื่น นั่นคือ ให้คะแนนสูงกว่าความเป็นจริง ตลอดช่วงความสามารถ ทำให้มีความผิดพลาดในการกำหนดความสามารถผู้สอบเช่นกัน การตรวจสอบอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมินด้วยโมเดลของราส์ช (Rasch model) พิจารณาได้จากคะแนนดิบ (raw scores) และคะแนนในหน่วยโลจิทของฟาเซทผู้ประเมิน (rater logit) หากมีคะแนนต่ำ หมายถึง ผู้ประเมินมีลักษณะปล่อยคะแนน (leniency) แต่หากมีคะแนนสูง หมายถึง ผู้ประเมินมีลักษณะกดคะแนน (severity)

**5.2 อิทธิพลแนวโน้มการให้คะแนนค่ากลาง (central tendency effect)** เป็นการให้คะแนนส่วนใหญ่มีแนวโน้มอยู่ตรงกลางของมาตรประเมิน โดยผู้ประเมินจะแสดงแนวโน้มในการใช้ระดับตรงกลางของมาตรวัดมากเกินไป ซึ่งมักเกิดจากผู้ประเมินขาดประสบการณ์และขาดคุณสมบัติที่ดีพอในการตรวจประเมิน (Myford & Wolfe, 2004) การตรวจสอบอิทธิพลแนวโน้มการให้คะแนนค่ากลางด้วยโมเดลของราส์ช สามารถตรวจสอบได้ด้วยการพิจารณาค่าสถิติเบื้องต้น เช่น ค่าเฉลี่ยมัธยฐาน พิสัย และส่วนเบี่ยงเบนมาตรฐาน หรือการตรวจสอบด้วยโมเดลของราส์ชเพื่อประมาณค่า



ผู้ประเมิน (rater estimates) และพิจารณาค่า infit MNSQ และ outfit MNSQ ที่น้อยกว่า 0.70 (Myford & Wolfe, 2000: 42) หรือการตรวจสอบการประมาณค่าข้อสอบ (item estimate) โดยมียค่า outfit สูงกว่า 1 มาก และมีค่า outfit t สูงกว่า 0 มาก แสดงว่าเกิดแนวโน้มการให้คะแนนค่ากลาง นอกจากนี้ ยังสามารถตรวจสอบแนวโน้มการให้คะแนนค่ากลางได้ด้วยโค้งลักษณะข้อสอบ (item characteristic curves) โดยหากผู้ถูกตรวจประเมินมีความสามารถสูงมีความผิดพลาดอย่างคงเส้นคงวาในการบรรจุคะแนนหนึ่งบนแกน X และจุดตัดบนแกน X มีค่ามากกว่า 0 อย่างนัยสำคัญ (Barrett, 2005)

**5.3 อิทธิพลฮาโล (halo effect)** เป็นความผิดพลาดของผู้ประเมินในการจำแนกความแตกต่างระหว่างความคิดและคำตอบจากผู้ถูกตรวจประเมิน (Thorndike, 1920 อ้างถึงใน Barrett, 2005) ยกตัวอย่างเช่น ผู้ประเมินอาจให้คะแนนข้อสอบบนพื้นฐานของการประทับใจในคำตอบโดยภาพรวม ดังนั้น ผู้ประเมินอาจมีความผิดพลาดในการจำแนกความแตกต่างระหว่างความคิดที่จำเป็นหรือเนื้อหาที่ไม่จำเป็น ผู้ประเมินอาจไม่สามารถประเมินความสามารถในมิติหรือเกณฑ์ต่างๆ ที่ข้อสอบต้องการวัด (Engelhard, 1994 อ้างถึงใน Barrett, 2005) การให้คะแนนโดยมองจากคำตอบโดยภาพรวมนี้อาจเกิดจากการสร้างลักษณะที่ไม่แท้จริง (artificial) ที่สัมพันธ์กันระหว่างข้อ ดังนั้น การให้คะแนนข้อสอบข้อแต่ละข้ออาจไม่เป็นอิสระจากข้อสอบข้ออื่น การขาดความเป็นอิสระของการให้คะแนนระหว่างข้อสอบการตรวจสอบอิทธิพลฮาโลด้วยโมเดลของราส์ชเพื่อประมาณค่าผู้ประเมิน (rater estimates) และ พิจารณาค่า infit MNSQ และ outfit MNSQ ที่มีค่าต่ำ เช่น มีค่า infit MNSQ น้อยกว่า 0.60 (Barrett, 2005) หรือ มีค่า infit MNSQ และ outfit MNSQ น้อยกว่า 0.70 (Myford & Wolfe, 2003) จะแสดงว่าผู้ประเมินอาจจะมีการให้คะแนนข้อสอบแต่ละข้อ ไม่เป็นอิสระจากข้ออื่น

**5.4 อิทธิพลการจำกัดช่วง (restriction of range effect)** มีความสัมพันธ์กับแนวโน้มการให้คะแนนค่ากลาง โดยสามารถใช้ในสถานการณ์ที่การให้คะแนนกระจุกอยู่จุดใดจุดหนึ่งของช่วงต่อเนื่อง (continuum) บนมาตราประมาณค่า ไม่ว่าจะเป็นจุดที่ให้คะแนนสูง (ปล่อยคะแนน) จุดที่ให้คะแนนต่ำ (กดยคะแนน) หรือจุดกลางๆ (แนวโน้มให้คะแนนค่ากลาง) ดังนั้น แนวโน้มให้คะแนนค่ากลางจึงเป็นประเภทหนึ่งของอิทธิพลการจำกัดช่วง แต่อิทธิพลการจำกัดช่วงไม่ใช่สิ่งเดียวกับแนวโน้มให้คะแนนค่ากลาง เพราะอิทธิพลการจำกัดช่วงอาจเป็นได้ทั้งการกดยคะแนน ปล่อยคะแนน หรือแนวโน้มให้คะแนนค่ากลาง (Myford & Wolfe, 2003) อิทธิพลการจำกัดช่วงเป็นการวัดขอบเขตของการจำแนกระดับความสามารถที่แตกต่างของผู้ถูกตรวจ (Engelhard 1994; Engelhard & Stone, 1998 อ้างถึงใน Barrett, 2005) ซึ่งอาจเกิดจากผู้ประเมินขาดประสบการณ์และขาดคุณสมบัติที่ดีพอในการตรวจประเมิน การตรวจสอบอิทธิพลการจำกัดช่วงด้วยการวิเคราะห์ด้วยโมเดลของราส์ชเพื่อประมาณค่าข้อสอบ (item estimate) และพิจารณาค่า intfit MNSQ ที่มากกว่า 1.30

หรือน้อยกว่า 0.77 (Barrett, 2005) หรือมีค่า infit MNSQ มากกว่า 1.2 หรือน้อยกว่า 0.40 (Wright & Linacre, 1994 อ้างถึงใน Iramaneerat, Yudkowsky, Myford, & Downing, 2008) หรือการตรวจสอบโดยการประมาณค่าผู้ประเมิน (rater estimates) มีค่า infit MNSQ และ outfit MNSQ ต่ำกว่า 0.60 (Kassim, 2011) หรือมีค่า infit MNSQ และ outfit MNSQ ต่ำกว่า 0.40 (Iramaneerat & Yudkowsky, 2008) จะแสดงว่ามีความคลาดเคลื่อนของอิทธิพลการจำกัดช่วง หรือ อาจแสดงว่าข้อสอบนั้นที่ไม่สามารถจำแนกระหว่างผู้ถูกตรวจประเมินที่มีความสามารถต่างกันได้

**5.5 อิทธิพลของความไม่มีแบบแผน (randomness effect) หรือความไม่สอดคล้องกัน** ของผู้ประเมิน (inconsistency) เป็นแนวโน้มที่ผู้ประเมินใดๆ ใช้มาตราประเมินในลักษณะที่ไม่สอดคล้องกับผู้ประเมินคนอื่นๆ ใช้ (Iramaneerat & Yudkowsky, 2007; Myford & wolfe, 2004) กล่าวคือ ผู้ประเมินจะแสดงความแปรปรวนที่ไม่มีแบบแผนมากกว่าการประเมินที่คาดหวัง (expected in ratings) ผู้ประเมินที่แสดงอิทธิพลของความไม่มีแบบแผนจะให้ลำดับ (rank) ในการประเมินผู้ถูกประเมินในลำดับที่แตกต่างไปจากที่ผู้ประเมินคนอื่นให้ และอาจมีการตีความมาตรวัด (หรือระดับขั้นของมาตรวัดตั้งแต่ 1 ระดับขึ้นไป) ที่แตกต่างออกไป ซึ่งอิทธิพลของความไม่มีแบบแผนนี้อาจเนื่องมาจากการขาดคุณสมบัติที่เพียงพอของผู้ประเมิน และ/หรือขาดการอบรมผู้ประเมินที่เพียงพอในการใช้มาตรวัดเพื่อจำแนกความแตกต่างได้ละเอียดอย่างเหมาะสม (Myford & Wolfe, 2004) การตรวจสอบอิทธิพลของความไม่มีแบบแผนโดยใช้โมเดลของราตัสในการประมาณค่าผู้ประเมิน (rater estimate) และพิจารณาค่า infit MNSQ และ outfit MNSQ ที่มีค่าสูง เช่น มีค่า infit MNSQ และ outfit MNSQ มากกว่า 1 (Myford & Wolfe, 2004) หรือมีค่า infit MNSQ และ outfit MNSQ มากกว่า 1.2 (Wright & Linacre, 1994 อ้างถึงใน Iramaneerat & Yudkowsky, 2007) หรือมีค่า infit MNSQ และ outfit MNSQ มากกว่า 1.3 (Myford & Wolfe, 2004) จะแสดงว่าผู้ประเมินมีอิทธิพลของความไม่มีแบบแผน หรือความไม่สอดคล้องในการประเมิน

**5.6 การทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป (differential rater functioning over time: DRIFT)** ในการสอบที่ผู้เรียนต้องแสดงความสามารถที่มีความซับซ้อน เช่น การสอบการเขียน เรียงความนั้น ผู้ประเมินได้ให้คะแนนที่คำนึงถึงคุณภาพของการตอบของนักเรียนเป็นสำคัญ แต่การที่ใช้มนุษย์เป็นผู้ให้คะแนนผลกระทบผู้ให้คะแนน อาจจะมีอิทธิพลต่อความแม่นยำในการตรวจให้คะแนนได้ จากการศึกษาที่ผ่านมาพบว่า ระดับการกดคะแนนของผู้ให้คะแนนนั้นเปลี่ยนแปลงอย่างเป็นระบบเมื่อเวลาผ่านไป (Myford & Wolfe, 2009) นั่นคือ ผู้ให้คะแนนแสดงหลักฐานความแตกต่างกันของการกดคะแนนเมื่อเวลาผ่านไปนั่นเอง โมเดลราตัสแบบหลายองค์ประกอบสามารถตรวจหาและวัดการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไปได้ ซึ่งในการวิเคราะห์ด้วย MFRM อาจจะมีรูปแบบโมเดลเชิงคณิตศาสตร์ที่หลากหลายที่ใช้สำหรับ

ตรวจหาความเคลื่อนที่ของผลกระทบจากผู้ประเมิน (dynamic rater effects) หนึ่งในนั้นคือ MFRM ที่ประกอบด้วยองค์ประกอบของเวลา (time facet)

**5.7 ความแม่นยำในการให้คะแนน (accurate)** เป็นแนวโน้มที่ผู้ตรวจประเมินสามารถพิจารณาตัดสินหรือให้คะแนนได้อย่างถูกต้องแม่นยำ ซึ่งความสามารถดังกล่าวมาจากหลายปัจจัย ได้แก่ ประสบการณ์ของผู้ประเมิน ความรู้ความสามารถของผู้ประเมิน ลักษณะของเกณฑ์ในการให้คะแนน หรือสิ่งแวดล้อมของการประเมิน (เช่น ความเป็นอิสระจากสิ่งรบกวน หรือรูปแบบของปฏิสัมพันธ์ทางสังคม) โดยผู้ประเมินที่เหมาะสมที่สุดอาจเป็นผู้ที่มีรูปแบบการคิดและระบบความเชื่อที่สอดคล้องกับแนวทางการฝึกอบรมและเกณฑ์การให้คะแนน นอกจากนี้ การฝึกอบรมผู้ประเมิน และการติดตามตรวจสอบผู้ประเมินในการประเมินอาจช่วยให้เกิดความแม่นยำโดยการจัดให้มีระบบการให้คะแนนที่เหมาะสม (Wofle, 2004) การตรวจสอบความแม่นยำในการให้คะแนนทำได้โดยการวิเคราะห์ด้วย MFRM เช่น การตรวจสอบส่วนเบี่ยงเบนมาตรฐานของความแตกต่างระหว่างข้อมูลเชิงประจักษ์กับค่าที่ทำนายได้จากโมเดล (residual) ที่มีค่าต่ำ และความสัมพันธ์ระหว่างค่าที่คาดหวังจากโมเดลกับ residual มีค่าเข้าใกล้ 0 หรือตรวจสอบจากการประมาณค่าผู้ประเมิน (rater estimates) โดยผู้ประเมินที่มีค่า infit MNSQ และ outfit MNSQ ตั้งแต่ 0.70 – 1.30 จะแสดงถึงความแม่นยำในการให้คะแนน

โดยสรุป อิทธิพลของผู้ประเมิน (rater effect) เป็นผลกระทบของความแปรปรวนอย่างเป็นระบบในการประเมินความสามารถที่เกี่ยวข้องอย่างใดอย่างหนึ่งกับผู้ประเมิน และเป็นสิ่งที่ไม่ได้เกี่ยวข้องกับความสามารถที่แท้จริงของผู้ถูกประเมิน ซึ่งอิทธิพลของผู้ประเมินมีหลายประเภท งานวิจัยนี้ ผู้วิจัยได้เลือกศึกษาถึงอิทธิพลของผู้ประเมินที่มีสามารถเกิดขึ้นได้ในบริบทของการศึกษา ความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินผลการเรียนรู้ จำนวน 4 ประเภท ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ตลอดจนจนศึกษาถึงความแม่นยำในการให้คะแนนของผู้ประเมิน

## ตอนที่ 6 โมเดลหลายองค์ประกอบของราส์ช

ผู้จัดได้การทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับโมเดลหลายองค์ประกอบของราส์ช (many-facets Rasch measurement: MFRM) แบ่งการนำเสนอออกเป็น 4 ประเด็นหลัก ได้แก่

6.1 แนวคิดพื้นฐานของโมเดลหลายองค์ประกอบของราส์ช

6.2 โมเดลทางคณิตศาสตร์ของโมเดลหลายองค์ประกอบของราส์ช

6.3 สถิติที่สำคัญที่ใช้ในการพิจารณาอิทธิพลของผู้ประเมินจากการวิเคราะห์ด้วยโมเดลหลายองค์ประกอบของราส์ช

6.4 งานวิจัยที่เกี่ยวข้องกับการใช้โมเดลหลายองค์ประกอบของราส์ชในการตรวจสอบอิทธิพลของผู้ประเมิน

รายละเอียดของแต่ละประเด็นมีดังนี้

### 6.1 แนวคิดพื้นฐานของโมเดลหลายองค์ประกอบของราส์ช

โมเดลหลายองค์ประกอบของราส์ช (many-facets Rasch measurement: MFRM) เป็นโมเดลทางสถิติที่สามารถใช้สำหรับศึกษาและอธิบายอิทธิพลของผู้ประเมิน (Engelhard, 1994; Iramaneerat & Yudkowsky, 2007; Myford & Wolfe, 2000) Linacre ได้พัฒนาโมเดลหลายองค์ประกอบของราส์ชขึ้นในปี 1989 โดยได้ปรับขยายจากโมเดลของราส์ช ให้สามารถใช้กับข้อสอบแบบลำดับขั้น (order categories) หรือข้อสอบที่มีการตรวจให้คะแนนแบบบางส่วน (partial credit) ซึ่งจากเดิมโมเดลของราส์ชมุ่งเน้นไปที่ 2 องค์ประกอบคือ ข้อสอบและผู้สอบ แต่ MFRM สามารถเพิ่มองค์ประกอบอื่นๆ ที่สนใจเข้าไปในโมเดล เช่น ความยากของข้อสอบ การกตคะแนนของผู้ประเมิน หรือโครงสร้างของมาตรฐานค่า ซึ่งฟังก์ชันความสัมพันธ์ขององค์ประกอบทั้งหมดแสดงด้วยสมการทางคณิตศาสตร์ ทั้งนี้ MFRM สามารถบ่งชี้พฤติกรรมการตรวจให้คะแนนว่ามีการกตหรือปล่อยคะแนนมากเกินไปได้ การวิเคราะห์ข้อมูลโดยใช้ MFRM จึงเป็นการศึกษาอิทธิพลของแหล่งความคลาดเคลื่อนต่างๆ ที่มีต่อคุณภาพของการวัดประเมิน โดยผู้วิจัยสามารถใช้โครงสร้างทางสถิติในการวิเคราะห์ผลการประเมินเพื่อการสรุปรูปแบบ (pattern) ของผลหลักสำหรับผู้ตรวจหรือผู้ประเมิน (rater) ผู้สอบ (ratee) และคุณลักษณะ (trait) หรือข้อสอบ (item) และตัวแปรอื่นๆ ที่สนใจ นอกจากนี้ MFRM ยังสามารถแยกวิเคราะห์ในแต่ละองค์ประกอบหรือฟาเซท (facet) อย่างเป็นอิสระจากองค์ประกอบอื่น ตลอดจนสามารถประเมินระดับสมาชิก (individual – level effects) ในแต่ละฟาเซทว่ามีลักษณะใด เช่น ผู้ประเมินแต่ละคน ผู้สอบแต่ละคน คุณลักษณะหรือข้อสอบแต่ละข้อ (Myford & Wolfe, 2003)

## 6.2 โมเดลทางคณิตศาสตร์ของโมเดลหลายองค์ประกอบของราล์ช

โมเดลทางคณิตศาสตร์ (the mathematical model) ของ MFRM ที่สอดคล้องกับข้อมูลการให้คะแนนแบบหลายค่า สามารถเขียนเป็นสมการได้ดังนี้ (Linacre, 1989 อ้างถึงใน Myford & Wolfe, 2003)

$$\ln|P_{nij k}/P_{nij k-1}| = B_n - D_i - C_j - F_k$$

- เมื่อ  $P_{nij k}$  คือ ความน่าจะเป็นที่ผู้ถูกประเมินคนที่  $n$  ได้ระดับคะแนน  $k$  ของรายการประเมินที่  $i$  จากผู้ประเมินคนที่  $j$
- $P_{nij k-1}$  คือ ความน่าจะเป็นที่ผู้ถูกประเมินคนที่  $n$  ได้ระดับคะแนน  $k-1$  ของรายการประเมินที่  $i$  จากผู้ประเมินคนที่  $j$
- $B_n$  คือ ระดับความสามารถของผู้ถูกประเมินคนที่  $n$
- $D_i$  คือ ค่าความยากของคุณลักษณะ  $i$
- $C_j$  คือ ค่าการกตคะแนนของผู้ประเมิน (severity of rater)  $j$
- $F_k$  คือ ค่าความยากประจำระดับที่  $k$  สัมพันธ์กับขั้นที่  $k-1$
- โดยที่  $F_k$  ถือว่าพารามิเตอร์ไม่นับว่าเป็นองค์ประกอบของโมเดล

จากสมการเป็นโมเดลการวิเคราะห์ที่มีให้คะแนนแบบมาตราประมาณค่า (rating scale) ที่ประกอบด้วย 3 องค์ประกอบ ได้แก่ ความสามารถของผู้ถูกประเมิน ความยากของข้อรายการประเมิน และการกตคะแนนของผู้ประเมิน โดยมีคะแนนโลจิท เป็นตัวแปรตาม องค์ประกอบต่างๆ (facets) อันได้แก่ ผู้ถูกประเมิน คุณลักษณะที่ประเมิน และผู้ประเมิน เป็นตัวแปรอิสระ ความน่าจะเป็นของผู้ถูกประเมิน  $n$  ได้รับการให้คะแนน  $x$  ในรายการประเมินที่  $i$  จากผู้ประเมินคนที่  $j$  เขียนเป็นสมการได้ดังนี้ (Myford & Wolfe, 2003)

$$P_{nijx} = \frac{\exp \sum_{c=0}^k (B_n - D_i - C_j - F_k)}{\sum_{c=0}^k \exp \sum_{c=0}^k (B_n - D_i - C_j - F_k)}$$

- เมื่อ  $P_{nijx}$  คือ ความน่าจะเป็นของผู้ถูกประเมิน  $n$  ได้รับการให้คะแนน  $x$  ในรายการประเมินที่  $i$  จากผู้ประเมินคนที่  $j$
- $k$  คือ ระดับขั้นของมาตราประมาณค่า (a rating scale category)
- $c$  คือ ดัชนีแสดงจำนวน (a counting index) มีช่วงตั้งแต่ 0 ถึง  $k$
- $K$  คือ ระดับขั้นสูงสุดของมาตราประมาณค่า

$h$  คือ ดัชนีแสดงจำนวน (a counting index) มีช่วงตั้งแต่ 0 ถึง  $K$

$F_0$  มีค่าเท่ากับ 0

การวิเคราะห์ MFRM แต่ละองค์ประกอบจะถูกวิเคราะห์พร้อมกัน แต่มีความเป็นอิสระจากกัน และจะถูกปรับ (calibrate) ให้อยู่ในมาตรเชิงเส้นเดียวกัน เรียกว่า มาตรโลจิต (logit scale) การปรับแต่ละองค์ประกอบทำให้สามารถวัดการกดคะแนน (rater severity) ของผู้ประเมินที่อยู่บนมาตรเดียวกับความสามารถของผู้ถูกประเมิน (rate performance) และความยากของข้อรายการประเมิน (trait difficulty)

6.3 สถิติที่สำคัญที่ใช้ในการพิจารณาอิทธิพลของผู้ประเมินจากการวิเคราะห์ด้วยโมเดลหลายองค์ประกอบของราส์ซ

สถิติที่สำคัญที่ใช้ในการพิจารณาอิทธิพลของผู้ประเมินจากการวิเคราะห์ด้วย MFRM มีดังนี้

### 6.3.1 สถิติทดสอบไคสแควร์ (fixed effect chi-square test)

สถิติไคสแควร์ (chi-square statistic) ที่ได้จากโปรแกรม FACETS เป็นสถิติที่ใช้ทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติระหว่างสมาชิกต่างๆ ในองค์ประกอบ (Engelhard, 1996) โดยมีสมมติฐานสำหรับการทดสอบองค์ประกอบของผู้ประเมินว่าผู้ประเมินทุกคนมีระดับการกดคะแนนอยู่ในระดับเดียวกันภายหลังจากอธิบายความคลาดเคลื่อนแล้ว ในการทดสอบองค์ประกอบผู้สอบมีสมมติฐานว่า ผู้สอบทุกคนมีระดับความสามารถเดียวกันหลังจากอธิบายความคลาดเคลื่อนแล้ว และในองค์ประกอบของข้อสอบมีสมมติฐานว่า ข้อสอบทุกข้อมีระดับความยากที่เท่ากันหลังจากอธิบายความคลาดเคลื่อนแล้ว หากค่าไคสแควร์ มีนัยสำคัญแสดงว่ามีสมาชิกอย่างน้อย 2 ตัวในองค์ประกอบที่แตกต่างอย่างมีนัยสำคัญทางสถิติ แต่ค่าสถิตินี้เป็นเพียงค่าประมาณของการแจกแจงแบบไคสแควร์ เท่านั้น โดยมีสูตรทางสถิติสำหรับผลแบบกำหนดของข้อสอบเป็นดังนี้ (Linacre, 2001 อ้างถึงใน Myford & Wolfe, 2003)

$$\chi^2 = \sum (w_i D_i^2) - \frac{(\sum w_i D_i)^2}{\sum w_i}, \quad \text{d.f.} = L - 1$$

เมื่อ  $D$  คือ ความยากของข้อสอบ

$L$  คือ จำนวนของข้อสอบ และ

$$w_i = \frac{1}{SE_i^2} \quad \text{สำหรับ } i = 1, \dots, L$$

จากการศึกษาเอกสารเกี่ยวกับ MFRM สรุปได้ว่า MFRM ที่ได้รับการปรับขยายจากโมเดลของราส์ซ ให้สามารถใช้กับข้อสอบแบบลำดับชั้น (order categories) หรือข้อสอบที่มีการตรวจให้คะแนนแบบบางส่วน (partial credit test items) ซึ่งจากเดิมโมเดลของราส์ซจะมุ่งเน้นไปที่ 2 องค์ประกอบคือ ข้อสอบและผู้สอบ แต่ MFRM สามารถเพิ่มองค์ประกอบอื่นๆ ที่สนใจเข้าไปในโมเดล ทั้งนี้ MFRM สามารถบ่งชี้พฤติกรรมการตรวจให้คะแนนว่ามีการกดหรือปล่อยคะแนนมากเพียงใด การวิเคราะห์ข้อมูลโดยใช้ MFRM จึงเป็นการศึกษาอิทธิพลของแหล่งความคลาดเคลื่อนต่างๆ ที่มีต่อคุณภาพของการวัดประเมิน โดยผู้วิจัยสามารถใช้โครงสร้างทางสถิติในการวิเคราะห์ผลการประเมินเพื่อการสรุปรูปแบบ (pattern) ของผลหลักสำหรับผู้ตรวจหรือผู้ประเมิน (rater) ผู้สอบ (ratee) และคุณลักษณะ (trait) หรือข้อสอบ (item) และตัวแปรอื่นๆ ที่สนใจ นอกจากนี้ MFRM ยังสามารถแยกวิเคราะห์ในแต่ละองค์ประกอบหรือฟาเซท (facet) อย่างเป็นอิสระจากองค์ประกอบอื่น และแต่ละองค์ประกอบจะถูกปรับ (calibrate) ให้อยู่ในมาตรเชิงเส้นเดียวกัน เรียกว่า มาตรโลจิท (logit scale) (Myford & Wolfe, 2003) ตลอดจนสามารถประเมินระดับสมาชิก (individual – level effects) ในแต่ละฟาเซทว่ามีลักษณะใด เช่น ผู้ประเมินแต่ละคน ผู้สอบแต่ละคน คุณลักษณะหรือข้อสอบแต่ละข้อ

### 6.3.2 สถิติตรวจสอบความสอดคล้องระหว่างข้อมูลกับ (fit statistics) ของผู้ประเมิน เป็นรายบุคคล

สถิติตรวจสอบความสอดคล้องระหว่างข้อมูลกับโมเดล (fit statistics) เป็นสถิติที่แสดงถึงระดับความสอดคล้องของข้อมูลเชิงประจักษ์หรือค่าสังเกต (คะแนนที่ผู้ประเมินให้กำหนดกับผู้ถูกประเมิน) กับค่าที่คาดหวังจาก MFRM (คะแนนจากการทำนายการให้คะแนนของผู้ประเมินของโมเดล) แนวทางในการตรวจสอบความสอดคล้องของข้อมูลเชิงประจักษ์กับโมเดลที่นิยมคือการใช้ค่ามาตรฐานของความคลาดเคลื่อน (standardized residual:  $Z_{nij}$ ) ซึ่งเป็นดัชนีบ่งชี้คุณภาพของการให้คะแนน ความคลาดเคลื่อนของค่าการสังเกตกับค่าที่คาดหวังของผู้ถูกประเมิน (n) ที่ได้รับการประเมินจากผู้ประเมิน (j) ในคุณลักษณะ (i) หากค่าได้จากสมการ (Myford & Wolfe, 2003)

$$R_{nij} = E_{nij} - X_{nij}$$

เมื่อ  $R_{nij}$  คือ ค่าส่วนที่เหลือ (residual) ของค่าสังเกตได้กับค่าที่คาดหวังของผู้ถูกประเมิน (n) ที่ได้รับการประเมินจากผู้ประเมิน (j) ในคุณลักษณะ (i)  
 $E_{nij}$  คือ ค่าที่คาดหวังของการประเมินบนฐานของการประมาณค่าพารามิเตอร์ของราส์ซ  
 $X_{nij}$  คือ ค่าที่สังเกตได้ของการประเมิน

และสูตรสำหรับคำนวณ  $E_{nij}$  (Linacre, 1994; Myford & Wolfe, 2003) เป็นดังนี้

$$E_{nij} = \sum_{k=0}^m kP_{nij}$$

- เมื่อ  $E_{nij}$  คือ ค่าที่คาดหวังของการประเมินบนฐานของการประมาณค่าพารามิเตอร์ของราล์ซ
- $m$  คือ จำนวนของรายการ (category) ในมาตรการประเมิน
- $k$  คือ ดัชนีจำนวนแสดงถึงค่าของแต่ละรายการของมาตรประเมิน
- $P_{nij}$  คือ ความน่าจะเป็นของผู้ถูกประเมิน  $n$  ได้รับคะแนนจากผู้ประเมินคนที่  $j$  ในรายการประเมินที่  $i$

ค่ามาตรฐานส่วนที่เหลือคำนวณได้ดังนี้ (Myford & Wolfe, 2003)

$$Z_{nij} = \frac{R_{nij}}{\sqrt{V_{E_{nij}}}}$$

เมื่อ

$$V_{E_{nij}} = \sum_{k=0}^m (k - E_{nij})^2 P_{nij}$$

นั่นคือ

$$Z_{nij} = \frac{R_{nij}}{\left[ \sum_{k=0}^m (k - E_{nij})^2 P_{nij} \right]^{1/2}}$$

- เมื่อ  $Z_{nij}$  คือ ค่ามาตรฐานของส่วนที่เหลือ (standardized residuals)
- $R_{nij}$  คือ ค่าส่วนที่เหลือของค่าสังเกตได้กับค่าที่คาดหวังของผู้สอบ ( $n$ ) ที่ได้รับจากผู้ประเมิน ( $j$ ) ในข้อรายการประเมินหรือคุณลักษณะ ( $i$ )
- $X_{nij}$  คือ คะแนนจากผู้ประเมินคนที่  $j$  ประเมินผู้สอบคนที่  $n$  ในข้อ  $i$
- $E_{nij}$  คือ ค่าที่คาดหวังของการประเมินจากโมเดลการวัด
- $V_{E_{nij}}$  คือ ความแปรปรวนของค่าที่สังเกตได้ (observation)
- $P_{nij}$  คือ ความน่าจะเป็นที่ผู้ถูกประเมินคนที่  $n$  ได้ระดับคะแนน  $k$  ของข้อรายการประเมินที่  $i$  จากผู้ประเมินคนที่  $j$
- $k$  คือ ดัชนีจำนวนแสดงถึงค่าของแต่ละรายการของมาตรประเมิน

ค่ามาตรฐานของความคลาดเคลื่อน (standardized residuals) จะให้รายละเอียดเกี่ยวกับการตรวจสอบความสอดคล้องระหว่างคะแนนสังเกตกับรูปแบบการประเมินที่คาดหวัง (expecting rating patterns) ของโมเดล ค่ามาตรฐานของส่วนที่เหลือสามารถสรุปความแตกต่างของ



องค์ประกอบ (facets) และความแตกต่างของสมาชิก (elements) ในองค์ประกอบ ค่ามาตรฐานของส่วนที่เหลือจึงเป็นดัชนีที่แสดงถึงความสอดคล้องของข้อมูลเชิงประจักษ์กับโมเดลทางทฤษฎี (Engelhard, 1996) เมื่อนำค่ามาตรฐานส่วนที่เหลือยกกำลังสองแล้วหาค่าเฉลี่ย ทั้งผู้สอบและข้อสอบ เพื่อสรุปเป็นค่าสถิติของผู้ประเมินแต่ละคน แสดงถึงค่าความไม่สอดคล้อง (misfit) ทั้งหมดในการตรวจของผู้ประเมิน (Linacre, 1994) ซึ่งค่าเฉลี่ยนี้จะให้ค่าดัชนีใน 2 ลักษณะ ดังนี้

**6.3.2.1 สถิติ outfit mean - square (outfit MNSQ)** เป็นดัชนีของค่าเฉลี่ยยกกำลังสองของส่วนที่เหลือแบบไม่ถ่วงน้ำหนัก (unweighted mean - square residual) ที่แสดงถึงความแตกต่างระหว่างคะแนนสังเกตกับค่าความคาดหวัง สถิติ outfit MNSQ, MSU สำหรับองค์ประกอบผู้ประเมินแสดงถึงค่าเฉลี่ยยกกำลังสองของค่ามาตรฐานส่วนที่เหลือจากผู้สอบและข้อสอบทั้งหมดของผู้ประเมิน สามารถคำนวณได้จากสมการ (Myford & Wolfe, 2003)

$$MS_U = \frac{\sum_{n=1}^N \sum_{i=1}^I Z_{R_{nij}}^2}{NI}$$

เมื่อ  $MS_U$  คือ สถิติ outfit mean - square

$Z_{R_{nij}}$  คือ ค่ามาตรฐานของส่วนที่เหลือ

$N$  คือ จำนวนผู้สอบ

$I$  คือ จำนวนข้อสอบ

เนื่องจากสถิติ outfit mean - square เป็นสถิติที่มีค่าค่อนข้างไว (sensitivity) ต่อการระบุข้อความที่ไม่สอดคล้องกับโมเดล (misfit) (Engelhard, 1994; Myford & Wolfe, 2000) ดังนั้นค่าสถิตินี้จึงมีจุดอ่อนในบางลักษณะของการตอบของกลุ่มผู้ตอบ เช่น กรณีที่ผู้ตอบบางคนมีรูปแบบการตอบที่แตกต่างไปจากลักษณะการตอบคำถามของกลุ่มผู้ตอบโดยส่วนรวม เช่น คนที่มีคุณลักษณะ (trait) ในระดับสูงตอบข้อคำถามในบางรายการในระดับต่ำๆ หรือในทางตรงกันข้าม ผู้ตอบที่มีคุณลักษณะระดับต่ำกลับตอบข้อความบางรายการในระดับสูงๆ ซึ่งลักษณะการตอบดังกล่าวนี้ถือว่าไม่สอดคล้องกับค่าความน่าจะเป็นที่ผู้ตอบควรจะตอบ จึงทำให้ค่าสถิติปฏิเสธข้อความในบางรายการไป ทั้งที่ความจริงแล้วข้อความดังกล่าวยังใช้ได้ เพียงแต่ผู้ตอบบางคนตอบข้อความรายการนั้น ด้วยความเสแสร้งเท่านั้น (Wright & Master, 1982) ดังนั้น จึงมีสถิติอีกค่าหนึ่ง ซึ่งมีความละเอียดขึ้น โดยมี  $E_{nij}$  เป็นตัวถ่วงน้ำหนัก

**6.3.2.2 สถิติ infit mean - square (infit MNSQ)** เป็นค่าดัชนีของค่าเฉลี่ยยกกำลังสองแบบถ่วงน้ำหนัก (weighted mean - squares) กับสัดส่วนความแปรปรวนส่วนที่เหลือยกกำลังสองแบบถ่วงน้ำหนัก (weighted proportion to the variance of the square residuals) ที่สามารถคำนวณแต่ละสมาชิกขององค์ประกอบได้ สถิติ infit mean - square มีความไว (sensitivity)

ต่ำต่อผลการประเมินที่ไม่คาดหวัง (outlying unexpected) สถิติ infit MNSQ หรือ  $MS_W$  สำหรับองค์ประกอบผู้ประเมิน สามารถคำนวณได้ดังสมการ (Myford & Wolfe, 2003)

$$MS_W = \frac{\sum_{n=1}^N \sum_{i=1}^I Z_{R_{nij}}^2 V_{E_{nij}}}{\sum_{n=1}^N \sum_{i=1}^I V_{E_{nij}}}$$

เมื่อ  $MS_W$  คือ สถิติ infit mean - square

$Z_{R_{nij}}$  คือ ค่ามาตรฐานของส่วนที่เหลือ

$V_{E_{nij}}$  คือ ความแปรปรวนของค่าที่สังเกตได้ (observation)

$$\text{โดย } V_{E_{nij}} = \sum_{k=0}^m (k - E_{nij})^2 P_{nijk}$$

$E_{nij}$  คือ ค่าที่คาดหวังของการประเมินจากโมเดล

$N$  คือ จำนวนผู้สอบ

$I$  คือ จำนวนข้อสอบ

สถิติทดสอบความสอดคล้องของโมเดลสำหรับองค์ประกอบอื่นๆ และสมาชิกภายในองค์ประกอบนั้น นิยมคำนวณแล้วสรุปบนค่าที่เหมาะสม เมื่อโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ ค่าคาดหวังของดัชนีค่าเฉลี่ยยกกำลังสอง infit MNSQ และ outfit MNSQ มีค่าเท่ากับ 1 หากค่าทั้งสองมีค่าเข้าใกล้ 1 แสดงว่าคะแนนสังเกตกับค่าที่คาดหวังนั้น มีค่าใกล้เคียงกัน หากมีค่าน้อยกว่า 1 แสดงว่าค่าสังเกตมีค่าใกล้กับค่าที่คาดหวังมากกว่าที่โมเดลทำนายไว้ (overfit) และหากมีค่ามากกว่า 1 แสดงว่าค่าที่สังเกตอยู่ห่างจากค่าที่คาดหวังมากกว่าที่โมเดลทำนายไว้ (misfit) (Myford & Wolfe, 2003) ทั้งนี้ขอบเขตของการยอมรับได้ที่เหมาะสมของโมเดลที่สอดคล้องกับข้อมูลนั้น Lunz, Wright, และ Linacre (1990) แนะนำว่า ขอบเขตของการยอมรับได้ควรมีค่าอยู่ระหว่าง 0.6 – 1.5 ส่วน Wright, Benjamin, และ Master (1982) ได้เสนอขอบเขตที่ยอมรับได้ที่เหมาะสมของโมเดลที่สอดคล้องกับข้อมูลเมื่อคำนวณสถิติ infit MNSQ และ outfit MNSQ ว่าควรมีค่าอยู่ระหว่าง -2 ถึง +2 ทั้งนี้ Myford และ Wolfe (2003) ได้ให้ข้อเสนอแนะว่าขอบเขตที่เหมาะสมของ infit MNSQ และ outfit MNSQ นั้น ขึ้นอยู่กับความต้องการใช้สารสนเทศที่ได้จากการวิเคราะห์ MFRM เช่น หากนำไปใช้ตัดสินใจในกรณีที่มีผู้มีส่วนได้ส่วนเสียจำนวนมาก ขอบเขตอาจจะแคบลง โดยอาจอยู่ระหว่าง 0.8 – 1.2 แต่หากมีผู้มีส่วนได้ส่วนเสียไม่มาก อาจผ่อนปรนให้อยู่ระหว่าง 0.5 – 2.0

#### 6.4 งานวิจัยที่เกี่ยวข้องกับการใช้โมเดลหลายองค์ประกอบของราล์ชในการตรวจอติพิลของผู้ประเมิน

Engelhard (1992) ได้ศึกษาการใช้ MFRM ในการวัดความสามารถทางการเขียนเรียงความ โดยสุ่มกลุ่มตัวอย่างนักเรียนเกรด 8 จำนวน 1,000 คน ผู้ตรวจจำนวน 2 คน ทำการตรวจงานเขียนเรียงความจำนวน 2 ข้อของนักเรียน ตามเกณฑ์การให้คะแนน 5 ด้าน ด้วยมาตรประมาณค่า 4 ระดับ วิเคราะห์ ด้วยโปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 4 องค์ประกอบ ได้แก่ ความสามารถในการเขียน การกตคะแนของผู้ตรวจ ความยากของข้อสอบเขียน และระดับของเกณฑ์การให้คะแนน ผลการศึกษาพบว่า ความเที่ยงแยกส่วนของผู้ตรวจเป็น 0.87 และมีความแตกต่างอย่างมีนัยสำคัญของการกตคะแนของผู้ตรวจ โดยมีค่าสถิติ infit และ outfit MNSQ ของข้อสอบอยู่ระหว่าง 0.9 - 1.1 และมีค่าความเที่ยงแยกส่วนของข้อสอบเท่ากับ 0.46 ซึ่งค่อนข้างต่ำ แต่มีนัยสำคัญทางสถิติ แสดงว่าความยากของข้อสอบมีความแตกต่างกันเพียง ค่าความเที่ยงแยกส่วนของผู้สอบเป็น 0.93 ซึ่งแตกต่างกันอย่างมีนัยสำคัญ

Engelhard (1994) ได้ใช้ MFRM ในการตรวจสอบความคลาดเคลื่อนของผู้ตรวจในการวัดประเมินงานเขียน ได้แก่ การกตคะแนของผู้ตรวจ อติพิลฮาโล แนวโน้มการให้คะแนนค่ากลาง และอติพิลการจำกัด โดยสุ่มตัวอย่างนักเรียนเกรด 8 ในการสอบการเขียนของประเทศจอร์เจีย จำนวน 264 คน ผู้ตรวจ 15 คน ทำการตรวจให้คะแนนตามเกณฑ์การให้คะแนน 5 ด้าน ด้วยมาตรประมาณค่า 4 ระดับ วิเคราะห์ข้อมูล ด้วยโปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ความสามารถในการเขียน การกตคะแนของผู้ตรวจ และระดับของเกณฑ์การให้คะแนน ผลการศึกษาพบว่า มีความแตกต่างอย่างมีนัยสำคัญของการกตคะแนของผู้ตรวจ มีอติพิลฮาโลในการประเมินแบบรวมมากกว่าแบบย่อยของผู้ตรวจ 2 คน มีการประเมินที่อยู่ในตรงค่ากลางของมาตรประเมินประมาณ ร้อยละ 80 แสดงอติพิลแนวโน้มเข้าสู่ส่วนกลาง หากยังไม่ได้ปรับแก้การกระจายของคะแนนดิบจะมีอติพิลการจำกัดช่วงเกิดขึ้น และเมื่อปรับแก้การประมาณค่าความสามารถในการเขียนแล้วพบว่า มีความคลาดเคลื่อนของผู้ตรวจเพียงเล็กน้อยเท่านั้น

Zhu, Ennis, และ Chen (1998) ได้ทำการศึกษาเพื่อสร้างโมเดลการตัดสินของผู้เชี่ยวชาญ ในการพัฒนาการทดสอบและเครื่องมือในการทดสอบ โดยใช้ MFRM เป็นเครื่องมือที่ใช้ในการประเมินคุณค่าของเป้าหมายหลักสูตรการศึกษาทางกายภาพ ที่เรียกว่า value orientation inventory-2 (VOI-2) ประกอบด้วย 150 ข้อคำถาม ผู้เชี่ยวชาญทั้งหมด 128 คน ประเมินความสอดคล้องของข้อคำถามแต่ละข้อกับเป้าหมายของหลักสูตรใน 5 ด้าน ด้วยมาตรประมาณค่า 5 ระดับ โดยข้อคำถาม 150 ข้อ จะถูกสุ่มเข้าสู่แบบฟอร์มการประเมิน (rating form) ซึ่งมี 4 ฟอร์ม ได้แก่ A-D (ภายใต้เป้าหมายหลักสูตร 5 ด้าน) แต่ละฟอร์มประกอบด้วยข้อคำถาม 60 ข้อ โดยทั้ง 4 ฟอร์มจะมี

ข้อคำถามร่วมกัน (common item) จำนวน 30 ข้อ ข้อคำถามร่วมนี้จะใช้สำหรับเชื่อมโยง (form linkage) เป็นชุดของมาตราส่วน (common scale) วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS กำหนดองค์ประกอบของความแปรปรวน (facet) มี 6 องค์ประกอบ ได้แก่ เพศ เชื้อชาติ รูปแบบการจ้างงาน ผู้ประเมิน บริบทแวดล้อม และข้อคำถาม ทำการตรวจสอบและปรับแก้อิทธิพลของการกดคะแนนและความสอดคล้องในการตัดสินของผู้เชี่ยวชาญก่อนนำข้อคำถามไปใช้ในการประเมิน และตรวจสอบผลกระทบของสมาชิกในกลุ่มที่มีต่อการตัดสินของผู้เชี่ยวชาญคนอื่นๆ ในกลุ่ม ผลการศึกษาพบว่า ข้อคำถามในเครื่องมือ VOI-2 มีเนื้อหาที่เป็นตัวแทนที่ดี และผู้ตรวจประเมินเป็นผู้เชี่ยวชาญอย่างแท้จริง นอกจากนี้ ยังแสดงให้เห็นว่าโมเดลหลายองค์ประกอบของราส์ชเป็นเทคนิคทางจิตมิติที่เหมาะสมสำหรับประยุกต์ใช้เพื่อการตัดสินของผู้เชี่ยวชาญในการพัฒนาการทดสอบ

Myford และ Wolfe (2000) ศึกษาการใช้ MFRM ในสถาบันทดสอบทางการศึกษา (educational testing service: ETS) ด้านการพูดภาษาอังกฤษ (test of spoken English: TSE) เพื่อตรวจสอบแหล่งของความผันแปรในระบบการวัดประเมินแต่ละแหล่งที่ส่งผลต่อผู้สอบ และข้อเสนอแนะการเปลี่ยนแปลงระบบการวัดประเมิน ในการศึกษาครั้งนี้ใช้ข้อมูลจากผู้สอบในปี 1997 เดือนกุมภาพันธ์ (1,469 คน) และเมษายน (1,446 คน) จำนวนทั้งหมด 2,911 คน ที่เป็นเสียงตอบข้อสอบ 12 ข้อ ที่ถูกบันทึกเทปไว้ ผู้ตรวจ (เดือนกุมภาพันธ์ จำนวน 66 คน และเมษายน จำนวน 74 คน) ทำการตรวจให้คะแนนโดยใช้มาตรฐานค่า 5 ระดับ วิเคราะห์ข้อมูลโดยใช้ MFRM แบบตรวจให้คะแนนบางส่วน (partial credit form of the MFRM) ด้วยโปรแกรม FACETS กำหนดองค์ประกอบของความแปรปรวน (facet) มี 4 องค์ประกอบ ได้แก่ ความสามารถของผู้สอบ ความกดคะแนนของผู้ตรวจ ความยากของข้อสอบ และระดับของมาตรฐานค่าผลการศึกษา พบว่า แบบทดสอบสามารถจำแนกระดับความสามารถของผู้สอบตามสถิติได้เป็น 8 ระดับ ความสามารถของผู้สอบมีความเชื่อถือได้ในประเด็นของความแม่นยำและความคงที่ ค่าความเที่ยงแยกส่วนของผู้สอบมีค่าเป็น 0.98 ข้อสอบทั้ง 12 ข้อมีค่า infit และ outfit อยู่ในช่วงที่ยอมรับได้ (0.7 - 1.3) และมีความยากแตกต่างกันเล็กน้อย รายการต่างๆ ของมาตรวัดประเมิน TSE ที่มี 5 ระดับนั้น มีความชัดเจนในการแยกความแตกต่างในแต่ละระดับและสามารถใช้เกณฑ์เดียวกันได้กับข้อสอบทั้ง 12 ข้อ สำหรับผู้ตรวจนั้นมีระดับการกดคะแนนที่แตกต่างกัน หากปรับแก้ความแตกต่างของการกดคะแนนของผู้ตรวจจะทำให้คะแนน 2 ใน 3 ของผู้สอบทั้งหมดมีค่าเฉลี่ยแตกต่างจากคะแนนดิบเฉลี่ยประมาณ 0.5 - 3.6 คะแนนซึ่งมีผลอย่างมากกับผู้สอบที่อยู่ใกล้คะแนนจุดตัด นอกจากนี้ จากการตรวจสอบความแปรปรวนที่เกิดจากผู้ประเมินในการสอบเดือนกุมภาพันธ์ พบว่า มีผู้ตรวจที่แสดงความแม่นยำ (accuracy) ร้อยละ 70 อิทธิพลของความไม่มีแบบแผน (randomness effect) ร้อยละ 3 และอื่นๆ ร้อยละ 24 สำหรับการสอบเดือนเมษายน พบว่า มีผู้ตรวจที่แสดงความแม่นยำ ร้อยละ 49 อิทธิพลการสุ่ม ร้อยละ 5 อิทธิพลการให้คะแนนสุดโต่ง ร้อยละ 1 และอื่นๆ ร้อยละ 24

Myford (2002) ได้ประยุกต์ใช้ MFRM ในการศึกษาลักษณะการออกแบบคำอธิบายมาตรฐานค่าเกี่ยวกับกราฟิก (descriptive graphic rating scales) ผู้ประเมินจำนวน 11 คน จะได้รับการฝึกอบรมการใช้คำอธิบายมาตรฐานค่าเกี่ยวกับกราฟิก (ที่เป็นมาตรฐานแบบต่อเนื่อง) ในการประเมินการทดสอบความก้าวหน้าทางการศึกษาแห่งชาติ (NAEP) ด้านทัศนศิลป์ของผู้เรียน 50 คน มาตรฐานได้รวมลักษณะต่างๆ ไว้เข้าด้วยกัน (ได้แก่ การมี/ไม่มีการกำหนดจุดตัดคะแนนกลาง หรือ hatchmarks หรือทั้งสองอย่าง) จุดมุ่งหมายของการศึกษานี้คือ เพื่อตัดสินจำนวนจุดตัดคะแนน (point) ที่ควรจะต้องมีในแต่ละมาตรวัดแบบต่อเนื่อง (continuous scale) และเพื่อตัดสินว่าลักษณะการออกแบบที่แตกต่างกันจะส่งผลต่อความเที่ยงหรือไม่ การวิเคราะห์ข้อมูลทำโดยใช้โปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 2 องค์ประกอบ ได้แก่ ความสามารถของผู้เรียน และการกตคะแนนของผู้ประเมิน วิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (intraclass correlational analyses) ทุกมาตรวัดที่พิจารณาจะมีอย่างน้อย 5 ระดับคะแนน ซึ่งบางมาตรวัดอาจมี 7 – 10 ระดับคะแนน ผลการศึกษาพบว่า มาตรวัดที่พิจารณามี 5 ระดับคะแนน ความเที่ยงแยกส่วนของผู้เรียนจะมีค่าอยู่ในช่วง 0.79 – 0.91 และมีค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่มอยู่ระหว่าง 0.73 – 0.85 มาตรวัดที่มีการกำหนดจุดตัดคะแนนกลาง (midpoints) ไม่ได้มีค่าความเที่ยงแบบแยกส่วนของผู้เรียนหรือค่าสัมประสิทธิ์สหสัมพันธ์ภายในกลุ่มที่สูงกว่ามาตรวัดที่ไม่มีการกำหนดจุดตัดคะแนนกลาง เช่นเดียวกับมาตรวัดที่มีหลาย hatchmarks ก็ไม่จำเป็นว่าจะต้องมีค่าความเที่ยงสูงหรือต่ำกว่ามาตรวัดที่มีเพียงไม่กี่ hatchmarks

Nakamura (2002) ได้ประยุกต์ใช้ MFRM ในการวิเคราะห์แบบทดสอบการเขียน โดยศึกษาว่าหากใช้ผู้ตรวจ 2 คน จะสามารถลดภาระงานของผู้ตรวจได้หรือไม่ โดยที่ยังมีความน่าเชื่อถือทางสถิติ ผู้ตรวจทั้งหมด 4 คน ตรวจข้อสอบเรียงความจำนวน 2 ข้อของเป็นนักศึกษาจำนวน 36 คน ด้วยมาตรฐานประเมิน 4 ระดับ โดยให้ผู้ตรวจเป็นคู่ตรวจงานนักศึกษาแต่ละคน ตามเกณฑ์การประเมิน 8 เกณฑ์ วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ความสามารถของผู้สอบ ความกตคะแนนของผู้ตรวจ และความยากของข้อสอบ ผลการศึกษาพบว่า การใช้ระบบการตรวจเป็นคู่สามารถลดภาระงานของผู้ตรวจได้ มาตรฐานค่า 4 ระดับ ของเกณฑ์การประเมินทั้งหมดมีสถิติ outfit อยู่ในช่วงที่ยอมรับได้ (0.8 – 1.1) ช่วงความสามารถของผู้สอบค่อนข้างกว้าง (5 ถึง -1) มีนักศึกษา 5 คนที่ misfit และ 2 คนที่ overfit มีความแตกต่างอย่างมีนัยสำคัญของการกตคะแนนของผู้ตรวจ ความเที่ยงระหว่างผู้ตรวจ (inter-rater reliability) ทุกคนมีค่าสูง แม้ว่าจะมีระดับของการกตหรือปล่อยคะแนนแตกต่างกัน เกณฑ์ทั้ง 8 ที่ใช้ในการศึกษานั้น ผู้ตรวจทั้ง 4 คน ไม่จำเป็นต้องตรวจข้อสอบนักเรียนทุกคน เนื่องจากความสอดคล้องของการตรวจอยู่ในช่วงที่ยอมรับได้

Park (2004) ได้ใช้ MFRM ในการตรวจสอบอิทธิพลของแหล่งความแปรปรวนในการทดสอบเพื่อจัดลำดับผู้สอบจากการเขียนภาษาอังกฤษเป็นภาษาที่ (ESL placement test of writing) ผู้ตรวจจำนวน 17 คน ทำการตรวจให้คะแนนข้อสอบความเรียงของผู้สอบ จำนวน 99 คน โดยใช้รูปรีกตามเกณฑ์การประเมิน 5 ด้าน โดยแต่ละด้านจะถูกประเมินด้วยมาตราประมาณค่า 4 ระดับ วิเคราะห์ข้อมูลโดยใช้ MFRM แบบตรวจให้คะแนนบางส่วน (partial credit form) กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ผู้สอบ ผู้ตรวจ และเกณฑ์ในการประเมิน ผลการศึกษาพบว่า ผู้สอบมีความแปรปรวนของความสามารถของค่อนข้างกว้าง คือ ตั้งแต่ -8 ถึง 7 โลจิท ผู้สอบส่วนใหญ่ได้คะแนนสูง นั่นคือ ความสามารถของผู้สอบมีการแจกแจงเบ้ทางลบ และมีร้อยละ 4 ของผู้สอบที่ไม่สอดคล้องกับโมเดล (misfit) จึงต้องมีการทบทวนการประเมินอีกครั้งก่อนการรายงานคะแนน โดยเฉพาะผู้สอบที่อยู่ใกล้เขตวิกฤตของการตัดสินใจ ความเที่ยงแยกส่วนของผู้ตรวจมีค่าเป็น 0.97 แสดงว่าผู้ตรวจมีระดับการกตคะแนนที่แตกต่างกัน ค่าความสอดคล้อง (fit value) ของผู้ตรวจทุกคนอยู่ในช่วงที่ยอมรับได้ ส่วนค่าความยากของข้อสอบทั้ง 4 ข้อมีค่าใกล้ 0 ค่าความเที่ยงแยกส่วนของข้อสอบมีค่าเป็น 0.64 แสดงว่าข้อสอบทั้ง 4 ข้อ มีค่าความยากใกล้เคียงกัน และค่าความสอดคล้องอยู่ในช่วงที่ยอมรับได้ และมาตราประเมินทั้ง 4 ระดับ มีความเหมาะสมในการจัดลำดับและมีความชัดเจนในการจำแนกแต่ละระดับ เพราะผู้สอบมีค่าเฉลี่ยของคะแนนสูงขึ้นตามน้ำหนักของรายการการประเมินที่เพิ่มขึ้นในข้อสอบทุกข้อ

Smith และ Kulikowich (2004) ได้นำทฤษฎีการสรุปอ้างอิงและ MFRM ไปใช้ในการประเมินทักษะการแก้ปัญหาที่ซับซ้อน กลุ่มตัวอย่างเป็นนักเรียนเกรด 4 จำนวน 44 คน ใช้ข้อสอบจำนวน 5 ข้อ ผู้ตรวจ จำนวน 2 คน และทดสอบ 2 ครั้ง สถานการณ์ที่กำหนดเป็นกิจกรรมเตะบอล (kickball) โดยให้ผู้สอบจัดทีมที่เหมาะสมภายใต้ข้อมูลที่กำหนดให้ จากนั้น ให้นำเสนองาน ผู้ตรวจจะตั้งคำถามในด้านการสังเกต การเปรียบเทียบความเหมือน ความต่างการจำแนกและการวางแผน เมื่อวิเคราะห์ข้อมูลด้วยโมเดลการสรุปอ้างอิง พบว่า แหล่งของความแปรปรวนที่มากที่สุดคือ ผู้สอบ (0.325) แสดงว่าผู้สอบมีความแตกต่างทางด้านความคิดซับซ้อน รองลงมาคือ ข้อสอบ (0.257) ส่วนความแปรปรวนของผู้ตรวจและจำนวนครั้งของการสอบมีน้อย (0.001 และ 0.004 ตามลำดับ) นั่นคือ ไม่มี ความแตกต่างของผู้ตรวจ มีความแตกต่างเพียงเล็กน้อยในการสอบ 2 ครั้ง ในการศึกษา D-study พบว่าหากใช้ผู้ตรวจ 2 คนและตรวจ 2 ครั้ง จะมีค่าสัมประสิทธิ์การสรุปอ้างอิงเป็น 0.8012 และค่าสัมประสิทธิ์ความเชื่อถือได้เป็น 0.6977 ส่วนการวิเคราะห์ด้วย MFRM ที่กำหนดแหล่งความแปรปรวน (facet) 4 แหล่ง ได้แก่ ความสามารถของนักเรียน ความกตคะแนน ความยากของการทดสอบ และจำนวนครั้งในการสอบ พบว่า ผู้สอบมีความแตกต่างกันทางด้านความคิดซับซ้อน และผู้ตรวจมีระดับการกตคะแนนคล้ายคลึงกัน และจำนวนครั้งของการสอบไม่มีผลต่อค่าประมาณค่า นอกจากนี้ พบว่า ความยากของข้อสอบนั้นมีการกระจายอยู่บนตลอดช่วงความสามารถ (continuum) โดยการวางแผน

มีค่าความยากมากที่สุด คือ มีคะแนนโลจิทเป็น 0.80 ส่วนอีก 3 ข้อ นั้น มีค่าความยากอยู่ใกล้ๆ โลจิท 0 ข้อสอบที่ง่ายที่สุดคือ การจำแนก ซึ่งมีค่าความยากของคะแนนโลจิทเป็น -1.0

Wolfe (2004) ทำการศึกษาเพื่อแสดงตรวจสอบอิทธิพลของผู้ตรวจด้วย MFRM โดยทำการตรวจสอบอิทธิพลของผู้ตรวจประเภทต่างๆ ได้แก่ อิทธิพลด้านความถูกต้องแม่นยำ/ไม่ถูกต้องแม่นยำ อิทธิพลด้านการกด/ปล่อยคะแนน และอิทธิพลแนวโน้มการให้คะแนนค่ากลาง/คะแนนปลายขั้ว ของผู้ตรวจจำนวน 101 คน ในการตรวจให้คะแนนข้อสอบความเรียงของผู้สอบจำนวน 28 คน ค่าสถิติที่ใช้ในการพิจารณาอิทธิพลของผู้ตรวจได้แก่ ส่วนเบี่ยงเบนมาตรฐานของส่วนที่เหลือ ( $SD_{residual}$ ) และสหสัมพันธ์ระหว่างส่วนที่เหลือกับคะแนนที่คาดหวัง ( $r_{residual,expected}$ ) ของผู้ตรวจแต่ละคน ผลการศึกษาพบว่า ผู้ตรวจร้อยละ 6 แสดงอิทธิพลของการปล่อยคะแนน ร้อยละ 5 แสดงอิทธิพลของการกดคะแนน ร้อยละ 10 แสดง การตรวจให้คะแนนที่ความถูกต้องแม่นยำ (มีค่า  $SD_{residual}$  ต่ำ,  $r_{residual,expected}$  เข้าใกล้ 0) ร้อยละ 1 แสดงการตรวจให้คะแนนไม่ถูกต้องแม่นยำ (มีค่า  $SD_{residual}$  สูง,  $r_{residual,expected}$  เข้าใกล้ 0) ร้อยละ 4 แสดงอิทธิพลแนวโน้มการให้คะแนนค่ากลาง (มีค่า  $SD_{residual}$  ต่ำ,  $r_{residual,expected}$  ติดลบ) และร้อยละ 2 แสดงอิทธิพลแนวโน้มการให้คะแนนปลายขั้ว (มีค่า  $SD_{residual}$  สูง,  $r_{residual,expected}$  เป็นบวก)

Eckes (2005) ใช้ MFRM ในการทดสอบอิทธิพลของผู้ตรวจของการประเมินประสิทธิภาพในการเขียนและการพูดในการทดสอบภาษาเยอรมันในฐานะ โดยได้ศึกษาอิทธิพลหลักของผู้ตรวจ (rater main effect) และปฏิสัมพันธ์ระหว่างผู้ตรวจและองค์ประกอบที่เกี่ยวข้อง ได้แก่ ผู้สอบ เกณฑ์การตรวจให้คะแนน (ในส่วนของ การเขียน) และงาน (ในส่วนของ การพูด) นอกจากนี้ ยังมีการศึกษาการทำงานที่แตกต่างกันเกี่ยวกับเพศของผู้สอบ เครื่องมือที่ใช้ในการสอบที่เรียกว่า TestDaf ประกอบด้วย 4 ส่วน คือ การอ่านทำความเข้าใจ การฟังทำความเข้าใจ การแสดงความคิดและความรู้สึกด้วยการเขียน และการแสดงความคิดและความรู้สึกด้วยการพูด กลุ่มตัวอย่าง ผู้สอบ ในส่วนการเขียนมีผู้เข้าร่วม 1,359 คน ในส่วนการพูดผู้เข้าร่วม 1,348 คน ผู้ตรวจให้คะแนนการเขียนของผู้สอบจำนวน 29 คน และให้คะแนนการพูดของผู้สอบ จำนวน 31 คน โดยใช้มาตรฐานค่าที่มีตั้งแต่ 2 - 5 ระดับ ผลการศึกษาพบว่า (1) ผู้ตรวจมีการกดคะแนนที่ความแตกต่างกันสูง (2) มีความสอดคล้องในการตรวจให้คะแนนในภาพรวม (3) มีความสอดคล้องกันน้อยมากในความสัมพันธ์ระหว่างผู้ตรวจกับเกณฑ์การตรวจให้คะแนน เมื่อเทียบกับความสัมพันธ์ผู้ตรวจกับผู้สอบ และ (4) ในกลุ่มไม่พบว่าการเกิดความลำเอียงทางเพศ

Sudweeks, Reeve, และ Bradshaw (2005) ได้เปรียบเทียบการใช้ทฤษฎีการสุบอ้างอิง และ MFRM ในการวิเคราะห์งานเขียน ตัวอย่างการวิจัยเป็นนักศึกษาจำนวน 24 คน จากนักศึกษาทั้งหมด 497 คน ที่ลงทะเบียนเรียนวิชาประวัติศาสตร์โลกในภาคเรียนฤดูร้อน ปี 2002 ผู้ตรวจจำนวน 9 คน ตรวจความเรียงของนักศึกษาจำนวน 2 บทความ โดยใช้มาตรฐานประเมิน 9 ระดับ ใช้

กฎเกณฑ์การให้คะแนนแบบรวม (holistic rating scale) แบบแผนการศึกษาที่ใช้ ประกอบด้วย 2 รูปแบบ ได้แก่ แบบไขว้สมบูรณ และแบบสอดแทรก (nested) กำหนดองค์ประกอบความแปรปรวน (facet) 4 องค์ประกอบ ได้แก่ ผู้สอบ ข้อสอบ ผู้ตรวจ และจำนวนครั้งในการตรวจ มีการอบรมผู้ตรวจก่อนการให้คะแนนเป็นเวลา 2 ชั่วโมง และในการตรวจจะไม่ให้ผู้ตรวจทราบชื่อผู้สอบ ผลการศึกษาพบว่า ทั้งสองทฤษฎีต่างให้ผลการวิเคราะห์ที่สอดคล้องกัน โดยในทฤษฎีการสุปอ้างอิงความแปรปรวนที่เกิดจากข้อสอบและปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบนั้นมีค่าสูง และแตกต่างกันอย่างมีนัยสำคัญ ในขณะที่ความแปรปรวนของผู้ตรวจและจำนวนครั้งของการตรวจนั้นมีค่าต่ำ และมีความแปรปรวนที่ไม่สามารถอธิบายได้ประมาณร้อยละ 22 ค่าความเที่ยงเมื่อใช้งานเขียน 2 ขึ้น ผู้ตรวจ 9 คน และตรวจ 2 ครั้ง โดยแบบแผนการศึกษาแบบไขว้สมบูรณมีค่าความเที่ยงเท่ากับ 0.75 ซึ่งเป็นค่าที่ค่อนข้างต่ำ ค่าสัมประสิทธิ์ความเที่ยงมีค่าสูงขึ้น หากคำนวณจากค่าเฉลี่ยของการประเมินนักเรียนโดยใช้ข้อสอบอย่างน้อย 3 ข้อ จากผู้ตรวจอย่างน้อย 2 คน การใช้ข้อสอบหลายข้อและผู้ตรวจหลายคนจะให้ค่าความเที่ยงเพิ่มสูงขึ้นมากกว่าการตรวจหลายครั้ง การใช้แบบแผนแบบ nested design นั้น ค่าความเที่ยงแตกต่างจากรูปแบบแบบไขว้สมบูรณเพียงร้อยละ 1.5 - 2.0 เท่านั้น ในขณะที่การวิเคราะห์ด้วย MFRM พบว่า ดัชนีแยก (separation index) และความเที่ยงแยกส่วน (separation reliability) ของข้อสอบมีค่าสูงและแตกต่างกันอย่างมีนัยสำคัญตามความยากของข้อสอบ และปฏิสัมพันธ์แบบสองทางระหว่างผู้สอบกับข้อสอบมีร้อยละสูงสุดอย่างมีนัยสำคัญ ความแปรปรวนของผู้ตรวจและจำนวนครั้งของการตรวจนั้นมีค่าต่ำ ความเที่ยงแยกส่วนของผู้สอบ ผู้ตรวจ และข้อสอบมีค่าสูง แต่ความเที่ยงแยกส่วนของจำนวนครั้งของการตรวจมีค่าต่ำ แสดงว่ามีความสอดคล้องกันของผู้ตรวจจากการตรวจเพียงครั้งเดียว

Iramaneerat และ Yudkowsky (2007) ได้ศึกษาการใช้ MFRM ในการวิเคราะห์ข้อมูลจากการประเมินทักษะทางคลินิกของนักศึกษาแพทย์ เพื่อตรวจสอบอิทธิพลการกดหรือปล่อยคะแนนความไม่สอดคล้องในการให้คะแนน (inconsistency) อิทธิพลฮาโล และอิทธิพลการจำกัดช่วงตัวอย่างในการศึกษา คือนักศึกษาแพทย์ชั้นปีที่ 4 จำนวน 173 คน แต่ละคนจะถูกทดสอบด้วยงานทางคลินิก (clinical tasks) หรือกรณีศึกษา จำนวน 6 งาน/กรณี ผู้ป่วยมาตรฐาน (standardized patients: SPs) จำนวน 6 คน ที่คัดเลือกจากทั้งหมด 17 คน SPs ทำการให้คะแนนความสามารถของนักศึกษาแต่ละคนใน 6 ทักษะ โดยใช้แบบตรวจสอบรายการ (checklist) และมาตรฐานค่า 4 - 5 ระดับ กำหนดองค์ประกอบความแปรปรวน (facet) 4 องค์ประกอบ ได้แก่ ทักษะทางคลินิกของนักศึกษา ความยากของงานทางคลินิกหรือกรณีศึกษาความกดดันของ SPs และความยากของข้อสอบ (งาน) ทางคลินิก ผลการศึกษาพบว่า SPs แสดงให้เห็นความแตกต่างอย่างมีนัยสำคัญทางสถิติของการกดคะแนน มี SPs จำนวน 4 คนที่แสดงความไม่สอดคล้องกันในการให้คะแนน



(inconsistency) มี SPs จำนวน 4 คน แสดงอิทธิพลของการจำกัดช่วง และมี SPs เพียง 1 คน ที่แสดงอิทธิพลฮาโล

Iramaneerat, Yudkowsky, Myford, และ Downing (2008) ได้ศึกษาการใช้ MFRM และ G-theory ในการตรวจสอบการควบคุมคุณภาพของการสอบทางคลินิก ที่เรียกว่า Objective Structured Clinical Examination (OSCE) ในการประเมินทักษะด้านการสื่อสารของผู้สอบเป็นแพทย์ประจำบ้าน จำนวน 79 คน แต่ละคนจะได้รับประเมินทักษะการสื่อสารด้วยสถานการณ์ทางคลินิก จำนวน 6 สถานการณ์ ผู้ป่วยมาตรฐาน (standardized patients: SPs) จำนวน 6 คน ที่ใช้ในการทำหน้าที่ประเมินทักษะการสื่อสาร โดยใช้ข้อรายการในการประเมิน 18 ข้อ ที่มีลักษณะเป็นมาตรฐานประมาณค่า 5 ระดับ วิเคราะห์ข้อมูลด้วย G-theory โดยใช้รูปแบบ fully crossed random effects design ( $p \times c \times i$ ) เพื่อตรวจสอบแหล่งความแปรปรวน และวิเคราะห์ MFRM โดยใช้โปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ทักษะทางคลินิกของผู้สอบ ความกตัญญูของ SPs หรือความยากของกรณีศึกษา และความยากของข้อรายการประเมิน ผลจากการศึกษา G-study พบว่า แหล่งความแปรปรวนของความคลาดเคลื่อนที่ใหญ่ที่สุดคือแหล่งความแปรปรวนของ SPs หรือกรณีศึกษา สำหรับผลการวิเคราะห์ด้วย MFRM พบว่า SPs หรือกรณีศึกษามีระดับความกตัญญูของ SPs หรือความยากของกรณีศึกษาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ( $\chi^2 = 2272.8, p < .05$ ) นอกจากนี้ พบว่า SPs จำนวน 2 คน แสดงอิทธิพลความไม่สอดคล้องในการให้คะแนน ซึ่งแสดงถึงปัญหาที่เกี่ยวข้องกับแนวทางในการบรรยายกรณีศึกษา ความเข้าใจในการใช้มาตรฐานประมาณค่า และ/หรือเนื้อหาของกรณีศึกษา ทั้งนี้ SPs มีการตีความข้อรายการ จำนวน 2 ข้อ ที่ไม่สอดคล้องกัน และมาตรฐานประมาณค่าของ 2 รายการทำหน้าที่ไม่เป็น 5 ระดับ ในขณะที่ SPs หรือกรณีศึกษาไม่แสดงอิทธิพลของการจำกัดช่วง เนื่องจากทุก SPs หรือกรณีศึกษา มีค่า infit MS สูงกว่า 0.40

Basturk (2008) ได้ประยุกต์ใช้ MFRM ในการประเมินการนำเสนองานด้วยโปรแกรม PowerPoint ของนักศึกษาในรายวิชาวิชาชีวเคมีเบื้องต้น ตัวอย่างเป็นนักศึกษาห้องหนึ่งจำนวน 30 คน ที่ทำการสุ่มจาก 5 ห้องเรียน แบ่งนักศึกษาเป็น 6 กลุ่มย่อย ผู้ประเมินการนำเสนอ จำนวน 7 คน ทำการประเมินด้วยรูบริกที่มีชื่อว่า “A+Powerpoint rubrics” ประกอบด้วยข้อรายการในการประเมิน 9 ข้อ มีระดับในการให้คะแนน 6 ระดับ (0 = ไม่มี, 6 = ดีเยี่ยม) วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ กลุ่มความสามารถของนักศึกษา ความยากของข้อรายการ/งาน และการตัดสินการกดหรือปล่อยคะแนน ผลการศึกษาพบว่า ค่าความเที่ยงของกลุ่มความสามารถในการนำเสนอของนักศึกษา การกดคะแนนของผู้ประเมิน และความยากของข้อรายการ/งาน เท่ากับ 0.86 0.88 และ 0.82 ตามลำดับ แสดงว่ามี

ความยุติธรรมในการจำแนกกลุ่มความสามารถของนักศึกษา การกตคะแนนของผู้ประเมิน และข้อรายการตามระดับคุณภาพ

Eckes (2008) ได้ใช้ MFRM ในการศึกษาอิทธิพลของผู้ตรวจในการประเมินความสามารถทางภาษา ผู้ตรวจจำนวน 65 คน ทำแบบสอบถามเกี่ยวกับเกณฑ์ที่ใช้ในการประเมินความสามารถในการเขียนของข้อสอบส่วนหนึ่งในการทดสอบ TestDaF แล้วพิจารณาว่าแต่ละเกณฑ์มีความสำคัญต่อการประเมินความสามารถในการเขียนของผู้สอบมากน้อยเพียงใด โดยแบบสอบถามมีลักษณะเป็นมาตราประมาณค่า 4 ระดับ (สำคัญน้อย สำคัญ สำคัญมาก และสำคัญที่สุด) เพื่อบ่งบอกว่าการใช้มาตราประมาณค่า 4 ระดับนี้ มีความสำคัญมากเพียงใดที่จะยึดเป็นเครื่องมือในการประเมินตามเกณฑ์ 9 ข้อ ที่ครอบคลุมความสามารถในการเขียนด้านต่างๆ วิเคราะห์ด้วยโปรแกรม FACETS 3.59 กำหนดองค์ประกอบความแปรปรวน (facet) 2 องค์ประกอบ ได้แก่ ผู้ตรวจ และเกณฑ์การตรวจให้คะแนน ผลการศึกษาพบว่า ผู้ตรวจมีความแตกต่างกันอย่างมีนัยสำคัญเกี่ยวกับแนวคิดความสำคัญของเกณฑ์ประเมิน จากการวิเคราะห์ด้วยเทคนิคการจัดกลุ่มสองรูปแบบ (two-mode clustering technique) พบว่า ให้การจำแนกกลุ่มของผู้ตรวจและกลุ่มของเกณฑ์การตรวจให้คะแนน โดยแบ่งผู้ตรวจได้เป็น 6 กลุ่ม แต่ละกลุ่มมีลักษณะของรูปแบบการให้คะแนนที่แตกต่างกัน แสดงว่าผู้ตรวจไม่ถูกแบ่งอย่างเท่าเทียมกันตามชุดของเกณฑ์การตรวจให้คะแนน

Myford และ Wolfe (2009) ได้ใช้ MFRM ในการตรวจสอบความสามารถของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป กรอบการทำงานสำหรับการตรวจสอบความแตกต่างในความแม่นยำ (differential accuracy) และความแตกต่างของระดับคะแนนในมาตรวัดที่ใช้ (differential scale category use) มีจุดประสงค์เพื่อแสดงดัชนีทางสถิติต่างๆ เพื่อระบุการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป ตัวอย่างในการศึกษาเป็นนักเรียน 51,233 คน ผู้ตรวจให้คะแนน 101 คน ทำการตรวจให้คะแนนข้อสอบแบบเขียนตอบแบบอิสระในการสอบ advanced placement English literature and composition examination (AP ELC) โดยใช้มาตราประมาณค่า 9 ระดับ กำหนดแนวทางการให้คะแนนแบบองค์รวม (single holistic rating) ในโปรแกรม AP วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ความสามารถของผู้สอบ การกตคะแนนของผู้ตรวจ และเวลา โดยองค์ประกอบของเวลา (time facet) ถูกแบ่งเป็น 8 ช่วงเวลา จากเวลา 4 วัน จากการวิเคราะห์ข้อมูลโดยใช้ MFRM แบบแยกส่วน (separate model) เพื่อคำนวณค่าสหสัมพันธ์ระหว่างผู้ตรวจแต่ละคนกับกลุ่มผู้ตรวจ (single rater-rest of the raters (SR-ROR) correlation:  $r_{SR-ROR}$ ) และค่าสหสัมพันธ์ระหว่างส่วนเหลือกับคะแนนที่คาดหวัง (residuals-expected ratings (R-ER) correlation:  $r_{res,exp}$ ) ของผู้ตรวจแต่ละคนในแต่ละช่วงเวลา (8 ช่วง) และปรับค่าสหสัมพันธ์ให้เป็นค่ามาตรฐานด้วย Fisher's z transformation พิจารณานัยสำคัญของความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์ของผู้ตรวจแต่ละคนที่ระดับ .05 เพื่อ

ตรวจสอบการเปลี่ยนแปลงความถูกต้องและการใช้ระดับการให้คะแนนมาตรวัดของผู้ตรวจ ผลการศึกษาพบว่า ผู้ตรวจทำหน้าที่แตกต่างกันในการให้คะแนนเมื่อเวลาผ่านไป นั่นคือ เกิดการเปลี่ยนแปลงระดับความถูกต้องหรือระดับคะแนน มาตรวัดที่ใช้ โดยผู้ตรวจบางคนแสดงการเปลี่ยนแปลงอย่างมีนัยสำคัญทางสถิติของระดับความถูกต้องแม่นยำตามเกณฑ์การให้คะแนน ในขณะที่ผู้ตรวจคนอื่นๆ แสดงหลักฐานว่ามีการทำหน้าที่ต่างกันของระดับในมาตรประมาณที่ใช้เมื่อเวลาผ่านไป

Farrokhi และ Esfandiari (2011) ได้ใช้ MFRM ในการตรวจสอบอิทธิพลฮาโลของผู้ประเมินงานเขียนความเรียงภาษาอังกฤษใน 3 รูปแบบ ได้แก่ ผู้ประเมินตนเอง ผู้ประเมินเพื่อน และครูผู้ประเมินจำนวน 194 คน แบ่งเป็นผู้ประเมินตนเองและเพื่อนผู้ประเมินจำนวน 188 คน และครูผู้ประเมินจำนวน 6 คน ทำการตรวจประเมินงานเขียนความเรียงภาษาอังกฤษของนักศึกษาปริญญาตรี วิชาเอกภาษาอังกฤษ มหาวิทยาลัยรัฐบาลในประเทศอิหร่าน จำนวน 2 แห่ง โดยใช้มาตรประมาณค่าแบบแยกองค์ประกอบ (analytic rating scale) 6 ระดับ ซึ่งประกอบด้วยข้อรายการประเมินจำนวน 15 รายการ ที่ใช้ในการประเมินงานเขียนความเรียงของนักศึกษา วิเคราะห์ข้อมูล ด้วยโปรแกรม FACETS 3.68.0 กำหนดให้มีองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ผู้สอบ ผู้ประเมิน และข้อสอบ ผลการวิเคราะห์พบว่า ไม่มีอิทธิพลฮาโลในระดับกลุ่ม (group level) แต่พบในระดับบุคคล (individual level) นอกจากนี้ ยังพบว่ารูปแบบของผู้ประเมินมีความชัดเจนว่ามีอิทธิพลฮาโลในข้อสอบ 4 ข้อ และพบว่าผู้ประเมินตนเองแสดงอิทธิพล ฮาโลมากกว่าผู้ประเมินรูปอื่น

Farrokhi, Esfandiari, และ Dalili (2011) ใช้ MFRM ในการตรวจสอบอิทธิพลของแนวโน้มการให้คะแนนค่า ในการประเมินตนเอง การประเมินโดยเพื่อน และการประเมินโดยครูผู้ประเมินทั้งหมด 194 คน แบ่งเป็นผู้ประเมินตนเองและเพื่อนผู้ประเมินจำนวน 188 คน และครูผู้ประเมินจำนวน 6 คน ทำการตรวจประเมินงานเขียนความเรียงภาษาอังกฤษของนักศึกษาวิชาเอกภาษาอังกฤษ มหาวิทยาลัยรัฐบาลในประเทศอิหร่าน จำนวน 2 แห่ง โดยใช้มาตรประมาณค่า 6 ระดับ ประกอบด้วยข้อคำถาม 15 ข้อ วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS 3.68.1 กำหนดให้มีองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ผู้สอบ ผู้ประเมิน และข้อสอบ ตรวจสอบอิทธิพลของแนวโน้มการให้คะแนนค่ากลางของผู้ประเมินทั้งในระดับบุคคล และระดับกลุ่ม ผลการศึกษาพบว่า การประเมินโดยผู้ประเมินทั้ง 3 รูปแบบ ได้แก่ การประเมินตนเอง การประเมินโดยเพื่อน และการประเมินโดยครู ไม่แสดงให้เห็นว่ามีแนวโน้มการให้คะแนนค่ากลางของผู้ประเมินทั้งในระดับกลุ่มและระดับบุคคล

Kassim (2011) ได้ประยุกต์ใช้ MFRM ในการศึกษาพฤติกรรมการตัดสินใจและความคลาดเคลื่อนของผู้ตรวจประเมิน ได้แก่ การกตคะแนนของผู้ตรวจ อิทธิพลการจำกัดช่วง แนวโน้มการให้คะแนนค่ากลาง และความสอดคล้องภายใน (internal consistency) ผู้ตรวจจำนวน 34 คน ทำการตรวจให้คะแนนงานเขียน ด้วยมาตรประมาณค่า 10 ระดับ วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS กำหนด

องค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ผู้สอบ ผู้ตรวจ และข้อสอบ ผลการวิเคราะห์พบว่า ผู้ตรวจมีความแตกต่างกันอย่างมากของระดับการคาดคะเน และแสดงให้เห็นผลของการคาดคะเนที่มีต่อความถูกต้องในการประมาณค่าความสามารถของผู้เรียน และพบว่า มีผู้ตรวจจำนวน 3 คนที่มีค่า infit และ outfit ต่ำกว่า 0.60 แสดงว่าผู้ตรวจ over-fitting หรือผู้ตรวจมีความเป็นไปได้สูงที่จะแสดงลักษณะอิทธิพลการจำกัด

Wiseman (2012) ได้ทำการศึกษาเปรียบเทียบสมรรถนะของรูบริกแบบแยกส่วน (analytic scoring rubric) รูบริกแบบภาพรวม (holistic scoring rubric) ในการประเมินการเขียนภาษาอังกฤษเป็นภาษาที่สอง โดยใช้ MFRM เพื่อตรวจสอบว่ารูบริกทั้ง 2 แบบสามารถจำแนกผู้สอบได้หรือไม่ โดยผู้ตรวจให้คะแนนจำนวน 5 คน ทำการประเมินงานเขียนของผู้สอบจำนวน 60 คน ที่สุ่มจากผู้สอบเข้าศึกษาในวิทยาลัยชุมชน (community college) ซึ่งเป็นผู้ที่ไม่ใช่เจ้าของภาษา (non-native speakers of English) โดยในรอบแรกผู้ตรวจจะประเมินงานเขียนโดยใช้รูบริกแบบภาพรวม ที่มีระดับคะแนน 6 ระดับ และหลังจากนั้น 2 สัปดาห์ ผู้ตรวจจะประเมินงานเขียนอีกครั้งโดยใช้รูบริกแบบแยกส่วน ที่มีระดับคะแนน 6 ระดับ วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ผู้สอบ ผู้ตรวจ และข้อสอบผลการศึกษาพบว่า ในรูบริกแบบภาพรวม ผู้ตรวจแทบไม่ได้ใช้รูบริกในระดับต่ำเลย จึงเป็นหลักฐานที่สนับสนุนว่าอาจจะยุบระดับการให้คะแนนจาก 6 ระดับ ให้เหลือเพียง 5 ระดับ ในขณะที่รูบริกแบบแยกส่วนสามารถจำแนกความสามารถในการเขียนภาษาอังกฤษเป็นภาษาที่สองได้หลายช่วงความสามารถมากกว่ารูบริกแบบภาพรวม ซึ่งแสดงว่ารูบริกแบบแยกส่วนเป็นเครื่องมือในการประเมินเพื่อวินิจฉัยและจัดตำแหน่งของผู้สอบได้ดีกว่ารูบริกแบบภาพรวม

Hung, Chen, และ Chen (2012) ใช้ MFRM ในการตรวจสอบอิทธิพลของผู้ตรวจประเมินผลงานสร้างสรรค์ ตัวอย่างในการวิจัยเป็นนักศึกษาปริญญาตรีชั้นปีที่ 1 – 3 สาขาการออกแบบจำนวน 113 คน ซึ่งถูกแบ่งเป็น 4 กลุ่มทดลอง ด้วยวิธีการสุ่มแบบแบ่งชั้น ใช้รูปแบบการทดลองแบบ  $2 \times 2$  แฟกทอเรียล ประกอบด้วย 2 รูปแบบการสอน (ได้แก่ สอบแบบให้ตัวอย่าง และสอนแบบไม่ให้ตัวอย่าง) และ 2 เงื่อนไขจำกัดของงาน (ได้แก่ ให้เลือกใช้เพียง 3 วัสดุจากวัสดุทั้งหมด 36 อย่างในการออกแบบ และสามารถเลือกใช้วัสดุในการออกแบบทั้ง 36 อย่างได้โดยไม่จำกัด) ดังนั้น 4 กลุ่มทดลองจะได้รับเงื่อนไขที่แตกต่างกันดังนี้ กลุ่มที่ 1 ได้รับการสอนแบบไม่ให้ตัวอย่างและจำกัดวัสดุในการออกแบบ กลุ่มที่ 2 ได้รับการสอนแบบไม่ให้ตัวอย่างและสามารถเลือกใช้วัสดุได้ไม่จำกัด กลุ่มที่ 3 ได้รับการสอนแบบให้ตัวอย่างและจำกัดวัสดุในการออกแบบ และกลุ่มที่ 4 ได้รับการสอนแบบให้ตัวอย่างและสามารถเลือกใช้วัสดุได้ไม่จำกัด นักศึกษาทุกคนในแต่ละกลุ่มเงื่อนไขจะได้รับเวลา 20 นาที ในการออกแบบผลิตภัณฑ์งานสร้างสรรค์ มีผู้ประเมินผลงานจำนวน 6 คน โดยแต่ละกลุ่มจะได้รับการประเมินอย่างอิสระจากกันทางออนไลน์จากผู้ประเมิน 4 คน ที่ได้รับการสุ่มจากผู้ประเมิน

ทั้งหมด 6 คน ตามเกณฑ์ในการให้คะแนน 8 เกณฑ์ ด้วยเป็นมาตรฐานค่า 4 ระดับ วิเคราะห์ข้อมูลโดยใช้ MFRM เพื่อศึกษาอิทธิพลของกตหรือปล่อยคะแนน อิทธิพลการให้คะแนนค่ากลาง อิทธิพลฮาโล และอิทธิพลการให้คะแนนแบบสุ่ม กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ความสามารถของนักศึกษา ความยากของงาน และการกตคะแนนของผู้ตรวจ ผลการศึกษาพบว่า ไม่พบอิทธิพลของผู้ตรวจประเมิน และพบว่า มีปฏิสัมพันธ์ระหว่างผู้ตรวจประเมินกับเกณฑ์ในการประเมิน นั่นคือ ผู้ตรวจประเมินบางคนแสดงพฤติกรรมการกตคะแนนมากกว่าสิ่งทีระบุในเกณฑ์

Farrokhi, Esfandiari, และ Schaefer (2012) ใช้ MFRM ในการตรวจสอบอิทธิพลการทำหน้าที่แตกต่างกันของการกตหรือปล่อยคะแนนของผู้ตรวจ ของรูปแบบผู้ตรวจประเมิน 3 รูปแบบ ได้แก่ ผู้ประเมินตนเอง ผู้ประเมินเพื่อน และครูผู้ประเมิน จำนวนทั้งหมด 194 คน ทำการตรวจประเมินงานเขียนความเรียงภาษาอังกฤษของมหาวิทยาลัยรัฐบาลที่มีชื่อเสียง 2 แห่งในประเทศอิหร่าน จำนวน 188 คน ด้วยมาตรฐานค่า 6 ระดับ รูปแบบการประเมินเป็นแบบ fully crossed design วิเคราะห์ข้อมูลด้วยโปรแกรม Facets 3.68.1 กำหนดให้มีองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ผู้สอบ รูปแบบของผู้ตรวจ และข้อสอบ ผลการวิเคราะห์ปฏิสัมพันธ์ของความลำเอียงระหว่างรูปแบบการประเมิน 3 รูปแบบ กับข้อสอบ พบว่า ผู้ประเมินที่เป็นนักศึกษา (ทั้งผู้ประเมินตนเอง และผู้ประเมินเพื่อน) จะแสดงรูปแบบของการกตหรือปล่อยคะแนนที่ตรงข้ามกับการประเมินจากครู โดยผู้ประเมินที่เป็นนักศึกษาระดับมัธยมศึกษาจะแสดงลักษณะการปล่อยคะแนน ในขณะที่ครูผู้ประเมินจะแสดงการกตคะแนน ในข้อสอบข้อที่ 2, 3 และ 4 ส่วนข้อสอบข้อที่ 7 และ 10 นักศึกษาผู้ประเมินจะแสดงลักษณะการกตคะแนน แต่ครูผู้ประเมินจะแสดงการปล่อยคะแนน และจากการวิเคราะห์ปฏิสัมพันธ์ของความลำเอียงระหว่างรูปแบบการประเมิน 3 รูปแบบกับความสามารถของนักศึกษา พบว่าผู้ประเมินตนเองและครูผู้ประเมินจะแสดงรูปแบบการกตหรือปล่อยคะแนนที่ตรงกันข้ามกับผู้ประเมินเพื่อน เมื่อนักศึกษาอยู่ในกลุ่มที่มีความสามารถสูงมากหรือต่ำมาก

Anderson, Irvin, Alonzo, และ Tindal (2015) ประยุกต์ใช้ MFRM เพื่อควบคุมอิทธิพลของผู้ตรวจ ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับมาตรฐานผ่านระบบออนไลน์ ข้อสอบที่ใช้เป็นข้อสอบคณิตศาสตร์สำหรับผู้เรียนระดับ 6 – 8 จำนวน 1,345 ข้อ จากคลังข้อสอบขนาดใหญ่ (ทั้งหมด 2,700 ข้อ) ผู้ประเมินความสอดคล้องเป็นครูคณิตศาสตร์ที่มีประสบการณ์เกี่ยวกับมาตรฐานแกนกลาง (Common Core State Standards: CCSS) และการสอนคณิตศาสตร์ระดับมัธยมศึกษาตอนต้น จาก 15 โรงเรียน จำนวน 25 คน ทำการประเมินความสอดคล้องระหว่างข้อสอบกับมาตรฐานที่กำหนดผ่านทางระบบออนไลน์ที่เรียกว่า Distributed Item Review (DIR) ข้อสอบแต่ละข้อจะถูกประเมินด้วยผู้ประเมินจำนวน 3 คน และแต่ละคนประเมินข้อสอบคนละประมาณ 270 ข้อ โดยใช้มาตรฐานค่าแบบเรียงอันดับ 4 อันดับ (4-point

ordinal alignment scale) การวิเคราะห์ข้อมูลเพื่อตรวจสอบความสอดคล้องของผู้ประเมิน (rater consistency) และควบคุมการคาดคะเนของผู้ประเมิน และให้ค่าสถิติการวินิจฉัยในระดับข้อสอบด้วย MFRM โดยมีการปรับแก้เล็กน้อยเพื่อให้เหมาะสมกับข้อมูลการศึกษาความสอดคล้องในแนวเดียวกัน โดยที่ให้ข้อสอบเป็นสิ่งที่ถูกประเมินแทนที่จะเป็นผู้ถูกประเมิน และผู้ตรวจจะเป็นสิ่งที่แสดงถึงคุณลักษณะแฝง (latent trait) ของข้อคำถาม เนื่องจากเงื่อนไขของการวัดเป็นการปรับเทียบให้อยู่บนมาตรเดียว โดยตำแหน่งของข้อคำถามอยู่บนเงื่อนไขของการคาดคะเนของผู้ตรวจให้คะแนนแต่ละคน ความสอดคล้องภายในผู้ตรวจถูกประเมินด้วยการทดสอบส่วนเหลือ (residual) ระหว่างค่าที่คาดหวังจากโมเดลและค่าที่สังเกตได้ ซึ่งค่าที่สังเกตได้ขึ้นอยู่กับ (1) ผู้ตรวจให้คะแนนตัดสินข้อคำถามทั้งหมดที่เหลืออย่างไร (เช่น การคาดคะเนของผู้ตรวจให้คะแนน) และ (2) ผู้ตรวจให้คะแนนทั้งหมดที่เหลือตัดสินข้อคำถามอย่างไร (เช่น ความยากในการเลือกข้อคำถามที่สอดคล้อง) ถ้าค่าที่คาดหวังมีค่าใกล้เคียงกับค่าที่สังเกตได้ ถือว่าผู้ตรวจได้ให้คะแนนที่สอดคล้องกัน ผลการศึกษาพบว่า มีความแตกต่างอย่างมีนัยสำคัญในการศึกษาความสอดคล้องระหว่างข้อสอบกับมาตรฐาน เมื่อมีการใช้สถิติในการควบคุมอิทธิพลของผู้ตรวจ งานวิจัยนี้ยังได้เสนอแนะการประยุกต์ใช้วิธี Latent trait กับการศึกษาความสอดคล้องในแนวเดียวกันในวิธีการอื่น เช่น วิธีการของ Webb (1999) ในเกณฑ์ความสอดคล้องด้านความลึกของความรู้ (depth of knowledge) ระหว่างเนื้อหาในมาตรฐาน โดยเสนอแนะการให้คะแนนแบบมาตรเรียงอันดับจึงจะสามารถประยุกต์ใช้กับข้อสอบวิธี Latent trait ได้ ในกรณีนี้ความลึกของความรู้จะเป็นเงื่อนไขของการประเมิน และสามารถให้ข้อมูลเกี่ยวกับแต่ละเกณฑ์ประเมินความสอดคล้องตามวิธีการของ Webb ได้

Liu และ Xie (2014) ได้ใช้ MFRM ในการศึกษาอิทธิพลการปล่อย/กดยคะแนน แนวโน้มการให้คะแนนค่ากลาง อิทธิพลฮาโล และความแตกต่างของปล่อย/กดยคะแนนในตรวจประเมินงานเขียน (Written Discourse Completion Task: WDCT) จากการทดสอบปฏิบัติเพื่อวัดความรู้ด้านภาษาอังกฤษของผู้เรียนที่ใช้ภาษาอังกฤษเป็นภาษาต่างประเทศ (EFL) ในการทดสอบ WDCT ตัวอย่างในการศึกษาเป็นนักศึกษาสาขาวิชาภาษาอังกฤษที่ใช้ภาษาอังกฤษเป็นภาษาต่างประเทศ ของมหาวิทยาลัยในสาธารณรัฐประชาชนจีน จำนวน 38 คน ผู้ตรวจจำนวน 6 คน ทำการตรวจประเมินโดยใช้รูปกริดที่มีลักษณะเป็นมาตรฐานค่าแบบลิเคิร์ต ที่ประเมินใน 4 มิติ ได้แก่ ความถูกต้องในการพูด ความเหมาะสมในการใช้คำ จำนวนข้อมูลสารสนเทศที่ให้ และความเหมาะสมของรูปแบบและความเรียบร้อย วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ผู้สอบ ผู้ตรวจ และคุณลักษณะที่วัด ผลการศึกษาพบว่ามีความแตกต่างอย่างมีนัยสำคัญในการคาดคะเนของผู้ตรวจ โดยส่วนใหญ่มีแนวโน้มที่จะกดยคะแนน และพบว่ามีอิทธิพลฮาโลในผู้ตรวจบางคน นอกจากนี้ ผู้ตรวจส่วนใหญ่ยังแสดง ความลำเอียง/ปฏิสัมพันธ์กับทั้งคุณลักษณะ (trait) และกับผู้สอบ

Esfandiari (2015) ใช้ MFRM ในการตรวจสอบอติพิลฮาโลและอติธินวโน้มนำการให้คะแนนเข้าสู่ค่ากลางของผู้ประเมินเพื่อน (peer-assessors) ในตรวจประเมินการเขียนความเรียงภาษาอังกฤษของผู้เรียนที่ใช้ภาษาอังกฤษเป็นภาษาที่สอง (EFL) และเปรียบเทียบผลการตรวจประเมินของผู้ประเมินเพื่อน จำนวน 39 คน กับครูผู้ประเมิน (teacher assessors) จำนวน 8 คน เพื่อนผู้ประเมินทำการตรวจประเมินงานเขียนของนักศึกษาปริญญาโทที่เรียนวิชาการเขียนภาษาอังกฤษชั้นสูง จาก 2 มหาวิทยาลัยเอกชน จำนวน 39 คน ด้วยมาตรประมาณค่าแบบแยกองค์ประกอบ (analytic rating scale) 6 ระดับ ประกอบด้วยข้อรายการประเมินจำนวน 15 รายการ ออกแบบการตรวจประเมินแบบ full crossed design วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS องค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ เพื่อนผู้ประเมิน เกณฑ์การประเมิน และข้อสอบ ผลการศึกษาพบว่า เพื่อนผู้ประเมินไม่แสดงอติธินวโน้มนำการให้คะแนนเข้าสู่ค่ากลางและอติพิลฮาโลของผู้ประเมินในระดับกลุ่ม แต่พบในระดับบุคคล และจากการวิเคราะห์ความแตกต่างของการตรวจประเมินระหว่างเพื่อนผู้ประเมินและครูผู้ประเมินด้วย Mann Whitney U test พบว่า เพื่อนผู้ประเมินและครูผู้ประเมินมีการตรวจให้คะแนนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

Karakaya (2015) ใช้ MFRM ในการวิเคราะห์เปรียบเทียบการประเมินแฟ้มสะสมงานของนักศึกษาครูฝึกหัดระหว่างการประเมินโดยตนเอง โดยเพื่อน และโดยอาจารย์ผู้สอน ทำการศึกษาในนักศึกษาครูฝึกหัด จำนวน 74 คน ที่กำลังเรียนวิชาการวัดและประเมินและกำลังจัดเตรียมแฟ้มสะสมงานตามที่รายวิชากำหนด เครื่องมือในการประเมิน ได้แก่ ระบุริกที่ประกอบด้วยเกณฑ์การให้คะแนน 7 เกณฑ์ วิเคราะห์ข้อมูลด้วยโปรแกรม MINIFAC กำหนดองค์ประกอบความแปรปรวน (facet) 4 องค์ประกอบ ได้แก่ ผลสัมฤทธิ์ของนักศึกษาครู เพศของนักศึกษาครูฝึกหัด การกตคะแนนของผู้ประเมินตนเองผู้ประเมินเพื่อนและอาจารย์ผู้ประเมิน และเกณฑ์ที่ใช้ในการประเมินแฟ้มสะสมงาน ผลการศึกษาพบว่า นักศึกษาที่มีระดับผลสัมฤทธิ์สูงจะมีการจัดทำแฟ้มสะสมงานสูง และนักศึกษาหญิงจะมีความสามารถในการจัดทำแฟ้มสะสมงานสูงกว่านักศึกษาชาย และพบว่า การประเมินโดยตนเองจะมีลักษณะของการปล่อยคะแนนมากที่สุด ในขณะที่การประเมินโดยเพื่อนจะมีลักษณะของการกตคะแนนมากที่สุด นอกจากนี้ ยังพบความแตกต่างอย่างมีนัยสำคัญทางสถิติใน การให้คะแนนระหว่างผู้ประเมินแต่ละประเภท

Toffolia, Andradeb, และ Borniac (2015) ใช้ MFRM ในการวิเคราะห์คุณภาพของการให้คะแนนข้อคำถามแบบให้สร้างคำตอบ (open-response item) ในการเขียนความเรียงภาษาโปรตุเกส จำนวน 2 ข้อ ของ ผู้สอบจำนวน 350 คน มีการตรวจให้คะแนนจากผู้ตรวจ 2 คน และกลุ่มผู้ตรวจจำนวน 42 คน ประเมินตามเกณฑ์การให้คะแนน 5 ข้อ ที่ด้วยมาตรประมาณค่า 5 ระดับ วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS 3.71.4 กำหนดองค์ประกอบความแปรปรวน (facet) 4 องค์ประกอบ ได้แก่ ความสามารถของผู้เรียน ความยากของงาน ความกตคะแนนของผู้ตรวจ และ

ความยากของข้อสอบ ผลการศึกษาพบว่า มีความแตกต่างอย่างมีนัยสำคัญระหว่างผู้ตรวจที่มีการกตคะแนนต่างกัน โดยมีค่าอยู่ในช่วง -1.53 ถึง 0.91 กล่าวคือ ผู้ตรวจมีลักษณะการกตคะแนนที่แตกต่างกัน แสดงให้เห็นว่าผู้ตรวจแต่ละคนไม่สามารถทำหน้าที่แทนกันได้

Wesolowski, Wind, และ Engelhard (2015) ทำการศึกษาความเป็นธรรมในการประเมินความสามารถทางดนตรี โดยใช้โมเดลหลายองค์ประกอบของราส์ชแบบตรวจแบบให้คะแนนบางส่วน (many-facet Rasch partial credit (MFR-PC) measurement model) ในการตรวจสอบความสอดคล้องระหว่างโมเดลกับข้อมูล (model-data fit) และตรวจสอบการทำหน้าที่ต่างของผู้ประเมินใน 4 ระดับโรงเรียน ได้แก่ ระดับมัธยมต้น ระดับมัธยมปลาย ระดับมหาวิทยาลัย และระดับมืออาชีพ ผู้เชี่ยวชาญด้านดนตรีแจ๊ส จำนวน 24 คน ทำการประเมินความสามารถทางดนตรีแจ๊สจากเสียงบันทึกดนตรี 4 อัน ที่ไม่ได้ระบุให้ผู้เชี่ยวชาญทราบว่าเสียงดนตรีแต่ละอันอยู่ในระดับใด เครื่องมือที่ใช้ในการประเมินได้แก่ The Jazz Big Band Performance Rating Scale (JBBPRS) มีลักษณะเป็นมาตรประมาณค่า 4 ระดับ ประกอบด้วยข้อรายการในการประเมิน จำนวน 22 ข้อ วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS โดยกำหนดองค์ประกอบความแปรปรวน (facet) 4 องค์ประกอบ ได้แก่ ความสามารถ ข้อสอบ ผู้ประเมิน และระดับโรงเรียน ผลศึกษาพบว่า ผู้ประเมินมีการทำหน้าที่ต่างกันทั้งในระดับกลุ่มและระดับบุคคล โดยในภาพรวมพบว่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติของการกตคะแนนในกลุ่มผู้ประเมิน ( $\chi^2=141.10$ ,  $p<0.50$ ) และจากการวิเคราะห์ปฏิสัมพันธ์ระหว่างผู้ประเมินกับระดับโรงเรียน ทั้งหมด 96 คู่ พบว่ามี 14 คู่ (ร้อยละ 15.20) ที่พบนัยสำคัญทางสถิติ และพบว่ามีผู้ประเมินจำนวน 10 คน ที่แสดงการกตคะแนนที่แตกต่างกันอย่างน้อยใน 1 ระดับโรงเรียน

Wu และ Dou (2015) ได้ใช้ MFRM ในการตรวจสอบความถูกต้อง (validity) ของการทดสอบภาษาอังกฤษแบบปากเปล่า โดยการสอบปากเปล่าแบ่งเป็น 3 ส่วน ได้แก่ การอ่านออกเสียงการตอบคำถาม และการพูดสนทนา ผู้ประเมินจำนวน 10 คน จะทำหน้าที่ในการให้คะแนนผู้สอบที่เป็นนักเรียนมัธยมศึกษาตอนปลาย จำนวน 300 คน โดยผู้สอบแต่ละคนจะได้รับการประเมินจากผู้ประเมินจำนวน 2 คน ใน 5 ด้าน ได้แก่ เนื้อหา น้ำเสียงและการออกเสียง คำศัพท์ ไวยากรณ์ และกลยุทธ์การสื่อสาร ด้วยมาตรประมาณค่า วิเคราะห์ข้อมูล ด้วยโปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facet) 3 องค์ประกอบ ได้แก่ ผู้สอบ ผู้ประเมิน และความยากของข้อสอบ ผลการศึกษาพบว่า มีความแตกต่างกันอย่างมีนัยสำคัญของความสามารถในการพูดปากเปล่า และผู้ประเมินมีความสอดคล้องภายในตนเอง (internal self-consistency) ในการให้คะแนน และพบว่ามี ความแตกต่างอย่างมีนัยสำคัญของการกตคะแนน และข้อสอบมีระดับความยากที่แตกต่างอย่างมี



นัยสำคัญ ซึ่งความแตกต่างนี้มีเพียงพอที่จะจำแนกความสามารถผู้สอบได้ โดยสรุปคือการทดสอบภาษาอังกฤษแบบปากเปล่านี้มีความถูกต้อง

สำหรับในประเทศไทย มีการศึกษาวิจัยการประยุกต์ใช้ MFRM ในการศึกษาอิทธิพลของผู้ประเมินค่อนข้างน้อย จากการค้นคว้าพบว่า ชาญวิทย์ จรัสสุทธิอิศร (2550) ได้ประยุกต์ใช้ MFRM ในการแสดงหลักฐานความตรงเชิงโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์ โดยมีวัตถุประสงค์การวิจัยเพื่อแสดงหลักฐานความตรงเชิงโครงสร้างของการให้คะแนนทักษะกระบวนการทางวิทยาศาสตร์ชั้นบูรณาการ กลุ่มตัวอย่างที่ใช้ คือ นักเรียนชั้นมัธยมศึกษาปีที่ 4 สังกัดเขตพื้นที่การศึกษานครราชสีมา เขต 3 จำนวน 280 คน ที่สุ่มแบบ 2 ชั้นตอน ตรวจสอบให้คะแนนโดยผู้ตรวจจำนวน 14 คน ภายใต้เงื่อนไขการจับคู่ตรวจให้คะแนน การแสดงหลักฐานความตรงเชิงโครงสร้างของการให้คะแนนโดยการวิเคราะห์ด้วย MFRM ที่มี 3 องค์ประกอบ ตามแนวคิดของทฤษฎีการตอบข้อสอบ พบว่า องค์ประกอบทั้ง 3 ให้ค่าพารามิเตอร์อยู่ในเกณฑ์ที่ยอมรับได้ และจากการวิเคราะห์ค่าสถิติความสอดคล้องกับโมเดล พบว่า โมเดลการวิจัยมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ทั้งสามองค์ประกอบ

น้ำผึ้ง อินทเนตร (2554) ได้ศึกษาคุณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ เมื่อจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่างกัน โดยใช้โมเดลการสรุปอ้างอิง และ MFRM เพื่อศึกษาคุณลักษณะของแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ โดยใช้โมเดลการสรุปอ้างอิงและ MFRM ภายใต้เงื่อนไขจำนวนผู้ตรวจต่างกัน และรูปแบบการตรวจให้คะแนนแตกต่างกัน 3 ลักษณะ เครื่องมือที่ใช้คือแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ ระดับมัธยมศึกษาตอนต้น ตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จำนวน 12 ข้อ ตัวอย่างในการศึกษาคือ นักเรียนชั้นมัธยมศึกษาที่ 4 จังหวัดน่าน จำนวน 180 คน ซึ่งได้มาจากการสุ่มแบบสองชั้นตอน ผลการวิจัยพบว่า เมื่อวิเคราะห์ด้วยโมเดลการสรุปอ้างอิง เมื่อใช้รูปแบบการตรวจให้คะแนนเดียวกันในทุกเงื่อนไข จำนวนผู้ตรวจ ความแปรปรวนขององค์ประกอบเดียวกัน มีค่าใกล้เคียงกัน คะแนนในทุกเงื่อนไขที่แตกต่างกันมีความตรงตามสภาพสูงและแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อวิเคราะห์ด้วย MFRM ในทุกเงื่อนไขจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนน ความแปรปรวนของผู้สอบมีค่าสูงสุด รองลงมาคือ ความแปรปรวนของข้อสอบ และความแปรปรวนของผู้ตรวจมีค่าต่ำสุด

ศุภรัตน์ อิงชาติเจริญ (2557) ได้พัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานโดยประยุกต์ใช้ MFRM เพื่อวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ตัวอย่างในการศึกษา คือ อาจารย์และนักศึกษาในสถาบันอุดมศึกษาที่จัดการเรียนการสอนในรูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐาน เครื่องมือที่ใช้ในการวิจัย คือ แบบสอบถามคุณลักษณะของผู้ประเมินและแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยที่มีลักษณะเป็นมาตรวัดประมาณค่า 4 ระดับ (rating

scale) จำนวน 19 ข้อกลุ่มตัวอย่างที่ได้รับการประเมินกระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐาน (PBL) ประกอบด้วย 120 กลุ่มย่อย ในแต่ละกลุ่มย่อย PBL ประกอบด้วยผู้ประเมิน 3 กลุ่ม ได้แก่ อาจารย์ เพื่อน และตนเอง อย่างละ 1 คน โดยทำการสุ่มอย่างง่ายให้นักศึกษาในแต่ละกลุ่มย่อย จำนวน 2 คน สำหรับเป็นผู้ประเมินเพื่อน และผู้ประเมินตนเอง วิเคราะห์ข้อมูลด้วยโปรแกรม FACETS กำหนดองค์ประกอบความแปรปรวน (facets) 3 องค์ประกอบ ได้แก่ ผู้รับการประเมิน (ratee) กลุ่มผู้ประเมิน (rater) รายการประเมิน (items) ผลการศึกษาพบว่า ในภาพรวมผู้ประเมินทั้งสามกลุ่มมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนเพียงเล็กน้อย (measure in logits อยู่ระหว่าง -0.75 ถึง 0.47) โดยผู้ประเมินกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน ส่วนผู้ประเมินกลุ่มอาจารย์มีแนวโน้มกดคะแนนเป็นอันดับสองรองจากผู้ประเมินกลุ่มตนเอง และทั้งหมดถือว่าการให้คะแนนที่มีความแม่นยำ (infit MNSQ อยู่ระหว่าง 0.97 ถึง 1.05)

จากการศึกษาวิจัยที่เกี่ยวข้องกับ MFRM ทั้งในด้านตัวแปร/ประเด็นที่ศึกษา จำนวน องค์ประกอบความแปรปรวนหรือฟาเซท (facet) เครื่องมือที่ใช้ในการตรวจให้คะแนน จำนวน ตัวอย่าง จำนวนผู้ตรวจหรือผู้ประเมิน ตลอดจนอิทธิพลของผู้ตรวจ สามารถสรุปได้ดังตาราง 8

ตาราง 8 งานวิจัยที่เกี่ยวข้องกับการใช้โมเดลหลายองค์ประกอบของราร์ชในการตรวจสอบ  
อิทธิพลของผู้ตรวจ

ที่	ชื่อผู้วิจัย	ตัวแปรประเด็นใน/ การประเมิน	โมเดล	จำนวน facets	เครื่องมือ		จำนวน ตัวอย่าง	จำนวน ผู้ ประเมิน	อิทธิพลของผู้ตรวจ						
					rating scale	rubrics			severity/lenient	central tendency	halo effect	restriction of range	randomness	อื่นๆ	
1	Engelhard (1992)	ความสามารถในการเขียน	MFRM	4	✓		1,000	120	✓						
2	Engelhard (1994)	ความสามารถในการเขียน	MFRM	3	✓		264	15	✓	✓	✓	✓			
3	Zhu, Ennis, Chen (1998)	ความสอดคล้องระหว่างข้อคำถามกับเป้าหมายหลักสูตร	MFRM	6	✓		128	231	✓					consistency of experts' judgments	
4	Myford & Wolfe (2000)	ความสามารถในการพูดภาษาอังกฤษ	MFR-PCM	4	✓		1,469, 1,446	66, 74	✓	✓	✓	✓	✓	accuracy	
5	Myford (2002)	ความสามารถด้านทัศนศิลป์	MFRM	2	✓		50	11	✓						
6	Nakamura (2002)	ความสามารถในการเขียนความเรียง	MFRM	3	✓		36	4	✓						
7	Park (2004)	ความสามารถในการเขียน	MFR-PCM	3	✓	✓	99	17	✓						
8	Smith & Kulikowich (2004)	ทักษะการแก้ปัญหาที่ซับซ้อน	GT, MFRM	3	✓		44	2	✓						
9	Wolfe (2004)	ความสามารถในการเขียนความเรียง	MFRM	2	✓		28	101	✓	✓				accuracy/inaccuracy	
10	Eckes (2005)	ประสิทธิภาพในการเขียนและการพูด	MFRM	4	✓		1,359, 1,348	29	✓					gender bias	
11	Sudweeks, Reeve, & Bradshaw (2005)	ความสามารถในการเขียนความเรียง	GT, MFRM	4	✓		24	9	✓						
12	Iramaneerat & Yudkowsky (2007)	ทักษะทางคลินิก	MFRM	4	✓		173	17	✓		✓	✓	✓		
13	Basturk (2008)	ความสามารถในการนำเสนองานด้วยโปรแกรม PowerPoint	MFRM	3		✓	30	7	✓						
14	Eckes (2008)	ความสำคัญของเกณฑ์การประเมินการเขียน	MFRM	2	✓		ไม่ระบุ	65	✓						

ตาราง 8 (ต่อ)

ที่.	ชื่อผู้วิจัย	ตัวแปรประเด็นใน/การประเมิน	โมเดล	จำนวน facets	เครื่องมือ		จำนวน ตัวอย่าง	จำนวน ผู้ประเมิน	อิทธิพลของผู้ตรวจ					
					rating scale	rubrics			severity/lenient	central tendency	halo effect	restriction of range	randomness	อื่นๆ
15	Iramaneerat, Yudkowsky, Myford, & Downing (2008)	ทักษะทางคลินิก	GT, MFRM	3	✓		79	6	✓			✓	✓	
16	Myford & Wolfe (2009)	ความสามารถในการเขียนความเรียง	MFRM	3	✓		51,233	101						differential accuracy, differential scale category use
17	Farrokhi & Esfandiari (2011)	ความสามารถในการเขียนความเรียง	MFRM	3	✓		188	194			✓			
18	Farrokhi, Esfandiari, & Dalili (2011)	ความสามารถในการเขียนความเรียง	MFRM	3	✓		188	194		✓				
19	Kassim (2011)	ความสามารถในการเขียน	MFRM	3	✓		ไม่ระบุ	34	✓	✓		✓		internal consistency
20	Wiseman (2011)	ความสามารถในการเขียน	MFRM	3		✓	60	5	✓					
21	Hung, Chen, & Chen (2012)	ความสามารถในการออกแบบชิ้นงาน	MFRM	3	✓		113	4	✓	✓	✓		✓	interaction
22	Farrokhi, Esfandiari, & Schaefer (2012)	ความสามารถในการเขียนความเรียง	MFRM	3	✓		188	194	✓					differential leniency/severity
23	Anderson, Irvin, Alonzo, & Tindal (2015)	ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับมาตรฐาน	MFRM	2	✓		1,345	25	✓					
24	Liu & Xie (2014)	ความสามารถในการเขียน	MFRM	3		✓	38	6	✓	✓	✓			differential leniency/severity
25	Esfandiari (2015)	ความสามารถในการเขียนความเรียง	MFRM	3	✓		20	39, 8		✓	✓			
26	Karakaya (2015)	การประเมินแฟ้มสะสมงาน	MFRM	4	✓		74	74	✓					
27	Toffolia, Andradeb, & Borniac (2015)	ความสามารถในการเขียนความเรียง	MFRM	4	✓		350	44	✓					

ตาราง 8 (ต่อ)

ที่	ชื่อผู้วิจัย	ตัวแปรประเด็นใน/การประเมิน	โมเดล	จำนวน facets	เครื่องมือ		จำนวน ตัวอย่าง	จำนวน ผู้ประเมิน	อิทธิพลของผู้ตรวจ						
					rating scale	rubrics			severity/lenient	central tendency	halo effect	restriction of range	randomness	อื่นๆ	
28	Wesolowski, Wind, & Engelhard (2015)	ความสามารถทางดนตรีแจ๊ส	MFR-PCM	4	✓		4	24	✓						differential leniency/severity
29	Wu & Dou (2015)	ความสามารถในการสอบภาษาอังกฤษแบบปากเปล่า	MFRM	3	✓		300	10	✓						consistency
30	ชาญวิทย์ จรัสสุทธิอิศร (2550)	ความตรงเชิงโครงสร้างในการประเมินทักษะกระบวนการทางวิทยาศาสตร์	MFRM	3		✓	280	14							construct validity
31	นำผึ้ง อินทะเนตร (2554)	ศึกษาความตรงตามสภาพของแบบสอบปลายเปิด	GT, MFRM	3		✓	180	2-4							concurrent validity, raters variance
32	ศุภรัตน์ อิงชาติเจริญ (2557)	กระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐาน	MFRM	3	✓		120	360	✓						accuracy

จากการศึกษางานวิจัยที่เกี่ยวข้องกับ MFRM ตั้งแต่ปี ค.ศ. 1992 จนถึงปัจจุบัน พบว่างานวิจัยส่วนใหญ่เป็นการศึกษาทางด้านภาษา ได้แก่ ความสามารถในการเขียน Eckes, 2005; Eckes, 2005; Engelhard, 1992; Esfandiari, 2015; Farrokhi & Esfandiari, 2011; Farrokhi, Esfandiari, & Dalili, 2011; Farrokhi, Esfandiari, & Schaefer, 2012; Kassim, 2011; Liu & Xie, 2014; Myford & Wolfe, 2000, 2009; Nakamura, 2002; Park, 2004; Sudweeks, Reeve, & Bradshaw, 2005; Wiseman, 2011) ความสามารถในการพูด (Eckes, 2005; Myford & Wolfe, 2000; Wu & Dou, 2015) MFRM ได้ถูกนำมาใช้ในการวิเคราะห์ลักษณะของผู้ตรวจให้คะแนน (rater's characteristic) เป็นรายบุคคล และรายละเอียดของลักษณะผู้ตรวจในกระบวนการให้คะแนน เช่น การศึกษาเพื่อระบุถึงอิทธิพลการกดหรือปล่อยคะแนน (Engelhard, 1992; Linacre, 1994; Mnamara & Adams, 1991) ต่อมาการประยุกต์ใช้ MFRM ในการศึกษาด้านอื่นๆ ได้แก่ ทักษะทางคลินิก (Iramaneerat & Yudkowsky, 2007; Iramaneerat, Yudkowsky, Myford, & Downing, 2008) ความสร้างสรรค์ของการออกแบบชิ้นงาน (Hung, Chen, & Chen, 2012)

ความสามารถด้านทัศนศิลป์ (Myford, 2002) ความสามารถทางดนตรีแจ๊ส (Wesolowski, Wind, & Engelhard (2015) ความสามารถในการนำเสนองานด้วยโปรแกรม PowerPoint (Basturk, 2008) ความสำคัญของเกณฑ์การประเมินการเขียน (Eckes, 2008) ทักษะการแก้ปัญหาที่ซับซ้อน (Smith & Kulikowich, 2004) กระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐาน (ศุภรัตน์ อิงชาติเจริญ, 2557) ทักษะกระบวนการทางวิทยาศาสตร์ (ชาญวิทย์ จรัสสุทธิอิสร์, 2550) รวมถึงการประยุกต์ใช้โมเดลในการตรวจสอบคุณภาพของเครื่องมือ (ชาญวิทย์ จรัสสุทธิอิสร์, 2550; น้ำผึ้ง อินทะเนตร, 2554)

งานวิจัยเกือบทั้งหมดเป็นการศึกษาผลการวิเคราะห์โดยใช้ MFRM (Anderson, Irvin, Alonzo, & Tindal, 2015; Basturk, 2008; Eckes, 2005; Eckes, 2008; Engelhard, 1992, 1994; Esfandiari, 2015; Farrokhi & Esfandiari, 2011; Farrokhi, Esfandiari, & Dalili, 2011; Farrokhi, Esfandiari, & Schaefer, 2012; Hung, Chen, & Chen, 2012; Iramaneerat & Yudkowsky, 2007; Karakaya, 2015; Kassim, 2011; Liu & Xie, 2014; Toffolia, Andradeb, & Borniac, 2015; Myford & Wolfe, 2009; Myford, 2002; Nakamura, 2002; Wiseman, 2011; Wolfe, 2004; Wu & Dou, 2015; Zhu, Ennis, & Chen, 1998; ชาญวิทย์ จรัสสุทธิอิสร์, 2550; ศุภรัตน์ อิงชาติเจริญ, 2557) รองลงมาคือ งานวิจัยที่ใช้ MFRM แบบตรวจให้คะแนนบางส่วน (many-facet Rasch partial credit (MFR-PC) measurement model) (Myford & Wolfe, 2000; Park, 2004; Wesolowski, Wind, & Engelhard, 2015) นอกจากนี้ ยังพบว่ามีการวิจัยที่เป็นการศึกษาเปรียบเทียบผลการวิเคราะห์ระหว่าง MFRM กับอื่นๆ ได้แก่ เปรียบเทียบกับการใช้ทฤษฎีการสรุปอ้างอิง (generalizability theory) (Iramaneerat, Yudkowsky, Myford, & Downing, 2008; Smith & Kulikowich, 2004; Sudweeks, Reeve, & Bradshaw, 2005; น้ำผึ้ง อินทะเนตร, 2554)

งานวิจัยที่พบเป็นการศึกษาแหล่งองค์ประกอบความแปรปรวน (facet) ตั้งแต่ 2-6 องค์ประกอบ โดยส่วนใหญ่เป็นการศึกษาองค์ประกอบความแปรปรวน จำนวน 3 องค์ประกอบ (Engelhard, 1994; Basturk, 2008; Esfandiari, 2015; Farrokhi & Esfandiari, 2011; Farrokhi, Esfandiari, & Dalili, 2011; Farrokhi, Esfandiari, & Schaefer, 2012; Hung, Chen, & Chen, 2012; Iramaneerat, Yudkowsky, Myford, & Downing, 2008; Kassim, 2011; Liu & Xie, 2014; Myford & Wolfe, 2009; Nakamura, 2002; Park, 2004; Smith & Kulikowich, 2004; Wiseman, 2011; Wu & Dou, 2015; ชาญวิทย์ จรัสสุทธิอิสร์, 2550; น้ำผึ้ง อินทะเนตร, 2554; ศุภรัตน์ อิงชาติเจริญ, 2557) รองลงมาคือ 4 องค์ประกอบ (Engelhard, 1992; Eckes, 2005; Iramaneerat & Yudkowsky, 2007; Karakaya, 2015; Toffolia, Andradeb, & Borniac, 2015; Myford & Wolfe, 2000; Sudweeks, Reeve, & Bradshaw, 2005; Wesolowski, Wind, &

Engelhard, 2015) 2 องค์ประกอบ (Anderson, Irvin, Alonzo, & Tindal, 2015; Myford, 2002; Wolfe, 2004; Eckes, 2008) และจำนวน 6 องค์ประกอบ (Zhu, Ennis, & Chen, 1998)

เมื่อพิจารณาจำนวนตัวอย่างในการวิจัยซึ่งอาจเป็นผู้สอบ ผู้ถูกประเมิน หรือข้อมูลที่ถูกนำมาประเมิน มีจำนวนหลากหลายตามระดับของการวิจัย เนื่องจากการวิเคราะห์โดยใช้ MFRM มีตั้งแต่เป็นงานวิจัยระดับเล็กไปจนถึงงานวิจัยระดับใหญ่ เช่น การประเมินในระดับใหญ่ (large-scale assessment) โดยมีตัวอย่างวิจัย มีจำนวนตั้งแต่ 4 – 5,919 ตัวอย่าง สามารถจำแนกเป็นงานวิจัยที่มีตัวอย่างจำนวนไม่เกิน 50 ตัวอย่าง (Basturk, 2008; Liu & Xie, 2014; Esfandiari, 2015; Myford, 2002; Nakamura, 2002; Sudweeks, Reeve, & Bradshaw, 2005; Smith & Kulikowich, 2004; Wesolowski, Wind, & Engelhard, 2015; Wolfe, 2004) จำนวน 51-100 ตัวอย่าง (Iramaneerat, Yudkowsky, Myford, & Downing, 2008; Karakaya, 2015; Park, 2004; Wiseman, 2011) จำนวน 100-500 ตัวอย่าง (Engelhard, 1994; Zhu, Ennis, & Chen, 1998; Farrokhi & Esfandiari, 2011; Farrokhi, Esfandiari, & Dalili, 2011; Farrokhi, Esfandiari, & Schaefer, 2012; Hung, Chen, & Chen, 2012; Iramaneerat & Yudkowsky, 2007; Toffolia, Andradeb, & Borniac, 2015; Wu & Dou, 2015; ชาญวิทย์ จรัสสุทธิอิสร์, 2550; น้ำผึ้ง อินทเนตร, 2554; ศุภรัตน์ อิงชาติเจริญ, 2557) และจำนวนมากกว่า 1,000 ตัวอย่าง (Anderson, Irvin, Alonzo, & Tindal, 2015; Eckes, 2005; Engelhard, 1992; Myford & Wolfe, 2000, 2009)

เมื่อพิจารณาจำนวนผู้ตรวจหรือผู้ประเมินพบว่า มีจำนวนหลากหลายตามระดับของการวิจัยเช่นกัน โดยมีจำนวนตั้งแต่ 2 – 1,420 คน โดยงานส่วนใหญ่มีจำนวนผู้ตรวจหรือผู้ประเมินไม่เกิน 20 คน (Basturk, 2008; Engelhard, 1994; Hung, Chen, & Chen, 2012; Iramaneerat & Yudkowsky, 2007; Iramaneerat, Yudkowsky, Myford, & Downing, 2008; Liu & Xie, 2014; Myford, 2002; Nakamura, 2002; Park, 2004; Smith & Kulikowich, 2004; Sudweeks, Reeve, & Bradshaw, 2005; Wiseman, 2011; Wu & Dou, 2015; ชาญวิทย์ จรัสสุทธิอิสร์, 2550; น้ำผึ้ง อินทเนตร, 2554) รองลงมา คือ มีจำนวนผู้ตรวจหรือผู้ประเมินมากกว่า 20 คน แต่ไม่เกิน 100 คน (Anderson, Irvin, Alonzo, & Tindal, 2015; Eckes, 2005, 2008; Kassim, 2011; Esfandiari, 2015; Karakaya, 2015; Toffolia, Andradeb, & Borniac, 2015; Myford & Wolfe, 2000; Wesolowski, Wind, & Engelhard, 2015) และมีจำนวนผู้ตรวจหรือผู้ประเมินมากกว่า 100 คน ซึ่งมักเป็นงานวิจัยที่เป็น การประเมินในระดับใหญ่ (large-scale assessment) (Engelhard, 1992; Myford & Wolfe, 2009; Wolfe, 2004; Zhu, Ennis, & Chen, 1998) หรือเป็นการเปรียบเทียบผลของผู้ประเมินหลายกลุ่ม เช่น อาจารย์ เพื่อน และตนเอง (Farrokhi & Esfandiari, 2011; Farrokhi, Esfandiari, & Dalili, 2011; Farrokhi, Esfandiari, & Schaefer, 2012; ศุภรัตน์ อิงชาติเจริญ, 2557)

งานวิจัยส่วนใหญ่เครื่องมือที่ใช้ในการตรวจให้คะแนนที่มีลักษณะเป็นมาตราประมาณค่า (rating scale) (Engelhard, 1992, 1994; Anderson, Irvin, Alonzo, & Tindal, 2015; Eckes, 2005, 2008; Esfandiari, 2015; Farrokhi & Esfandiari, 2011; Farrokhi, Esfandiari, & Dalili, 2011; Farrokhi, Esfandiari, & Schaefer, 2012; Hung, Chen, & Chen, 2012; Iramaneerat & Yudkowsky, 2007; Iramaneerat, Yudkowsky, Myford, & Downing, 2008; Karakaya, 2015; Kassim, 2011; Toffolia, Andradeb, & Borniac, 2015; Myford & Wolfe, 2000, 2009; Myford, 2002; Nakamura, 2002; Park, 2004; Smith & Kulikowich, 2004; Sudweeks, Reeve, & Bradshaw, 2005; Wesolowski, Wind, & Engelhard, 2015; Wolfe, 2004; Wu & Dou, 2015; Zhu, Ennis, & Chen, 1998) และมีงานวิจัยบางส่วนที่ใช้เกณฑ์การให้คะแนนแบบ รุปริก (scoring rubric) (Basturk, 2008; Wiseman, 2011; Liu & Xie, 2014; ชาญวิทย์ จรัสสุทธิ อิศร, 2550; น้ำผึ้ง อินทเนตร, 2554)

งานวิจัยส่วนใหญ่เป็นการใช้ MFRM ในการตรวจสอบอิทธิพลของผู้ตรวจด้านอิทธิพลการกด หรือปล่อยคะแนน (severity/leniency effect) (Anderson, Irvin, Alonzo, & Tindal, 2015; Eckes, 2005, 2008; Engelhard, 1992, 1994; Farrokhi, Esfandiari, & Schaefer, 2012; Hung, Chen, & Chen, 2012; Iramaneerat & Yudkowsky, 2007; Iramaneerat, Yudkowsky, Myford, & Downing, 2008; Karakaya, 2015; Kassim, 2011; Toffolia, Andradeb, & Borniac, 2015; Myford & Wolfe, 2000; Myford, 2002; Nakamura, 2002; Park, 2004; Wolfe, 2004; Wesolowski, Wind, & Engelhard, 2015) รองลงมาคือ อิทธิพลแนวโน้มการให้คะแนนค่ากลาง (centrality tendency effect) (Engelhard, 1992, 1994; Esfandiari, 2015; Farrokhi, Esfandiari, & Dalili, 2011; Hung, Chen, & Chen, 2012; Kassim, 2011; Liu & Xie, 2014; Myford & Wolfe, 2000; Wolfe, 2004) อิทธิพลฮาโล (halo effect) (Engelhard, 1994; Esfandiari, 2015; Farrokhi & Esfandiari, 2011; Hung, Chen, & Chen, 2012; Iramaneerat & Yudkowsky, 2007; Liu & Xie, 2014; Myford & Wolfe, 2000; Wolfe, 2004) อิทธิพลการจำกัด ช่วง (restrictive of range) (Engelhard, 1994; Kassim, 2011; Iramaneerat & Yudkowsky, 2007; Iramaneerat, Yudkowsky, Myford, & Downing, 2008; Myford & Wolfe, 2000) อิทธิพลของความไม่มีแบบแผน (randomness effect) (Esfandiari, 2015; Farrokhi, Esfandiari, & Dalili, 2011; Iramaneerat & Yudkowsky, 2007; Iramaneerat, Yudkowsky, Myford, & Downing, 2008; Hung, Chen, & Chen, 2012; Myford & Wolfe, 2000) นอกจากนี้ ยังพบว่ามีการศึกษาอิทธิพลของผู้ตรวจในลักษณะอื่นๆ ได้แก่ ความไม่แม่นยำ (inaccuracy) (Wolfe, 2004) ความลำเอียงด้านเพศ (gender bias) (Eckes, 2005) ปฏิสัมพันธ์ระหว่างผู้ตรวจกับผู้สอบ และ ผู้ตรวจกับเกณฑ์การให้คะแนน (Hung, Chen, & Chen, 2012) การทำหน้าที่ต่างกันของผู้ให้คะแนน



เมื่อเวลาผ่านไป (differential rater functioning over time: DRIFT) ได้แก่ การทำหน้าที่ต่างกันของการกดหรือปล่อยคะแนน (differential leniency/ severity) (Liu & Xie, 2014; Wesolowski, Wind, & Engelhard, 2015) การทำหน้าที่ต่างกันของผู้ตรวจในด้านความแม่นยำ (differential accuracy) (Wolfe, 2004) การทำหน้าที่ต่างกันของผู้ตรวจในด้านการใช้มาตรวัด (differential scale category use) (Wolfe, 2004) นอกจากนี้ ยังพบว่ามีการตรวจสอบความสอดคล้องของผู้ประเมิน (consistency) (Kassim, 2011; Zhu, Ennis, & Chen, 1998; Wu & Dou, 2015) ความแม่นยำ (accuracy) (Myford & Wolfe, 2000; Wolfe, 2004; ศุภรัตน์ อิงชาติเจริญ, 2557)

จากการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ด้วย MFRM พบว่า โมเดลหลายองค์ประกอบของราส์สามารถนำมาประยุกต์ใช้กับการวิเคราะห์ข้อมูลจากการศึกษาความสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบทางการศึกษา ซึ่งในการวิจัยนี้ผู้วิจัยจะดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางขั้นพื้นฐาน พุทธศักราช 2551 กับการประเมินผลการเรียนรู้ อันได้แก่ ข้อสอบที่ครูในโรงเรียนสร้างขึ้น เพื่อประเมินผลผู้เรียน ในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาชั้นปีที่ 3 โดยผู้ประเมินใช้มาตรประมาณค่าเป็นเครื่องมือการประเมินความสอดคล้อง วิเคราะห์ข้อมูลโดยประยุกต์ใช้ MFRM กำหนดองค์ประกอบความแปรปรวน 2 องค์ประกอบ ได้แก่ ข้อสอบ (items) และผู้ตรวจประเมิน (raters) ซึ่งเป็นโมเดลราส์ที่มีการปรับแก้เล็กน้อยเพื่อให้เหมาะสมกับข้อมูลการศึกษาความสอดคล้องในแนวเดียวกัน โดยกำหนดที่ให้ข้อสอบ (items) เป็นสิ่งที่ถูกประเมินแทนที่จะเป็นผู้ถูกประเมิน (persons) และผู้ตรวจ (raters) จะเป็นสิ่งที่แสดงถึงคุณลักษณะแฝง (latent trait) ของข้อสอบ ความคลาดเคลื่อนหรืออิทธิพลของผู้ตรวจในการศึกษานี้ ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน

## ตอนที่ 7 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (generalizability theory) หรือ G-theory เป็นทฤษฎีทางสถิติของการวิเคราะห์ความน่าเชื่อถือของผลการวัดในสถานการณ์ของการวัดลักษณะต่างๆ ที่เป็นเป้าหมายของการนำเครื่องมือไปใช้ ซึ่งทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดเป็นทั้งวิธีการประเมินที่น่าเชื่อถือ หรือความเที่ยงของเครื่องมือและกลยุทธ์ของการออกแบบการวัดให้ผลของการวัดมีความน่าเชื่อถือ หรือมีความเที่ยงสูงถึงระดับที่ต้องการ เพื่อนำผลไปใช้เป็นสารสนเทศสำหรับการตัดสินใจอย่างมีประสิทธิภาพ

## 7.1 แนวคิดพื้นฐานของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ครอนบาค และคณะ (1972 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) ได้เสนอทฤษฎีสำหรับการวิเคราะห์ความน่าเชื่อถือของผลการวัดพฤติกรรม (the dependability of behavioral measurement) สำหรับสถานการณ์ของการวัดผลลักษณะต่างๆ ซึ่งต่อมาเป็นที่รู้จักในชื่อของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ทฤษฎีนี้ได้ขยายแนวคิดของความเที่ยงในทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory) ให้สามารถสรุปความเที่ยงในสถานการณ์หรือเงื่อนไขการวัดในลักษณะต่างๆ ได้ ทฤษฎีการทดสอบแบบดั้งเดิมไม่ได้ให้ความสนใจต่อสถานการณ์หรือเงื่อนไขการวัดและปฏิสัมพันธ์ระหว่างผู้สอบกับเงื่อนไขของการวัดที่สามารถส่งผลต่อความน่าเชื่อถือของผลการวัดและเสนอว่ามีความคลาดเคลื่อนจากการวัดแหล่งเดียว แต่ G-theory ได้เสนอว่ามีความคลาดเคลื่อนมากกว่าหนึ่งแหล่ง (multiple error source) โดยแยกส่วนความคลาดเคลื่อนของการวัดเป็นสองแหล่งประกอบด้วยความคลาดเคลื่อนอย่างมีระบบ (systematic source) และความคลาดเคลื่อนสุ่ม (random source)

7.1.1 โมเดล ข้อตกลงเบื้องต้น แหล่งความคลาดเคลื่อนหรือแหล่งความแปรปรวน และจำนวนพาเซทหรือองค์ประกอบ ดังนี้

### 7.1.1.1 โมเดล

โมเดลของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดเป็นโมเดลเชิงสุ่มถือว่าเงื่อนไขการวัดที่นำมาศึกษานั้นได้มาจากการสุ่มแบบง่ายหรือสุ่มแบบแบ่งชั้น จึงเป็นทฤษฎีที่ให้สารสนเทศที่สามารถนำไปสรุปอ้างอิงไปยังเงื่อนไขการวัดอื่นๆ ในเอกภพเดียวกันได้ การประยุกต์ใช้ทฤษฎีนี้โดยการวิเคราะห์ความแปรปรวนศึกษารูปแบบที่เหมาะสมต่อการวัด โดยทำการประมาณค่าความแปรปรวนจากแหล่งต่างๆ ด้วยการวิเคราะห์ความแปรปรวน (analysis of variance: ANOVA)

### 7.1.1.2 ข้อตกลงเบื้องต้น

การประมาณค่าความเที่ยงด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดไม่มีข้อตกลงว่าการวัดมีความเท่าเทียมกัน ในลักษณะของความเป็นคู่ขนาน เป็นทฤษฎีที่ยินยอมให้ออกแบบเก็บรวบรวมข้อมูลให้เหมาะสมกับปัญหาของการวัดผล

### 7.1.1.3 แหล่งความคลาดเคลื่อนหรือแหล่งความแปรปรวน

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดประมาณค่าความคลาดเคลื่อนของการวัดได้หลายแหล่ง โดยประมาณค่าความแปรปรวนได้จากการวิเคราะห์ครั้งเดียว จากนั้นประมาณค่าความแปรปรวนจากแหล่งต่างๆ ที่เป็นไปได้ทั้งหมดทั้งจากผลหลัก (main effect) และผลร่วมหรือปฏิสัมพันธ์ (interaction effect) ระหว่างองค์ประกอบด้วย เพื่อให้ผลการวิเคราะห์มีความถูกต้อง แม่นยำ และได้สารสนเทศที่เพียงพอสำหรับการตัดสินใจเลือกแบบวัดที่มีประสิทธิภาพ

#### 7.1.1.4 จำนวนพาเซทหรือองค์ประกอบ

ทฤษฎีการสุ่มอ้างอิงความน่าเชื่อถือของผลการวัดสามารถศึกษาได้หลายพาเซทพร้อมกัน รวมทั้งสามารถจำแนกแหล่งความแปรปรวนที่เป็นไปได้ทั้งหมด ได้ค่าความแปรปรวนของความคลาดเคลื่อนที่ถูกต้อง ทำให้ประมาณค่าความเที่ยงได้ถูกต้องแม่นยำ สามารถปรับรูปแบบการวัดเพื่อให้ได้ค่าความเที่ยงที่ดีขึ้นได้ในหลายสถานการณ์วัด โดยการเพิ่มหรือลดจำนวนพาเซทต่างๆ เช่น จำนวนข้อสอบ จำนวนครั้งในการวัด จำนวนผู้ตรวจหรือจำนวนผู้ประเมิน เป็นต้น

7.1.2 ความรู้พื้นฐานเกี่ยวกับโครงสร้างของการศึกษาทฤษฎีการสุ่มอ้างอิงความน่าเชื่อถือของผลการวัด ได้เสนอคำศัพท์ที่ใช้ในทฤษฎีการสุ่มอ้างอิง ดังนี้

7.1.2.1 ประชากร (population) หมายถึง สิ่งที่เป็นเป้าหมายของการวัดหรือสิ่งที่มุ่งวัดทั้งหมด (object of measurement) ในสถานการณ์ของการทดสอบโดยทั่วไป สิ่งที่มีมุ่งวัดได้แก่ ผู้สอบ บุคคล นักเรียนรายบุคคล หรือนักเรียนทั้งชั้น

7.1.2.2 องค์ประกอบ (facet) หมายถึง ชุดของเงื่อนไขของการวัดที่มีลักษณะคล้ายกันหรือกลุ่มเงื่อนไขของการวัด เป็นองค์ประกอบที่คาดว่าจะมีผลต่อการวัดความคลาดเคลื่อน เช่น องค์ประกอบของจำนวนข้อสอบ (item facet) องค์ประกอบของจำนวนครั้งของการสอบ (occasion facet) องค์ประกอบของผู้ตรวจข้อสอบ (rater facet)

7.1.2.3 เงื่อนไขของการวัด (condition of measurement) หมายถึง ระดับขององค์ประกอบที่ทำให้ได้ค่าสังเกตแต่ละค่าในการวัดครั้งหนึ่ง เช่น องค์ประกอบหรือพาเซทของจำนวนผู้ตรวจ กำหนดระดับเงื่อนไขการวัดเป็น 1, 2 และ 3 คน องค์ประกอบ/พาเซทความยาวของแบบสอบ (จำนวนข้อของแบบสอบ) กำหนดระดับเงื่อนไขการวัดเป็น 10, 20 และ 30 ข้อ

7.1.2.4 รูปแบบความสัมพันธ์ระหว่างองค์ประกอบ หมายถึง องค์ประกอบเป็นสิ่งที่ชี้ให้เห็นแหล่งความแปรปรวนของรูปแบบการวัดในแบบจำลองการวิเคราะห์ความแปรปรวน รูปแบบความสัมพันธ์แบ่งออกเป็น 3 รูปแบบ ดังนี้

(1) ความสัมพันธ์แบบไขว้ (crossed) คือ ลักษณะความสัมพันธ์ที่แต่ละระดับของสิ่งที่ถูกวัดนั้น ถูกวัดภายใต้เงื่อนไขเดียวกันทั้งหมด ใช้สัญลักษณ์ “x” คือ “crossed with” เช่น กำหนดรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น  $p \times i \times r$  หมายถึง ผู้สอบ (p) ทำข้อสอบ (i) ทุกข้อ และผู้ตรวจ (r) ตรวจข้อสอบทุกข้อ ของผู้สอบทุกคน

(2) ความสัมพันธ์แบบแฝง (nested) คือ ลักษณะความสัมพันธ์ที่แต่ละระดับของสิ่งที่ถูกวัดนั้น ถูกวัดภายใต้เงื่อนไขที่แตกต่างกัน ใช้สัญลักษณ์ “:” คือ “nested within” เช่น กำหนดรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น  $i : r$  หมายถึง ผู้ตรวจ (r) ตรวจข้อสอบ (i) ต่างข้อกัน

(3) ความสัมพันธ์แบบผสม (confounded) คือ ลักษณะความสัมพันธ์ที่มีทั้งความสัมพันธ์แบบไขว้และความสัมพันธ์แบบแฝงรวมกันอยู่ เช่น กำหนดรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น  $p \times (i : t)$  หมายถึง ผู้สอบ (p) ทำข้อสอบ (i) บางข้อที่อยู่ในแบบสอบ (t)

7.1.2.5 เอกภพ (universe) หมายถึง เงื่อนไขการวัด (condition of measurement) ทั้งหมด โดยกลุ่มเงื่อนไขการวัดเรียกว่าฟาเซท (facet) ซึ่งเป็นองค์ประกอบที่คาดว่า มีผลต่อความคลาดเคลื่อนของการวัด เช่น ความยาวของแบบสอบ รูปแบบของข้อสอบ จำนวนครั้งของการสอบ จำนวนผู้ตรวจ

7.1.2.6 คะแนนเอกภพ (universe Score) หมายถึง คะแนนของเงื่อนไขการวัดทั้งหมดของทุกองค์ประกอบซึ่งคล้ายกับคะแนนจริง (Ideal score) ของสิ่งที่วัดในทฤษฎีการทดสอบแบบดั้งเดิม

7.1.2.7 เอกภพของการสังเกตที่ยอมรับได้ (university of admission observation) หมายถึง กลุ่มเงื่อนไขของการวัดที่เป็นไปได้ ซึ่งสามารถวัดหรือสังเกตได้ในแต่ละฟาเซท เช่น ฟาเซท ข้อสอบ และฟาเซทของผู้ตรวจ เป็นต้น

7.1.2.8 เอกภพการสรุปอ้างอิง (universe of generalization) หมายถึง สถานการณ์หรือเงื่อนไขการวัดอันเป็นเป้าหมายของการสรุปอ้างอิง กล่าวได้ว่า เป็นการวัดครอบคลุมเงื่อนไขที่สนใจทั้งหมด หรือเงื่อนไขในเอกภพของการสังเกตที่ยอมรับได้ทั้งหมด ซึ่งอาจประกอบด้วย เซตย่อยของเงื่อนไขในเอกภพของการสังเกตได้

7.1.3 การศึกษาเพื่อการสรุปอ้างอิง (G study) และการศึกษาเพื่อการตัดสินใจ (D study)

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ประกอบด้วยการศึกษา 2 ขั้นตอนที่สำคัญคือ การศึกษาเพื่อการสรุปอ้างอิง (generalizability study: G study) การศึกษาเพื่อการตัดสินใจ (decision study: D study) มีรายละเอียดดังนี้

7.1.3.1 การศึกษาเพื่อการสรุปอ้างอิง (generalizability study: G study) หมายถึง การวิเคราะห์สรุปอ้างอิงเพื่อประมาณค่าความแปรปรวนจากแหล่งต่างๆ ของรูปแบบการวัด โดยเริ่มจากการกำหนดองค์ประกอบต่างๆ กำหนดความสัมพันธ์ขององค์ประกอบ และกำหนดเอกภพของการสังเกตที่ยอมรับได้ และการประมาณค่าความแปรปรวนจากแหล่งต่างๆ

7.1.3.2 การศึกษาเพื่อการตัดสินใจ (decision study: D study) หมายถึง การศึกษาเพื่อประเมินวิธีการวัดที่เหมาะสม และตัดสินใจ ซึ่งเป็นขั้นสุดท้ายของการวิเคราะห์การสรุปอ้างอิง เริ่มจากระบุสิ่งที่ต้องการวัด เอกภพการสรุปอ้างอิง จำนวนเงื่อนไขของการวัด หรือขนาดกลุ่มตัวอย่างของแต่ละองค์ประกอบ และรูปแบบของการวัด จากนั้นคำนวณค่าความแปรปรวนจากการศึกษาสัมประสิทธิ์หรือความเที่ยงของแบบสอบ ขนาดกลุ่มตัวอย่างเพื่อการตัดสินใจไม่จำเป็นต้อง

เท่ากับขนาดกลุ่มตัวอย่างในการศึกษาเพื่อการสรุปอ้างอิง แต่ขึ้นอยู่กับนำไปใช้เพื่อการตัดสินใจเกี่ยวกับสิ่งที่วัดซึ่งกำหนดโดยมีกลุ่มตัวอย่างในการศึกษาเพื่อการสรุปอ้างอิงเป็นฐาน

#### 7.1.4 ความแปรปรวนของคะแนนความคลาดเคลื่อน

ความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่างๆ ของการวัดแบ่งออกเป็น 2 ประเภท คือ ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (absolute error variance:  $\sigma_{Abs}^2$  หรือ  $\sigma_{\Delta}^2$ ) และ ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ (relative error variance:  $\sigma_{Rel}^2$  หรือ  $\sigma_{\delta}^2$ ) มีรายละเอียด ดังนี้

7.1.4.1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (absolute error variance :  $\sigma_{Abs}^2$  หรือ  $\sigma_{\Delta}^2$ ) หมายถึง ความแปรปรวนของผลต่างระหว่างคะแนนสังเกตได้กับคะแนนเอกภพ ( $\mu_p - X_p$ ) ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่างๆ ยกเว้นความแปรปรวนของแหล่งบุคคล ( $\sigma_{\mu_p}^2$  หรือ  $\sigma_p^2$ )

7.1.4.2 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ (relative error variance:  $\sigma_{Rel}^2$  หรือ  $\sigma_{\delta}^2$ ) หมายถึง ความแปรปรวนของผลต่างระหว่างคะแนนสังเกตได้กับคะแนนเอกภพ ( $\mu_p - X_p$ ) ซึ่งคำนวณจากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่างๆ ที่มีปฏิสัมพันธ์ร่วมกับผู้สอบ (p)

7.1.5 สัมประสิทธิ์การสรุปอ้างอิง (generalizability coefficient:  $\rho^2$ ) หมายถึง อัตราส่วนของความแปรปรวนของคะแนนเอกภพ และความแปรปรวนของคะแนนที่สังเกตที่คาดหวัง ซึ่งค่าสัมประสิทธิ์การสรุปอ้างอิงคำนวณได้จากกำลังสองของค่าสหสัมพันธ์ระหว่างคะแนนเอกภพกับคะแนนสังเกตได้ โดย สัมประสิทธิ์การสรุปอ้างอิงแบ่งเป็น 2 ประเภทคือสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์กับสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ดังนี้

7.1.5.1 สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์หรือการวัดแบบอิงกลุ่ม เป็นความคลาดเคลื่อนในการวัดแบบสัมพัทธ์ ค่าสัมประสิทธิ์การสรุปอ้างอิงประมาณค่าได้จากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพและผลบวกของความแปรปรวนของคะแนนเอกภพกับความแปรปรวนของความคลาดเคลื่อนสัมพัทธ์

7.1.5.2 สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์หรือการวัดแบบอิงเกณฑ์ เป็นความคลาดเคลื่อนในการวัดแบบสัมบูรณ์ ค่าสัมประสิทธิ์การสรุปอ้างอิงประมาณค่าได้จากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพและผลบวกของความแปรปรวนของคะแนนเอกภพกับความแปรปรวนของความคลาดเคลื่อนสัมบูรณ์

## 7.2 การวิเคราะห์สัมประสิทธิ์การสรุปอ้างอิง

สัมประสิทธิ์การสรุปอ้างอิง (Cronbach, 1972 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) ได้เสนอขั้นตอนการวิเคราะห์สัมประสิทธิ์การสรุปอ้างอิงออกเป็นสองขั้นตอน ดังนี้ ขั้นตอนแรกเป็นการศึกษาเพื่อ การสรุปอ้างอิงหรือการศึกษา G (generalizability study: G study) และขั้นตอนที่สองเป็นการศึกษาเพื่อการตัดสินใจหรือการศึกษา (decision study: D study) สำหรับหลักการดำเนินงานแต่ละขั้นตอนมีดังต่อไปนี้

7.2.1 การศึกษาเพื่อการสรุปอ้างอิง (generalizability study: G study) การศึกษาเพื่อการสรุปอ้างอิง เป็นกระบวนการที่มุ่งประมาณค่าความแปรปรวนจากแหล่งต่างๆ ภายใต้เอกภพของคะแนนสังเกตที่ยอมรับได้ โดยมีขั้นตอนดังนี้

1. กำหนดสิ่งที่จะวัด/กำหนดสิ่งที่ต้องการวัด เช่น ความสามารถด้านการคิดวิเคราะห์ของนักเรียน

2. กำหนดพาเซทของการวัด เช่น กำหนดให้มี 2 พาเซท คือ จำนวนข้อสอบและจำนวนผู้ตรวจให้คะแนน

3. กำหนดเอกภพของการสังเกตที่ยอมรับได้ประกอบด้วยพาเซทข้อสอบและพาเซทผู้ตรวจให้คะแนน

4. กำหนดความสัมพันธ์ระหว่างพาเซทว่าเป็นแบบ crossed แทนด้วย “x” หรือ nested แทนด้วย “:” ถ้าต้องการศึกษาให้ผู้ตรวจทุกคนตรวจข้อสอบทุกข้อจะเป็นการศึกษาแบบ crossed คือ  $i \times r$  ถ้าให้ผู้ตรวจแต่ละคนตรวจชุดข้อสอบต่างกันจะเป็นการศึกษาแบบ nested design คือ  $i : r$  ทั้งนี้ แล้วแต่ความสนใจของผู้ศึกษา ซึ่งโดยหลักการแล้วถ้าเป็นแบบ crossed design จะทำให้ได้จำนวนแหล่งความแปรปรวนมากกว่าแบบ nested design ซึ่งจะช่วยให้ทราบถึงความแปรปรวนแหล่งใดมีผลต่อสัมประสิทธิ์การสรุปอ้างอิง และช่วยเลือกรูปแบบที่เหมาะสมในขั้น D study และได้ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีค่ามากกว่า nested design ในกรณีของ 1- facet design แต่ในการปฏิบัตินั้นสามารถใช้ได้ง่ายกว่า

5. การเก็บรวบรวมข้อมูล โดยการสุ่มตัวอย่างมาศึกษา เช่น การศึกษาหนึ่งสุ่มตัวอย่างนักเรียน  $n_p$  คน สุ่มตัวอย่างข้อสอบ  $n_i$  ข้อ และสุ่มตัวอย่างผู้ตรวจ  $n_r$  คน จะเป็นการศึกษาแบบ G study ในกรณีนี้ประกอบด้วยแหล่งความแปรปรวน 7 แหล่ง เป็นผลหลัก (main effect) 3 แหล่ง คือ ผลจากนักเรียน ผลจากข้อสอบ และผลจากผู้ตรวจ และเป็นผลของปฏิสัมพันธ์ (interaction effect) 4 แหล่ง คือ ปฏิสัมพันธ์ระหว่างนักเรียนกับข้อสอบ นักเรียนกับผู้ตรวจข้อสอบ ข้อสอบกับข้อสอบนักเรียนกับแบบสอบและผู้ตรวจ จะเห็นว่ามีผลหลัก (main effect) อยู่ 3 ค่า คือ ผลของนักเรียน (p) ผลของข้อสอบ (i) และผลของผู้ตรวจ (r) ส่วนผลร่วมหรือปฏิสัมพันธ์ (interaction effect) 2 ระดับ 3 ค่า คือ ผลร่วมของผู้สอบและข้อสอบ (pi) ผลร่วมของผู้สอบและ

ผู้ตรวจ (pr) ผลรวมของข้อสอบและผู้ตรวจ (ir) และผลรวมทั้งหมดของผู้สอบ ข้อสอบ และผู้ตรวจ (pir) ค่าความแปรปรวนของแหล่งต่างๆ คำนวณจากค่ากำลังสองเฉลี่ย (mean square) โดยวิเคราะห์ความแปรปรวนแบบ factorial design  $p \times i \times r$  ในการประมาณค่าความแปรปรวนจากค่าเฉลี่ยกำลังสองสูตรคำนวณใช้ค่าเฉลี่ยของกำลังสองของความเบี่ยงเบน (mean square of deviation) หรือ  $MS(\alpha)$  โดยคำนวณจาก

$$MS(\alpha) = \frac{SS(\alpha)}{df(\alpha)}$$

เมื่อ $df(\alpha)$	แทนชั้นแห่งความเป็นอิสระที่สอดคล้องกับผลขององค์ประกอบ
$SS(\alpha)$	แทน ผลขององค์ประกอบที่ต้องการศึกษา
$MS(\alpha)$	แทน ค่าเฉลี่ยกำลังสองของความเบี่ยงเบน
Effect ( $\alpha$ )	แทน ผลรวมกำลังสองของความเบี่ยงเบน
$\sigma^2_{(\alpha)}$	แทน ค่าประมาณความแปรปรวน

6. การวิเคราะห์ข้อมูล หลังจากผ่านขั้นตอนในการพิจารณารูปแบบการวัดและเก็บรวบรวมข้อมูลจากรูปแบบที่กำหนดแล้ว ข้อมูลที่ได้จะนำไปสู่กระบวนการวิเคราะห์ทางสถิติเพื่อประมาณขนาดของความคลาดเคลื่อนของการวัดจากแหล่งต่างๆ กระบวนการทางสถิติที่ใช้ในทฤษฎีการสรุปอ้างอิง คือการวิเคราะห์ความแปรปรวน (analysis of variance procedure) โดยวิเคราะห์ในขั้นตอนของการศึกษา G study

#### 7.2.2 การศึกษาเพื่อการตัดสินใจ (decision study: D study)

การศึกษาเพื่อการตัดสินใจ (D study) เป็นการนำผลที่ได้จาก G study มาวิเคราะห์เพื่อการตัดสินใจเลือกสถานการณ์ของการวัด โดยมีขั้นตอน ดังนี้

1. กำหนดเอกภพของการสรุปอ้างอิงที่ผู้ตัดสินใจต้องการสรุปอ้างอิงใน D study ซึ่งเอกภพของการสรุปอ้างอิงประกอบด้วยเงื่อนไขทั้งหมดหรือเป็นเพียงเงื่อนไขย่อยในเอกภพการสังเกตที่ยอมรับได้ ผู้วิจัยต้องเลือกว่าจะใช้โมเดลรูปแบบใดเป็นโมเดลสุ่ม (random model) โมเดลกำหนด (fixed model) หรือโมเดลแบบผสม (mixed model) ซึ่งโดยหลักการแล้วโมเดลแบบสุ่มจะสามารถสรุปอ้างอิงได้กว้างกว่าโมเดลแบบกำหนดและโมเดลแบบผสม

2. การกำหนดขนาดกลุ่มตัวอย่าง (sample size) จำนวนเงื่อนไขของตัวอย่างพาเซทใน G study ไม่จำเป็นต้องเป็นตัวอย่างใน D study ทั้งนี้การกำหนดขนาดกลุ่มตัวอย่างขึ้นกับความสนใจของผู้วิจัย

3. โครงสร้างของแบบที่ศึกษา (design structure) ใน D study อาจใช้โครงสร้างของแบบที่ศึกษาเหมือนกันหรือต่างจากใน G study ก็ได้ เช่น ใน D study ตัดสินใจให้

นักเรียนทุกคนทำข้อสอบ  $n_i$  ข้อ เหมือนกันจะมีโครงสร้างเป็น D study ในลักษณะ  $p \times i \times r$  เหมือนโครงสร้างใน G study ที่กล่าวมาแล้ว แต่ถ้าตัดสินใจให้นักเรียนทุกคนทำข้อสอบทุกข้อ โดยผู้ตรวจแต่ละคนทำการตรวจชุดของข้อสอบต่างกัน จะได้โครงสร้างเป็น D study ในลักษณะ  $p \times (i : r)$

4. การประมาณ ค่าความแปรปรวน (estimate D study variance component) ในการศึกษาเพื่อตัดสินใจขึ้นอยู่กับนักวิจัย ดังนั้นต้องประมาณค่าความแปรปรวน โดยอาศัยผลจากการประมาณค่าในชั้นการศึกษาเพื่อการสรุปอ้างอิง ซึ่งสอดคล้องกับแบบแผนและขนาดตัวอย่างที่ต้องการตัดสินใจ

5. ประมาณค่าความแปรปรวนของความคลาดเคลื่อน นอกจากการประมาณค่าองค์ประกอบความแปรปรวน (variance component) ตามรูปแบบและขนาดกลุ่มตัวอย่างที่ต้องการแล้ว จากนั้นประมาณค่าความแปรปรวนของความคลาดเคลื่อน โดยความแปรปรวนของความคลาดเคลื่อน แบ่งออกเป็น 2 ประเภท คือ

5.1 การประมาณค่าความแปรปรวนของความคลาดเคลื่อนแบบสัมบูรณ์ (absolute error variance:  $\sigma_{Abs}^2$  หรือ  $\sigma_{\Delta}^2$ ) เป็นความแปรปรวนของผลต่างระหว่างคะแนนสังเกตได้กับคะแนนเอกภพ ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่างๆ ยกเว้นความแปรปรวนของแหล่งบุคคล ( $\sigma_{\mu p}^2$  หรือ  $\sigma_p^2$ ) เช่น การออกแบบ  $p \times i \times r$  design โดยมีสูตรดังนี้

$$\sigma_{Abs}^2 = \sigma_i^2 + \sigma_r^2 + \sigma_{pi}^2 + \sigma_{pr}^2 + \sigma_{ir}^2 + \sigma_{pir,e}^2$$

5.2 การประมาณค่าความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ (relative error variance:  $\sigma_{Rel}^2$  หรือ  $\sigma_{\delta}^2$ ) หมายถึง ความแปรปรวนของผลต่างระหว่างคะแนนสังเกตได้กับคะแนนเอกภพ ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่างๆ ที่มีปฏิสัมพันธ์ร่วมกับผู้สอบ ( $p$ ) โดยมีสูตรดังนี้

$$\sigma_{Rel}^2 = \sigma_{pi}^2 + \sigma_{pr}^2 + \sigma_{pir,e}^2$$

6. การประมาณค่าสัมประสิทธิ์การสรุปอ้างอิง (generalizability coefficient) เป็นขั้นสุดท้ายในการศึกษาเพื่อตัดสินใจ เป็นดัชนีบอกความน่าเชื่อถือของผลการวัด ซึ่งคำนวณจากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพและความแปรปรวนของคะแนนคาดหวังที่สังเกตได้ ซึ่งค่าสัมประสิทธิ์การสรุปอ้างอิง (G Coefficient) สำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho_{Abs}^2$  หรือ  $\rho_{\Delta}^2$ ) และสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho_{Rel}^2$  หรือ  $\rho_{\delta}^2$ ) สามารถคำนวณได้จากสูตร ดังนี้

6.1 สูตรคำนวณสัมประสิทธิ์การสรุปอ้างอิง สำหรับการตัดสินใจเชิงสัมบูรณ์

$$\rho_{Abs}^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{Abs}^2}$$



## 6.2 สูตรคำนวณสัมประสิทธิ์การสุร่อ้างอิง สำหรับการตัดสินใจเชิงสัมพัทธ์

$$\rho_{\text{Rel}}^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{\text{Rel}}^2}$$

### ตอนที่ 8 มาตรฐานของการประเมิน

คณะกรรมการมาตรฐานสำหรับการประเมินทางการศึกษา หรือ Joint Committee on Standards for Educational Evaluation ได้พัฒนามาตรฐานของการประเมิน เพื่อใช้ประเมินคุณภาพทางการศึกษา จำแนกเป็น 4 ด้าน ได้แก่ ด้านอรรถประโยชน์ (utility) มาตรฐานด้านความเป็นไปได้ (feasibility) ด้านความเหมาะสม (propriety standard) และด้านความถูกต้อง (accuracy) (The Joint Committee on Standards for Educational Evaluation, 1994) มีรายละเอียดดังนี้

1. มาตรฐานด้านอรรถประโยชน์ (utility standard) เป็นมาตรฐานที่กำหนดขึ้นเพื่อเป็นหลักประกันว่าการประเมินจะก่อให้เกิดสารสนเทศตรงกับความต้องการของผู้ใช้ผลการประเมิน ประกอบด้วย 7 เกณฑ์ ได้แก่ การระบุผู้มีส่วนได้ส่วนเสีย ความน่าเชื่อถือของการประเมิน ความครอบคลุมและตอบสนองความต้องการใช้สารสนเทศของผู้เกี่ยวข้อง การแปลความหมายและการตัดสินคุณค่ามีความชัดเจน รายงานการประเมินมีความชัดเจน รายงานประเมินดำเนินการและเผยแพร่ผลประเมินทันกำหนดเวลา และการประเมินส่งผลกระทบต่อผู้มีส่วนได้ส่วนเสียเพื่อเพิ่มโอกาสการนำผลประเมินไปใช้ประโยชน์

2. มาตรฐานด้านความเป็นไปได้ (feasibility standard) เป็นมาตรฐานที่กำหนดขึ้นเพื่อเป็นหลักประกันว่าการดำเนินการประเมินอยู่บนพื้นฐานของความเป็นจริง ประกอบด้วย 3 เกณฑ์ ได้แก่ วิธีการประเมินสามารถปฏิบัติได้จริง การเป็นที่ยอมรับของผู้ได้รับผลกระทบจากการประเมิน ความคุ้มค่าของการประเมิน

3. มาตรฐานด้านความเหมาะสม (propriety standard) เป็นมาตรฐานที่กำหนดขึ้นเพื่อเป็นหลักประกันว่าการดำเนินการประเมินส่งผลต่อผู้เกี่ยวข้อง โดยคำนึงถึงเรื่องกฎหมายและจริยธรรม ประกอบด้วย 8 เกณฑ์ ได้แก่ การกำหนดเป้าหมายของการประเมิน การกำหนดข้อตกลงอย่างเป็นทางการ การคำนึงถึงสิทธิส่วนบุคคลของผู้ให้ข้อมูล การเคารพสิทธิในการมีปฏิสัมพันธ์ของผู้เกี่ยวข้อง ความสมบูรณ์และความเป็นธรรมของการประเมิน การเปิดเผยข้อค้นพบของการประเมิน ผลประโยชน์ทับซ้อน และความรับผิดชอบด้านการเงิน

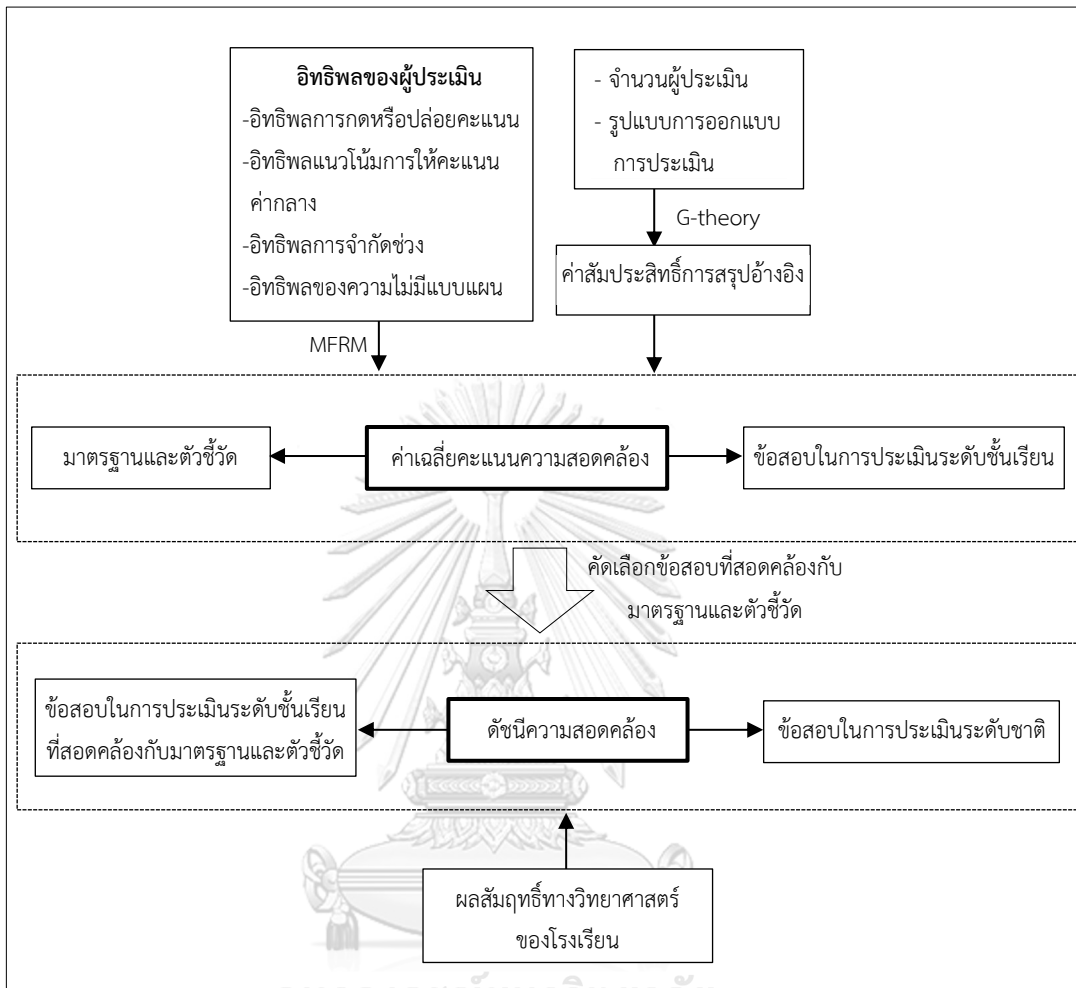
4. มาตรฐานด้านความถูกต้อง (accuracy standard) เป็นมาตรฐานที่กำหนดขึ้นเพื่อเป็นหลักประกันว่าการประเมินจะเปิดเผยและถ่ายทอดข้อมูลที่เพียงพอ เกี่ยวกับการตัดสินมูลค่าหรือ

คุณค่าของโปรแกรมที่ได้รับการประเมิน ประกอบด้วย 12 เกณฑ์ ได้แก่ การระบุวัตถุประสงค์ของการประเมินอย่างชัดเจน การวิเคราะห์บริบทของการประเมิน การอธิบายถึงจุดประสงค์และกระบวนการประเมินอย่างชัดเจน การอธิบายถึงแหล่งข้อมูลอย่างชัดเจน ความถูกต้องของสารสนเทศ ความน่าเชื่อถือของสารสนเทศ ความเป็นระบบของสารสนเทศ การวิเคราะห์ข้อมูลเชิงปริมาณ การวิเคราะห์ข้อมูลเชิงคุณภาพ การลงข้อสรุปที่สมเหตุสมผล การรายงานอย่างเป็นปรนัย และการประเมินอภิमान

สรุปได้ว่า มาตรฐานการประเมินเป็นการประมาณค่าหรือการตัดสินคุณค่าของสิ่งที่ต้องการประเมิน ในการประเมินทางการศึกษา ควรมีการออกแบบการประเมินให้ได้คุณภาพตามมาตรฐานการประเมินทั้ง 4 ด้าน คือ ด้านการใช้ประโยชน์ ด้านความเป็นไปได้ ด้านความเหมาะสม และด้านความถูกต้อง ทั้งนี้ ภายหลังจากดำเนินการและสรุปผลการวิเคราะห์ในงานวิจัยนี้ ผู้วิจัยจะจัดทำคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติกับข้อสอบในการประเมินระดับชั้นเรียน ซึ่งอยู่บนพื้นฐานของขั้นตอนการดำเนินการวิจัย และผลการวิจัยในส่วนของการประมาณค่าความเที่ยงในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินแตกต่างกัน เพื่อให้ผู้ที่สนใจใช้เป็นแนวทางในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบทางการศึกษาต่อไป ซึ่งคู่มือฯ ดังกล่าวจะถูกประเมินคุณภาพด้วยผู้คุณทรงคุณวุฒิ เพื่อตรวจสอบความเหมาะสมของคู่มือฯ ภายใต้อาณาเขตการประเมิน 4 ด้าน ได้แก่ ด้านการใช้ประโยชน์ ด้านความเป็นไปได้ ด้านความเหมาะสม และด้านความถูกต้อง ก่อนจัดทำเป็นคู่มือฉบับสมบูรณ์ต่อไป

## ตอนที่ 9 กรอบแนวคิดการวิจัย

กรอบแนวคิดเบื้องต้นในการวิจัย เรื่อง ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราล์ช และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ผู้วิจัยได้ศึกษาแนวคิดทฤษฎีและงานวิจัยที่เกี่ยวข้อง กรอบแนวคิดเบื้องต้นในการวิจัย เรื่อง ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราล์ช และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ผู้วิจัยได้ศึกษาแนวคิดทฤษฎีและงานวิจัยที่เกี่ยวข้อง เพื่อออกแผนการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น โดยประยุกต์ใช้ MFRM ในการตรวจสอบอิทธิพลของผู้ประเมิน 4 ประเภท ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน และตัดสินผลการประเมินความสอดคล้องๆ โดยทำการควบคุมอิทธิพลการกดหรือปล่อยคะแนน เพื่อคัดเลือกข้อสอบไปใช้ในศึกษาความสอดคล้องกับข้อสอบในการประเมินระดับชาติ ซึ่งการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ใช้แนวคิดตามวิธีการของ Porter (2001) โดยเปรียบเทียบดัชนีความสอดคล้องระหว่างโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน นอกจากนี้ ได้มีการศึกษาถึงจำนวนผู้เชี่ยวชาญและรูปแบบการออกแบบการประเมินที่จะทำให้ผลการประเมินความสอดคล้องในแนวเดียวกันมีความน่าเชื่อถือ โดยการใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดในประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของผลการประเมินความสอดคล้องในแนวเดียวกัน โดยกำหนดให้ข้อสอบ (item) และจำนวนผู้เชี่ยวชาญ (rater) เป็นองค์ประกอบความแปรปรวนความแปรปรวน และออกแบบรูปแบบการประเมินความสอดคล้องใน 2 ลักษณะ คือ ผู้เชี่ยวชาญแต่ละคนประเมินข้อสอบทุกข้อ (i x r design) และผู้เชี่ยวชาญแต่ละคนประเมินข้อสอบบางข้อ (r : i design) กรอบแนวคิดการวิจัยแสดงดังภาพ 2



ภาพ 2 กรอบแนวคิดการวิจัย

### บทที่ 3

#### วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงบรรยาย (descriptive research) มีวัตถุประสงค์เพื่อ (1) ตรวจสอบอิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น (2) เปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน (3) ศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ของโรงเรียนที่มีคะแนนการทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (O-NET) วิทยาศาสตร์แตกต่างกัน และ (4) ประเมินค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้เชี่ยวชาญและรูปแบบการออกแบบการประเมินต่างกัน แบ่งการดำเนินการวิจัยเป็น 4 ส่วน ได้แก่ (1) ประชากรและตัวอย่างในการวิจัย (2) เครื่องมือในการวิจัย (3) การเก็บรวบรวมข้อมูล และ (4) การวิเคราะห์ข้อมูลมีรายละเอียดดังนี้

#### 1. ประชากรและตัวอย่างในการวิจัย

1.1 ประชากรในการวิจัยครั้งนี้ แบ่งเป็น 2 ส่วน ดังนี้

1.1.1 ข้อสอบที่ใช้ในการวิจัย แบ่งเป็น 2 ส่วน คือ ข้อสอบในการประเมินระดับชั้นเรียน และข้อสอบในการประเมินระดับชาติ

1.1.1.1 ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 – 3) ของโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร ที่ถูกจัดอยู่ในกลุ่มโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูงสุดและต่ำสุด ซึ่งประกอบด้วยข้อสอบที่ครูสร้างขึ้น (teacher-made test) เพื่อ

ประเมินผลการเรียนรู้ของผู้เรียนในระดับชั้นเรียน ทั้งข้อสอบระหว่างภาคและข้อสอบประจำภาค ปีการศึกษา 2559

1.1.1.2 ข้อสอบในการประเมินระดับชาติ ได้แก่ ข้อสอบในการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (ordinary national educational test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำนวนทั้งสิ้น 45 ข้อ

1.1.2 ผู้เชี่ยวชาญในการตรวจสอบความสอดคล้องในแนวเดียวกัน ได้แก่ ครูสอนวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น และ/หรือนักวิชาการ และ/หรืออาจารย์ระดับอุดมศึกษาที่สอนในรายวิชาวิทยาศาสตร์ และ/หรือสาขาที่เกี่ยวข้องกับการวัดและประเมินผลการศึกษา

1.2 ตัวอย่างในการวิจัยนี้ แบ่งเป็น 2 ส่วน ดังนี้

1.2.1 ข้อสอบที่ใช้ในการวิจัย แบ่งเป็น 2 ส่วน คือ ข้อสอบในการประเมินระดับชั้นเรียน และข้อสอบในการประเมินระดับชาติ

1.2.1.1 ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 – 3) ของโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร ที่ถูกจัดอยู่ในกลุ่มโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูงสุดและต่ำสุด จำนวน 4 โรงเรียน ซึ่งประกอบด้วยข้อสอบที่ครูสร้างขึ้น (teacher-made test) เพื่อประเมินผลการเรียนรู้ของผู้เรียน ทั้งข้อสอบระหว่างภาคและข้อสอบประจำภาค ในปีการศึกษา 2559 การได้มาซึ่งตัวอย่างข้อสอบใช้วิธีการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) มีรายละเอียดดังต่อไปนี้

ขั้นที่ 1 จำแนกโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร ตามค่าเฉลี่ยคะแนน O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 ออกเป็น 2 กลุ่ม ได้แก่ กลุ่มโรงเรียนที่มีค่าเฉลี่ยคะแนน O-NET วิทยาศาสตร์สูงสุด (10 อันดับแรก) จำนวน 10 โรงเรียน และกลุ่มโรงเรียนที่มีค่าเฉลี่ยคะแนน O-NET วิทยาศาสตร์ต่ำสุด (10 อันดับท้ายสุด) จำนวน 10 โรงเรียน

ขั้นที่ 2 ทำการสุ่มอย่างง่ายโรงเรียนจากกลุ่มโรงเรียนที่มีค่าเฉลี่ยคะแนน O-NET วิทยาศาสตร์สูงสุด และกลุ่มโรงเรียนที่มีค่าเฉลี่ยคะแนน O-NET วิทยาศาสตร์ต่ำสุด อย่างละจำนวน 2 โรงเรียน

ขั้นที่ 3 ขอความอนุเคราะห์แต่ละโรงเรียนที่ได้จากการสุ่ม เพื่อเก็บรวบรวมข้อมูลในการศึกษาค้นคว้า โดยแบ่งเป็น (1) ข้อมูลพื้นฐานของโรงเรียนและครูผู้สอน และ (2) ข้อสอบที่ครูสร้างขึ้นเพื่อใช้ประเมินผลการเรียนรู้ระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 – 3) ปีการศึกษา 2559 ทั้งข้อสอบระหว่างภาคและข้อสอบ

ประจำภาค จะได้แบบสอบทั้งหมด 48 ชุด (2 แบบสอบกลางภาคและประจำภาค x 2 ภาคการศึกษา x 3 ระดับชั้น x 4 โรงเรียน)

ชั้นที่ 4 คัดเลือกเฉพาะข้อสอบที่วัดตัวชี้วัด จำนวน 40 ตัวชี้วัด ซึ่งเป็นตัวชี้วัดที่ สทศ. ใช้ในการออกข้อสอบ O-NET ได้ข้อสอบในการศึกษาทั้งหมดประมาณ 1,089 ข้อ

1.2.1.2 ข้อสอบในการประเมินระดับชาติ ได้แก่ ทุกหน่วยของประชากรข้อสอบ ในการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (ordinary national educational test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำนวน 45 ข้อ ประกอบด้วยข้อสอบปรนัย 4 ตัวเลือก 1 คำตอบ จำนวน 40 ข้อ และข้อสอบปรนัย 6 ตัวเลือก 2 คำตอบ จำนวน 5 ข้อ ซึ่งมีการระบุคุณลักษณะจำเพาะของข้อสอบแต่ละข้อ (test specification) โดย สทศ. ทั้งนี้ ผู้วิจัยได้จัดทำหนังสือขอความอนุเคราะห์ไปยัง สทศ. ในการเก็บรวบรวมข้อมูล คุณลักษณะจำเพาะของข้อสอบด้วยแบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบในการทดสอบ O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 ที่ผู้วิจัย สร้างขึ้น

1.2.2 ผู้เชี่ยวชาญในการตรวจสอบความสอดคล้องในแนวเดียวกัน จำนวน 20 ท่าน ประกอบด้วย ครูสอนวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น และ/หรือนักวิชาการ และ/หรือ อาจารย์ระดับอุดมศึกษาที่สอนในรายวิชาวิทยาศาสตร์ และ/หรือสาขาวิชาที่เกี่ยวข้องกับการวัดและ ประเมินผลการศึกษา ซึ่งได้มาจากการเลือกแบบเจาะจง (purposive sampling) โดยกำหนดเกณฑ์ ในการคัดเลือกคุณสมบัติของผู้เชี่ยวชาญ คือ เป็นผู้ที่มีความรู้ความเชี่ยวชาญด้านมาตรฐานและ ตัวชี้วัดกลุ่มสาระวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น และด้านการวัดและประเมินผลทางการศึกษา โดยสำเร็จการศึกษาอย่างน้อยในระดับปริญญาตรีทางการสอนวิทยาศาสตร์ และสำเร็จการศึกษา อย่างน้อยในระดับปริญญาโททางการวัดและประเมินผลทางการศึกษา และ/หรือมีประสบการณ์ สอนกลุ่มสาระวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ไม่นต่ำกว่า 3 ปี

## 2. เครื่องมือในการวิจัย

เครื่องมือในการวิจัยครั้งนี้ ได้แก่ (1) แบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (test-item specification) ในการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้ วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 และ (2) แบบประเมินความสอดคล้องใน แนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระ- การเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ขึ้นตอนในการสร้างและตรวจสอบคุณภาพ เครื่องมือในการวิจัยมีรายละเอียดดังต่อไปนี้

### 1. แบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (test-item specification) ในการทดสอบทางการศึกษาระดับขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559

แบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (test-item specification) ในการทดสอบ O-NET เป็นแบบบันทึกที่ให้ สทศ. กรอกข้อมูลลักษณะเฉพาะของข้อสอบ O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 จำนวน 45 ข้อ ข้อมูลนี้ถูกนำไปสร้างเป็นเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) เพื่อให้ในการคำนวณดัชนีความสอดคล้องในแนวเดียวกัน (alignment index) ตามแนวคิดของ Porter ต่อไป ลักษณะและการขั้นตอนสร้างแบบบันทึกฯ มีรายละเอียดดังต่อไปนี้

#### ขั้นที่ 1 กำหนดตัวแปรในการศึกษา และกำหนดนิยามเชิงปฏิบัติการ

ระดับความซับซ้อนทางปัญญา หมายถึง ระดับการวัดด้านพุทธิปัญญา ซึ่งเป็นลำดับขั้นของกระบวนการทางสมองหรือความลึกของความรู้ในเนื้อหาตัวชี้วัด ภายใต้มาตรฐานต่างๆ และข้อสอบในการประเมินผลการเรียนรู้ ซึ่งวัดได้จากการพิจารณาของผู้เชี่ยวชาญ ความซับซ้อนทางปัญญาจำแนกตามแนวคิด revised Bloom's taxonomy (Anderson et al., 2001) ตามลำดับขั้นของกระบวนการทางปัญญา (cognitive process dimension) ซึ่งจำแนกได้เป็น 6 ชั้น เรียงลำดับจากความรู้ระดับต่ำไปยังความรู้ระดับสูง ดังนี้

1. จำ (remember) เป็นความสามารถในการระลึกหรือจำความรู้หรือสารสนเทศที่เก็บไว้ในสมอง ซึ่งเป็นความจำระยะยาว

2. เข้าใจ (understand) เป็นความสามารถในการสร้างความหมายหรือความรู้จากสื่อหรือเครื่องมือทางการศึกษาด้วยตนเอง เช่น จากการอ่าน การอธิบายของครู ทักษะย่อยของความสามารถในขั้นนี้ ได้แก่ การแปลความหมาย (interpreting) การให้ตัวอย่าง (exemplifying) การจัดจำแนก (classifying) การสรุป (summarizing) การสรุปอ้างอิง (inferring) การเปรียบเทียบ (comparing) และการอธิบาย (explaining)

3. ประยุกต์ใช้ (apply) เป็นความสามารถในการใช้กระบวนการที่ได้เรียนรู้มาในสถานการณ์ใหม่หรือสถานการณ์ที่คล้ายคลึงกัน

4. วิเคราะห์ (analyze) เป็นความสามารถในการแยกความรู้ออกเป็นส่วนๆ โดยสามารถให้เหตุผลว่า ความรู้ส่วนย่อยที่แยกแต่ละส่วนมีความเกี่ยวข้องกับโครงสร้างของความรู้ทั้งหมดอย่างไร สามารถจำแนกความแตกต่าง จัดระบบความรู้ และบอกที่มาของความรู้หรือองค์ประกอบแต่ละส่วนได้

5. ประเมินค่า (evaluate) เป็นความสามารถในการตรวจสอบ วิพากษ์ และตัดสินใจ

6. สร้างสรรค์ (create) เป็นความสามารถการสร้างสิ่งใหม่จากสิ่งที่เคยเรียนรู้ หรือสิ่งที่พบเห็นในบริบทต่างๆ สามารถสร้างสรรค์งาน แผนงาน หรือผลิตภัณฑ์ หรือชิ้นงานที่แปลกใหม่



**ขั้นที่ 2** ร่างแบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (test-item Specification) ในการทดสอบทางการศึกษาระดับขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3

การร่างแบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ O-NET มีขั้นตอนดังนี้

1. ศึกษามาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 โดยมีการกำหนดสาระสำคัญที่เกี่ยวข้องกับระดับชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 - 3) ซึ่งประกอบด้วยตัวชี้วัดจำนวน 119 ตัวชี้วัด ภายใต้ม.13 มาตรฐาน จาก 8 สาระ ดังตาราง 9 (รายละเอียดของแต่ละสาระ มาตรฐานและตัวชี้วัดได้แสดงไว้ในตาราง 1 ในเอกสารและงานวิจัยที่เกี่ยวข้อง ตอนที่ 1)

**ตาราง 9** สรุปจำนวนสาระ มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จำแนกตามระดับชั้น

สาระที่	มาตรฐานที่	จำนวนตัวชี้วัด		
		ม.1	ม.2	ม.3
1	1.1	13	6	-
	1.2	-	-	6
2	2.1	-	-	4
	2.2	-	-	6
3	3.1	4	3	-
	3.2	3	4	-
4	4.1	2	2	3
	4.2	-	-	3
5	5.1	4	3	5
6	6.1	7	10	-
7	7.1	-	-	3
	7.2	-	-	1
8	8.1	9	9	9
<b>รวม</b>	<b>13</b>	<b>42</b>	<b>37</b>	<b>40</b>

2. ศึกษาารูปแบบข้อสอบ O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 สถาบันทดสอบทางการศึกษาแห่งชาติ (2559: เว็บไซต์) จำนวน 45

ข้อ ซึ่งพบว่า ตัวชี้วัดที่ สทศ. ใช้ในการออกข้อสอบ O-NET ปีการศึกษา 2559 ดังกล่าว มีจำนวน 40 ตัวชี้วัด ซึ่งเป็นตัวชี้วัดที่อยู่ภายใต้สาระที่ 1 - 7 ทั้งนี้ สำหรับสาระที่ 8 นั้น สทศ. ให้เหตุผลว่ามีการแทรกอยู่ในสาระที่ 1 - 7 แล้ว ดังนั้น การศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ในงานวิจัยนี้ จึงยึดตัวชี้วัดตามที่ สทศ. ใช้ในการออกข้อสอบ คือ ประกอบด้วย 40 ตัวชี้วัด ซึ่งอยู่ภายใต้สาระที่ 1 - 7 รายละเอียดดังตาราง 10

ตาราง 10 สรุปรูปตัวชี้วัดตามรูปแบบของข้อสอบ O-NET วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 ของ สทศ.

สาระ	มาตรฐาน	ตัวชี้วัด						สรุปจำนวน ตัวชี้วัด	รูปแบบข้อสอบ (จำนวนข้อ)	คะแนน
		ม.1		ม.2		ม.3				
1	ว 1.1	ม.1/1	-	ม.2/1	-	-	-	6	4 ตัวเลือก 1 คำตอบ (6 ข้อ)	12
		ม.1/2	-	ม.2/2	✓	-	-			
		ม.1/3	-	ม.2/3	✓	-	-			
		ม.1/4	✓	ม.2/4	✓	-	-			
		ม.1/5	✓	ม.2/5	✓	-	-			
		ม.1/6	-	ม.2/6	-	-	-			
		ม.1/7	-	-	-	-	-			
		ม.1/8	-	-	-	-	-			
		ม.1/9	-	-	-	-	-			
		ม.1/10	-	-	-	-	-			
		ม.1/11	-	-	-	-	-			
		ม.1/12	-	-	-	-	-			
		ม.1/13	-	-	-	-	-			
	1	ว 1.2	-	-	-	ม.3/1	-	2	4 ตัวเลือก 1 คำตอบ (1 ข้อ) 6 ตัวเลือก 2 คำตอบ (1 ข้อ)	6
-			-	-	ม.3/2	-				
-			-	-	ม.3/3	✓				
-			-	-	ม.3/4	-				
-			-	-	ม.3/5	-				
-			-	-	ม.3/6	✓				
2	ว 2.1	-	-	-	ม.3/1	✓	2	4 ตัวเลือก 1 คำตอบ (2 ข้อ) 6 ตัวเลือก 2 คำตอบ (1 ข้อ)	8	
		-	-	-	ม.3/2	-				
		-	-	-	ม.3/3	✓				
		-	-	-	ม.3/4	-				
	ว 2.2	-	-	-	ม.3/1	-	2	4 ตัวเลือก 1 คำตอบ (2 ข้อ)	4	
		-	-	-	ม.3/2	✓				
		-	-	-	ม.3/3	-				
		-	-	-	ม.3/4	✓				
		-	-	-	ม.3/5	-				
		-	-	-	ม.3/6	-				

ตาราง 10 (ต่อ)

สาระ	มาตรฐาน	ตัวชี้วัด						สรุปจำนวน ตัวชี้วัด	รูปแบบข้อสอบ (จำนวนข้อ)	คะแนน
		ม.1		ม.2		ม.3				
3	ว 3.1	ม.1/1	✓	ม.2/1	-	-		5	4 ตัวเลือก 1 คำตอบ (4 ข้อ)	8
		ม.1/2	✓	ม.2/2	✓					
		ม.1/3	-	ม.2/3	✓					
		ม.1/4	✓							
	ว 3.2	ม.1/1	✓	ม.2/1	✓	-		3	4 ตัวเลือก 1 คำตอบ (2 ข้อ)	4
		ม.1/2	✓	ม.2/2	-					
		ม.1/3	-	ม.2/3	-					
				ม.2/4	-					
4	ว 4.1	ม.1/1	-	ม.2/1	-	ม.3/1	✓	4	4 ตัวเลือก 1 คำตอบ (7 ข้อ) 6 ตัวเลือก 2 คำตอบ (1 ข้อ)	18
		ม.1/2	✓	ม.2/2	✓	ม.3/2	✓			
						ม.3/3	-			
	ว 4.2	-		-		ม.3/1	✓	3	4 ตัวเลือก 1 คำตอบ (4 ข้อ) 6 ตัวเลือก 2 คำตอบ (1 ข้อ)	12
						ม.3/2	✓			
						ม.3/3	✓			
5	ว 5.1	ม.1/1	-	ม.2/1	✓	ม.3/1	✓	7	4 ตัวเลือก 1 คำตอบ (6 ข้อ) 6 ตัวเลือก 2 คำตอบ (1 ข้อ)	16
		ม.1/2	✓	ม.2/2	✓	ม.3/2	-			
		ม.1/3	-	ม.2/3	✓	ม.3/3	✓			
		ม.1/4	-			ม.3/4	✓			
						ม.3/5	-			
6	ว 6.1	ม.1/1	-	ม.2/1	-	-		3	4 ตัวเลือก 1 คำตอบ (1 ข้อ)	2
		ม.1/2	-	ม.2/2	✓					
		ม.1/3	-	ม.2/3	-					
		ม.1/4	-	ม.2/4	✓					
		ม.1/5	-	ม.2/5	-					
		ม.1/6	✓	ม.2/6	-					
		ม.1/7	-	ม.2/7	-					
				ม.2/8	-					
				ม.2/9	-					
				ม.2/10	-					
7	ว 7.1	-		-		ม.3/1	✓	2	4 ตัวเลือก 1 คำตอบ (5 ข้อ)	10
						ม.3/2	-			
						ม.3/3	✓			
	ว 7.2	-		-		ม.3/1	✓	1		
<b>รวม</b>							<b>40</b>	<b>45</b>	<b>100</b>	

3. ร่างแบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (test-item Specification) ในการทดสอบ O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 มีรายละเอียดดังนี้




1) ข้อสอบ เป็นส่วนที่แสดงถึงข้อสอบ O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 ที่ออกและเผยแพร่โดย สทศ. จำนวนทั้งหมด 45 ข้อ

2) มาตรฐาน/ตัวชี้วัด เป็นส่วนที่ให้ สทศ. ระบุมาตรฐานการเรียนรู้และตัวชี้วัดตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ที่ สทศ. ใช้ในการออกข้อสอบ O-NET ข้อนี้ๆ

3) ระดับความซับซ้อนทางปัญญา เป็นส่วนที่ให้ สทศ. ระบุความซับซ้อนทางปัญญา หรือระดับการวัดทางพุทธิปัญญาของข้อสอบแต่ละข้อ ที่ผู้สอบจำเป็นต้องใช้ในการทำข้อสอบให้ถูกต้อง โดยจำแนกระดับความซับซ้อนทางปัญญาตามลำดับขั้นของกระบวนการทางปัญญาของ revised Bloom's taxonomy ซึ่งแบ่งเป็น 6 ชั้น ได้แก่ จำ เข้าใจ ประยุกต์ใช้ วิเคราะห์ ประเมินค่า และสร้างสรรค์ ตัวอย่างแบบบันทึกระดับความซับซ้อนทางปัญญาของข้อสอบในการประเมินระดับชาติแสดงดังตาราง 11

4. นำ (ร่าง) แบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ ONET เสนออาจารย์ที่ปรึกษา ตรวจสอบความถูกต้องเหมาะสมของแบบบันทึก และปรับแก้ตามข้อเสนอแนะ

ตาราง 11 ตัวอย่างแบบบันทึกระดับความซับซ้อนทางปัญญาของข้อสอบในการประเมินระดับชาติ

ข้อ ที่	ข้อสอบ	มาตรฐาน	ตัวชี้วัด	ระดับความซับซ้อนทางปัญญา						คะแนน รายข้อ (คะแนน)												
				จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์													
1	<p>นำชดสัณฐานเดียวกัน ขนาดเท่ากัน ไปใส่ในสารละลาย A B และ C ที่มีความเข้มข้นแตกต่างกันเป็นเวลานานเท่าๆ กัน ได้ผลดังภาพ</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>สารละลาย A</p> </div> <div style="text-align: center;">  <p>สารละลาย B</p> </div> <div style="text-align: center;">  <p>สารละลาย C</p> </div> </div> <p>จงเรียงลำดับความเข้มข้นของสารละลาย A B และ C จากความเข้มข้นน้อยที่สุดไปมากที่สุด</p> <ol style="list-style-type: none"> <li>1. A B C</li> <li>2. A C B</li> <li>3. B A C</li> <li>4. C A B</li> </ol>	.....	.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2												
2	<p>การทดลองเพื่อศึกษาการแพร่ของต่างับทีม 3 ชุดทดลอง ได้ผลการสังเกตดังนี้</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>ชุดทดลองที่</th> <th>ตัวกลาง</th> <th>ผลการสังเกต</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>A</td> <td>สีม่วงของต่างับทีมกระจายไปทุกทิศทุกทางอย่างรวดเร็ว</td> </tr> <tr> <td>2</td> <td>B</td> <td>สีม่วงของต่างับทีมค่อยๆ กระจายไปทุกทิศทุกทาง อย่างช้าๆ</td> </tr> <tr> <td>3</td> <td>ไม่มีสารตัวกลาง</td> <td>อนุภาคของต่างับทีมไม่มีการเคลื่อนที่</td> </tr> </tbody> </table> <p>ตัวแปรต้นของการทดลองนี้ คืออะไร</p> <ol style="list-style-type: none"> <li>1. ปริมาณของต่างับทีม</li> <li>2. ขนาดอนุภาคของต่างับทีม</li> <li>3. ชนิดของตัวกลาง</li> <li>4. อัตราเร็วในการแพร่ของต่างับทีม</li> </ol>	ชุดทดลองที่	ตัวกลาง	ผลการสังเกต	1	A	สีม่วงของต่างับทีมกระจายไปทุกทิศทุกทางอย่างรวดเร็ว	2	B	สีม่วงของต่างับทีมค่อยๆ กระจายไปทุกทิศทุกทาง อย่างช้าๆ	3	ไม่มีสารตัวกลาง	อนุภาคของต่างับทีมไม่มีการเคลื่อนที่	.....	.....	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
ชุดทดลองที่	ตัวกลาง	ผลการสังเกต																				
1	A	สีม่วงของต่างับทีมกระจายไปทุกทิศทุกทางอย่างรวดเร็ว																				
2	B	สีม่วงของต่างับทีมค่อยๆ กระจายไปทุกทิศทุกทาง อย่างช้าๆ																				
3	ไม่มีสารตัวกลาง	อนุภาคของต่างับทีมไม่มีการเคลื่อนที่																				

2. แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น

ขั้นตอนในการสร้างและตรวจสอบคุณภาพแบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน มีรายละเอียดดังต่อไปนี้

ขั้นที่ 1 กำหนดตัวแปรในการศึกษา และกำหนดนิยามเชิงปฏิบัติการ

ระดับความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน หมายถึง ระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างผลการเรียนรู้ที่คาดหวัง ได้แก่ มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตร

แกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กับการประเมิน ได้แก่ ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ของโรงเรียนมัธยมศึกษา การวัดความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนทำโดยให้ผู้เชี่ยวชาญเป็นผู้ประเมิน ด้วยแบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น ที่ผู้วิจัยสร้างขึ้น

## ขั้นที่ 2 ร่างแบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียนฯ

การร่างประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนฯ มีขั้นตอนดังนี้

1. ศึกษามาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ของระดับชั้นมัธยมศึกษาตอนต้น โดยมีการกำหนดสาระสำคัญไว้ จำนวน 8 สาระ 13 มาตรฐาน ตัวชี้วัดจำแนกตามระดับชั้นมัธยมศึกษาปีที่ 1 2 และ 3 มีจำนวนเท่ากับ 42 37 และ 40 ตัวชี้วัด ตามลำดับ สรุปจำนวนตัวชี้วัดดังตาราง 9 (รายละเอียดของสาระ มาตรฐานและตัวชี้วัดดังที่ได้กล่าวมาแล้วในตาราง 3 ที่กล่าวมาแล้วในบทที่ 2)

2. ร่างแบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ซึ่งประกอบด้วยข้อรายการที่ให้ผู้เชี่ยวชาญประเมินระดับความซับซ้อนทางปัญญาและระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด แต่ละข้อรายการประเมินประกอบด้วย 2 ส่วนหลัก แสดงตัวอย่างข้อรายการดังภาพ 3 มีรายละเอียดดังนี้

ส่วนที่ 1 เป็นรายละเอียดของข้อสอบที่ครูสร้างขึ้นเพื่อใช้ในการประเมินผลการเรียนรู้ระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น มีรายละเอียดดังนี้

1) ข้อสอบที่ต้องการประเมิน ซึ่งประกอบด้วย ข้อคำถามและตัวเลือก (กรณีเป็นข้อสอบหลายตัวเลือก)

2) คะแนนรายข้อ เป็นคะแนนเต็มรายข้อของข้อสอบที่ครูผู้ออกข้อสอบเป็นผู้กำหนด

3) ระดับชั้น แบ่งได้ 3 ระดับ คือ ระดับชั้นมัธยมศึกษาปีที่ 1 – 3

4) มาตรฐานและตัวชี้วัด เป็นมาตรการเรียนรู้และตัวชี้วัดสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ที่ครูผู้สร้างข้อสอบระบุว่าต้องการวัดด้วยข้อสอบข้อนั้นๆ ทั้งนี้เพื่อให้สอดคล้องกับวัตถุประสงค์และขอบเขตของการวิจัยนี้ ผู้วิจัยจึงได้คัดเลือกเฉพาะมาตรฐานและตัวชี้วัดที่เหมาะสมสำหรับการประเมินผลด้วยข้อสอบเท่านั้น โดยศึกษาเฉพาะตัวชี้วัดที่อยู่ในสาระที่ 1 – 7 เนื่องจากตัวชี้วัดในมาตรฐาน ว 8.1

ของสาระที่ 8 เป็นมาตรฐานและตัวชี้วัดที่มุ่งเน้นการใช้กระบวนการทางวิทยาศาสตร์และจิตวิทยาศาสตร์ ในการสืบเสาะหาความรู้ การแก้ปัญหา โดยครูผู้สอนจำเป็นต้องใช้การประเมินทักษะการปฏิบัติต่างๆ ที่มีความเหมาะสมกว่าการประเมินด้วยข้อสอบ ซึ่งสอดคล้องกับรูปแบบของข้อสอบ O-NET ที่ สทศ. มีการออกข้อสอบโดยใช้ตัวชี้วัดที่อยู่ภายใต้สาระที่ 1 - 7 โดย สทศ. ให้เหตุผลว่าสาระที่ 8 มีการแทรกอยู่ในสาระที่ 1 - 7 แล้ว ดังนั้น สรุปลักษณะข้อสอบในงานวิจัยนี้ มีจำนวนทั้งหมด 40 ตัวชี้วัด ภายใต้ 12 มาตรฐาน ของ 7 สาระ (สาระที่ 1 - 7) โดยไม่รวมตัวชี้วัดของสาระที่ 8 รายละเอียดดังตาราง 17 ใน บทที่ 4

ส่วนที่ 2 เป็นส่วนที่ให้ผู้เชี่ยวชาญพิจารณาประเมินหรือให้ข้อเสนอแนะต่อข้อสอบเป็น รายข้อ ใน 3 ประเด็น ได้แก่

1) ระดับความซับซ้อนทางปัญญา เป็นส่วนที่ให้ผู้เชี่ยวชาญพิจารณาระดับ ความ-ซับซ้อนทางปัญญาหรือความลึกของความรู้ในเนื้อหาข้อสอบ ที่ผู้เรียนต้องใช้เพื่อให้สามารถทำข้อสอบ ได้ถูกต้อง จำแนกตามแนวคิด revised Bloom's taxonomy แบ่งเป็น 6 ชั้น ได้แก่ จำ เข้าใจ ประยุกต์ใช้ วิเคราะห์ ประเมินค่า และสร้างสรรค์ ทั้งนี้ ผู้เชี่ยวชาญสามารถระบุระดับความซับซ้อน ทางปัญญาสูงสุดเพียง 1 ระดับ ข้อมูลในส่วนนี้จะถูกนำไปสร้างเมทริกซ์ข้อสอบ (เมทริกซ์ Y) และนำไป คำนวณค่าดัชนีความสอดคล้อง (alignment index) ต่อไป

2) ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด มีลักษณะเป็นมาตรประมาณค่า 5 ระดับ (0 - 4) ซึ่งให้ผู้เชี่ยวชาญพิจารณาตัดสินว่าเนื้อหาของข้อสอบกับเนื้อหาตัวชี้วัดที่กำหนด มีความสอดคล้องกันในระดับใด โดยรายละเอียดของระดับความสอดคล้องมีดังนี้

0 คะแนน หมายถึง เนื้อหาของข้อสอบไม่สอดคล้องกับเนื้อหาของมาตรฐานและ ตัวชี้วัดที่กำหนด

1 คะแนน หมายถึง เนื้อหาของข้อสอบค่อนข้างไม่สอดคล้องกับเนื้อหาของมาตรฐาน และตัวชี้วัดที่กำหนด

2 คะแนน หมายถึง ไม่แน่ใจว่าเนื้อหาของข้อสอบสอดคล้องกับเนื้อหาของมาตรฐาน และตัวชี้วัดที่กำหนด

3 คะแนน หมายถึง เนื้อหาของข้อสอบค่อนข้างสอดคล้องกับเนื้อหาของมาตรฐาน และตัวชี้วัดที่กำหนด

4 คะแนน หมายถึง เนื้อหาของข้อสอบสอดคล้องโดยตรงกับเนื้อหาของมาตรฐาน และตัวชี้วัดที่กำหนด

ทั้งนี้ กำหนดเกณฑ์ที่ข้อสอบจะมีความสอดคล้องกับมาตรฐานและตัวชี้วัดก็ต่อเมื่อมี ค่าเฉลี่ยคะแนนความสอดคล้องตั้งแต่ 3.00 ขึ้นไป





3. นำ (ร่าง) แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนฯ เสนออาจารย์ที่ปรึกษาตรวจสอบความถูกต้องเหมาะสม และปรับแก้ตามข้อเสนอแนะ

4. นำ (ร่าง) แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนฯ ที่ปรับแก้ตามข้อเสนอแนะของอาจารย์ที่ปรึกษาแล้ว ไปให้ผู้ทรงคุณวุฒิ จำนวน 7 คน ตรวจสอบความถูกต้องและความเหมาะสม ทำการปรับแก้ตามข้อเสนอแนะ ทั้งนี้ กำหนดเกณฑ์ในการคัดเลือกคุณสมบัติของผู้ทรงคุณวุฒิ คือ เป็นผู้ที่มีความรู้ความเชี่ยวชาญด้าน มาตรฐานและตัวชี้วัดกลุ่มสาระวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น และด้านการวัดและประเมินผลทางการศึกษา โดยสำเร็จการศึกษาอย่างน้อยในระดับปริญญาตรีทางการสอนวิทยาศาสตร์ และสำเร็จการศึกษาอย่างน้อยในระดับปริญญาโทในสาขาการวัดและประเมินผลทางการศึกษา

### ขั้นที่ 3 ทดลองใช้แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนฯ

การทดลองใช้แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนฯ มีขั้นตอนดังต่อไปนี้

1) สุ่มข้อสอบในการประเมินระดับชั้นเรียน จำนวน 40 ข้อ (จากทั้งหมด 1,089 ข้อ) โดย สุ่มตัวชี้วัดละ 1 ข้อ เพื่อจัดทำเป็นแบบประเมินฯ ฉบับทดลองใช้

2) ทดลองใช้แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนฯ กับผู้ทรงคุณวุฒิจำนวน 7 คน เพื่อนำข้อมูลมาวิเคราะห์ ความเที่ยงระหว่างผู้ประเมิน (inter-rater reliability: IRR) ซึ่งเป็นการตรวจสอบความสอดคล้องกัน ระหว่างผู้ประเมิน โดยแบ่งเป็น 2 ส่วน ดังนี้

ส่วนที่ 1 ความเที่ยงระหว่างผู้ประเมินของการประเมินระดับความซับซ้อนทางปัญญา ผู้วิจัยเลือกใช้สถิติแคปปาของฟลีส (Fleiss's kappa statistic:  $K_f$ ) (Fleiss, 1971) เนื่องจากเป็น การตรวจสอบความเที่ยงระหว่างผู้ประเมินที่มีตั้งแต่ 3 คนขึ้นไป สำหรับข้อมูลที่อยู่ในระดับ นามบัญญัติ (Gisev, Bell, & Chen, 2013) มีสูตรในการคำนวณสถิติแคปปาของฟลีสดังนี้

$$K_f = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

$$\text{เมื่อ } \bar{P} = \frac{1}{N} \sum_{i=1}^N p_i$$

$$\text{โดย } p_i = \frac{1}{n(n-1)} \sum_{j=1}^k n_{ij}^2 - n$$

$$\text{ดังนั้น } \bar{P} = \frac{1}{Nn(n-1)} \sum_{i=1}^N \sum_{j=1}^k n_{ij}^2 - n \quad \text{และ} \quad \bar{P}_e = \sum_{j=1}^k p_j^2$$

โดยที่ N แทน จำนวนข้อสอบที่ถูกประเมิน

n แทน จำนวนผู้ประเมินข้อสอบแต่ละข้อ

k แทน จำนวนประเภทของการประเมิน (number of categories of scale)

i แทน ข้อสอบที่ถูกประเมิน

j แทน ประเภทที่ถูกประเมิน (categories of scale)

กำหนดเกณฑ์การพิจารณาระดับความสอดคล้องตามแนวทาง Fleiss, Levin, และ Paik (2003) ดังนี้

$0.00 \leq K_f \leq 0.39$  หมายถึง ผู้ประเมินมีความสอดคล้องกันในการประเมินระดับต่ำ

$0.40 \leq K_f \leq 0.74$  หมายถึง ผู้ประเมินมีความสอดคล้องกันในการประเมินระดับดี

$0.75 \leq K_f \leq 1.00$  หมายถึง ผู้ประเมินมีความสอดคล้องกันในการประเมินระดับดีมาก

คำนวณค่าความคลาดเคลื่อนมาตรฐานของสถิติแคปปาของฟลีส หรือ  $SE(K_f)$  ได้จากรากที่สองของค่าความแปรปรวนของสถิติแคปปา ( $Var(K_f)$ ) (Fleiss, 1971) สูตรในการคำนวณมี ดังนี้

$$SE(K_f) = \frac{1}{\sqrt{Var(K_f)}}$$

$$SE(K_f) = \frac{1}{\sqrt{\frac{2}{Nn(n-1)} \times \frac{\sum_j p_j^2 - (2n-3)(\sum_j p_j^2)^2 + 2(n-2) \sum_j p_j^3}{(1 - \sum_j p_j^2)^2}}}$$

ส่วนที่ 2 ความเที่ยงระหว่างผู้ประเมินของการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ (0 - 4) ผู้วิจัยเลือกใช้สถิติสหสัมพันธ์ภายในชั้น (intra-class correlation: ICC) เนื่องจากเป็นข้อมูลต่อเนื่อง (continuous data) (Gisev, Bell, & Chen, 2013) โดยวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้นที่วิเคราะห์จากโมเดลอิทธิพลผสมแบบสองทาง (two-way mixed-effects model) มีผู้ประเมินจำนวน 7 คน (ใช้สัญลักษณ์ คือ ICC(3,7)) ที่ระดับความเชื่อมั่น 95% ด้วยโปรแกรม SPSS กำหนดเกณฑ์การพิจารณาระดับความสอดคล้องตามแนวทางของ Portney & Watkins (2015) ดังนี้

$ICC > 0.50$  หมายถึง ผู้ประเมินมีความสอดคล้องในการประเมินระดับต่ำ

$0.50 \leq ICC < 0.75$  หมายถึง ผู้ประเมินมีความสอดคล้องในการประเมินระดับพอใช้

$0.75 \leq ICC \leq 0.90$  หมายถึง ผู้ประเมินมีความสอดคล้องในการประเมินระดับดี

$ICC > 0.90$  หมายถึง ผู้ประเมินมีความสอดคล้องในการประเมินระดับดีมาก

คำนวณค่าความคลาดเคลื่อนมาตรฐานในการวัด (standard error of measurement: SEM) มีสูตรดังนี้

$$SEM = S \times \sqrt{1 - ICC}$$

ผลการวิเคราะห์ความเที่ยงระหว่างผู้ประเมิน (inter-rater reliability) จากการทดลองให้ผู้ทรงคุณวุฒิจำนวน 7 คน ด้วยใช้แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนฯ พบว่า ความเที่ยงระหว่างผู้ประเมินในการประเมินระดับความซับซ้อนทางปัญญา ที่วิเคราะห์ด้วยสถิติแคปปาของฟลีส มีค่าเท่ากับ 0.596 แสดงว่ามีความสอดคล้องระหว่างผู้ประเมินอยู่ในระดับดี (Fleiss, Levin, & Paik, 2003) และความเที่ยงระหว่างผู้ประเมินในการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ที่วิเคราะห์ด้วยสถิติสหสัมพันธ์ภายในชั้น (intra-class correlation: ICC) ได้ค่าสหสัมพันธ์ภายในชั้นที่วิเคราะห์จากโมเดลอิทธิพลผสมแบบสองทาง โดยมีผู้ประเมินจำนวน 7 คน (ICC(3,7)) เท่ากับ 0.943 ที่ระดับความเชื่อมั่น 95% แสดงว่ามีความสอดคล้องระหว่างผู้ประเมินอยู่ในระดับดีมาก (Portney & Watkins, 2015)

### 3. การเก็บรวบรวมข้อมูล

เก็บรวบรวมข้อมูลในการวิจัยครั้งนี้แบ่งเป็น 3 ส่วน คือ (1) การเก็บข้อมูลคุณลักษณะจำเพาะของข้อสอบ (item-test specification) ในการทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 (2) การเก็บรวบรวมข้อมูลการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น และ (3) การเก็บรวบรวมข้อมูลที่เกี่ยวข้องหาประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น รายละเอียดมีดังนี้

#### 3.1 การเก็บข้อมูลคุณลักษณะจำเพาะของข้อสอบ (item-test specification) ในการทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 2559 ประจำปีการศึกษา 3

การเก็บข้อมูลคุณลักษณะจำเพาะของข้อสอบ (test specification) ในการทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (ordinary national educational test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 ดำเนินการโดยขอความอนุเคราะห์ข้อมูลคุณลักษณะจำเพาะของข้อสอบแต่ละข้อ ไปยัง สทศ. ได้แก่ ตัวชี้วัดที่ใช้ในการออกข้อสอบระดับความซับซ้อนทางปัญญาที่ต้องการวัด โดยใช้แบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบฯ ที่

ผู้วิจัยสร้างขึ้น ข้อมูลที่ได้จะถูกนำไปสร้างเป็นเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) เพื่อใช้ในการวิเคราะห์ค่าดัชนีความสอดคล้อง (alignment index) ต่อไป

### 3.2 การเก็บรวบรวมข้อมูลการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น

การเก็บรวบรวมข้อมูลการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ดำเนินการโดยการทำหนังสือขอความอนุเคราะห์ไปยังผู้อำนวยการโรงเรียนเพื่อเก็บรวบรวมข้อมูลดังนี้

3.2.1 ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 - 3) ทั้งข้อสอบระหว่างภาคและข้อสอบประจำภาค ปีการศึกษา 2559 พร้อมทั้งแผนผังในการออกข้อสอบ (test blueprint) (ถ้ามี) ของแต่ละโรงเรียนที่ได้รับการคัดเลือก จำนวน 4 โรงเรียน ซึ่งเป็นข้อสอบที่วัด 40 ตัวชี้วัดที่ สทศ. ใช้ในการออกข้อสอบ O-NET ได้ข้อสอบในการประเมินระดับชั้นเรียนที่ใช้ในการศึกษาทั้งหมด 1,089 ข้อ โดยข้อสอบแต่ละข้อจะต้องมีข้อมูลที่โรงเรียนระบุมาตรฐานและตัวชี้วัดที่ต้องการวัดด้วยข้อสอบข้อนั้นๆ ทั้งนี้ ในกรณีที่ข้อสอบไม่มี test blueprint หรือไม่มีการระบุมาตรฐานและตัวชี้วัดของข้อสอบแต่ละข้อ ผู้วิจัยได้ทำการเก็บข้อมูลเพิ่มเติม โดยนำข้อสอบกลับไปให้โรงเรียนช่วยระบุว่าข้อสอบแต่ละข้อใช้วัดมาตรฐานและตัวชี้วัดใด ก่อนนำมาคัดเลือกข้อสอบที่วัด 40 ตัวชี้วัดดังกล่าวไปใช้ในการศึกษาต่อไป

3.2.2 ข้อมูลเบื้องต้นที่เกี่ยวข้องกับการประเมินผลการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 - 3) ของโรงเรียน โดยใช้แบบสอบถามที่ให้โรงเรียนระบุข้อมูลเบื้องต้นในเรื่องจำนวนห้องเรียนโดยเฉลี่ยต่อระดับชั้น จำนวนนักเรียนโดยเฉลี่ยต่อห้อง จำนวนครูผู้สอน จำนวนครูผู้ออกข้อสอบ ข้อมูลของครูผู้ออกข้อสอบ ได้แก่ อายุ อายุงาน ประวัติการศึกษา ภาระงานสอน ภาระงานอื่นๆ การได้รับการอบรมด้านการวัดและประเมินผล ตลอดจนข้อมูลเกี่ยวกับกระบวนการในการออกข้อสอบของโรงเรียน

### 3.3 การเก็บรวบรวมข้อมูลจากผู้เชี่ยวชาญประเมินระดับความสอดคล้องในแนวเดียวกัน ระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น

การประเมินความสอดคล้องระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ดำเนินการโดยให้ผู้เชี่ยวชาญพิจารณาเพื่อประเมินความสอดคล้องระหว่างเนื้อหาของข้อสอบกับเนื้อหาของตัวชี้วัดที่โรงเรียนระบุ ข้อมูลที่ได้จะถูกนำไปสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (Y) เพื่อนำไปวิเคราะห์ค่าดัชนีความสอดคล้อง (alignment index) ตามวิธีการของ Porter ต่อไป รายละเอียดการดำเนินการเก็บรวบรวมข้อมูลจากผู้เชี่ยวชาญประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน มีดังนี้

1) จัดประชุมผู้เชี่ยวชาญทั้ง 20 ท่าน ใช้ระยะเวลาครึ่งวัน โดยมีจุดประสงค์เพื่อสร้างความเข้าใจที่ตรงกันระหว่างผู้เชี่ยวชาญเกี่ยวกับแนวทางในการดำเนินงาน ทดลองการฝึกปฏิบัติ และทำการปฏิบัติจริง เนื้อหาหรือกิจกรรมในประชุมมีดังนี้

1.1) ทบทวนนิยามของระดับความซับซ้อนทางปัญญาตามแนวคิด revised Bloom's taxonomy เพื่อให้ผู้เชี่ยวชาญสามารถพิจารณากำหนดระดับความซับซ้อนทางปัญญาที่ข้อสอบต้องการวัด ทั้งนี้ ได้มีการเชิญวิทยากรมาบรรยายให้ความรู้เกี่ยวกับการจำแนกระดับความซับซ้อนทางปัญญาตามแนวคิด revised Bloom's taxonomy ของข้อสอบ พร้อมทั้งมีเอกสารคู่มือที่มีคำอธิบายและยกตัวอย่างประกอบ

1.2) ให้ผู้เชี่ยวชาญทดลองฝึกปฏิบัติในกิจกรรมต่างๆ ได้แก่ การฝึกปฏิบัติกระบวนการค้นหามติเพื่อพิจารณาตัดสินการระบุระดับความซับซ้อนทางปัญญาของข้อสอบ การฝึกปฏิบัติการพิจารณาเพื่อประเมินความสอดคล้องระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบเป็นรายบุคคล และเปิดโอกาสให้ผู้เชี่ยวชาญได้อภิปรายในภายในกลุ่มเพื่อให้ได้ผลการพิจารณาที่ถูกต้องและเข้าใจตรงกัน

1.3) หลังจากผู้เชี่ยวชาญได้ฝึกปฏิบัติแล้ว ให้ผู้เชี่ยวชาญแต่ละคนทดลองทำหน้าที่จริงในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบฯ อย่างอิสระ โดยยังไม่ปรึกษากัน ประมาณ 10 - 15 ข้อ

2) หลังจากการประชุมผู้เชี่ยวชาญแล้ว ผู้เชี่ยวชาญแต่ละคนจะทำหน้าที่ประเมินความสอดคล้องระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ได้อย่างอิสระ โดยใช้แบบประเมินความสอดคล้องระหว่างในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบฯ และผู้ประเมินจะได้รับเอกสารคู่มือในการดำเนินงานที่ผู้วิจัยจัดทำขึ้น ทั้งนี้ผู้เชี่ยวชาญสามารถดำเนินการประเมินได้ 2 รูปแบบ ได้แก่ รูปแบบเอกสาร หรือรูปแบบออนไลน์ ขึ้นอยู่กับความสะดวกในการดำเนินงานของผู้เชี่ยวชาญแต่ละท่าน และเนื่องจากข้อสอบในการประเมินระดับชั้นเรียนที่ใช้ในการศึกษานี้มีจำนวนมากถึง 1,089 ข้อ ผู้วิจัยจึงออกแบบให้ข้อสอบแต่ละข้อถูกประเมินโดยผู้เชี่ยวชาญจำนวน 5 คน โดยข้อสอบทั้งหมดจะถูกสุ่มเข้าสู่ชุดข้อสอบย่อย (set) จำนวน 25 ชุด (ชุด A - Y) ชุดข้อสอบย่อยละประมาณ 40 - 46 ข้อ ผู้เชี่ยวชาญแต่ละคนจะได้พิจารณาข้อสอบอย่างสุ่ม คนละประมาณ 287 - 312 ข้อ (7 ชุดย่อย) แต่จะมีข้อสอบ จำนวน 40 ข้อ (ชุด A) ที่สุ่มจากตัวชี้วัดละ 1 ข้อ จะถูกประเมินด้วยผู้เชี่ยวชาญทั้ง 20 ท่าน ข้อมูลในส่วนนี้จะถูกนำไปใช้ในการวิเคราะห์ค่าความเที่ยงระหว่างผู้ประเมินต่อไป ทั้งนี้ ผู้วิจัยกำหนดกรอบเวลา 1 เดือน ที่ให้ผู้เชี่ยวชาญใช้ในการประเมินข้อสอบ รายละเอียดแผนผังการจัดข้อสอบที่ใช้ในการศึกษาความสอดคล้องในแนวเดียวกันแสดงดังตาราง 12

3) รวบรวมข้อมูลที่ได้จากผู้เชี่ยวชาญทุกคน ตรวจสอบความครบถ้วนของข้อมูล ก่อนบันทึกลงในตารางเพื่อใช้ในการวิเคราะห์ต่อไป

ตาราง 12 ตัวอย่างแผนผังการจัดข้อสอบที่ใช้ในการศึกษาความสอดคล้องในแนวเดียวกัน

ผู้ เชี่ยวชาญ	จำนวนข้อสอบจำแนกตามชุดข้อสอบย่อย (ข้อ)																				รวม					
	ชุด A	ชุด B	ชุด C	ชุด D	ชุด E	ชุด F	ชุด G	ชุด H	ชุด I	ชุด J	ชุด K	ชุด L	ชุด M	ชุด N	ชุด O	ชุด P	ชุด Q	ชุด R	ชุด S	ชุด T		ชุด U	ชุด V	ชุด W	ชุด X	ชุด Y
	(ข้อ 1-40)	(ข้อ 41-84)	(ข้อ 85-128)	(ข้อ 129-172)	(ข้อ 173-216)	(ข้อ 217-259)	(ข้อ 260-302)	(ข้อ 303-345)	(ข้อ 346-387)	(ข้อ 388-428)	(ข้อ 429-469)	(ข้อ 470-510)	(ข้อ 511-551)	(ข้อ 552-592)	(ข้อ 593-638)	(ข้อ 639-684)	(ข้อ 685-729)	(ข้อ 730-774)	(ข้อ 775-819)	(ข้อ 820-864)	(ข้อ 865-909)	(ข้อ 910-954)	(ข้อ 955-999)	(ข้อ 1000-1044)	(ข้อ 1045-1089)	
1	40	44	44	44	44		43														45					304
2	40	44	44	44	44																45	45				306
3	40	44	44	44	44																	45	45			306
4	40	44	44			43		43															45	45		304
5	40	44				43	43	43																45	45	303
6	40		44			43	43	43						41											45	299
7	40			44	44	43	43	43	42																	299
8	40			44	44	43	43	43		41																298
9	40								42	41	41	41	41	41												287
10	40								42	41	41	41	41	41												287
11	40								42	41	41	41	41	41												287
12	40								42	41	41	41	41	41												287
13	40												41								45	45	45	45	45	306
14	40											41									45	45	45	45	45	306
15	40														46	46	45	45	45	45						312
16	40														46	46	45	45	45	45						312
17	40														46	46	45	45	45	45						312
18	40														46	46	45	45	45	45						312
19	40														46	46	45	45	45	45						312
20	40										41										45	45	45	45	45	306

#### 4. การวิเคราะห์ข้อมูล

4.1 การตรวจสอบอิทธิพลของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน

การตรวจสอบอิทธิพลของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน โดยพิจารณาผลการวิเคราะห์จาก MFRM ด้วยโปรแกรม FACETS 3.80.3 (Linacre, 2017) มีรายละเอียดดังนี้

4.1 นำผลการประเมินที่ผู้เชี่ยวชาญในส่วนองระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ (0 - 4) ของข้อสอบทั้ง 1,089 ข้อ มาวิเคราะห์ค่าเฉลี่ยความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด เพื่อตัดสินระดับความสอดคล้องฯ โดยประยุกต์ใช้ MFRM ด้วยโปรแกรม FACETS 3.80.3 (Linacre, 2017) เพื่อควบคุมอิทธิพลการกดหรือปล่อยคะแนน ซึ่งจะให้ผลการประเมินที่ถูกต้องมากกว่าการใช้ค่าเฉลี่ยคะแนนดิบในการตัดสินความสอดคล้องฯ เนื่องจากค่าเฉลี่ยคะแนนความสอดคล้องฯ ที่ได้จาก MFRM เป็นคะแนนโลจิทที่ถูกรับ (calibrate) การกดหรือปล่อยคะแนนของผู้ประเมินแล้ว

โมเดลในการศึกษาครั้งนี้ ได้มีการปรับพารามิเตอร์ในการศึกษาจาก MFRM ทั่วไป (ซึ่งปกติจะประกอบด้วย 4 พารามิเตอร์ ได้แก่ ผู้สอบ (person) ข้อสอบ (item) ระดับความยากของข้อสอบ (item threshold) และผู้ตรวจหรือผู้ประเมิน (rater)) แต่โมเดลในการศึกษาครั้งนี้ประกอบด้วย 3 พารามิเตอร์ ได้แก่ ข้อสอบ (item) ผู้ประเมิน (rater) และระดับความกดคะแนนของผู้ประเมิน (rater threshold) โดยที่ให้พารามิเตอร์ข้อสอบ (item) เป็นสิ่งที่ถูกประเมินแทนที่จะเป็น ผู้สอบ (person) และพารามิเตอร์ผู้ประเมิน (raters) แทนข้อสอบ (item) กล่าวคือ กำหนดให้แถว (row) ของเมทริกซ์ข้อมูล (subscript n) เป็นพารามิเตอร์ข้อสอบ (items) และคอลัมน์ (column) ของเมทริกซ์ข้อมูล (subscript i) เป็นพารามิเตอร์ผู้ประเมิน (raters) (Anderson, Irvin, Alonzo, & Tindal, 2015) ซึ่งอยู่บนพื้นฐานของโมเดลมาตรประมาณค่าของ Andrich (Andrich's rating scale model) (1978) จึงมีข้อตกลงเบื้องต้น 2 ประการ คือ (1) ผู้ประเมินทุกคนมีความสามารถในการจำแนกระหว่างข้อสอบที่สอดคล้องและไม่สอดคล้องกับมาตรฐานและตัวชี้วัด และ (2) ระดับความสอดคล้องในแนวเดียวกันที่ผู้ประเมินจะเปลี่ยนจากลำดับขั้นหนึ่งไปยังอีกขั้นหนึ่งของมาตรประมาณค่ามีค่าเท่ากัน (เช่น จาก “ค่อนข้างสอดคล้อง (3)” ไป “ค่อนข้างสอดคล้องโดยตรง (4)”) ทั้งนี้ ผู้ประเมินจะเป็นผู้ตัดสินคุณลักษณะแฝง (latent trait) ของข้อสอบ โดยที่ข้อสอบจะอยู่บนเงื่อนไขของการกดหรือปล่อยคะแนนของผู้ประเมินให้คะแนนแต่ละคน สมการในการวิเคราะห์เขียนได้ดังนี้

$$\ln\left(\frac{P_{nik}}{P_{nik-1}}\right) = B_n - D_i - F_k$$

เมื่อ  $P_{nik}$  คือ ความน่าจะเป็นที่ผู้ประเมิน i จะประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบข้อที่ n โดยให้ระดับคะแนน k  
 $P_{nik-1}$  คือ ความน่าจะเป็นที่ผู้ประเมิน i จะประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบข้อที่ n โดยให้ระดับคะแนน k-1

$B_n$  คือ ความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบ (latent alignment) ข้อที่  $n$

$D_i$  คือ อิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมินคนที่  $i$

$F_k$  คือ พารามิเตอร์ประจำชั้นที่  $k$  สัมพันธ์กับชั้นที่  $k-1$

โดยที่  $F_k$  ถือว่าพารามิเตอร์ไม่นับว่าเป็นองค์ประกอบของโมเดล

จากสมการข้างต้น เป็นโมเดล 2 องค์ประกอบ ซึ่งตัวแปรตาม คือ คะแนนโลจิท ตัวแปรอิสระ คือ องค์ประกอบ (facets) ได้แก่ ข้อสอบ และผู้ประเมิน

4.2 ตรวจสอบความสอดคล้อง (fit) ระหว่างข้อมูลชุดนี้กับโมเดลของราส์ซ (Rasch model) โดยพิจารณาจากค่าเฉลี่ยของความคลาดเคลื่อน (mean residuals: Resd) และค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐาน (mean of the standardized residuals: StRes) ที่มีค่าเข้าใกล้ 0.00 และส่วนเบี่ยงเบนมาตรฐาน (sample standard deviation: SD) มีค่าเข้าใกล้ 1.00 จะแสดงว่าข้อมูลมีความสอดคล้องกับโมเดลของราส์ซ (Linacre, 2018)

4.3 ตรวจสอบเพื่อระบุผู้ประเมินที่แสดงอิทธิพลการกดหรือปล่อยคะแนน โดยพิจารณาคะแนนในหน่วยโลจิทของผู้ประเมิน (rater logit) หากมีคะแนนต่ำ หมายถึง ผู้ประเมินมีลักษณะปล่อยคะแนน (leniency) แต่หากมีคะแนนสูง หมายถึง ผู้ประเมินมีลักษณะกดคะแนน (severity) (Barrett, 2005: 164-5; Myford & Wolfe, 2000: 42)

4.4 คำนวณร้อยละของผู้ประเมินที่แสดงอิทธิพลของผู้ประเมิน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนอีก 3 ชนิด ได้แก่ อิทธิพลแนวโน้มการให้คะแนนค่ากลาง และอิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ตลอดจนตรวจสอบความแม่นยำในการให้คะแนน จากการประมาณค่าผู้ประเมิน (rater estimate) ด้วย MFRM จากฟาทของผู้ประเมิน มีรายละเอียดของเกณฑ์การพิจารณาดังนี้

1. อิทธิพลแนวโน้มการให้คะแนนค่ากลาง: พิจารณา ค่า infit MNSQ และ outfit MNSQ ของผู้ประเมินที่มีค่าน้อยกว่า 0.70 (Myford & Wolfe, 2000: 42)

2. อิทธิพลการจำกัดช่วง: พิจารณา ค่า infit MNSQ และ outfit MNSQ ของผู้ประเมินที่มีค่าน้อยกว่า 0.60 (Kassim, 2011)

3. อิทธิพลของความไม่มีแบบแผน: พิจารณา ค่า infit MNSQ และ outfit MNSQ ที่มีค่าสูง คือ มีค่า infit MNSQ และ outfit MNSQ มากกว่า 1.30 (Myford & Wolfe, 2000: 42)

4. ความแม่นยำในการให้คะแนน: พิจารณา infit MNSQ และ outfit MNSQ ที่มีค่าอยู่ระหว่าง 0.70 – 1.30 (Myford & Wolfe, 2000: 42)

สรุปเกณฑ์ในการพิจารณาดังนี้



ตาราง 13 ค่าสถิติในการตรวจสอบอิทธิพลของผู้ประเมินและความแม่นยำในการให้คะแนน  
ด้วย MFRM

ลักษณะของผู้ประเมิน	ค่าสถิติที่พิจารณาจากการประมาณค่าผู้ประเมินจากฟาเซทผู้ประเมิน		
	infit MNSQ	out MNSQ	คะแนนในหน่วยโลจิทของผู้ประเมิน
อิทธิพลการกดหรือ ปล่อยคะแนน	-	-	-มีค่าสูง --> ลักษณะการกดคะแนน -มีค่าต่ำ --> ลักษณะการปล่อยคะแนน
อิทธิพลแนวโน้ม การให้คะแนนค่ากลาง	infit < 0.70	outfit < 0.70	-
อิทธิพลการจำกัดช่วง	infit < 0.60	outfit < 0.60	-
อิทธิพลของความไม่มี แบบแผน	infit > 1.30	outfit > 1.30	-
มีความแม่นยำใน การให้คะแนน	$0.70 \leq \text{infit} \leq 1.30$	$0.70 \leq \text{outfit} \leq 1.30$	-

4.2 การเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

การเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ทำโดยการนำผลจากการวิเคราะห์ในข้อ 4.1 มาใช้ในการพิจารณา โดยมีวัตถุประสงค์เพื่อตัดสินความสอดคล้องในแนวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน โดยตรวจสอบว่าเมื่อมีการควบคุมอิทธิพลการกดหรือปล่อยคะแนนแล้ว ข้อสอบแต่ละข้อมีความสอดคล้องหรือวัดตัวชี้วัดที่ครูระบุจริงหรือไม่ ก่อนคัดเลือกข้อสอบไปใช้ในการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) และนำไปคำนวณดัชนีความสอดคล้อง (alignment index) ต่อไป การพิจารณาเปรียบเทียบมีรายละเอียดดังนี้

4.2.1 พิจารณาผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยพิจารณาค่า Observed Average (Obs. Avge) ซึ่งเป็นค่าเฉลี่ยคะแนนดิบ (raw mean) ของ

ข้อสอบเป็นรายข้อ ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยกำหนดเกณฑ์ความสอดคล้องของข้อสอบกับมาตรฐานและตัวชี้วัดที่ระบุ คือ ต้องมีค่า Obs. Avge ตั้งแต่ 3.00 คะแนนขึ้นไป

4.2.2 พิจารณาผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยพิจารณาค่า Fair-mean Average (Fair-M Avge) จากฟาเซทข้อสอบ (item facet) ซึ่งเป็นค่าเฉลี่ยคะแนนที่ปรับใหม่ (adjusted mean) หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน กำหนดเกณฑ์ที่ข้อสอบจะมีความสอดคล้องในแนวเดียวกันกับมาตรฐานและตัวชี้วัดที่ระบุก็ต่อเมื่อ

1) มีค่า Fair-M Avge ตั้งแต่ 3.00 คะแนนขึ้นไป

2) มีข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล คือ มีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.50 – 1.50

4.2.3 เปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ซึ่งมีการพิจารณาเปรียบเทียบผลการประเมินความสอดคล้องฯ ทั้งในภาพรวมและพิจารณาเป็นรายข้อ ดังนี้

4.2.3.1 ทดสอบความแตกต่างในภาพรวม ระหว่างค่าเฉลี่ยคะแนนการประเมินความสอดคล้องในแนวเดียวกันฯ ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน คือ การทดสอบความแตกต่างระหว่างค่า Obs. Avge กับค่า Fair-M Avge จากฟาเซทข้อสอบ ด้วยการทดสอบสถิติที (paired-samples t-test) และวิเคราะห์ขนาดอิทธิพล (effect size) ที่คำนวณจากสูตร Cohen's d ของ Cohen (1988) ดังนี้

$$d = \frac{t}{\sqrt{n}}$$

เมื่อ d คือ ขนาดอิทธิพลของ Cohen

t คือ ค่าสถิติการทดสอบที

n คือ จำนวนข้อสอบในการทดสอบ 1,089 ข้อ

กำหนดเกณฑ์ในการตีความหมายขนาดอิทธิพล (Cohen, 1988) ดังนี้

d = 0.20 คือ มีอิทธิพลขนาดเล็ก

d = 0.50 คือ มีอิทธิพลปานกลาง

d = 0.80 คือ มีอิทธิพลใหญ่

4.2.3.2 พิจารณาการเปลี่ยนแปลงผลการประเมินความสอดคล้องในแนวเดียวกันของข้อสอบฯ ระหว่างก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนนเป็นรายข้อ โดยการคำนวณร้อยละของข้อสอบที่มีการเปลี่ยนแปลงผลการประเมินความสอดคล้องระหว่างคะแนนก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Obs. Avge) กับคะแนนหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Fair-M Avge) ซึ่งการเปลี่ยนแปลงผลการประเมินมี 2 กรณี ดังนี้

1) การเปลี่ยนแปลงผลการประเมินของข้อสอบจาก “สอดคล้องกับตัวชี้วัดที่ระบุ (Obs. Avge  $\geq$  3.00)” เป็น “ไม่สอดคล้องกับตัวชี้วัดที่ระบุ (Fair-M Avge  $<$  3.00)”

2) การเปลี่ยนแปลงผลการประเมินของข้อสอบจาก “ไม่สอดคล้องกับตัวชี้วัดที่ระบุ (Obs. Avge  $<$  3.00)” เป็น “สอดคล้องกับตัวชี้วัดที่ระบุ (Fair-M Avge  $\geq$  3.00)”

**4.3 การวิเคราะห์ผลการศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน**

การวิเคราะห์ดัชนีความสอดคล้อง (alignment index) ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ตามแนวคิดของ Porter ทำโดยการสร้างเมทริกซ์ 2 เมทริกซ์ ได้แก่ เมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) แล้วคำนวณค่าดัชนีความสอดคล้อง (alignment index) โดยทำการวิเคราะห์ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ทั้งโดยภาพรวม และจำแนกแต่ละโรงเรียน รายละเอียดดังต่อไปนี้

1) สร้างเมทริกซ์เนื้อหา (context matrix) จำนวน 2 เมทริกซ์ ซึ่งประกอบด้วยเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) โดยเมทริกซ์ X และ Y ในการประเมินความสอดคล้องในแนวเดียวกันแต่ละมาตรฐานนั้น จะกำหนดให้แถวของเมทริกซ์ (แนวนอน) เป็นตัวชี้วัดภายใต้มาตรฐาน และกำหนดให้คอลัมน์ (แนวตั้ง) เป็นระดับความซับซ้อนทางปัญญาตามลำดับชั้นของกระบวนการทางปัญญา ภายใต้แนวคิด revised Bloom's taxonomy ซึ่งแบ่งเป็น 6 ชั้น ได้แก่ จำ เข้าใจ ประยุกต์ใช้ วิเคราะห์ ประเมินค่า และสร้างสรรค์ ตัวอย่างเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) แสดงดังภาพ 4

เมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X)						
มาตรฐาน/ตัวชี้วัด	ระดับความซับซ้อนทางปัญญา					
	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
สาระที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต 1						
มาตรฐาน ว 1.1 เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้าง และหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต						
ตัวชี้วัดที่ 4 ทดลองและอธิบายกระบวนการสารผ่านเซลล์ โดยการแพร่และออสโมซิส						
ตัวชี้วัดที่ 5 ทดลองหาปัจจัยบางประการที่จำเป็นต่อการสังเคราะห์ด้วยแสงของพืชและอธิบายว่าแสง คลอโรฟิลล์ แก๊ส คาร์บอนไดออกไซด์ น้ำ เป็นปัจจัยที่จำเป็น ต้องใช้ในการสังเคราะห์ด้วยแสง						
.....						

เมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y)						
มาตรฐาน/ตัวชี้วัด	ข้อที่	ระดับความซับซ้อนทางปัญญา				
		จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า
สาระที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต						
มาตรฐาน ว 1.1 เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้าง และหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต						
ตัวชี้วัดที่ 4 ทดลองและอธิบายกระบวนการสารผ่านเซลล์โดยการแพร่และออสโมซิส	1					
	2					
	...					
	...					
	n					
	รวม					
ตัวชี้วัดที่ 5 ทดลองหาปัจจัยบางประการที่จำเป็นต่อการสังเคราะห์ด้วยแสงของพืชและอธิบายว่าแสง คลอโรฟิลล์ แก๊ส คาร์บอนไดออกไซด์ น้ำ เป็นปัจจัยที่จำเป็นต้องใช้ในการสังเคราะห์ด้วยแสง	1					
	2					
	...					
	...					
	n					
	รวม					
.....						

ภาพ 4 ตัวอย่างเมทริกซ์ข้อสอบในการประเมินตัวอย่างเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y)

### 1.1) การสร้างเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X)

โดยทั่วไปการสร้างเมทริกซ์ข้อสอบตามแนวคิดของ Porter ดำเนินการโดยจำแนกข้อสอบตามเนื้อหาแต่ละหัวข้อ (topic) ที่ต้องการศึกษา (ซึ่งในงานวิจัยนี้จำแนกข้อสอบตามตัวชี้วัดและสรุปข้อมูลตามสาระ 7 สาระ) จากนั้นจึงให้ผู้เชี่ยวชาญพิจารณาระดับความซับซ้อนทางปัญญาของข้อสอบแต่ละข้อ แต่เนื่องจากข้อสอบ O-NET เป็นข้อสอบในการประเมินระดับชาติที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัด มีกระบวนการสร้างที่มีมาตรฐานโดยผู้เชี่ยวชาญที่ สทศ. พิจารณาแล้วว่ามีความรู้ความเชี่ยวชาญตามสาระการเรียนรู้และระดับชั้นนั้นๆ ดังนั้น ในการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) ในงานวิจัยนี้ ผู้วิจัยจึงได้ปรับการดำเนินงานจากแนวคิดของ Porter โดยไม่มีขั้นตอนที่ให้ผู้เชี่ยวชาญประเมินระดับความซับซ้อนทางปัญญาของข้อสอบ O-NET แต่ดำเนินการโดยการนำข้อมูลจากแบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (test-item specification) ของข้อสอบในการประเมินระดับชาติ มาลงข้อมูลและจัดทำเป็นเมทริกซ์ X ได้เลย โดยให้แต่ละช่องในเมทริกซ์ X เป็น ความถี่ที่ สทศ. กำหนดระบุว่าตัวชี้วัดนั้นๆ มีข้อสอบที่วัดระดับความซับซ้อนทางปัญญาในระดับใด เช่น สทศ. ระบุว่าข้อสอบข้อที่ 1 มีคะแนนเต็มรายข้อเท่ากับ 2 คะแนน ข้อสอบข้อนี้ใช้วัดมาตรฐาน 1.1 ของระดับมัธยมศึกษาปีที่ 1 ตัวบ่งชี้ที่ 1 ในระดับวิเคราะห์ ผู้วิจัยจะใส่เลข 2 ในช่องตัวชี้วัดที่ 1 ของมาตรฐาน 1.1 ระดับมัธยมศึกษาปีที่ 1 ที่ตัดกับระดับวิเคราะห์ และเติมเลข 0 ลงในช่องที่ข้อสอบข้อนี้ไม่ได้วัด ทำเช่นเดียวกันในข้อสอบข้ออื่นๆ

### 1.2) การสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) ดำเนินการดังนี้

1.2.1) การนำข้อมูลที่ผู้เชี่ยวชาญกำหนดระดับความซับซ้อนทางปัญญาของข้อสอบฯ จากแบบประเมินความสอดคล้องระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ในส่วนการประเมินระดับความซับซ้อนทางปัญญา ซึ่งผู้เชี่ยวชาญแต่ละคนจะพิจารณาข้อสอบและระบุระดับความซับซ้อนทางปัญญาที่ข้อสอบแต่ละข้อต้องการวัด โดยตามวิธีการของ Porter (2002) ผู้เชี่ยวชาญจะสามารถระบุความซับซ้อนทางปัญญาของข้อสอบข้อหนึ่งๆ ได้มากกว่า 1 ระดับ โดยไม่เกิน 3 ระดับ แต่ในงานวิจัยนี้ ได้ปรับการให้คะแนนประเมินเป็นแบบที่ให้ผู้เชี่ยวชาญสามารถระบุระดับความซับซ้อนทางปัญญาของข้อสอบแต่ละข้อได้เพียง 1 ระดับเท่านั้น โดยให้ผู้เชี่ยวชาญเติมเครื่องหมาย  ลงในช่อง  ที่แสดงระดับความซับซ้อนทางปัญญาที่ต้องการระบุ ซึ่งเป็นระดับความซับซ้อนทางปัญญาที่ผู้เชี่ยวชาญเห็นว่า เป็นระดับสูงสุดที่ผู้เรียนต้องใช้ในการทำข้อสอบข้อนั้นๆ ได้อย่างถูกต้อง เพื่อให้สอดคล้องกับข้อมูลของข้อสอบ O-NET ที่ สทศ. มีการระบุระดับความซับซ้อนทางปัญญาของข้อสอบแต่ละเพียงระดับเดียว

1.2.2) หลังจากเก็บข้อมูลจากผู้เชี่ยวชาญครบทุกคนแล้ว ผู้วิจัยทำการรวบรวมและบันทึกข้อมูลการประเมินข้อสอบแต่ละข้อจากผู้เชี่ยวชาญแต่ละคน เช่น ข้อสอบข้อที่ 1 มีคะแนนเต็มรายข้อเท่ากับ 0.50 คะแนน ผู้เชี่ยวชาญพิจารณาว่าข้อสอบข้อนี้ใช้วัดมาตรฐาน 2.1 ของระดับชั้น

มัธยมศึกษาปีที่ 3 ตัวบ่งชี้ที่ 1 ในระดับเข้าใจ ผู้วิจัยจะใส่เลข 0.50 ช่องตัวบ่งชี้ที่ 1 ของมาตรฐาน 2.1 ระดับชั้นมัธยมศึกษาปีที่ 3 ตัดกับระดับเข้าใจ และเติมเลข 0 ลงในช่องระดับอื่นๆ ข้อสอบข้อสอบข้อนี้ ไม่ได้วัดทำเช่นเดียวกันในผู้ประเมินคนอื่นๆ ที่ประเมินข้อสอบข้อที่ 1 และเช่นเดียวกันในข้อสอบ ข้ออื่นๆ จากนั้นทำการรวมคะแนนของข้อสอบทุกข้อในแต่ละสาระ จำแนกตามระดับความซับซ้อนทางปัญญา

1.2.3) นำผลรวมในเมทริกซ์ Y (ซึ่งเป็นตัวเลขในบรรทัดสุดท้ายของแต่ละสาระ ในเมทริกซ์) ของผู้เชี่ยวชาญแต่ละคนรวมกัน การหาผลรวมนี้เพื่อให้ได้ภาพรวมการประเมินของผู้เชี่ยวชาญแต่ละคนได้เป็นเมทริกซ์ผลรวม Y

2) ทำค่าในเมทริกซ์ผลรวม X และเมทริกซ์ผลรวม Y ให้เป็นเมทริกซ์สัดส่วน ซึ่งเป็นการทำให้ค่าในเมทริกซ์ผลรวม X และเมทริกซ์ผลรวม Y อยู่มาตรฐานเดียวกัน เพื่อให้สามารถเปรียบเทียบค่าระหว่าง 2 เมทริกซ์ได้ โดยนำค่าผลรวมของแต่ละเมทริกซ์ไปหารสมาชิกในเมทริกซ์เดียวกัน ผลของการหารสมาชิก (แต่ละช่อง) ทุกตัวของเมทริกซ์ด้วยผลรวม จะได้เป็นเมทริกซ์สัดส่วน และทำให้ ผลรวมของค่าสัดส่วนในเมทริกซ์จะเท่ากับ 1.00 ดังนั้น จากกระบวนการนี้ จะทำให้ได้เมทริกซ์สัดส่วนจำนวน 2 เมทริกซ์ คือ เมทริกซ์สัดส่วน X และเมทริกซ์สัดส่วน Y

3) นำเมทริกซ์สัดส่วน X และเมทริกซ์สัดส่วน Y จากข้อ 2) มาเปรียบเทียบกัน และคำนวณดัชนีความสอดคล้อง (alignment index) โดยการหาผลต่างของเมทริกซ์ X และ Y หรือ  $X - Y$  ถ้าหาก 2 เมทริกซ์มีความสอดคล้องกัน ผลลัพธ์ของผลต่างควรมีค่าเป็นศูนย์หรือเข้าใกล้ศูนย์ สูตรการคำนวณความสอดคล้องของ Porter เป็นดังนี้

$$\text{alignment index} = 1 - \frac{\sum |X_i - Y_i|}{2}$$

เมื่อ  $X_i$  คือ ค่าสมาชิกของเมทริกซ์ข้อสอบในการประเมินระดับชาติ ตำแหน่งที่  $i$

$Y_i$  คือ ค่าสมาชิกของเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ตำแหน่งที่  $i$

ทั้งนี้ ค่าดัชนีความสอดคล้อง (alignment index) มีค่าตั้งแต่ 0.00 ถึง 1.00 โดย 0.00 หมายถึง ไม่มีความสอดคล้องกัน และ 1.00 หมายถึง มีความสอดคล้องกันอย่างสมบูรณ์

**4.4 การประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้เชี่ยวชาญและรูปแบบการออกแบบการประเมินต่างกัน**

การประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจของผลการประเมินความสอดคล้องในแนวเดียวกัน ดำเนินการโดยใช้ข้อมูลการประเมินของผู้เชี่ยวชาญจากแบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบใน

การประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการประเมินต่างกัน โดยใช้การวิเคราะห์ความแปรปรวนแบบสองทาง (two-way ANOVA) มีขั้นตอนการดำเนินการดังต่อไปนี้

4.4.1 นำผลการประเมินความสอดคล้องในแนวเดียวกันๆ ของข้อสอบ (item : i) จำนวน 40 ข้อ (ชุดย่อย A) ที่ถูกประเมินด้วยผู้ประเมิน (rater: r) จำนวน 20 คน มาจัดกระทำตามเงื่อนไขจำนวนผู้ประเมิน ตั้งแต่ 1 – 20 คน ตามรูปแบบการประเมิน 2 รูปแบบ ดังนี้

รูปแบบที่ 1 ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ (i x r design)

รูปแบบที่ 2 ผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อ (r : i design) หรือที่เรียกว่า ผู้ประเมินสอดแทรก (nested) ภายในข้อสอบ

4.4.2 ทำการวิเคราะห์องค์ประกอบความแปรปรวน โดยใช้โมเดลการสรุปอ้างอิง เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน ซึ่งกำหนดให้ข้อสอบ (item: i) เป็นสิ่งที่ถูกประเมิน และผู้ประเมิน (rater: r) เป็นแหล่งความแปรปรวน รายละเอียดดังภาพ 5 ใช้การวิเคราะห์ความแปรปรวนแบบสองทาง (two-way ANOVA) ด้วยโปรแกรม SPSS โดยจำแนกผลการวิเคราะห์ตามลักษณะของแบบประเมินความสอดคล้องในแนวเดียวกันๆ แบ่งเป็น 2 ส่วน ได้แก่ การประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ที่มีลักษณะเป็นมาตรฐานค่า 5 ระดับ



เมื่อ  $\sigma^2_{X_{ir}}$  คือ ความแปรปรวนของคะแนนที่สังเกตได้

$\sigma^2_i$  คือ ความแปรปรวนของคะแนนเอกภาพของข้อสอบ ( $\mu_i$ )

$\sigma^2_r$  คือ ความแปรปรวนของคะแนนเฉลี่ยอันเกิดจากผู้ประเมิน ( $\mu_r$ )

$\sigma^2_{ir,e}$  คือ ความแปรปรวนของคะแนนความคลาดเคลื่อนหรือส่วนที่เหลือ ( $e_{ir}$ )

เมื่อ  $\sigma^2_{X_{r:i}}$  คือ ความแปรปรวนของคะแนนที่สังเกตได้

$\sigma^2_r$  คือ ความแปรปรวนของคะแนนเฉลี่ยอันเกิดจากผู้ประเมิน ( $\mu_r$ )

$\sigma^2_{r,ir,e}$  คือ ความแปรปรวนของคะแนนความคลาดเคลื่อนหรือส่วนที่เหลือ ( $e_{r,ir}$ )

โดยที่  $\sigma^2_{r,ir,e} = \sigma^2_r + \sigma^2_{ir,e}$

ภาพ 5 Venn diagram แสดงแหล่งความแปรปรวนของรูปแบบการประเมิน i x r design และ r : i design

4.4.3 ประมวลค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจ ( $\rho^2$ ) เมื่อจำนวนผู้เชี่ยวชาญเท่ากัน แต่มีรูปแบบการตรวจให้คะแนนต่างกัน โดยจำแนกเป็น 2 ส่วน ได้แก่ การประเมินความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ที่มีลักษณะเป็นมาตรฐานค่า สูตรในการคำนวณสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจ ดังนี้

(1) สัมประสิทธิ์การสรุปอ้างอิง

(2) สัมประสิทธิ์การสรุปอ้างอิง

สำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho_{\text{Rel}}^2$ )

สำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho_{\text{Abs}}^2$ )

$$\rho_{\text{Rel}}^2 = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_{\text{Rel}}^2}$$

$$\rho_{\text{Abs}}^2 = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_{\text{Abs}}^2}$$

เมื่อ  $\sigma_i^2$  คือ ความแปรปรวนของคะแนนความสอดคล้องในแนวเดียวกันของข้อสอบ

$\sigma_{\text{Rel}}^2$  คือ ความแปรปรวนของความคลาดเคลื่อนแบบสัมพัทธ์ (relative error variance)

$\sigma_{\text{Abs}}^2$  คือ ความแปรปรวนของความคลาดเคลื่อนแบบสัมบูรณ์ (absolute error variance)

ทั้งนี้ กำหนดเกณฑ์ความน่าเชื่อถือของการประเมินในระดับที่ยอมรับได้ คือ มีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับตัดสินเชิงสัมบูรณ์ ( $\rho_{\text{Abs}}^2$ ) ตั้งแต่ 0.70 ขึ้นไป (Nunnally, 1978)

4.4.4 ประมวลค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจ ( $\rho^2$ ) เมื่อรูปแบบการออกแบบการประเมินเหมือนกัน แต่จำนวนประเมินต่างกัน ตามสูตรในข้อ 4.4.3 โดยจำแนกเป็น 2 ส่วน คือ การประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ที่มีลักษณะเป็นมาตรฐานค่า

## 5. การจัดทำคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

การจัดทำคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียนซึ่งอยู่บนพื้นฐานของขั้นตอนการดำเนินการวิจัย และผลการวิจัยในส่วนของการประมวลค่าความเที่ยงในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินแตกต่างกัน มีรายละเอียดดังนี้

5.1 ผู้วิจัยร่างคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน โดยนำข้อมูลวิจัยมาใช้ในการร่างคู่มือฯ เพื่อเป็นแนวทางในการดำเนินงานให้กับครู ผู้บริหาร เขตพื้นที่การศึกษา และผู้ที่เกี่ยวข้อง



นำไปใช้ในการออกแบบการศึกษาความสอดคล้องในแนวเดียวกันต่อไป รายละเอียดของคู่มือประกอบด้วย 3 ส่วน ดังนี้

#### บทที่ 1 หลักการและวัตถุประสงค์ของคู่มือ

- 1.1 หลักการและแนวคิด
- 1.2 วัตถุประสงค์ของคู่มือ
- 1.3 ประโยชน์ที่คาดว่าจะได้รับจากการใช้คู่มือ
- 1.4 คำจำกัดความที่ใช้ในคู่มือ

#### บทที่ 2 แนวคิดและทฤษฎีที่เกี่ยวข้อง

- 2.1 แนวคิดการจำแนกพฤติกรรมการเรียนรู้ด้านพุทธิพิสัยของ Bloom ที่ได้รับการปรับปรุงใหม่ (revised Bloom's taxonomy)
- 2.2 แนวคิดพื้นฐานของโมเดลหลายองค์ประกอบของราล์ฟ
- 2.3 แนวคิดของวิธีการศึกษาความสอดคล้องในแนวเดียวกันของ Porter

บทที่ 3 ขั้นตอนในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter

- ขั้นตอนที่ 1 ขั้นเตรียมการ
- ขั้นตอนที่ 2 ขั้นดำเนินการ
- ขั้นตอนที่ 3 ขั้นสรุปผล

5.2 นำ (ร่าง) คู่มือฯ ที่ร่างขึ้นให้อาจารย์ที่ปรึกษาตรวจสอบความเหมาะสม และปรับแก้ตามข้อเสนอแนะ

5.3 นำ (ร่าง) คู่มือฯ ให้ผู้ทรงคุณวุฒิ จำนวน 5 ท่าน ตรวจสอบความเหมาะสม และปรับแก้ตามข้อเสนอแนะ โดยกำหนดเกณฑ์ในการคัดเลือกผู้ทรงคุณวุฒิ คือ เป็นผู้ทรงคุณวุฒิด้านการศึกษา ความสอดคล้องในแนวเดียวกันและมีความรู้ความสามารถด้านการวัดและประเมินผลการศึกษา โดยเป็นผู้ที่มีประสบการณ์ในการศึกษาวิจัยที่เกี่ยวกับสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบทางการศึกษา และสำเร็จการศึกษาในสาขาวิชาที่เกี่ยวกับการวัดและประเมินผลการศึกษา อย่างน้อยในระดับปริญญาโทขึ้นไป

ผู้วิจัยให้ผู้ทรงคุณวุฒิทั้ง 5 ท่าน ประเมินคู่มือฯ ด้วยแบบประเมินแบบตรวจสอบความเหมาะสมของคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันฯ ที่ผู้วิจัยสร้างขึ้น ประกอบด้วย 2 ส่วน คือ 1) การตรวจสอบความสอดคล้องความเหมาะสมของคู่มือฯ กับวัตถุประสงค์ของคู่มือฯ และ 2) การตรวจสอบความเหมาะสมของคู่มือฯ ตามมาตรฐานการประเมิน 4 ด้าน (The Joint Committee on Standards for Educational Evaluation, 1994) ได้แก่ ด้าน

อรรถประโยชน์ (utility) ด้านความเป็นไปได้ (feasibility) ด้านความเหมาะสม (propriety) และด้านความถูกต้อง (accuracy)

5.4 วิเคราะห์ข้อมูลผลการประเมินจากผู้ทรงคุณวุฒิ มีรายละเอียดดังนี้

5.4.1 วิเคราะห์ผลการตรวจสอบด้านความตรงตามเนื้อหา ซึ่งเป็นการตรวจสอบความสอดคล้องความเหมาะสมของคู่มือฯ กับวัตถุประสงค์ของคู่มือฯ ด้วยการวิเคราะห์ค่า index of item – objective congruence (ค่าดัชนี IOC) มีลักษณะการให้คะแนน คือ

1 หมายถึง รายการที่ถูกประเมินมีความเหมาะสม/สอดคล้องกับวัตถุประสงค์ของคู่มือฯ

0 หมายถึง ไม่แน่ใจว่ารายการที่ถูกประเมินมีความเหมาะสม/สอดคล้องกับวัตถุประสงค์ของคู่มือฯ

-1 หมายถึง รายการที่ถูกประเมินไม่เหมาะสม/สอดคล้องกับวัตถุประสงค์ของคู่มือฯ

การคำนวณค่าดัชนี IOC มีสูตรดังนี้

$$IOC = \frac{\sum R}{N}$$

เมื่อ IOC คือ ดัชนีความสอดคล้องระหว่างรายการที่ถูกประเมินกับวัตถุประสงค์ของคู่มือฯ

$\sum R$  คือ ผลรวมของคะแนนผลการตัดสินของผู้ทรงคุณวุฒิ

N คือ จำนวนผู้ทรงคุณวุฒิ

ทั้งนี้ กำหนดเกณฑ์ที่ใช้ในการตัดสินความตรงเชิงเนื้อหา คือ รายการที่ถูกประเมินต้องมีค่าดัชนี IOC มากกว่า 0.50 จึงจะถือว่ารายการที่ถูกประเมินนั้นมีความเหมาะสม/สอดคล้องกับวัตถุประสงค์ของคู่มือฯ (ศิริชัย กาญจนวาสี, 2556)

5.4.1 วิเคราะห์ผลการตรวจสอบความเหมาะสมของคู่มือฯ ตามมาตรฐานการประเมิน 4 ด้าน ได้แก่ ด้านอรรถประโยชน์ (utility) ด้านความเป็นไปได้ (feasibility) ด้านความเหมาะสม (propriety) และด้านความถูกต้อง (accuracy) โดยแปลความหมายของคะแนนจากแบบประเมินเทียบกับเกณฑ์ (นพวรรณ ศรีเกตุ, 2557) ดังนี้

คะแนนเฉลี่ย 4.51 – 5.00 หมายถึง คู่มือฯ มีความเหมาะสมในระดับมากที่สุด

คะแนนเฉลี่ย 3.51 – 4.50 หมายถึง คู่มือฯ มีความเหมาะสมในระดับมาก

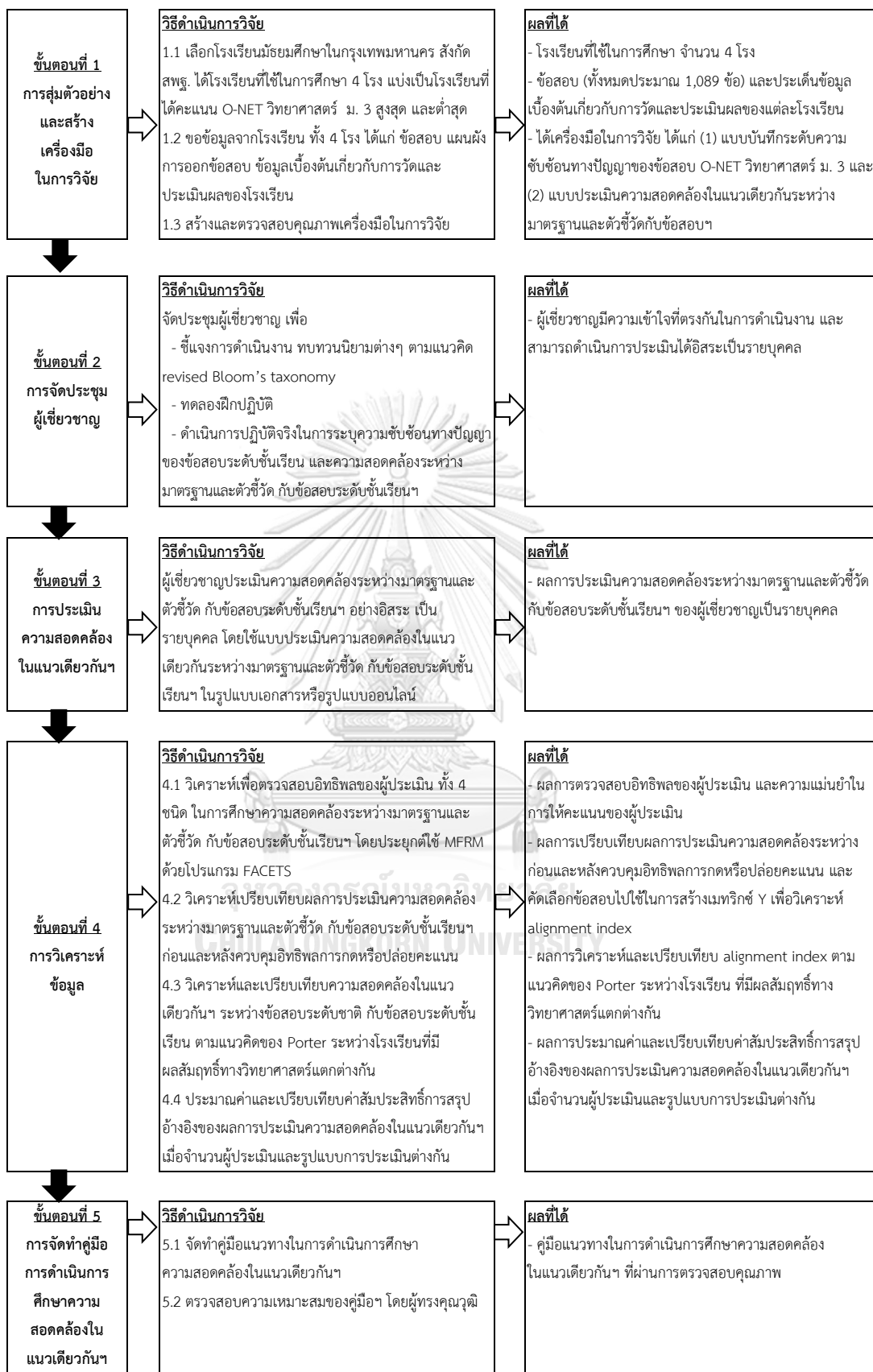
คะแนนเฉลี่ย 2.51 – 3.50 หมายถึง คู่มือฯ มีความเหมาะสมในระดับปานกลาง

คะแนนเฉลี่ย 1.51 – 2.50 หมายถึง คู่มือฯ มีความเหมาะสมในระดับน้อย

คะแนนเฉลี่ย 1.00 – 1.50 หมายถึง คู่มือฯ มีความเหมาะสมในระดับน้อยที่สุด

5.5 ผู้วิจัยปรับแก้ตามข้อเสนอแนะของผู้ทรงคุณวุฒิ และจัดทำเป็นคู่มือฯ ฉบับสมบูรณ์

แผนผังสรุปขั้นตอนการดำเนินการวิจัย แสดงดังภาพ 6



ภาพ 6 แผนผังแสดงขั้นตอนการดำเนินการวิจัย

## บทที่ 4

### ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ตรวจสอบอิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น 2) เปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน 3) ศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน และ 4) ประเมินค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน ผู้วิจัยได้แบ่งการนำเสนอผลการวิเคราะห์ข้อมูลออกเป็น 5 ตอน ดังนี้

ตอนที่ 1 ข้อมูลพื้นฐานของตัวอย่างในการวิจัย

ตอนที่ 2 ผลการตรวจสอบอิทธิพลของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน

ตอนที่ 3 ผลการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

ตอนที่ 4 ผลการศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน

ตอนที่ 5 ผลการประเมินค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน

ตอนที่ 6 ผลการตรวจสอบคุณภาพคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

มีรายละเอียดของผลการวิเคราะห์ข้อมูล มีดังต่อไปนี้

## ตอนที่ 1 ข้อมูลพื้นฐานของตัวอย่างในการวิจัย

ข้อมูลพื้นฐานของตัวอย่างในการวิจัยเป็นข้อมูลเบื้องต้นจากการเก็บรวบรวมข้อมูลตัวอย่างที่ใช้ในการวิจัย ได้แก่ ข้อสอบในการประเมินระดับชาติ ข้อสอบในการประเมินผลระดับชั้นเรียน และผู้เชี่ยวชาญในการตรวจสอบความสอดคล้องในแนวเดียวกัน รายละเอียดข้อมูลเบื้องต้นของตัวอย่างในการวิจัยมีรายละเอียดดังนี้

### 1.1 ข้อมูลพื้นฐานข้อสอบในการประเมินระดับชาติ

ข้อสอบในการประเมินระดับชาติ ได้แก่ ข้อสอบในการทดสอบทางการศึกษาระดับชาตินิยมขั้นพื้นฐาน (ordinary national educational test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำนวน 45 ข้อ ประกอบด้วยข้อสอบปรนัย 4 ตัวเลือก 1 คำตอบ จำนวน 40 ข้อ และข้อสอบปรนัย 6 ตัวเลือก 2 คำตอบ จำนวน 5 ข้อ ซึ่งผู้วิจัยได้ขอข้อมูลคุณลักษณะจำเพาะของข้อสอบแต่ละข้อ (item-test specification) จาก สทศ. ทั้งนี้ จากข้อสอบทั้งหมด 45 ข้อ ที่ สทศ. ใช้ตัวชี้วัดจำนวน 40 ตัวชี้วัด พบว่า ข้อสอบวัดมาตรฐานและตัวชี้วัดภายใต้สาระที่ 1 จำนวน 12 ข้อ สาระที่ 2 จำนวน 4 ข้อ สาระที่ 3 จำนวน 7 ข้อ สาระที่ 4 จำนวน 9 ข้อ สาระที่ 5 จำนวน 7 ข้อ สาระที่ 6 จำนวน 2 ข้อ สาระที่ 7 จำนวน 4 ข้อ และพบว่า ข้อสอบส่วนใหญ่วัดความซับซ้อนทางปัญญาระดับเข้าใจและระดับวิเคราะห์ จำนวนระดับละ 16 ข้อ คิดเป็นร้อยละ 35.56 เท่ากัน รองลงมาคือ ระดับประยุกต์ใช้ จำนวน 12 ข้อ คิดเป็นร้อยละ 26.67 และระดับประเมินค่า จำนวน 1 ข้อ คิดเป็นร้อยละ 2.22 รายละเอียดดังตาราง 14

ตาราง 14 คุณลักษณะจำเพาะของข้อสอบ (item-test specification) ในการทดสอบ O-NET วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559

สาระ	ข้อที่	มาตรฐานและตัวชี้วัด	ระดับความซับซ้อนทางปัญญา				คะแนนรายข้อ	
			จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์		
1	1	ว 1.1 ม .1/4			✓		2	
	2	ว 1.1 ม .1/4		✓			2	
	3	ว 1.1 ม .1/5				✓	2	
	4	ว 1.1 ม .1/5		✓			2	
	5	ว 1.1 ม .2/2, ว 1.1 ม .2/3		✓			2	
	6	ว 1.1 ม .2/2, ว 1.1 ม .2/3				✓	2	
	7	ว 1.1 ม .2/4		✓			2	
	8	ว 1.1 ม .2/5			✓		2	
	9	ว 1.1 ม .2/5				✓	2	
	10	ว 1.2 ม .3/3			✓		2	
	11	ว 1.2 ม .3/6					✓	2
	43	ว 1.2 ม .3/3				✓	4	
	รวม	12	ข้อ					

ตาราง 14 (ต่อ)

สาระ	ข้อที่	มาตรฐานและตัวชี้วัด	ระดับความซับซ้อนทางปัญญา					คะแนน รายข้อ
			จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	
2	12	ว 2.1 ม .3/1		✓				2
	13	ว 2.2 ม .3/2				✓		2
	14	ว 2.2 ม .3/4				✓		2
	45	ว 2.1 ม .3/3				✓		4
<b>รวม</b>	<b>4</b>	<b>ข้อ</b>						
3	15	ว 3.1 ม .1/1		✓				2
	16	ว 3.1 ม .1/2				✓		2
	17	ว 3.1 ม .1/4				✓		2
	18	ว 3.1 ม .2/2			✓			2
	19	ว 3.1 ม .2/3			✓			2
	20	ว 3.2 ม .1/1			✓			2
	21	ว 3.2 ม .1/2, ว 3.2 ม .2/1				✓		2
<b>รวม</b>	<b>7</b>	<b>ข้อ</b>						
4	22	ว 4.1 ม .1/2			✓			2
	23	ว 4.1 ม .1/2			✓			2
	24	ว 4.1 ม .2/2		✓				2
	25	ว 4.1 ม .3/1		✓				2
	26	ว 4.1 ม .3/2		✓				2
	27	ว 4.1 ม .3/1				✓		2
	28	ว 4.2 ม .3/2, ว 4.2 ม .3/3			✓			2
	41	ว 4.2 ม .3/1				✓		4
	42	ว 4.1 ม .1/2			✓			4
<b>รวม</b>	<b>9</b>	<b>ข้อ</b>						
5	29	ว 5.1 ม .1/2		✓				2
	30	ว 5.1 ม .1/2		✓				2
	31	ว 5.1 ม .2/3				✓		2
	32	ว 5.1 ม .2/3		✓				2
	33	ว 5.1 ม .3/1		✓				2
	34	ว 5.1 ม .3/3, ว 5.1 ม .3/4			✓			2
	44	ว 5.1 ม .2/1, ว 5.1 ม .2/2				✓		4
<b>รวม</b>	<b>7</b>	<b>ข้อ</b>						
6	35	ว 6.1 ม .1/6				✓		2
	36	ว 6.1 ม .2/2, ว 6.1 ม .2/4		✓				2
<b>รวม</b>	<b>2</b>	<b>ข้อ</b>						
7	37	ว 7.1 ม .3/1		✓				2
	38	ว 7.1 ม .3/3				✓		2
	39	ว 7.2 ม .3/1			✓			2
	40	ว 7.2 ม .3/1		✓				2
<b>รวม</b>	<b>4</b>	<b>ข้อ</b>						
		รวม (ข้อ)	0	16	12	16	1	0
		ร้อยละ	0.00%	35.56%	26.67%	35.56%	2.22%	0.00%

## 1.2 ข้อมูลพื้นฐานของโรงเรียนและข้อสอบในการประเมินระดับชั้นเรียน

### 1.2.1 ข้อมูลพื้นฐานของโรงเรียน

เมื่อพิจารณาข้อมูลพื้นฐานของโรงเรียนที่เป็นตัวอย่างในการวิจัย จากโรงเรียนทั้งหมด 4 โรงเรียน จำแนกเป็นโรงเรียนที่มีคะแนน O-NET สูง จำนวน 2 โรงเรียน และโรงเรียนที่มีคะแนน O-NET ต่ำ จำนวน 2 โรงเรียน พบว่า โรงเรียนที่มีคะแนน O-NET สูง จำนวน 2 โรงเรียน ได้แก่ โรงเรียน<sub>ONETสูง1</sub> และโรงเรียน<sub>ONETสูง2</sub> มีค่าเฉลี่ยคะแนน O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 เท่ากับ 60.24 และ 49.59 และโรงเรียนที่มีคะแนน O-NET ต่ำ จำนวน 2 โรงเรียน ได้แก่ โรงเรียน<sub>ONETต่ำ1</sub> และ โรงเรียน<sub>ONETต่ำ2</sub> มีค่าเฉลี่ยคะแนน O-NET เท่ากับ 33.88 และ 31.05 โรงเรียนมีจำนวนห้องเรียนในแต่ละระดับชั้น (ม. 1 – 3) โดยเฉลี่ย 6 – 12 ห้อง มีจำนวนนักเรียนโดยเฉลี่ย 35 – 50 คนต่อห้อง ต่อระดับชั้น โรงเรียนส่วนใหญ่มีครูผู้สอนวิทยาศาสตร์ในแต่ละระดับชั้นจำนวน 2 – 4 คน และส่วนใหญ่ในแต่ละระดับชั้นครูเกือบทุกคนที่สอนจะเป็นผู้ออกข้อสอบ ซึ่งมีจำนวน 2 – 4 คน รายละเอียดแสดงดังตาราง 15

เมื่อพิจารณาอายุของครูผู้ออกข้อสอบวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น จากโรงเรียนทั้ง 4 โรงเรียน พบว่า โดยภาพรวมครูผู้ออกข้อสอบมีอายุระหว่าง 26 – 60 ปี โดยมีอายุเฉลี่ยอยู่ระหว่าง 36.50 – 45.83 ปี มีอายุงานอยู่ระหว่าง 6 – 36 ปี มีอายุงานโดยเฉลี่ยอยู่ระหว่าง 8.78 – 20.67 ปี โดยครูผู้ออกข้อสอบของโรงเรียน<sub>ONETสูง1</sub> มีอายุระหว่าง 30 – 35 ปี อายุโดยเฉลี่ยเท่ากับ 36.50 ปี มีอายุงานอยู่ระหว่าง 6 – 30 ปี อายุงานโดยเฉลี่ยเท่ากับ 20.00 ปี ครูผู้ออกข้อสอบของโรงเรียน<sub>ONETสูง2</sub> มีอายุอยู่ระหว่าง 26 – 60 ปี อายุโดยเฉลี่ยเท่ากับ 35.44 ปี มีอายุงานอยู่ระหว่าง 1 – 36 ปี อายุงานโดยเฉลี่ยเท่ากับ 8.78 ปี ครูผู้ออกข้อสอบของโรงเรียน<sub>ONETต่ำ1</sub> มีอายุอยู่ระหว่าง 28 – 58 ปี อายุโดยเฉลี่ยเท่ากับ 40.16 ปี มีอายุงานอยู่ระหว่าง 1 – 36 ปี อายุงานโดยเฉลี่ยเท่ากับ 15.50 ปี และครูผู้ออกข้อสอบวิทยาศาสตร์ของโรงเรียน<sub>ONETต่ำ2</sub> มีอายุอยู่ระหว่าง 25 – 59 ปี อายุโดยเฉลี่ยเท่ากับ 45.83 ปี มีอายุงานอยู่ระหว่าง 1 – 35 ปี อายุงานโดยเฉลี่ยเท่ากับ 20.67 ปี

เมื่อพิจารณาประวัติการศึกษาของครูผู้ออกข้อสอบวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น จากโรงเรียนทั้ง 4 โรงเรียน พบว่า โดยภาพรวมครูผู้ออกข้อสอบส่วนใหญ่สำเร็จการศึกษาระดับปริญญาตรี โดยครูผู้ออกข้อสอบของโรงเรียน<sub>ONETสูง1</sub> โรงเรียน<sub>ONETต่ำ1</sub> และโรงเรียน<sub>ONETต่ำ2</sub> สำเร็จการศึกษาระดับปริญญาตรีด้านการศึกษาศาสตร์ ร้อยละ 100.00 และโรงเรียนทั้ง 3 โรงเรียนมีครูผู้ออกข้อสอบที่สำเร็จการศึกษาระดับปริญญาโททางการศึกษาร้อยละ 16.67 สำหรับครูผู้ออกข้อสอบของโรงเรียน<sub>ONETต่ำ1</sub> ส่วนใหญ่สำเร็จการศึกษาระดับปริญญาตรีด้านวิทยาศาสตร์ ร้อยละ 55.56 สำเร็จการศึกษาระดับปริญญาตรีทางการศึกษาศาสตร์ ร้อยละ 44.44 และมีครูที่สำเร็จการศึกษาระดับปริญญาโททางการศึกษาร้อยละ 44.44

เมื่อพิจารณาภาระงานของครูผู้ออกข้อสอบวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น พบว่า โดยภาพรวมมีจำนวนชั่วโมงสอนโดยเฉลี่ยเท่ากับ 9.00 – 18.00 ชั่วโมงต่อคน ต่อสัปดาห์ โดยครูผู้ออกข้อสอบของโรงเรียน<sub>ONETสูง1</sub> โรงเรียน<sub>ONETสูง2</sub> และโรงเรียน<sub>ONETต่ำ1</sub> มีจำนวนชั่วโมงสอนโดยเฉลี่ยใกล้เคียงกัน เท่ากับ 16.00 18.00 และ 17.67 ชั่วโมงต่อคน ต่อสัปดาห์ ตามลำดับ และครูผู้ออกข้อสอบของโรงเรียน<sub>ONETต่ำ2</sub> มีจำนวนชั่วโมงสอน โดยเฉลี่ยที่ค่อนข้างน้อย คือเท่ากับ 9.00 ชั่วโมงต่อคน ต่อสัปดาห์ และพบว่า โดยภาพรวมครูผู้ออกข้อสอบจำนวนชั่วโมงภาระงานอื่นที่ได้รับมอบหมายโดยเฉลี่ยเท่ากับ 3.50 – 18.67 ชั่วโมงต่อคน ต่อสัปดาห์ โดยครูผู้ออกข้อสอบวิทยาศาสตร์ของโรงเรียน<sub>ONETต่ำ2</sub> มีจำนวนชั่วโมงภาระงานอื่นที่ได้รับมอบหมายโดยเฉลี่ยมากที่สุด เท่ากับ 18.67 ชั่วโมงต่อคน ต่อสัปดาห์ รองลงมาคือ ครูผู้ออกข้อสอบของโรงเรียน<sub>ONETต่ำ1</sub> มีจำนวนชั่วโมงภาระงานอื่นที่ได้รับมอบหมายโดยเฉลี่ย เท่ากับ 6.50 ชั่วโมงต่อคน ต่อสัปดาห์ และครูผู้ออกข้อสอบของโรงเรียน<sub>ONETสูง1</sub> โรงเรียน<sub>ONETสูง2</sub> มีจำนวนชั่วโมงภาระงานอื่นที่ได้รับมอบหมายโดยเฉลี่ยใกล้เคียงกัน เท่ากับ 3.67 และ 3.50 ชั่วโมงต่อคน ต่อสัปดาห์ ตามลำดับ

เมื่อพิจารณาการได้รับการอบรมด้านการวัดและประเมินผลของครูผู้ออกข้อสอบวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น ในรอบ 5 ปีที่ผ่านมา พบว่า โดยภาพรวมครูผู้ออกข้อสอบได้รับการอบรมด้านการวัดและประเมินผล จำนวน 0 – 15 ครั้ง โดยโดยครูผู้ออกข้อสอบของโรงเรียน<sub>ONETต่ำ2</sub> ได้รับการอบรมด้านการวัดและประเมินผลมากที่สุด จำนวน 10 – 15 ครั้งต่อคน รองลงมาคือ ครูผู้ออกข้อสอบของโรงเรียน<sub>ONETต่ำ1</sub> ได้รับการอบรมด้านการวัดและประเมินผล จำนวน 0 - 2 ครั้งต่อคน ครูผู้ออกข้อสอบของโรงเรียน<sub>ONETสูง2</sub> ได้รับการอบรมด้านการวัดและประเมินผล จำนวน 0 – 1 ครั้งต่อคน และครูผู้ออกข้อสอบของโรงเรียน<sub>ONETสูง1</sub> ไม่ได้รับการอบรมด้านการวัดและประเมินผล ในรอบ 5 ปีที่ผ่านมา

**ตาราง 15 ข้อมูลเกี่ยวกับการออกข้อสอบในการประเมินระดับชั้นเรียน จำแนกตามโรงเรียน**

ข้อมูล	ร.ร. ONETสูง1	ร.ร. ONETสูง2	ร.ร. ONETต่ำ1	ร.ร. ONETต่ำ2
1. ค่าเฉลี่ยคะแนน O-NET วิทยาศาสตร์ ปีการศึกษา 2559	60.24	49.59	33.88	31.05
2. จำนวนห้องเรียนโดยเฉลี่ย (ห้อง/ระดับชั้น)	8	12	12	6
3. จำนวนนักเรียนโดยเฉลี่ย (คน/ห้อง/ระดับชั้น)	50	50	45	35
4. จำนวนครูที่สอนวิทยาศาสตร์ในแต่ละระดับชั้น (คน)	2	3 – 4	3 – 4	2
5. จำนวนครูที่ออกข้อสอบวิทยาศาสตร์ ในแต่ละระดับชั้น (คน)	2	3 – 4	2	2
6. อายุและ (อายุโดยเฉลี่ย) ของครูผู้ออกข้อสอบ (ปี)	30 - 35 (36.50)	26 - 60 (35.44)	28 - 58 (40.16)	25 - 59 (45.83)



ตาราง 15 (ต่อ)

ข้อมูล	ร.ร. ONET สูง 1	ร.ร. ONET สูง 2	ร.ร. ONET ต่ำ 1	ร.ร. ONET ต่ำ 2
7. อายุงานและ (อายุงานโดยเฉลี่ย) ของ ครูผู้ออกข้อสอบ (ปี)	6 - 30 (20.00)	1 - 36 (8.78)	1 - 36 (15.50)	1 - 35 (20.67)
8. ประวัติการศึกษาของครูผู้ออกข้อสอบ				
8.1 ปริญญาตรี				
8.1.1 ปริญญาตรีด้านการศึกษาวิทยาศาสตร์	100.00	44.44	100.00	100.00
8.1.2 ปริญญาตรีด้านอื่นๆ เช่น วิทยาศาสตร์	0.00	55.56	0.00	0.00
8.2 ปริญญาโท (ด้านการศึกษา)				
8.2.1 ปริญญาโท (ด้านการศึกษา)	100.00	100.00	100.00	100.00
8.2.2 ปริญญาโทด้านอื่นๆ	0.00	0.00	0.00	0.00
9. จำนวนชั่วโมงสอนของครูผู้ออกข้อสอบ โดยเฉลี่ยต่อ 1 สัปดาห์ (ชั่วโมง/สัปดาห์/คน)	16.00	18.00	17.67	9.00
10. จำนวนชั่วโมงภาระงานอื่นที่ครูผู้ออกข้อสอบได้รับ มอบหมาย โดยเฉลี่ยต่อ 1 สัปดาห์ (ชั่วโมง/สัปดาห์/ คน)	3.67	3.50	6.50	18.67
11. จำนวนครั้งในการอบรมด้านการวัดและประเมินผล ของครูผู้ออกข้อสอบ (ในรอบ 5 ปีที่ผ่านมา) (ครั้ง/คน)	0	0 - 1	0 - 2	10 - 15

เมื่อพิจารณาข้อมูลของโรงเรียนเกี่ยวกับข้อมูลเกี่ยวกับกระบวนการในการออกข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น จากโรงเรียนทั้ง 4 โรงเรียน ได้แก่ โรงเรียน ONET สูง 1 โรงเรียน ONET สูง 2 โรงเรียน ONET ต่ำ 1 และโรงเรียน ONET ต่ำ 2 พบว่าทุกโรงเรียนมีการประชุมร่วมกันระหว่างครูผู้สอนเพื่อออกข้อสอบ เท่ากับร้อยละ 100.00 100.00 66.67 และ 100.00 ตามลำดับ ทุกโรงเรียนมีการระบุเนื้อหา และตัวชี้วัดที่ต้องการวัด เท่ากับร้อยละ 100.00 66.67 100.00 และ 100.00 ตามลำดับ ทุกโรงเรียนมีการระบุระดับพฤติกรรมการเรียนรู้ที่ต้องการวัดของข้อสอบแต่ละข้อ เท่ากับร้อยละ 100.00 100.00 66.67 และ 33.33 โรงเรียนที่มีการจัดทำแผนผังในการออกข้อสอบ (test blueprint) มีจำนวน 3 โรงเรียน ได้แก่ โรงเรียน ONET สูง 1 และโรงเรียน ONET ต่ำ 2 เท่ากับร้อยละ 100.00 เท่ากัน และโรงเรียน ONET สูง 2 เท่ากับร้อยละ 33.33 ทุกโรงเรียนมีการพิจารณา/ตรวจสอบความถูกต้องและความเหมาะสมของแบบทดสอบโดยผู้อื่นที่ไม่ใช่ครูผู้ออกข้อสอบ ก่อนนำไปใช้จริง เท่ากับร้อยละ 100.00 66.67 100.00 และ 100.00 ตามลำดับ ทุกโรงเรียน

มีการวิเคราะห์ข้อสอบหลังจากนำข้อสอบไปใช้จริง เช่น การวิเคราะห์ค่าความยากค่าอำนาจจำแนก ประสิทธิภาพตัวลอง เท่ากับร้อยละ 100.00 เท่ากัน มีโรงเรียนที่มีการนำผลการวิเคราะห์ข้อสอบไปใช้ในการปรับปรุงการออกข้อสอบในครั้งถัดไป จำนวน 3 โรง ได้แก่ โรงเรียน<sub>ONET</sub>ต่ำ2 โรงเรียน<sub>ONET</sub>สูง2 และโรงเรียน<sub>ONET</sub>ต่ำ1 เท่ากับร้อยละ 100.00 66.67 และ 33.33 ตามลำดับ และมีโรงเรียน<sub>ONET</sub>สูง2 เพียงโรงเรียนเดียวที่มีการคัดเลือกข้อสอบที่มีคุณภาพจัดเก็บเป็นฐานข้อมูลหรือคลังข้อสอบ รายละเอียดดังตาราง 16

ตาราง 16 ข้อมูลเกี่ยวกับกระบวนการในการออกข้อสอบของโรงเรียน

ข้อมูล	ร้อยละ			
	ร.ร. <sub>ONET</sub> สูง1	ร.ร. <sub>ONET</sub> สูง2	ร.ร. <sub>ONET</sub> ต่ำ1	ร.ร. <sub>ONET</sub> ต่ำ2
1. ประชุมร่วมกันระหว่างครูผู้สอนเพื่อออกข้อสอบ (กรณีที่มีผู้สอนมากกว่า 1 คน)	100.00	100.00	66.67	100.00
2. ระบุเนื้อหา และตัวชี้วัดที่ต้องการวัด	100.00	66.67	100.00	100.00
3. ระบุระดับพฤติกรรมการเรียนรู้ที่ต้องการวัดของข้อสอบ แต่ละข้อ (เช่น รู้/จำ เข้าใจ ประยุกต์ใช้ วิเคราะห์ ประเมินค่า สร้างสรรค์)	100.00	33.33	66.67	100.00
4. จัดทำแผนผังในการออกข้อสอบ (test blueprint)	100.00	33.33	0.00	100.00
5. มีการพิจารณา/ตรวจสอบความถูกต้องและความเหมาะสมของแบบทดสอบ โดยผู้อื่นที่ไม่ใช่ครูผู้ออกข้อสอบ ก่อนนำไปใช้จริง	100.00	100.00	66.67	100.00
6. วิเคราะห์ข้อสอบหลังจากนำข้อสอบไปใช้จริง เช่น ค่าความยาก ค่าอำนาจจำแนก ประสิทธิภาพตัวลอง	100.00	100.00	100.00	100.00
7. นำผลการวิเคราะห์ข้อสอบไปใช้ในการปรับปรุงการออกข้อสอบในครั้งถัดไป	0.00	66.67	33.33	100.00
8. คัดเลือกข้อสอบที่มีคุณภาพจัดเก็บเป็นฐานข้อมูลหรือคลังข้อสอบ	0.00	100.00	0.00	0.00

### 1.2.2 ข้อมูลพื้นฐานของข้อสอบในการประเมินระดับชั้นเรียน

ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 – 3) ของโรงเรียนมัธยมศึกษา สังกัด สพฐ. ในกรุงเทพมหานคร จำนวน 4 โรง ซึ่งประกอบด้วยข้อสอบที่ครูสร้างขึ้นเพื่อประเมินผลการเรียนรู้ของผู้เรียน ทั้งข้อสอบระหว่างภาคและข้อสอบประจำภาค ในปีการศึกษา 2559 ที่ครูระบุว่าใช้ในการวัดตัวชี้วัด 40 ตัวชี้วัดที่ สทศ. ใช้ในการออกข้อสอบ O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ปี

การศึกษา 2559 ที่ได้มาด้วยวิธีการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) ได้ข้อสอบในการศึกษาทั้งหมดประมาณ 1,089 ข้อ จำแนกตามระดับชั้น มาตรฐานและตัวชี้วัด และผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียน ที่แบ่งเป็นโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียนสูง จำนวน 2 โรงเรียน (โรงเรียน<sub>ONETสูง1</sub> และ โรงเรียน<sub>ONETสูง2</sub>) และโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียนต่ำ จำนวน 2 โรงเรียน (โรงเรียน<sub>ONETต่ำ1</sub> และ โรงเรียน<sub>ONETต่ำ2</sub>) ดังตาราง 17 รายละเอียดมีดังนี้

ข้อสอบในการประเมินผลระดับชั้นเรียนจำนวนทั้งหมด 1,089 ข้อ จำแนกเป็นข้อสอบระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 256 ข้อ ประกอบด้วยข้อสอบที่ครูระบุว่าใช้วัดมาตรฐาน ว 1.1 จำนวน 20 ข้อ มาตรฐาน ว 3.1 จำนวน 101 ข้อ มาตรฐาน ว 3.2 จำนวน 69 ข้อ มาตรฐาน ว 4.1 จำนวน 29 ข้อ มาตรฐาน ว 5.1 จำนวน 14 ข้อ และมาตรฐาน ว 6.1 จำนวน 23 ข้อ เป็นข้อสอบระดับชั้นมัธยมศึกษาปีที่ 2 จำนวน 330 ข้อ ประกอบด้วยข้อสอบที่ครูระบุว่าใช้วัดมาตรฐาน ว 1.1 จำนวน 141 ข้อ มาตรฐาน ว 3.1 จำนวน 60 ข้อ มาตรฐาน ว 3.2 จำนวน 25 ข้อ มาตรฐาน ว 4.1 จำนวน 9 ข้อ มาตรฐาน ว 5.1 จำนวน 48 ข้อ และมาตรฐาน ว 6.1 จำนวน 47 ข้อ และเป็นข้อสอบระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 503 ข้อ ประกอบด้วยข้อสอบที่ครูระบุว่าใช้วัดมาตรฐาน ว 1.2 จำนวน 44 ข้อ มาตรฐาน ว 2.1 จำนวน 38 ข้อ มาตรฐาน ว 2.2 จำนวน 7 ข้อ มาตรฐาน ว 4.1 จำนวน 64 ข้อ มาตรฐาน ว 4.2 จำนวน 73 ข้อ มาตรฐาน ว 5.1 จำนวน 128 ข้อ มาตรฐาน ว 7.1 จำนวน 124 ข้อ และมาตรฐาน ว 7.2 จำนวน 25 ข้อ

เมื่อพิจารณาจำแนกรายโรงเรียนทั้ง 4 โรงเรียน มีรายละเอียดดังนี้

โรงเรียนที่ 1 ซึ่งเป็นโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียนสูง โรงเรียนที่ 1 (โรงเรียน<sub>ONETสูง1</sub>) พบว่า มีข้อสอบในการประเมินผลระดับชั้นเรียน จำนวน 295 ข้อ จำแนกเป็นข้อสอบระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 78 ข้อ ประกอบด้วยข้อสอบที่ครูระบุว่าใช้วัดมาตรฐาน ว 1.1 จำนวน 8 ข้อ มาตรฐาน ว 3.1 จำนวน 33 ข้อ มาตรฐาน ว 3.2 จำนวน 20 ข้อ มาตรฐาน ว 4.1 จำนวน 3 ข้อ มาตรฐาน ว 5.1 จำนวน 5 ข้อ และมาตรฐาน ว 6.1 จำนวน 9 ข้อ เป็นข้อสอบระดับชั้นมัธยมศึกษาปีที่ 2 จำนวน 78 ข้อ ประกอบด้วยข้อสอบที่ครูระบุว่าใช้วัดมาตรฐาน ว 1.1 จำนวน 24 ข้อ มาตรฐาน ว 3.1 จำนวน 22 ข้อ มาตรฐาน ว 3.2 จำนวน 10 ข้อ มาตรฐาน ว 4.1 จำนวน 3 ข้อ มาตรฐาน ว 5.1 จำนวน 9 ข้อ และมาตรฐาน ว 6.1 จำนวน 10 ข้อ และเป็นข้อสอบระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 139 ข้อ ประกอบด้วยข้อสอบที่ครูระบุว่าใช้วัดมาตรฐาน ว 1.2 จำนวน 13 ข้อ มาตรฐาน ว 2.1 จำนวน 14 ข้อ มาตรฐาน ว 2.2 จำนวน 2 ข้อ มาตรฐาน ว 4.1 จำนวน 15 ข้อ มาตรฐาน ว 4.2 จำนวน 17 ข้อ มาตรฐาน ว 5.1 จำนวน 36 ข้อ มาตรฐาน ว 7.1 จำนวน 39 ข้อ และมาตรฐาน ว 7.2 จำนวน 3 ข้อ



มาตรฐาน ว 5.1 จำนวน 8 ข้อ และมาตรฐาน ว 6.1 จำนวน 18 ข้อ และเป็นข้อสอบระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 144 ข้อ ประกอบด้วยข้อสอบที่ครูระบุว่าใช้วัดมาตรฐาน ว 1.2 จำนวน 4 ข้อ มาตรฐาน ว 2.1 จำนวน 9 ข้อ มาตรฐาน ว 4.1 จำนวน 23 ข้อ มาตรฐาน ว 4.2 จำนวน 26 ข้อ มาตรฐาน ว 5.1 จำนวน 46 ข้อ มาตรฐาน ว 7.1 จำนวน 29 ข้อ มาตรฐาน ว 7.2 จำนวน 7 ข้อ และไม่มีข้อสอบที่วัดมาตรฐาน ว 2.2

**ตาราง 17 จำนวนข้อสอบในการศึกษาความสอดคล้องในแนวเดียวกันจำแนกตามโรงเรียนระดับชั้น มาตรฐานและตัวชี้วัด**

ระดับชั้น	มาตรฐาน	ตัวชี้วัด	จำนวนข้อสอบจำแนกตามโรงเรียน (ข้อ)					รวม รายมาตรฐาน	
			ร.ร. ONET สูง 1	ร.ร. ONET สูง 2	ร.ร. ONET ต่ำ 1	ร.ร. ONET ต่ำ 2	รวม		
ม.1	ว 1.1	ม. 1/4	6	4	2	0	12	20	
		ม. 1/5	2	3	3	0	8		
	ว 3.1	ม. 1/1	19	7	9	1	36		101
		ม. 1/2	6	2	13	10	31		
		ม. 1/4	8	6	8	12	34		
	ว 3.2	ม. 1/1	12	9	1	24	46		69
		ม. 1/2	8	8	2	5	23		
	ว 4.1	ม. 1/2	3	12	1	13	29		29
	ว 5.1	ม. 1/2	5	4	3	2	14		14
	ว 6.1	ม. 1/6	9	7	7	0	23		23
<b>รวม ระดับชั้น ม. 1</b>			<b>78</b>	<b>62</b>	<b>49</b>	<b>67</b>	<b>256</b>	<b>256</b>	
ม.2	ว 1.1	ม. 2/2	1	1	0	2	4	141	
		ม. 2/3	1	4	2	0	7		
		ม. 2/4	1	5	1	0	7		
		ม. 2/5	21	20	40	42	123		
	ว 3.1	ม. 2/2	14	4	2	7	27	60	
		ม. 2/3	8	13	12	0	33		
	ว 3.2	ม. 2/1	10	4	7	4	25	25	
	ว 4.1	ม. 2/2	3	1	2	3	9	9	
	ว 5.1	ม. 2/1	8	10	20	2	40	48	
		ม. 2/2	0	0	0	1	1		
		ม. 2/3	1	1	0	5	7		
	ว 6.1	ม. 2/2	2	5	1	6	14	47	
		ม. 2/4	8	5	8	12	33		
	<b>รวม ระดับชั้น ม. 2</b>			<b>78</b>	<b>73</b>	<b>95</b>	<b>84</b>	<b>330</b>	<b>330</b>
ม.3	ว 1.2	ม. 3/3	8	14	5	4	31	44	
		ม. 3/6	5	4	4	0	13		

ตาราง 17 (ต่อ)

ระดับชั้น	มาตรฐาน	ตัวชี้วัด	จำนวนข้อสอบจำแนกตามโรงเรียน (ข้อ)				รวม	รวม รายมาตรฐาน
			ร.ร.-ONETสูง1	ร.ร.-ONETสูง2	ร.ร.-ONETต่ำ1	ร.ร.-ONETต่ำ2		
ว 2.1	ม. 3/1	ม. 3/1	11	4	8	6	29	38
		ม. 3/3	3	1	2	3	9	
ว 2.2	ม. 3/2	ม. 3/2	1	2	1	0	4	7
		ม. 3/4	1	1	1	0	3	
ว 4.1	ม. 3/1	ม. 3/1	12	12	9	20	53	64
		ม. 3/2	3	2	3	3	11	
ว 4.2	ม. 3/1	ม. 3/1	8	4	8	5	25	73
		ม. 3/2	4	5	8	8	25	
		ม. 3/3	5	2	3	13	23	
ว 5.1	ม. 3/1	ม. 3/1	10	10	16	23	59	128
		ม. 3/3	8	1	5	5	19	
		ม. 3/4	18	12	2	18	50	
ว 7.1	ม. 3/1	ม. 3/1	19	20	21	13	73	124
		ม. 3/3	20	9	6	16	51	
ว 7.2	ม. 3/1	ม. 3/1	3	15	0	7	25	25
รวม ระดับชั้น ม. 3			139	118	102	144	503	503
รวม ระดับชั้น ม. 1 - 3			295	253	246	295	1,089	1,089

### 1.3 ข้อมูลพื้นฐานของผู้เชี่ยวชาญ

ผู้เชี่ยวชาญในการประเมินความสอดคล้องในแนวเดียวกัน จำนวน 20 ท่าน ประกอบด้วย ครูสอนวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น และ/หรือนักวิชาการ และ/หรืออาจารย์ระดับอุดมศึกษาที่สอนในรายวิชาวิทยาศาสตร์ และ/หรือสาขาวิชาที่เกี่ยวข้องกับการวัดและประเมินผลการศึกษา ซึ่งได้มาจากการเลือกแบบเจาะจง (purposive sampling) โดยกำหนดเกณฑ์ในการคัดเลือกคุณสมบัติของผู้เชี่ยวชาญ คือ เป็นผู้ที่มีความรู้ความเชี่ยวชาญด้านมาตรฐานและตัวชี้วัดกลุ่มสาระวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น และด้านการวัดและประเมินผลทางการศึกษา โดยสำเร็จการศึกษาอย่างน้อยในระดับปริญญาตรีทางการสอนวิทยาศาสตร์ และสำเร็จการศึกษาอย่างน้อยในระดับปริญญาโททางการวัดและประเมินผลทางการศึกษา และ/หรือมีประสบการณ์สอนกลุ่มสาระวิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ไม่นต่ำกว่า 3 ปี

เมื่อพิจารณาข้อมูลพื้นฐานของผู้เชี่ยวชาญที่ทำหน้าที่ประเมินความสอดคล้องในแนวเดียวกัน พบว่า ผู้เชี่ยวชาญส่วนใหญ่เป็นเพศหญิง จำนวน 15 คน คิดเป็นร้อยละ 75.00 และเป็นเพศชาย จำนวน 5 คน คิดเป็นร้อยละ 25.00 ส่วนใหญ่อายุ 30 - 34 ปี จำนวน 9 คน คิดเป็นร้อยละ

45.00 รองลงมาคือ อายุ 35 – 40 ปี จำนวน 7 คน คิดเป็นร้อยละ 35.00 และมีอายุต่ำกว่า 30 ปี และมากกว่า 40 ปี จำนวนละ 2 คน คิดเป็นร้อยละ 10.00 เท่ากัน ส่วนใหญ่เป็นครูโรงเรียนมัธยมศึกษา จำนวน 14 คน คิดเป็นร้อยละ 70.00 รองลงมาคือ อาจารย์มหาวิทยาลัย จำนวน 4 คน คิดเป็นร้อยละ 20.00 และนักวิชาการ จำนวน 2 คน คิดเป็นร้อยละ 10.00 ส่วนใหญ่มีประสบการณ์การสอนวิทยาศาสตร์ 6 – 8 ปี จำนวน 9 คน คิดเป็นร้อยละ 45.00 รองลงมาคือ มากกว่า 8 ปี จำนวน 8 คน คิดเป็นร้อยละ 40.00 และ 3 -5 ปี จำนวน 3 คน คิดเป็นร้อยละ 15.00 รายละเอียดดังตาราง 18

ตาราง 18 ข้อมูลพื้นฐานของผู้เชี่ยวชาญ

	ข้อมูล	จำนวน (คน)	ร้อยละ
เพศ	ชาย	5	25.00
	หญิง	15	75.00
อายุ	ต่ำกว่า 30 ปี	2	10.00
	30 - 34 ปี	9	45.00
	35 - 40 ปี	7	35.00
	มากกว่า 40 ปี	2	10.00
ตำแหน่ง	ครูโรงเรียนมัธยมศึกษา	14	70.00
	อาจารย์มหาวิทยาลัย	4	20.00
	นักวิชาการ	2	10.00
ประสบการณ์การสอนวิทยาศาสตร์	3 - 5 ปี	3	15.00
	6 - 8 ปี	9	45.00
	มากกว่า 8 ปี	8	40.00
รวม		20	100.00

#### 1.4 ผลการวิเคราะห์ความเที่ยงระหว่างผู้ประเมิน

ผู้วิจัยได้ทำการวิเคราะห์ความเที่ยงระหว่างผู้ประเมิน โดยใช้ข้อมูลจากที่ผู้เชี่ยวชาญจำนวน 20 คน ประเมินข้อสอบในการประเมินระดับชั้นเรียน จำนวน 40 ข้อ (ข้อสอบชุดย่อย A) ทั้งในการประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ ผลการวิเคราะห์มีรายละเอียดดังนี้

1.4.1 ผลการวิเคราะห์ความเที่ยงระหว่างผู้ประเมิน ในการประเมินระดับความซับซ้อนทางปัญญา ซึ่งผู้วิจัยเลือกใช้สถิติแคปปาของฟลีส (Fleiss's kappa statistic:  $K_p$ ) (Fleiss, 1971)

เนื่องจากการตรวจสอบความเที่ยงระหว่างผู้ประเมินที่มีตั้งแต่ 3 คนขึ้นไป สำหรับข้อมูลที่อยู่ในระดับนามบัญญัติ (Gisev, Bell, & Chen, 2013) พบว่า มีค่า  $K_f$  เท่ากับ 0.510 แสดงว่ามีค่าความสอดคล้องกันระหว่างผู้ประเมินอยู่ในระดับดี (Fleiss, Levin, & Paik, 2003) โดยมีค่าความคลาดเคลื่อนมาตรฐาน ( $SE(K_f)$ ) เท่ากับ 0.017 รายละเอียดดังตาราง 19

**ตาราง 19** ค่าสถิติแคปปาของฟลีสของการประเมินความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด

จำนวนข้อสอบ (ข้อ)	จำนวนผู้ประเมิน (คน)	Fleiss's kappa	Var ( $K_f$ )	SE ( $K_f$ )
40	20	0.510	0.0003	0.017

1.4.2 ผลการวิเคราะห์ความเที่ยงระหว่างผู้ประเมิน ในส่วนของการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ที่มีลักษณะเป็นมาตรฐานค่า 5 ระดับ (0 – 4) ซึ่งผู้วิจัยเลือกใช้สถิติสหสัมพันธ์ภายในชั้น (intra-class correlation: ICC) เนื่องจากเป็นข้อมูลเรียงอันดับ (ordinal data) (Gisev, Bell, Chen & 2013) โดยวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้นที่วิเคราะห์จากโมเดลอิทธิพลผสมแบบสองทาง (two-way mixed-effects model) มีผู้ประเมินจำนวน 7 คน (ใช้สัญลักษณ์ คือ ICC(3,7)) ที่ระดับความเชื่อมั่น 95% ด้วยโปรแกรม SPSS ผลการวิเคราะห์พบว่า มีค่าสหสัมพันธ์ภายในชั้น เท่ากับ 0.954 อย่างมีนัยสำคัญทางสถิติ ( $p = .000$ ) มีช่วงความเชื่อมั่นที่ 95% อยู่ระหว่าง 0.929 – 0.973 แสดงว่ามีความสอดคล้องระหว่างผู้ประเมินในระดับดีมาก (Porney & Watkins, 2015) มีค่าความคลาดเคลื่อนมาตรฐานในการวัด (standard error of measurement: SEM) เท่ากับ 0.02 ค่าเฉลี่ยการประเมินของผู้ประเมินอยู่ระหว่าง 2.90 – 3.88 โดยผู้ประเมินคนที่ 12 และ 13 มีค่าเฉลี่ยในการประเมินสูงสุดเท่ากัน ( $M=3.88$ ,  $SD=0.33$  เท่ากัน) รองลงมาคือ ผู้ประเมินคนที่ 14 ( $M=3.85$ ,  $SD=0.43$ ) ผู้ประเมินคนที่ 7 11 และ 15 ( $M=3.83$ ,  $SD=0.38$ ,  $M=3.83$ ,  $SD=0.38$  และ  $M=3.83$ ,  $SD=0.55$  ตามลำดับ) และผู้ประเมินคนที่ 9 ( $M=3.80$ ,  $SD=0.41$ ) ในขณะที่ผู้ประเมินคนที่ 4 มีค่าเฉลี่ยในการประเมินต่ำสุด ( $M=2.90$ ,  $SD=0.81$ ) รองลงมาคือ ผู้ประเมินคนที่ 6 ( $M=3.30$ ,  $SD=0.72$ ) ผู้ประเมินคนที่ 18 ( $M=3.33$ ,  $SD=0.83$ ) และผู้ประเมินคนที่ 2 ( $M=3.40$ ,  $SD=0.67$ ) รายละเอียดแสดงดังตาราง 20



ตาราง 20 สถิติสหสัมพันธ์ภายในชั้นของการประเมินระดับความสอดคล้องในแนวเดียวกัน ระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ

ผู้ประเมินคนที่	ค่าเฉลี่ย $\pm$ ส่วนเบี่ยงเบนมาตรฐาน	ICC (ช่วงความเชื่อมั่น 95%)	p	SEM
1	3.68 $\pm$ 0.57	0.954 (0.929 - 0.973)	.000	0.02
2	3.40 $\pm$ 0.67			
3	3.70 $\pm$ 0.72			
4	2.90 $\pm$ 0.81			
5	3.43 $\pm$ 0.98			
6	3.30 $\pm$ 0.72			
7	3.83 $\pm$ 0.38			
8	3.75 $\pm$ 0.59			
9	3.80 $\pm$ 0.41			
10	3.65 $\pm$ 0.48			
11	3.83 $\pm$ 0.38			
12	3.88 $\pm$ 0.33			
13	3.88 $\pm$ 0.33			
14	3.85 $\pm$ 0.43			
15	3.83 $\pm$ 0.55			
16	3.63 $\pm$ 0.90			
17	3.75 $\pm$ 0.59			
18	3.33 $\pm$ 0.83			
19	3.43 $\pm$ 0.96			
20	3.48 $\pm$ 0.75			

ตอนที่ 2 ผลการตรวจสอบอิทธิพลของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกัน ระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน

ผลการตรวจสอบอิทธิพลของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกัน ระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน โดยการพิจารณาอิทธิพลของผู้ประเมิน 4 ชนิด ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง และอิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ที่วิเคราะห์โดยประยุกต์ใช้ MFRM ด้วยโปรแกรม FACETS (Linacre, 2017) แบ่งการนำเสนอเป็น 2 ส่วน คือ 1) การตรวจสอบอิทธิพลการกดหรือปล่อยคะแนน และ 2) การตรวจสอบอิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และความแม่นยำในการให้คะแนน มีรายละเอียดดังนี้

## 2.1 การตรวจสอบอิทธิพลการกดหรือปล่อยคะแนน (severity/leniency effect)

จากการทดสอบสถิติไคสแควร์ (fixed chi-square test) ที่ได้จากโปรแกรม FACETS เพื่อทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติระหว่างสมาชิกต่างๆ ในฟาเซท (Engelhard, 1996) โดยมีสมมติฐานสำหรับการทดสอบฟาเซทผู้ประเมินว่า ผู้ประเมินทุกคนมีระดับการกดคะแนนอยู่ในระดับเดียวกันภายหลังอธิบายความคลาดเคลื่อนแล้ว ผลการทดสอบพบว่า ค่าไคสแควร์มีนัยสำคัญที่ระดับ .01 ( $\chi^2 = 975.70$ ,  $d.f. = 19$ ,  $p = .00$ ) แสดงว่ามีผู้ประเมินอย่างน้อย 2 คน ที่มีระดับการกดคะแนนแตกต่างอย่างมีนัยสำคัญทางสถิติ (Myford & Wolfe, 2003: 409)

เมื่อพิจารณาคะแนนโลจิทที่แสดงถึงค่าการกดหรือปล่อยคะแนนของผู้ประเมินเป็นรายบุคคล ได้แก่ ค่า Measure ของผลการวิเคราะห์ด้วย MFRM จากฟาเซทผู้ประเมิน แสดงดังตาราง 21 โดยหากคะแนนโลจิทยังมีค่ามาก แสดงถึงลักษณะการกดคะแนนของผู้ประเมิน และหากคะแนนโลจิทยังมีค่าน้อย แสดงถึงลักษณะการปล่อยคะแนนของผู้ประเมิน ผลการพิจารณาพบว่า โดยภาพรวมผู้ประเมินมีแนวโน้มที่จะมีลักษณะของการกดคะแนนมากกว่าปล่อยคะแนน คือ มีคะแนนโลจิทไปทางบวกมากกว่าทางลบ ทั้งนี้ ผู้ประเมินคนที่ 4 มีลักษณะกดคะแนนมากที่สุด (คะแนนโลจิท = 1.83) รองลงมาคือ ผู้ประเมินคนที่ 20 19 และ 16 ตามลำดับ (คะแนนโลจิท = 1.03, 0.93 และ 0.92 ตามลำดับ) ส่วนผู้ประเมินคนที่ 13 มีลักษณะ การปล่อยคะแนนมากที่สุด (คะแนนโลจิท = -3.24) รองลงมาคือ ผู้ประเมินคนที่ 12 15 และ 14 ตามลำดับ (คะแนนโลจิท = -1.83, -1.47 และ -1.15 ตามลำดับ) ทั้งนี้เมื่อพิจารณาแผนที่ตัวแปร (Variable map) จากภาพ 7 ในคอลัมน์ที่ 3 แสดงการแจกแจงของคะแนนโลจิท จะแสดงให้เห็นถึงระดับการกดหรือปล่อยคะแนนได้ชัดเจนมากยิ่งขึ้น โดยผู้ประเมินคนที่ 4 ที่แสดงลักษณะกดคะแนนมากที่สุด ซึ่งคะแนนโลจิทมีค่าเป็นบวก และมีคะแนนสูงสุด จะอยู่ด้านบนสุดของคอลัมน์ เมื่อเทียบกับผู้ประเมินคนอื่นๆ และผู้ประเมินคนที่ 13 ที่แสดงลักษณะปล่อยคะแนนมากที่สุด ซึ่งคะแนนโลจิทมีค่าเป็นลบและมีคะแนนต่ำสุด จะอยู่ด้านล่างสุดของคอลัมน์

เมื่อพิจารณาค่าสถิติ infit MNSQ และ outfit MNSQ ซึ่งเป็นดัชนีที่บ่งบอกถึงความสอดคล้องภายในของผู้ประเมิน (internal consistency of raters) หรือความเที่ยงภายในผู้ประเมิน (intra-rater reliability) พบว่า จากผู้ประเมินทั้งหมด 20 คน มีผู้ประเมินจำนวน 17 คน คิดเป็นร้อยละ 85.00 ที่มีคะแนนเชิงประจักษ์สอดคล้องกลมกลืนกับโมเดล โดยมีค่า infit MNSQ และ outfit MNSQ อยู่ในช่วง 0.50 – 1.30 และมีผู้ประเมินจำนวน 3 คน คือเป็นร้อยละ 15.00 ที่มีคะแนนเชิงประจักษ์ ไม่สอดคล้องกับโมเดล ได้แก่ ผู้ประเมินคนที่ 12, 13 และ 15 โดยผู้ประเมินคนที่ 13 มีค่า outfit MNSQ เท่ากับ 0.18 ซึ่งต่ำกว่า 0.50 กล่าวคือ ค่าที่สังเกตมีค่าใกล้กับค่าที่คาดหวังมากกว่าที่โมเดลทำนายไว้ (overfit) และผู้ประเมินคนที่ 12 และ 15 มีค่า infit MNSQ สูงกว่า 1.50

เท่ากับ 1.79 และ 1.69 ตามลำดับ คือ ค่าที่สังเกตอยู่ห่างจากค่าที่คาดหวังมากกว่าที่โมเดลทำนายไว้ (misfit)

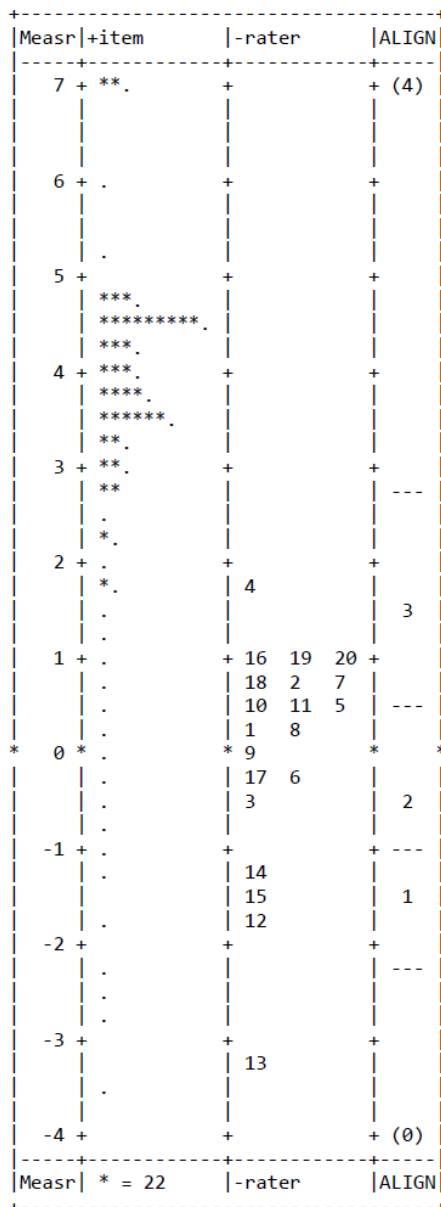
ตาราง 21 ผลการวิเคราะห์ฟาเซทผู้ประเมินด้วยโปรแกรม FACETS

rater	Count	Obs. Avge	Fair-M Avge	Measure	S.E.	infitMS	outfitMS
1	304	3.63	3.62	0.24	0.12	0.93	0.86
2	306	3.43	3.46	0.80	0.11	1.17	1.25
3	306	3.70	3.78	-0.47	0.14	1.20	0.79
4	304	3.07	3.09	1.83	0.09	0.71	0.73
5	303	3.53	3.58	0.40	0.11	1.10	1.09
6	299	3.71	3.74	-0.26	0.13	1.10	1.12
7	299	3.54	3.50	0.66	0.11	0.99	1.02
8	298	3.64	3.63	0.22	0.12	0.77	0.70
9	287	3.67	3.71	-0.09	0.13	0.98	0.90
10	287	3.54	3.55	0.49	0.11	0.99	0.96
11	287	3.53	3.55	0.51	0.11	0.99	0.98
12	287	3.90	3.94	-1.83	0.22	1.79	0.51
13	307	3.93	3.98	-3.24	0.27	0.98	0.18
14	307	3.75	3.88	-1.15	0.15	1.01	0.76
15	312	3.85	3.91	-1.47	0.18	1.69	0.55
16	312	3.40	3.42	0.92	0.10	1.27	1.12
17	312	3.68	3.74	-0.23	0.13	1.28	0.83
18	312	3.46	3.49	0.71	0.11	1.31	1.30
19	312	3.39	3.41	0.93	0.10	0.98	0.92
20	307	3.25	3.38	1.03	0.10	1.09	1.08

หมายเหตุ Obs. Avge: ค่าเฉลี่ยคะแนนดิบ (raw mean) ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Fair-M Avge: ค่าเฉลี่ยคะแนนที่ปรับใหม่ (adjusted mean) หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Measure: คะแนนในหน่วยโลจิทของฟาเซทผู้ประเมิน (rater logit)



ภาพ 7 แผนที่ตัวแปร (variables map) จากการวิเคราะห์ด้วยโปรแกรม FACETS

2.2 การตรวจสอบอิทธิพลแนวโน้มการให้คะแนนค่ากลาง (central tendency effect) อิทธิพลการจำกัดช่วง (restriction of range effect) อิทธิพลของความไม่มีแบบแผน (randomness effect) และความแม่นยำในการให้คะแนน (accurate)

เมื่อพิจารณาค่าสถิติ infit MNSQ และ outfit MNSQ ของผู้ประเมินจากตาราง 20 พบว่า ผู้ประเมินส่วนใหญ่แสดงรูปแบบของความแม่นยำในการให้คะแนน (accurate) คือ มีค่า infit MNSQ และ outfit MNSQ อยู่ในเกณฑ์ 0.70 – 1.30 (Myford & Wolfe, 2000) จำนวน 16 คน คิดเป็นร้อยละ 80.00 โดยมีค่า infit MNSQ อยู่ระหว่าง 0.71 – 1.28 และมีค่า outfit MNSQ อยู่

ระหว่าง 0.70 – 1.25 ทั้งนี้ไม่มีผู้ประเมินที่แสดงอิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน สำหรับผู้ประเมินอีกจำนวน 4 คน ได้แก่ ผู้ประเมินคนที่ 12 13 15 และ 18 คิดเป็นร้อยละ 20.00 ไม่ถูกจัดให้อยู่ในรูปแบบของผู้ประเมินใดๆ โดยผู้ประเมินคนที่ 12 และ 15 มีค่า infit MNSQ มากกว่า 1.30 ตามลำดับ และมีค่า outfit MNSQ น้อยกว่า 0.70 โดยผู้ประเมินคนที่ 12 มีค่า infit MNSQ เท่ากับ 1.79 และมีค่า outfit MNSQ เท่ากับ 0.51 และผู้ประเมินคนที่ 15 มีค่า infit MNSQ เท่ากับ 1.69 และมีค่า outfit MNSQ เท่ากับ 0.55 ผู้ประเมินคนที่ 13 มีค่า infit MNSQ เท่ากับ 0.98 ซึ่งอยู่ในเกณฑ์ 0.70 – 1.30 แต่มีค่า outfit MNSQ เท่ากับ 0.18 ซึ่งน้อยกว่าเกณฑ์ 0.70 และผู้ประเมินคนที่ 18 ค่า infit MNSQ เท่ากับ 1.31 ซึ่งสูงกว่าเกณฑ์ 1.30 เล็กน้อย และมีค่า outfit MNSQ เท่ากับ 0.30 น้อยกว่าเกณฑ์ 0.70 สรุปจำนวนรายละเอียดลักษณะของผู้ประเมินที่แสดงความแม่นยำในการให้คะแนนและแสดงอิทธิพลของผู้ประเมินแสดงดังตาราง 22

ตาราง 22 ลักษณะของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน

ลักษณะของผู้ประเมิน	infit & outfit MNSQ	จำนวนผู้ประเมิน (ร้อยละ)	ผู้ประเมินคนที่
อิทธิพลแนวโน้มการให้คะแนน ค่ากลาง (central tendency effect)	infit < 0.70 outfit < 0.70	-	-
อิทธิพลการจำกัดช่วง (restriction of range effect)	infit < 0.60 outfit < 0.60	-	-
อิทธิพลของความไม่มีแบบแผน (randomness effect)	infit > 1.30 outfit > 1.30	-	-
มีความแม่นยำในการให้คะแนน (accurate)	0.70 ≤ infit ≤ 1.30 0.70 ≤ outfit ≤ 1.30	16 (80.00)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 16, 19, 20
อื่นๆ		4 (20.00)	12, 13, 15, 18

หมายเหตุ: อื่นๆ หมายถึง ผู้ประเมินที่มีค่า infit MNSQ และ outfit MNSQ ไม่สอดคล้องกับรูปแบบใดๆ

### ตอนที่ 3 ผลการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐาน และตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพล การกดหรือปล่อยคะแนน

การเปรียบเทียบผลประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน มีวัตถุประสงค์เพื่อตรวจสอบความสอดคล้องในแนวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน โดยมีการควบคุมความคลาดเคลื่อนที่เกิดจากอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน และเพื่อตัดสินว่าข้อสอบแต่ละข้อสอดคล้องหรือตัวชี้วัดที่ครูระบุจริงหรือไม่ ก่อนคัดเลือกข้อสอบไปใช้ในการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) ก่อนนำไปคำนวณค่าดัชนีความสอดคล้อง (alignment index) ต่อไป การวิเคราะห์ดำเนินการโดยการนำข้อมูลที่ผู้เชี่ยวชาญจำนวน 20 คน ประเมินความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรา-ประมาณค่า 5 ระดับ ของข้อสอบจำนวน 1,089 ข้อ มาวิเคราะห์โดยประยุกต์ใช้ MFRM ด้วยโปรแกรม FACETS 3.80.3 (Linacre, 2017) และพิจารณาตัดสินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบ จากผลการวิเคราะห์ฟาเซทข้อสอบ (item facet) โดยข้อสอบที่มีความสอดคล้องกับตัวชี้วัดที่ระบุจะต้องมีค่าเฉลี่ยคะแนนการประเมินหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Fair-M Avege) ตั้งแต่ 3.00 คะแนนขึ้นไป และมีข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล โดยมีค่าดัชนีตรวจสอบความสอดคล้องระหว่างข้อมูลกับโมเดล คือมีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.50 – 1.50

ก่อนการพิจารณาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนเป็นรายข้อ ผู้วิจัยได้ทำการตรวจสอบความสอดคล้อง (fit) ระหว่างข้อมูลชุดนี้กับโมเดลของราส์ช (Rasch model) ผลการตรวจสอบพบว่า ข้อมูลชุดนี้มีความสอดคล้องกับโมเดลของราส์ช โดยพิจารณาจากค่าเฉลี่ยของความคลาดเคลื่อน (mean residuals: Resd) มีค่าเท่ากับ 0.00 ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐาน (mean of the standardized residuals: StRes) มีค่าเท่ากับ 0.02 และส่วนเบี่ยงเบนมาตรฐาน (sample standard deviation: SD) มีค่าเท่ากับ 0.94 ซึ่งถ้าค่าเฉลี่ยของ Resd และ StRes มีค่าเข้าใกล้ 0.00 และ SD มีค่าเข้าใกล้ 1.00 จะแสดงว่าข้อมูลมีความสอดคล้องกับโมเดลของราส์ช (Linacre, 2018) ผลการวิเคราะห์ฟาเซทข้อสอบ (item facet) เป็นดังนี้

### 3.1 ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน

ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยพิจารณาค่า Obs. Avge ซึ่งเป็นค่าเฉลี่ยคะแนนดิบ (raw mean) ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน พบว่า จากข้อสอบ 1,089 ข้อ มีข้อสอบที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัดที่ระบุ คือมีค่า Obs. Avge ตั้งแต่ 3.00 คะแนนขึ้นไป จำนวน 1,012 ข้อ คิดเป็นร้อยละ 92.93 รายละเอียดดังตาราง 23

ตาราง 23 จำนวนและร้อยละของข้อสอบจากการวิเคราะห์ด้วย MFRM จำแนกตามค่าเฉลี่ยคะแนนความสอดคล้องฯ ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน

ค่าเฉลี่ยคะแนน	จำนวน (ข้อ)	ร้อยละ
Obs. Avge $\geq$ 3.00	1,012	92.93
Obs. Avge $<$ 3.00	77	7.07
<b>รวม</b>	<b>1,089</b>	<b>100.00</b>

### 3.2 ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยพิจารณาค่า Fair-M Avge ซึ่งเป็นค่าเฉลี่ยคะแนนที่ปรับใหม่ (adjusted mean) ของการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน พบว่า จากข้อสอบ 1,089 ข้อ มีข้อสอบที่มีค่า Fair-M Avge ตั้งแต่ 3.00 คะแนนขึ้นไป จำนวน 1,013 ข้อ คิดเป็นร้อยละ 93.02 มีคะแนนโลจิท (Measure) อยู่ระหว่าง 1.59 – 6.25 รายละเอียดดังตาราง 24 และมีข้อสอบที่มีค่า Fair-M Avge ต่ำกว่า 3.00 คะแนน จำนวน 76 ข้อ คิดเป็นร้อยละ 6.89 มีคะแนนโลจิทอยู่ระหว่าง -3.43 - 1.45 ซึ่งรายละเอียดข้อสอบที่มีค่า Fair-M Avge ต่ำกว่า 3.00 คะแนน

ตาราง 24 จำนวนและร้อยละของข้อสอบจากการวิเคราะห์ด้วย MFRM จำแนกตามค่าดัชนีความสอดคล้องระหว่างข้อมูลเชิงประจักษ์กับโมเดลและค่าเฉลี่ยคะแนนความสอดคล้องๆ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

ค่าดัชนีความสอดคล้อง ระหว่างข้อมูลเชิงประจักษ์กับโมเดล	Fair-M Avge $\geq$ 3.00		Fair-M Avge $<$ 3.00	
	จำนวน (ข้อ)	ร้อยละ*	จำนวน (ข้อ)	ร้อยละ*
infit MNSQ $<$ 0.50	23	2.11	9	0.83
0.05 $\leq$ infit MNSQ $\leq$ 1.50	973	89.35	34	3.12
infit MNSQ $>$ 1.50	17	1.56	33	3.03
<b>รวม</b>	<b>1,013</b>	<b>93.02</b>	<b>76</b>	<b>6.98</b>
outfit MNSQ $<$ 0.50	76	6.98	9	0.83
0.05 $\leq$ outfit MNSQ $\leq$ 1.50	908	83.38	31	2.85
outfit MNSQ $>$ 1.50	29	2.66	36	3.31
<b>รวม</b>	<b>1,013</b>	<b>93.02</b>	<b>76</b>	<b>6.98</b>

\*เทียบร้อยละจากข้อสอบที่ใช้ในการศึกษาทั้งหมด จำนวน 1,089 ข้อ

จากตารางที่ 24 เมื่อพิจารณาความสอดคล้องกลมกลืนระหว่างข้อมูลเชิงประจักษ์ของข้อสอบแต่ละข้อกับโมเดลด้วยค่าสถิติ infit MNSQ และ outfit MNSQ ของข้อสอบที่จำแนกตามค่าเฉลี่ยคะแนนความสอดคล้องๆ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน เป็นดังนี้

ข้อสอบที่มีค่าเฉลี่ยคะแนนความสอดคล้องในแนวเดียวกันๆ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Fair-M Avge) มากกว่า 3.00 คะแนนขึ้นไป จำนวน 1,013 ข้อ เมื่อพิจารณาค่าสถิติ infit MNSQ พบว่า ข้อสอบส่วนใหญ่มีค่า infit MNSQ อยู่ระหว่าง 0.50 – 1.50 จำนวน 973 ข้อ คิดเป็นร้อยละ 89.38 ของข้อสอบทั้งหมด รองลงมาคือ มีค่า infit MNSQ น้อยกว่า 0.50 จำนวน 23 ข้อ คิดเป็นร้อยละ 2.11 ของข้อสอบทั้งหมด (1,089 ข้อ) และมีค่า infit MNSQ มากกว่า 1.50 จำนวน 17 ข้อ คิดเป็นร้อยละ 1.56 ของข้อสอบทั้งหมด และเมื่อพิจารณาค่าสถิติ outfit MNSQ พบว่า ข้อสอบส่วนใหญ่มีค่า outfit MNSQ อยู่ระหว่าง 0.50 – 1.50 จำนวน 908 ข้อ คิดเป็นร้อยละ 83.38 ของข้อสอบทั้งหมด รองลงมาคือ มีค่า outfit MNSQ น้อยกว่า 0.50 จำนวน 76 ข้อ คิดเป็นร้อยละ 6.98 ของข้อสอบทั้งหมด มีค่า outfit MNSQ มากกว่า 1.50 จำนวน 29 ข้อ คิดเป็นร้อยละ 2.66 ของข้อสอบทั้งหมด

เมื่อพิจารณาค่าสถิติ infit MNSQ ของข้อสอบที่มีค่าเฉลี่ยคะแนนความสอดคล้องในแนวเดียวกันๆ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Fair-M Avge) น้อยกว่า 3.00 คะแนน จำนวน 76 ข้อ พบว่า ข้อสอบส่วนใหญ่มีค่า infit MNSQ อยู่ระหว่าง 0.50 – 1.50 จำนวน 34 ข้อ



คิดเป็นร้อยละ 3.12 ของข้อสอบทั้งหมด (1,089 ข้อ) และมีค่า infit MNSQ มากกว่า 1.50 จำนวน 33 ข้อ คิดเป็นร้อยละ 3.03 ของข้อสอบทั้งหมด รองลงมาคือ มีค่า infit MNSQ น้อยกว่า 0.50 จำนวน 9 ข้อ คิดเป็นร้อยละ 0.83 ของข้อสอบทั้งหมด และเมื่อพิจารณาค่าสถิติ outfit MNSQ พบว่า ข้อสอบส่วนใหญ่มีค่า outfit MNSQ มากกว่า 1.50 จำนวน 36 ข้อ คิดเป็นร้อยละ 3.31 ของข้อสอบทั้งหมด รองลงมาคือ มีค่า outfit MNSQ อยู่ระหว่าง 0.50 – 1.50 จำนวน 31 ข้อ คิดเป็นร้อยละ 2.85 ของข้อสอบทั้งหมด และมีค่า outfit MNSQ น้อยกว่า 0.50 จำนวน 9 ข้อ คิดเป็นร้อยละ 0.83 ของข้อสอบทั้งหมด รายละเอียดข้อสอบที่มี Fair-M Avge น้อยกว่า 3.00 คะแนน จำนวน 76 ข้อ แสดงดังตาราง 25

ตาราง 25 ข้อสอบที่มีค่าเฉลี่ยคะแนนความสอดคล้องในแนวเดียวกันฯ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Fair-M Avge) ต่ำกว่า 3.00 คะแนน (จำนวน 76 ข้อ)

ข้อที่	โรงเรียน	ระดับชั้น	ตัวชี้วัด	Obs.Avge	FairMAvge	Measure	S.E.	InfitMS	OutfitMS
4	4	1	3112	2.20	2.32	0.08	0.28	0.67	0.65
18	1	2	4122	2.30	2.41	0.23	0.28	0.61	0.63
28	1	3	2232	2.90	2.96	1.36	0.34	1.20	1.52
30	2	3	4131	2.30	2.41	0.23	0.28	1.18	1.46
32	4	3	4231	2.15	2.27	0.00	0.27	0.79	0.79
67	4	3	1233	1.20	1.64	-0.84	0.52	0.14	0.16
136	1	2	1125	2.20	2.41	0.22	0.55	1.34	1.28
138	2	2	1125	2.20	2.41	0.22	0.55	1.21	1.35
180	1	2	1125	2.80	2.93	1.31	0.66	0.91	0.96
182	2	2	1125	2.80	2.93	1.31	0.66	0.91	0.96
187	3	2	1125	2.80	2.93	1.31	0.66	0.91	0.96
252	2	1	6116	2.40	2.75	0.88	0.58	3.13	3.14
282	1	3	1233	2.60	2.73	0.84	0.62	0.65	0.68
293	2	3	2234	2.60	2.73	0.84	0.62	1.07	1.15
313	3	2	1125	2.20	2.59	0.55	0.56	1.44	1.72
328	4	3	1233	2.60	2.90	1.24	0.62	1.02	0.91
370	3	2	3123	2.00	2.06	-0.30	0.52	0.47	0.47
374	2	1	3211	2.80	2.82	1.03	0.65	0.75	0.78
428	3	2	3221	3.00	2.97	1.39	0.70	0.64	0.68
445	1	2	3122	2.60	2.67	0.72	0.61	0.51	0.56
480	4	1	3112	2.60	2.45	0.29	0.61	0.77	0.83
486	1	2	3122	2.60	2.45	0.29	0.61	1.03	1.04
495	3	2	3123	1.80	1.55	-0.93	0.52	1.57	1.86
516	4	1	3111	3.00	2.74	0.86	0.70	0.74	1.16

ตาราง 25 (ต่อ)

ข้อที่	โรงเรียน	ระดับชั้น	ตัวชี้วัด	Obs.Avg	FairMAvg	Measure	S.E.	InfitMS	OutfitMS
520	4	1	3112	2.20	1.83	-0.61	0.55	3.19	4.61
521	4	1	3112	2.40	2.06	-0.31	0.57	3.86	5.37
535	3	2	3123	2.20	1.83	-0.61	0.55	3.78	5.05
540	4	1	3211	3.00	2.74	0.86	0.70	0.70	0.68
541	4	1	3211	3.00	2.74	0.86	0.70	1.07	1.00
561	4	1	3112	3.00	2.93	1.29	0.70	1.03	1.01
577	3	2	3123	2.80	2.73	0.84	0.65	1.75	1.69
585	1	1	3212	2.80	2.73	0.84	0.65	1.75	1.69
586	2	1	3212	2.80	2.73	0.84	0.65	2.02	1.95
598	2	3	4131	2.00	2.22	-0.08	0.52	2.96	3.75
600	4	3	4131	0.60	0.53	-2.25	0.68	1.95	3.42
621	1	3	5134	2.80	2.91	1.26	0.65	0.37	0.38
641	4	1	4112	2.80	2.91	1.26	0.65	0.38	0.40
643	1	3	4131	1.40	1.60	-0.88	0.53	2.59	2.74
644	2	3	4131	2.40	2.57	0.51	0.57	2.19	3.12
646	4	3	4131	1.60	1.83	-0.61	0.52	3.25	3.49
656	1	2	5121	2.20	2.40	0.20	0.54	2.13	2.09
666	4	3	5133	2.80	2.91	1.26	0.65	2.90	3.14
669	3	3	5134	1.20	1.35	-1.17	0.54	3.03	3.48
673	2	3	7131	2.80	2.91	1.26	0.65	4.07	3.71
689	1	3	4131	1.40	1.60	-0.88	0.53	0.59	0.60
690	2	3	4131	2.80	2.91	1.26	0.65	0.48	0.47
705	4	2	5123	2.20	2.40	0.20	0.54	2.65	2.71
711	1	3	5133	2.80	2.91	1.26	0.65	0.38	0.38
733	1	3	4131	2.20	2.40	0.20	0.54	1.84	2.70
767	4	3	7131	2.60	2.74	0.86	0.61	3.03	2.70
779	2	3	4131	2.60	2.74	0.86	0.61	1.89	2.11
826	4	3	4131	1.80	2.03	-0.35	0.51	2.41	2.58
827	4	3	4131	1.20	1.35	-1.17	0.54	3.52	3.47
840	4	2	5123	2.00	2.22	-0.08	0.52	3.55	3.82
851	4	3	5134	1.20	1.35	-1.17	0.54	1.92	1.82
867	3	2	4122	3.00	2.92	1.27	0.70	1.33	1.06
868	1	3	4131	3.00	2.92	1.27	0.70	1.33	1.06
872	4	3	4131	2.80	2.72	0.82	0.65	4.57	3.81
878	1	3	4233	1.20	0.79	-1.83	0.63	6.30	7.04
910	2	1	4112	0.80	0.27	-2.87	0.71	2.37	6.19
911	4	1	4112	1.00	0.44	-2.42	0.65	0.95	0.92
912	3	2	4122	2.60	2.45	0.29	0.60	0.94	0.99

ตาราง 25 (ต่อ)

ข้อที่	โรงเรียน	ระดับชั้น	ตัวชี้วัด	Obs.Avg	FairMAvg	Measure	S.E.	InfitMS	OutfitMS
914	2	3	4131	0.60	0.15	-3.43	0.79	0.21	0.17
916	4	3	4131	3.00	2.86	1.12	0.70	0.25	0.23
917	4	3	4131	1.80	1.49	-1.00	0.56	5.27	4.67
959	2	3	4131	2.60	2.62	0.61	0.61	4.41	4.79
960	3	3	4131	3.00	2.99	1.45	0.69	0.31	0.35
961	4	3	4131	1.00	0.47	-2.37	0.66	1.09	3.04
962	4	3	4131	2.80	2.80	1.00	0.64	4.93	5.40
1005	3	3	4131	2.20	2.32	0.08	0.57	2.42	2.06
1007	4	3	4131	2.80	2.87	1.16	0.65	2.04	1.79
1029	2	3	5134	2.60	2.70	0.77	0.61	1.23	1.35
1047	2	2	4122	3.00	2.83	1.06	0.70	1.15	0.91
1050	3	3	4131	2.60	2.41	0.22	0.61	0.88	1.12
1051	4	3	4131	2.40	2.19	-0.13	0.58	0.61	0.63
1088	2	3	7231	1.60	1.18	-1.36	0.56	0.64	0.59

หมายเหตุ Obs. Avg: ค่าเฉลี่ยคะแนนดิบ (raw mean) ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Fair-M Avg: ค่าเฉลี่ยคะแนนที่ปรับใหม่ (adjusted mean) หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Measure: คะแนนในหน่วยลอจิทของฟาเซทข้อสอบ (item logit)

โรงเรียน: 1 = โรงเรียน<sub>ONETสูง</sub>1, 2 = โรงเรียน<sub>ONETสูง</sub>2, 3 = โรงเรียน<sub>ONETต่ำ</sub>1, 4 = โรงเรียน<sub>ONETต่ำ</sub>2

เมื่อพิจารณาข้อสอบที่มีค่า Fair-M Avg มากกว่า 3.00 คะแนนขึ้นไป จำนวน 1,013 ข้อ พบว่า เป็นข้อสอบที่มีคะแนนเชิงประจักษ์สอดคล้องกับโมเดล คือ มีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.50 – 1.50 จำนวน 902 ข้อ คิดเป็นร้อยละ 82.83 ของข้อสอบทั้งหมด (1,089 ข้อ) โดยมีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.51 – 1.49 ตัวอย่างข้อสอบแสดงรายละเอียด ดังตาราง 26

ดังนั้น ภายหลังจากการควบคุมอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมินแล้ว มีข้อสอบที่มีความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนที่ระบุ จำนวน 902 ข้อ ที่จะถูกนำไปใช้ในการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) ก่อนนำไปวิเคราะห์ดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียนตามแนวคิดของ Porter ต่อไป

ตาราง 26 ตัวอย่างข้อสอบที่มีค่าเฉลี่ยคะแนนความสอดคล้องในแนวเดียวกันฯ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ตั้งแต่ 3.00 คะแนนขึ้นไป และมีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.50 – 1.50 (ตัวอย่างจำนวน 30 ข้อ จาก 902 ข้อ)

ข้อที่	โรงเรียน	ระดับชั้น	ตัวชี้วัด	Obs.Avg	FairMAvg	Measure	S.E.	InfitMS	OutfitMS
1	1	1	1114	3.85	3.89	4.79	0.64	0.82	0.66
39	3	3	7133	3.85	3.89	4.79	0.64	0.92	0.98
75	1	1	6116	3.20	3.42	2.59	0.73	1.11	1.13
111	4	3	1233	3.80	3.89	4.75	1.14	1.13	0.87
134	1	2	1125	3.60	3.68	3.52	0.86	0.59	0.55
146	4	2	1125	3.80	3.85	4.44	1.10	0.86	0.72
174	1	1	1114	3.60	3.68	3.52	0.86	1.02	0.99
192	4	2	1125	3.60	3.68	3.52	0.86	0.78	0.75
202	1	3	2131	3.80	3.85	4.44	1.10	0.86	0.72
223	1	2	1125	3.80	3.90	4.83	1.13	1.27	1.33
235	4	2	1125	3.40	3.59	3.16	0.78	1.13	1.03
286	1	3	1236	3.80	3.84	4.37	1.10	0.80	0.67
294	1	1	6116	3.60	3.67	3.46	0.86	0.75	0.74
302	4	2	6124	3.80	3.84	4.37	1.10	0.99	0.99
334	4	3	2131	3.80	3.90	4.83	1.13	1.17	1.04
391	2	1	3111	3.80	3.81	4.16	1.11	0.98	0.84
413	3	2	3123	3.80	3.81	4.16	1.11	0.85	0.65
438	4	1	3112	3.60	3.66	3.41	0.89	1.10	0.97
494	3	2	3123	3.60	3.53	2.96	0.90	1.12	1.25
501	4	1	3211	3.80	3.77	3.94	1.13	0.78	0.53
527	1	2	3122	3.40	3.22	2.03	0.84	1.17	0.90
571	4	2	3122	3.80	3.80	4.06	1.12	1.09	1.03
781	4	3	4131	3.00	3.10	1.71	0.70	0.90	0.90
796	1	3	5131	3.60	3.69	3.54	0.89	1.04	1.02
871	4	3	4131	3.40	3.37	2.45	0.84	1.39	1.04
930	1	3	5131	3.60	3.56	3.06	0.93	0.91	0.75
973	3	2	5121	3.40	3.42	2.60	0.84	1.02	0.79
996	4	3	7133	3.60	3.66	3.40	0.96	0.96	0.88
1054	2	3	4231	3.60	3.54	2.98	0.92	1.01	1.13
1089	4	3	7231	3.75	3.71	3.65	0.86	1.23	0.85

หมายเหตุ Obs. Avg: ค่าเฉลี่ยคะแนนดิบ (raw mean) ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Fair-M Avg: ค่าเฉลี่ยคะแนนที่ปรับใหม่ (adjusted mean) หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Measure: คะแนนในหน่วยลอจิทของฟาเซทข้อสอบ (item logit)

โรงเรียน: 1 = โรงเรียน<sub>ONETสูง</sub>1, 2 = โรงเรียน<sub>ONETสูง</sub>2, 3 = โรงเรียน<sub>ONETต่ำ</sub>1, 4 = โรงเรียน<sub>ONETต่ำ</sub>2

เมื่อพิจารณาข้อสอบที่มี ค่า Fair-M Avge มากกว่า 3.00 คะแนนขึ้นไป แต่ข้อมูลเชิงประจักษ์ไม่สอดคล้องกับโมเดล คือมีค่า infit MNSQ และ outfit MNSQ ไม่อยู่ช่วง 0.50 – 1.50 พบว่า จากข้อสอบ 1,013 ข้อ จำนวน 111 ข้อ คิดเป็นร้อยละ 10.28 ของข้อสอบทั้งหมด (1,089 ข้อ) ตัวอย่างรายข้อแสดงรายละเอียดดังตาราง 26 ซึ่งข้อสอบทั้ง 111 ข้อนี้ จะไม่ถูกนำไปใช้ในการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) และไม่ถูกนำไปใช้ในการวิเคราะห์ดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

**ตาราง 27** ตัวอย่างข้อสอบที่มีค่าเฉลี่ยคะแนนความสอดคล้องในแนวเดียวกันฯ หลังควบคุมอิทธิพล การกดหรือปล่อยคะแนน ตั้งแต่ 3.00 คะแนนขึ้นไป แต่มีข้อมูลเชิงประจักษ์ไม่สอดคล้องกับโมเดล (ตัวอย่างจำนวน 15 ข้อ จาก 111 ข้อ)

ข้อที่	โรงเรียน	ระดับชั้น	ตัวชี้วัด	Obs.Avge	FairMAvge	Measure	S.E.	InfitMS	OutfitMS
38	2	3	7131	3.85	3.89	4.79	0.64	0.74	0.49
69	3	3	1236	3.60	3.76	3.86	0.89	0.46	0.45
444	4	1	3114	3.40	3.45	2.71	0.80	1.57	1.39
504	1	1	3212	3.60	3.53	2.96	0.90	0.42	0.36
538	1	1	3211	3.60	3.49	2.81	0.93	1.34	1.85
544	1	1	3212	3.80	3.76	3.85	1.15	0.77	0.48
596	4	2	4122	3.00	3.10	1.71	0.70	0.40	0.39
599	3	3	4131	3.20	3.29	2.23	0.75	5.53	8.20
723	4	3	7131	3.40	3.49	2.83	0.80	1.60	1.40
743	1	3	4233	3.40	3.49	2.83	0.80	0.22	0.23
945	3	3	7131	3.80	3.80	4.12	1.16	0.65	0.38
1028	1	3	5134	3.60	3.70	3.59	0.95	0.31	0.25
1067	3	3	5131	3.20	3.05	1.59	0.76	2.53	2.04
1079	2	3	7131	3.80	3.79	4.02	1.16	0.55	0.33

**หมายเหตุ** Obs. Avge: ค่าเฉลี่ยคะแนนดิบ (raw mean) ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Fair-M Avge: ค่าเฉลี่ยคะแนนที่ปรับใหม่ (adjusted mean) หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Measure: คะแนนในหน่วยลอจิทของฟาเซทข้อสอบ (item logit)

โรงเรียน: 1 = โรงเรียน<sub>ONETสูง</sub>1, 2 = โรงเรียน<sub>ONETสูง</sub>2, 3 = โรงเรียน<sub>ONETต่ำ</sub>1, 4 = โรงเรียน<sub>ONETต่ำ</sub>2

### 3.3 ผลการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

ผลการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ซึ่งผลการเปรียบเทียบผลการประเมินความสอดคล้องฯ ทั้งในภาพรวมและเป็นรายข้อ มีรายละเอียดดังนี้

2.3.1 ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Obs. Avge) และหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Fair-M Avge) ด้วยการทดสอบสถิติที (paired-samples t-test) พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ( $t = 17.044$ ,  $p = .00$ ) โดยค่าเฉลี่ยหลังการควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Fair-M Avge) มีค่าสูงกว่าค่าเฉลี่ยก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Obs. Avge) เล็กน้อย โดยผลต่างของค่าเฉลี่ยเท่ากับ 0.046 รายละเอียดดังตาราง 26 ทั้งนี้ เมื่อคำนวณค่าขนาดอิทธิพล (effect size) จากสูตรของ Cohen พบว่า มีค่าขนาดอิทธิพลเท่ากับ 0.52 ซึ่งอยู่ในระดับปานกลาง (Cohen, 1988)

#### ตาราง 28 ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนการประเมินความสอดคล้องในแนวเดียวกัน ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Pair	Mean	SD	Std. Error Mean	99% Confidence Interval of the Difference		t	d.f.	p
				Lower	Upper			
FairM_Avge - Obs_Avge	0.046	0.090	0.003	0.039	0.053	17.044	1,088	.000

3.3.2 ผลการศึกษาการเปลี่ยนแปลงผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน พบว่า มีข้อสอบ จำนวน 21 ข้อ คิดเป็นร้อยละ 1.93 ที่มีการเปลี่ยนแปลงผลการประเมิน โดยพบว่า มีข้อสอบจำนวน 11 ข้อ ที่มีการเปลี่ยนแปลงผลการประเมินจาก “สอดคล้องกับตัวชี้วัดที่ระบุ” คือมีค่า Obs. Avge ตั้งแต่ 3.00 คะแนนขึ้นไป เปลี่ยนเป็น “ไม่สอดคล้องกับตัวชี้วัดที่ระบุ” คือมีค่า Fair-M Avge ต่ำกว่า 3.00 คะแนน ได้แก่ ข้อสอบข้อที่ 428

516 540 541 561 867 868 916 960 และ 1047 และมีข้อสอบจำนวน 10 ข้อ ที่เปลี่ยนแปลงผลการประเมินจาก “ไม่สอดคล้องกับตัวชี้วัดที่ระบุ” คือมีค่า Obs. Avge ต่ำกว่า 3.00 คะแนน เปลี่ยนเป็น “สอดคล้องกับตัวชี้วัดที่ระบุ” คือมีค่า Fair-M Avge ตั้งแต่ 3.00 คะแนนขึ้นไป ได้แก่ ข้อสอบข้อที่ 57 78 79 80 220 230 251 254 316 317 และ 339 รายละเอียดแสดงดังตาราง 27

**ตาราง 29 ข้อสอบที่มีการเปลี่ยนแปลงผลการประเมินความสอดคล้องในแนวเดียวกัน หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน (จำนวน 21 ข้อ)**

ข้อที่	ชั้น	ระดับชั้น	ตัวชี้วัด	Obs.Avge	FairMAvge	Measure	S.E.	InfitMS	OutfitMS
57	3	2	1125	2.80	3.07	1.64	0.65	0.69	0.81
78	3	1	6116	2.80	3.07	1.64	0.65	0.69	0.81
79	2	2	6122	2.80	3.07	1.64	0.65	0.56	0.47
80	3	2	6124	2.80	3.07	1.64	0.65	0.56	0.47
220	3	2	1123	2.80	3.07	1.64	0.65	1.22	1.38
230	3	2	1125	2.80	3.07	1.64	0.65	1.12	1.32
251	1	1	6116	2.80	3.07	1.64	0.65	0.59	0.51
254	2	2	6122	2.80	3.07	1.64	0.65	0.59	0.51
316	3	2	1125	2.80	3.07	1.64	0.65	1.12	1.32
317	3	2	1125	2.80	3.07	1.64	0.65	0.75	0.72
339	3	1	6116	2.80	3.07	1.64	0.65	0.75	0.72
428	3	2	3221	3.00	2.97	1.39	0.70	0.64	0.68
516	4	1	3111	3.00	2.74	0.86	0.70	0.74	1.16
540	4	1	3211	3.00	2.74	0.86	0.70	0.70	0.68
541	4	1	3211	3.00	2.74	0.86	0.70	1.07	1.00
561	4	1	3112	3.00	2.93	1.29	0.70	1.03	1.01
867	3	2	4122	3.00	2.92	1.27	0.70	1.33	1.06
868	1	3	4131	3.00	2.92	1.27	0.70	1.33	1.06
916	4	3	4131	3.00	2.86	1.12	0.70	0.25	0.23
960	3	3	4131	3.00	2.99	1.45	0.69	0.31	0.35
1047	2	2	4122	3.00	2.83	1.06	0.70	1.15	0.91

หมายเหตุ Obs. Avge: ค่าเฉลี่ยคะแนนดิบ (raw mean) ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Fair-M Avge: ค่าเฉลี่ยคะแนนที่ปรับใหม่ (adjusted mean) หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Measure: คะแนนในหน่วยลอจิทของฟาเซทข้อสอบ (item logit)

**ตอนที่ 4 ผลการศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน**

ผลการศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน แบ่งการนำเสนอเป็น 3 ส่วน ได้แก่ 1) ผลการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) 2) ผลการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) และ 3) ผลการเปรียบเทียบค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

**4.1 ผลการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X)**

การสร้างเมทริกซ์ข้อสอบในการประเมินระดับชาติ หรือ เมทริกซ์ X เริ่มต้นจากผู้วิจัยนำข้อมูลจากแบบบันทึกระดับความซับซ้อนทางปัญญาของข้อสอบในการประเมินระดับชาติมาสรุปข้อมูลความถี่ของข้อสอบ O-NET วิทยาศาสตร์ ที่วัดในแต่ละระดับความซับซ้อนทางปัญญา โดยจำแนกความถี่ตามรายมาตรฐานและตัวชี้วัด และสรุปความถี่ภายใต้ 7 สาระ รายละเอียดดังตาราง 30 มาสร้างเป็นเมทริกซ์ผลรวม X ภาพ 8

**ตาราง 30 สรุปข้อมูลความถี่ของข้อสอบ O-NET วิทยาศาสตร์ จำแนกตามสาระและระดับความซับซ้อนทางปัญญา**

สาระที่	มาตรฐาน	ตัวชี้วัด	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
1	ว 1.1	1114	0	2	2	0	0	0
		1115	0	2	0	2	0	0
		1122	0	1	0	1	0	0
		1123	0	1	0	1	0	0
		1124	0	2	0	0	0	0
		1125	0	0	2	2	0	0
		<b>รวม สาระที่ 1</b>	<b>0</b>	<b>8</b>	<b>6</b>	<b>10</b>	<b>2</b>	<b>0</b>
ว 1.2	1233	1233	0	0	2	4	0	0
		1236	0	0	0	0	2	0
		<b>รวม สาระที่ 2</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>8</b>	<b>0</b>	<b>0</b>
2	ว 2.1	2131	0	2	0	0	0	0
		2133	0	0	0	4	0	0
	ว 2.2	2232	0	0	0	2	0	0
		2234	0	0	0	2	0	0
<b>รวม สาระที่ 2</b>	<b>0</b>	<b>2</b>	<b>0</b>	<b>8</b>	<b>0</b>	<b>0</b>		



ตาราง 30 (ต่อ)

สาระที่	มาตรฐาน	ตัวชี้วัด	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	
3	ว 3.1	3111	0	2	0	0	0	0	
		3112	0	0	0	2	0	0	
		3114	0	0	0	2	0	0	
		3122	0	0	2	0	0	0	
		3123	0	0	2	0	0	0	
ว 3.2		3211	0	0	2	0	0	0	
		3212	0	0	0	1	0	0	
		3221	0	0	0	1	0	0	
<b>รวม สาระที่ 3</b>			<b>0</b>	<b>2</b>	<b>6</b>	<b>6</b>	<b>0</b>	<b>0</b>	
4	ว 4.1	4112	0	2	6	0	0	0	
		4122	0	2	0	0	0	0	
		4131	0	2	0	2	0	0	
		4132	0	2	0	0	0	0	
	ว 4.2		4231	0	0	0	4	0	0
			4232	0	0	1	0	0	0
			4233	0	0	1	0	0	0
	<b>รวม สาระที่ 4</b>			<b>0</b>	<b>6</b>	<b>10</b>	<b>6</b>	<b>0</b>	<b>0</b>
	5	ว 5.1	5112	0	4	0	0	0	0
			5121	0	0	0	2	0	0
5122			0	0	0	2	0	0	
5123			0	2	0	2	0	0	
5131			0	2	0	0	0	0	
5133			0	0	1	0	0	0	
5134			0	0	1	0	0	0	
<b>รวม สาระที่ 5</b>			<b>0</b>	<b>8</b>	<b>2</b>	<b>6</b>	<b>0</b>	<b>0</b>	
6	ว 6.1	6116	0	0	0	2	0	0	
		6122	0	1	0	0	0	0	
		6124	0	1	0	0	0	0	
		<b>รวม สาระที่ 6</b>			<b>0</b>	<b>2</b>	<b>0</b>	<b>2</b>	<b>0</b>

ตาราง 30 (ต่อ)

สาระที่	มาตรฐาน	ตัวชี้วัด	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
7	ว 7.1	7131	0	2	0	0	0	0
		7133	0	0	0	2	0	0
	ว 7.2	7231	0	2	2	0	0	0
รวม สาระที่ 7			0	4	2	2	0	0
รวม สาระที่ 1 - 7			0	32	26	40	2	0

เมื่อพิจารณาเมทริกซ์ผลรวมข้อสอบในการประเมินระดับชาติ หรือเมทริกซ์ผลรวม X ดังภาพ 8 พบว่า ข้อสอบในการประเมินระดับชาติ ส่วนใหญ่วัดความซับซ้อนทางปัญญาระดับวิเคราะห์ ร้อยละ 40.00 รองลงมาคือ ระดับเข้าใจ ร้อยละ 32.00 ระดับประยุกต์ใช้ ร้อยละ 26.00 ระดับประเมินค่า ร้อยละ 2.00 และระดับสร้างสรรค์ ร้อยละ 0.00

หลังจากได้เมทริกซ์ผลรวม X แล้ว ผู้วิจัยทำค่าในเมทริกซ์ผลรวม X ให้เป็นเมทริกซ์สัดส่วน X ซึ่งเป็นการทำให้ค่าในเมทริกซ์ X ให้อยู่ในมาตรเดียวกับเมทริกซ์ผลรวม Y เพื่อให้สามารถเปรียบเทียบค่าระหว่าง 2 เมทริกซ์ได้ โดยนำค่าผลรวมของเมทริกซ์ X คือ 100 ไปหารสมาชิกในเมทริกซ์ผลรวม X ผลของการหารสมาชิกทุกตัวของเมทริกซ์ด้วยผลรวม จะได้เป็นเมทริกซ์สัดส่วน และทำให้ผลรวมของค่าสัดส่วนในเมทริกซ์เท่ากับ 1.00 รายละเอียดดังภาพ 9 เมทริกซ์สัดส่วน X นี้ จะถูกนำไปเปรียบเทียบกับเมทริกซ์สัดส่วน Y ซึ่งจะกล่าวต่อไปในหัวข้อ 4.2

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	0	8	6	10	2	0	26
2	0	2	0	8	0	0	10
3	0	2	6	6	0	0	14
4	0	6	10	6	0	0	22
5	0	8	2	6	0	0	16
6	0	2	0	2	0	0	4
7	0	4	2	2	0	0	8
ผลรวม	0	32	26	40	2	0	100

ภาพ 8 เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชาติ (เมทริกซ์ผลรวม X)

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	0.000	0.080	0.060	0.100	0.020	0.000	0.260
2	0.000	0.020	0.000	0.080	0.000	0.000	0.100
3	0.000	0.020	0.060	0.060	0.000	0.000	0.140
4	0.000	0.060	0.100	0.060	0.000	0.000	0.220
5	0.000	0.080	0.020	0.060	0.000	0.000	0.160
6	0.000	0.020	0.000	0.020	0.000	0.000	0.040
7	0.000	0.040	0.020	0.020	0.000	0.000	0.080
<b>รวม</b>	<b>0.000</b>	<b>0.320</b>	<b>0.260</b>	<b>0.400</b>	<b>0.020</b>	<b>0.000</b>	<b>1.000</b>
ร้อยละ	0.00%	32.00%	26.00%	40.00%	2.00%	0.00%	100.00%

ภาพ 9 เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชาติ (เมทริกซ์สัดส่วน X)

#### 4.2 ผลการสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y)

การสร้างเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน หรือเมทริกซ์ Y แบ่งเป็นเมทริกซ์ Y โดยภาพรวม และเมทริกซ์ Y ที่จำแนกรายโรงเรียน 4 โรง ซึ่งจำแนกเป็นโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง (โรงเรียน<sub>ONETสูง</sub>) จำนวน 2 โรง ได้แก่ เมทริกซ์ Y โรงเรียน<sub>ONETสูง</sub>1 และเมทริกซ์ Y โรงเรียน<sub>ONETสูง</sub>2 และโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ (โรงเรียน<sub>ONETต่ำ</sub>) จำนวน 2 โรง ได้แก่ เมทริกซ์ Y โรงเรียน<sub>ONETต่ำ</sub>1 และเมทริกซ์ Y โรงเรียน<sub>ONETต่ำ</sub>2 มีรายละเอียดดังต่อไปนี้

##### 4.2.1 เมทริกซ์ Y โดยภาพรวม

เมื่อพิจารณาเมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียนของโดยภาพรวม หรือเมทริกซ์ผลรวม Y โดยภาพรวม แสดงดังภาพ 10 พบว่า ข้อสอบในการประเมินระดับชาติ ส่วนใหญ่วัดความซับซ้อนทางปัญญาระดับเข้าใจ ร้อยละ 46.99 รองลงมาคือ ระดับจำ ร้อยละ 25.21 ระดับประยุกต์ใช้ ร้อยละ 15.07 ระดับวิเคราะห์ ร้อยละ 11.64 ระดับประเมินค่า ร้อยละ 1.09 และระดับสร้างสรรค์ ร้อยละ 0.00

หลังจากได้เมทริกซ์ผลรวม Y แล้ว ผู้วิจัยทำค่าในเมทริกซ์ผลรวม Y ให้เป็นเมทริกซ์สัดส่วน ซึ่งเป็นการทำให้ค่าในเมทริกซ์ Y ให้อยู่ในมาตรฐานเดียวกับเมทริกซ์ผลรวม Y เพื่อให้สามารถเปรียบเทียบค่าระหว่าง 2 เมทริกซ์ได้ โดยนำค่าผลรวมของเมทริกซ์ X คือ 2,386.300 ไปหารสมาชิกในเมทริกซ์ผลรวม Y ผลของการหารสมาชิกทุกตัวด้วยผลรวม จะได้เป็นเมทริกซ์สัดส่วน และทำให้ผลรวมของค่าสัดส่วนในเมทริกซ์จะเท่ากับ 1.00 รายละเอียดดังภาพ 11 เมทริกซ์สัดส่วน Y โดยภาพรวมนี้จะถูกนำไปเปรียบเทียบกับเมทริกซ์สัดส่วน X ต่อไป

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	138.340	240.340	41.890	99.430	7.500	0.000	527.500
2	17.500	56.970	2.480	14.540	3.060	0.000	94.550
3	191.110	203.110	83.930	46.950	3.000	0.000	528.100
4	23.580	185.210	101.450	28.310	1.750	0.000	340.300
5	58.670	201.800	105.890	50.280	3.660	0.000	420.300
6	65.000	75.500	12.500	25.500	6.500	0.000	185.000
7	107.430	158.280	11.570	12.770	0.500	0.000	290.550
รวม	601.630	1,121.210	359.710	277.780	25.970	0.000	2,386.300

ร้อยละ 25.21% 46.99% 15.07% 11.64% 1.09% 0.00% 100.00%

ภาพ 10 เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียนโดยภาพรวม (เมทริกซ์ผลรวม  $Y_{ภาพรวม}$ )

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	0.058	0.101	0.018	0.042	0.003	0.000	0.221
2	0.007	0.024	0.001	0.006	0.001	0.000	0.040
3	0.080	0.085	0.035	0.020	0.001	0.000	0.221
4	0.010	0.078	0.043	0.012	0.001	0.000	0.143
5	0.025	0.085	0.044	0.021	0.002	0.000	0.176
6	0.027	0.032	0.005	0.011	0.003	0.000	0.078
7	0.045	0.066	0.005	0.005	0.000	0.000	0.122
รวม	0.252	0.470	0.151	0.116	0.011	0.000	1.000

ภาพ 11 เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียนโดยภาพรวม (เมทริกซ์สัดส่วน  $Y_{ภาพรวม}$ )

#### 4.2.2 เมทริกซ์ Y ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 1 (โรงเรียน<sub>ONETสูง</sub>1)

เมื่อพิจารณาเมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 1 (โรงเรียน<sub>ONETสูง</sub>1) หรือเมทริกซ์ผลรวม Y ของโรงเรียน<sub>ONETสูง</sub>1 แสดงดังภาพ 12 พบว่า ข้อสอบส่วนใหญ่วัดความซับซ้อนทางปัญญาระดับเข้าใจ ร้อยละ 50.39 รองลงมาคือ ระดับวิเคราะห์ ร้อยละ 16.62 ระดับจำ ร้อยละ 18.18 ระดับประยุกต์ใช้ ร้อยละ 13.75 ระดับประเมินค่า ร้อยละ 1.02 และระดับสร้างสรรค์ ร้อยละ 0.00 ทั้งนี้ เมื่อแปลงเมทริกซ์ผลรวมเป็น

เมทริกซ์สัดส่วน โดยนำค่าผลรวมของเมทริกซ์ คือ 627.90 ไปหารสมาชิกในเมทริกซ์ผลรวม แสดงดัง  
 ภาพ 13 เมทริกซ์สัดส่วน Y โรงเรียน<sub>ONETสูง1</sub> นี้จะถูกนำไปเปรียบเทียบกับเมทริกซ์สัดส่วน X ต่อไป

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	20.290	60.190	2.490	25.780	1.000	0.000	109.750
2	0.000	22.770	1.980	2.640	0.660	0.000	28.050
3	50.320	69.830	23.420	26.480	2.000	0.000	172.050
4	2.500	47.500	23.000	6.000	1.000	0.000	80.000
5	12.500	46.500	31.500	30.500	1.000	0.000	122.000
6	15.000	28.000	1.000	10.000	1.000	0.000	55.000
7	13.530	41.580	2.970	2.970	0.000	0.000	61.050
รวม	114.140	316.370	86.360	104.370	6.660	0.000	627.900
ร้อยละ	18.18%	50.39%	13.75%	16.62%	1.06%	0.00%	100.00%

ภาพ 12 เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน<sub>ONETสูง1</sub> (เมทริกซ์  
 ผลรวม Y โรงเรียน<sub>ONETสูง1</sub>)

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	0.032	0.096	0.004	0.041	0.002	0.000	0.175
2	0.000	0.036	0.003	0.004	0.001	0.000	0.045
3	0.080	0.111	0.037	0.042	0.003	0.000	0.274
4	0.004	0.076	0.037	0.010	0.002	0.000	0.127
5	0.020	0.074	0.050	0.049	0.002	0.000	0.194
6	0.024	0.045	0.002	0.016	0.002	0.000	0.088
7	0.022	0.066	0.005	0.005	0.000	0.000	0.097
รวม	0.182	0.504	0.138	0.166	0.011	0.000	1.000

ภาพ 13 เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน<sub>ONETสูง1</sub> (เมทริกซ์  
 สัดส่วน Y โรงเรียน<sub>ONETสูง1</sub>)

#### 4.2.3 เมทริกซ์ Y ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 2 (โรงเรียน<sub>ONETสูง2</sub>)

เมื่อพิจารณาเมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มี  
 มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 2 (โรงเรียน<sub>ONETสูง2</sub>) หรือเมทริกซ์ผลรวม Y ของโรงเรียน<sub>ONETสูง2</sub>  
 ดังภาพ 14 พบว่า ข้อสอบส่วนใหญ่วัดความซับซ้อนทางปัญญาระดับเข้าใจ ร้อยละ 46.80 รองลงมา  
 คือ ระดับจำ ร้อยละ 20.36 ระดับประยุกต์ใช้ ร้อยละ 20.36 ระดับวิเคราะห์ ร้อยละ 11.02 ระดับ  
 ประเมินค่า ร้อยละ 2.06 และระดับสร้างสรรค์ ร้อยละ 0.00 ทั้งนี้ เมื่อแปลงเมทริกซ์ผลรวมเป็น

เมทริกซ์สัดส่วน โดยนำค่าผลรวมของเมทริกซ์ คือ 528.000 ไปหารสมาชิกในเมทริกซ์ผลรวม แสดงดังภาพ 15 เมทริกซ์สัดส่วน Y โรงเรียน<sub>ONETสูง2</sub> นี้จะถูกนำไปเปรียบเทียบกับเมทริกซ์สัดส่วน X ต่อไป

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	26.900	61.200	13.400	19.500	3.500	0.000	124.500
2	0.000	9.200	0.000	2.400	2.400	0.000	14.000
3	26.000	51.000	40.000	14.500	1.000	0.000	132.500
4	5.000	24.000	29.500	4.000	0.000	0.000	62.500
5	10.500	47.000	20.000	6.500	1.000	0.000	85.000
6	9.500	17.500	1.000	6.500	3.000	0.000	37.500
7	26.400	37.200	3.600	4.800	0.000	0.000	72.000
รวม	104.300	247.100	107.500	58.200	10.900	0.000	528.000
ร้อยละ	19.75%	46.80%	20.36%	11.02%	2.06%	0.00%	100.00%

ภาพ 14 เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน<sub>ONETสูง2</sub> (เมทริกซ์ผลรวม Y โรงเรียน<sub>ONETสูง2</sub>)

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	0.051	0.116	0.025	0.037	0.007	0.000	0.236
2	0.000	0.017	0.000	0.005	0.005	0.000	0.027
3	0.049	0.097	0.076	0.027	0.002	0.000	0.251
4	0.009	0.045	0.056	0.008	0.000	0.000	0.118
5	0.020	0.089	0.038	0.012	0.002	0.000	0.161
6	0.018	0.033	0.002	0.012	0.006	0.000	0.071
7	0.050	0.070	0.007	0.009	0.000	0.000	0.136
รวม	0.198	0.468	0.204	0.110	0.021	0.000	1.000

ภาพ 15 เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน<sub>ONETสูง2</sub> (เมทริกซ์สัดส่วน Y โรงเรียน<sub>ONETสูง2</sub>)

#### 4.2.4 เมทริกซ์ Y ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 1 (โรงเรียน<sub>ONETต่ำ1</sub>)

เมื่อพิจารณาเมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 1 (โรงเรียน<sub>ONETต่ำ1</sub>) หรือเมทริกซ์ผลรวม Y ของโรงเรียน<sub>ONETต่ำ1</sub> แสดงดังภาพ 16 พบว่า ข้อสอบส่วนใหญ่วัดความซับซ้อนทางปัญญาระดับเข้าใจ ร้อยละ 48.83 รองลงมาคือ ระดับจำ ร้อยละ 35.32 ระดับประยุกต์ใช้ ร้อยละ 10.56 ระดับวิเคราะห์ ร้อยละ 4.68 ระดับประเมินค่า ร้อยละ 0.61 และระดับสร้างสรรค์ 0.00 ทั้งนี้ เมื่อแปลงเมทริกซ์ผลรวมเป็นเมท

ริกซ์สัดส่วน โดยนำค่าผลรวมของเมทริกซ์ คือ 577.500 ไปหารสมาชิกในเมทริกซ์ผลรวม แสดงดัง  
ภาพ 17 เมทริกซ์สัดส่วน Y โรงเรียน<sub>ONET</sub>ต่ำ 1 นี้จะถูกนำไปเปรียบเทียบกับเมทริกซ์สัดส่วน X ต่อไป

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	51.500	63.500	7.500	8.500	1.500	0.000	132.500
2	9.500	15.500	0.000	2.500	0.000	0.000	27.500
3	62.500	47.500	5.000	2.500	0.000	0.000	117.500
4	11.500	43.000	11.500	3.500	0.500	0.000	70.000
5	18.500	68.500	25.000	7.000	1.000	0.000	120.000
6	14.500	17.000	9.000	1.500	0.500	0.000	42.500
7	36.000	27.000	3.000	1.500	0.000	0.000	67.500
รวม	204.000	282.000	61.000	27.000	3.500	0.000	577.500
ร้อยละ	35.32%	48.83%	10.56%	4.68%	0.61%	0.00%	100.00%

ภาพ 16 เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน<sub>ONET</sub>ต่ำ 1 (เมทริกซ์  
ผลรวม Y โรงเรียน<sub>ONET</sub>ต่ำ 1)

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	0.089	0.110	0.013	0.015	0.003	0.000	0.229
2	0.016	0.027	0.000	0.004	0.000	0.000	0.048
3	0.108	0.082	0.009	0.004	0.000	0.000	0.203
4	0.020	0.074	0.020	0.006	0.001	0.000	0.121
5	0.032	0.119	0.043	0.012	0.002	0.000	0.208
6	0.025	0.029	0.016	0.003	0.001	0.000	0.074
7	0.062	0.047	0.005	0.003	0.000	0.000	0.117
รวม	0.353	0.488	0.106	0.047	0.006	0.000	1.000

ภาพ 17 เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน<sub>ONET</sub>ต่ำ 1 (เมทริกซ์  
สัดส่วน Y โรงเรียน<sub>ONET</sub>ต่ำ 1)

#### 4.2.5 เมทริกซ์ Y ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 2 (โรงเรียน<sub>ONET</sub>ต่ำ 2)

เมื่อพิจารณาเมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มี  
ผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 2 (โรงเรียน<sub>ONET</sub>ต่ำ 2) หรือเมทริกซ์ผลรวม Y ของโรงเรียน<sub>ONET</sub>ต่ำ 2  
แสดงดังภาพ 18 พบว่า ข้อสอบส่วนใหญ่วัดความซับซ้อนทางปัญญาระดับเข้าใจ ร้อยละ 42.23  
รองลงมาคือ ระดับจำ ร้อยละ 27.45 ระดับประยุกต์ใช้ ร้อยละ 16.06 ระดับวิเคราะห์ ร้อยละ 13.51

ระดับประเมินค่า ร้อยละ 0.75 และระดับสร้างสรรค์ ร้อยละ 0.00 ทั้งนี้ เมื่อแปลงเมทริกซ์ผลรวมเป็นเมทริกซ์สัดส่วน โดยนำค่าผลรวมของเมทริกซ์ คือ 652.900 ไปหารสมาชิกในเมทริกซ์ผลรวม แสดงดังภาพ 19 เมทริกซ์สัดส่วน Y โรงเรียน<sub>ONET</sub>ต่ำ2 นี้จะถูกนำไปเปรียบเทียบกับเมทริกซ์สัดส่วน X ต่อไป

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	39.650	55.450	18.500	45.650	1.500	0.000	160.750
2	8.000	9.500	0.500	7.000	0.000	0.000	25.000
3	52.290	34.780	15.510	3.470	0.000	0.000	106.050
4	4.580	70.710	37.450	14.810	0.250	0.000	127.800
5	17.170	39.800	29.390	6.280	0.660	0.000	93.300
6	26.000	13.000	1.500	7.500	2.000	0.000	50.000
7	31.500	52.500	2.000	3.500	0.500	0.000	90.000
รวม	179.190	275.740	104.850	88.210	4.910	0.000	652.900
ร้อยละ	27.45%	42.23%	16.06%	13.51%	0.75%	0.00%	100.00%

ภาพ 18 เมทริกซ์ผลรวมข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน<sub>ONET</sub>ต่ำ2 (เมทริกซ์ผลรวม Y โรงเรียน<sub>ONET</sub>ต่ำ2)

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	รวม
1	0.061	0.085	0.028	0.070	0.002	0.000	0.246
2	0.012	0.015	0.001	0.011	0.000	0.000	0.038
3	0.080	0.053	0.024	0.005	0.000	0.000	0.162
4	0.007	0.108	0.057	0.023	0.000	0.000	0.196
5	0.026	0.061	0.045	0.010	0.001	0.000	0.143
6	0.040	0.020	0.002	0.011	0.003	0.000	0.077
7	0.048	0.080	0.003	0.005	0.001	0.000	0.138
รวม	0.274	0.422	0.161	0.135	0.008	0.000	1.000

ภาพ 19 เมทริกซ์สัดส่วนข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน<sub>ONET</sub>ต่ำ2 (เมทริกซ์สัดส่วน Y โรงเรียน<sub>ONET</sub>ต่ำ2)

4.3 ผลการเปรียบเทียบค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน

การเปรียบเทียบค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน ทำโดยการวิเคราะห์และเปรียบเทียบค่าดัชนีความสอดคล้อง (alignment index) ระหว่างเมทริกซ์ X และ



เมทริกซ์ Y ผลการวิเคราะห์แบ่งเป็น 2 ส่วน คือ 1) ผลการวิเคราะห์ค่าดัชนีความสอดคล้องฯ และ 2) ผลการเปรียบเทียบค่าดัชนีความสอดคล้องฯ ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน

#### 4.3.1 ผลการวิเคราะห์ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน

(1) ผลการวิเคราะห์ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนโดยภาพรวม

ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนโดยภาพรวม มีค่าเท่ากับ 0.560 รายละเอียดของเมทริกซ์ สัดส่วนของข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) เมทริกซ์ในการประเมินระดับชั้นเรียน ของโรงเรียนโดยภาพรวม (เมทริกซ์ Y) และเมทริกซ์ผลต่างระหว่างทั้ง 2 เมทริกซ์ มีรายละเอียดดังภาพ 20

สาระ	เมทริกซ์สัดส่วน X (ข้อสอบระดับชาติ)						เมทริกซ์สัดส่วน Y (ข้อสอบระดับชั้นเรียน ของโรงเรียนโดยภาพรวม)					
	จำ	เข้าใจ	ประยุกต์ ใช้	วิเคราะห์	ประเมิน ค่า	สร้าง สรรค์	จำ	เข้าใจ	ประยุกต์ ใช้	วิเคราะห์	ประเมิน ค่า	สร้าง สรรค์
1	0.000	0.080	0.060	0.100	0.020	0.000	0.058	0.101	0.018	0.042	0.003	0.000
2	0.000	0.020	0.000	0.080	0.000	0.000	0.007	0.024	0.001	0.006	0.001	0.000
3	0.000	0.020	0.060	0.060	0.000	0.000	0.080	0.085	0.035	0.020	0.001	0.000
4	0.000	0.060	0.100	0.060	0.000	0.000	0.010	0.078	0.043	0.012	0.001	0.000
5	0.000	0.080	0.020	0.060	0.000	0.000	0.025	0.085	0.044	0.021	0.002	0.000
6	0.000	0.020	0.000	0.020	0.000	0.000	0.027	0.032	0.005	0.011	0.003	0.000
7	0.000	0.040	0.020	0.020	0.000	0.000	0.045	0.066	0.005	0.005	0.000	0.000

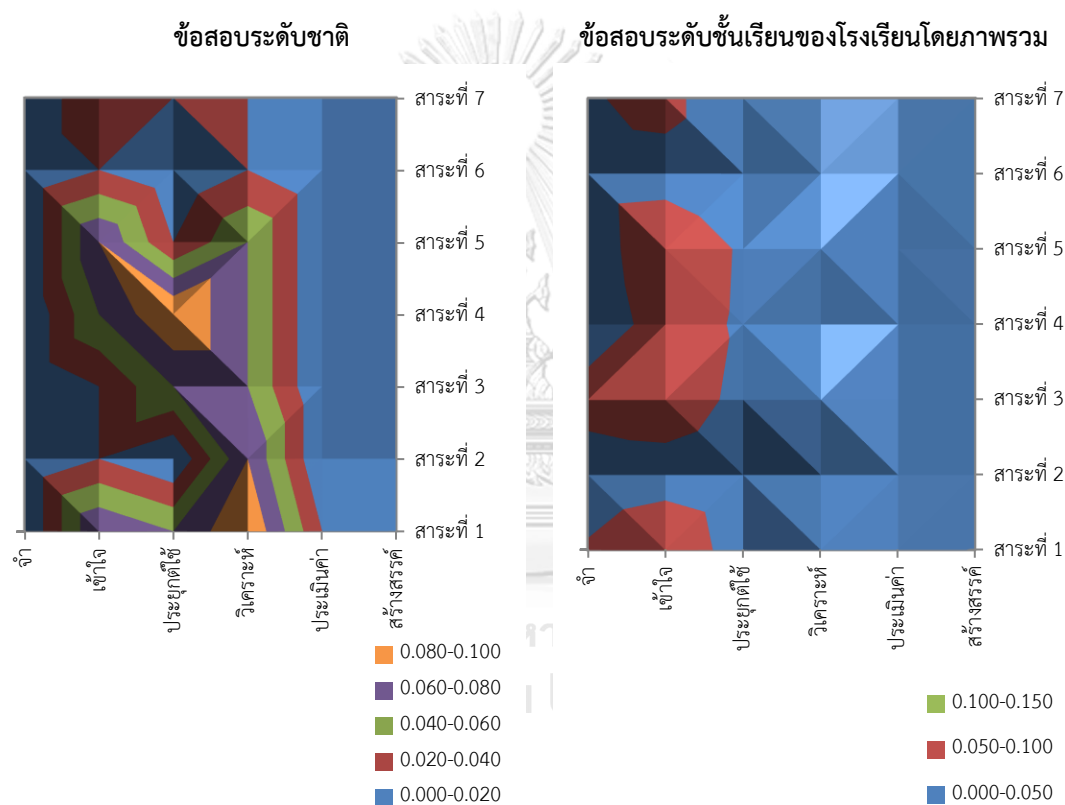
#### เมทริกซ์ผลต่าง IX- YI

จำ	เข้าใจ	ประยุกต์ ใช้	วิเคราะห์	ประเมิน ค่า	สร้าง สรรค์
0.058	0.021	0.042	0.058	0.017	0.000
0.007	0.004	0.001	0.074	0.001	0.000
0.080	0.065	0.025	0.040	0.001	0.000
0.010	0.018	0.057	0.048	0.001	0.000
0.025	0.005	0.024	0.039	0.002	0.000
0.027	0.012	0.005	0.009	0.003	0.000
0.045	0.026	0.015	0.015	0.000	0.000

alignment index = 0.560

ภาพ 20 เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนโดยภาพรวม

เมื่อนำเมทริกซ์สัดส่วน X เมทริกซ์สัดส่วน Y ของโรงเรียนโดยภาพรวม มาสร้างแผนที่เนื้อหา (content map) เพื่อให้เห็นภาพที่ชัดเจนขึ้น โดยมีการจำแนกตามสาระและระดับความซับซ้อนทางปัญญาของเนื้อหาข้อสอบที่นำมาเปรียบเทียบกัน พบว่า ข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียนโดยภาพรวม มีการวัดเนื้อหาใน 7 สาระ โดยข้อสอบในการประเมินระดับชาติเน้นการวัดระดับวิเคราะห์มากที่สุด และไม่มีข้อสอบที่วัดระดับจำ ซึ่งแตกต่างจากข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียนโดยภาพรวม ที่เน้นการวัดระดับเข้าใจและระดับจำ อย่างไรก็ตาม ทั้งข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียนโดยภาพรวม ไม่มีการวัดระดับสร้างสรรค์ แสดงดังภาพ 21



ภาพ 21 แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนโดยภาพรวม

(2) ผลการวิเคราะห์ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 1

ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 1 (โรงเรียน ONET สูง 1) มีค่าเท่ากับ 0.584 รายละเอียดของเมทริกซ์สัดส่วนของข้อสอบในการประเมินระดับชาติระดับชาติ

(เมทริกซ์ X) เมทริกซ์ในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 1 ระดับชาติ (เมทริกซ์ Y) และเมทริกซ์ผลต่างระหว่าง 2 เมทริกซ์ มีรายละเอียดดังภาพ 22

เมทริกซ์สัดส่วน X (ข้อสอบระดับชาติ)							เมทริกซ์สัดส่วน Y (ข้อสอบระดับชั้นเรียน ของโรงเรียน <sub>ONETสูง1</sub> )					
สาระ	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
1	0.000	0.080	0.060	0.100	0.020	0.000	0.032	0.096	0.004	0.041	0.002	0.000
2	0.000	0.020	0.000	0.080	0.000	0.000	0.000	0.036	0.003	0.004	0.001	0.000
3	0.000	0.020	0.060	0.060	0.000	0.000	0.080	0.111	0.037	0.042	0.003	0.000
4	0.000	0.060	0.100	0.060	0.000	0.000	0.004	0.076	0.037	0.010	0.002	0.000
5	0.000	0.080	0.020	0.060	0.000	0.000	0.020	0.074	0.050	0.049	0.002	0.000
6	0.000	0.020	0.000	0.020	0.000	0.000	0.024	0.045	0.002	0.016	0.002	0.000
7	0.000	0.040	0.020	0.020	0.000	0.000	0.022	0.066	0.005	0.005	0.000	0.000

#### เมทริกซ์ผลต่าง IX- YI

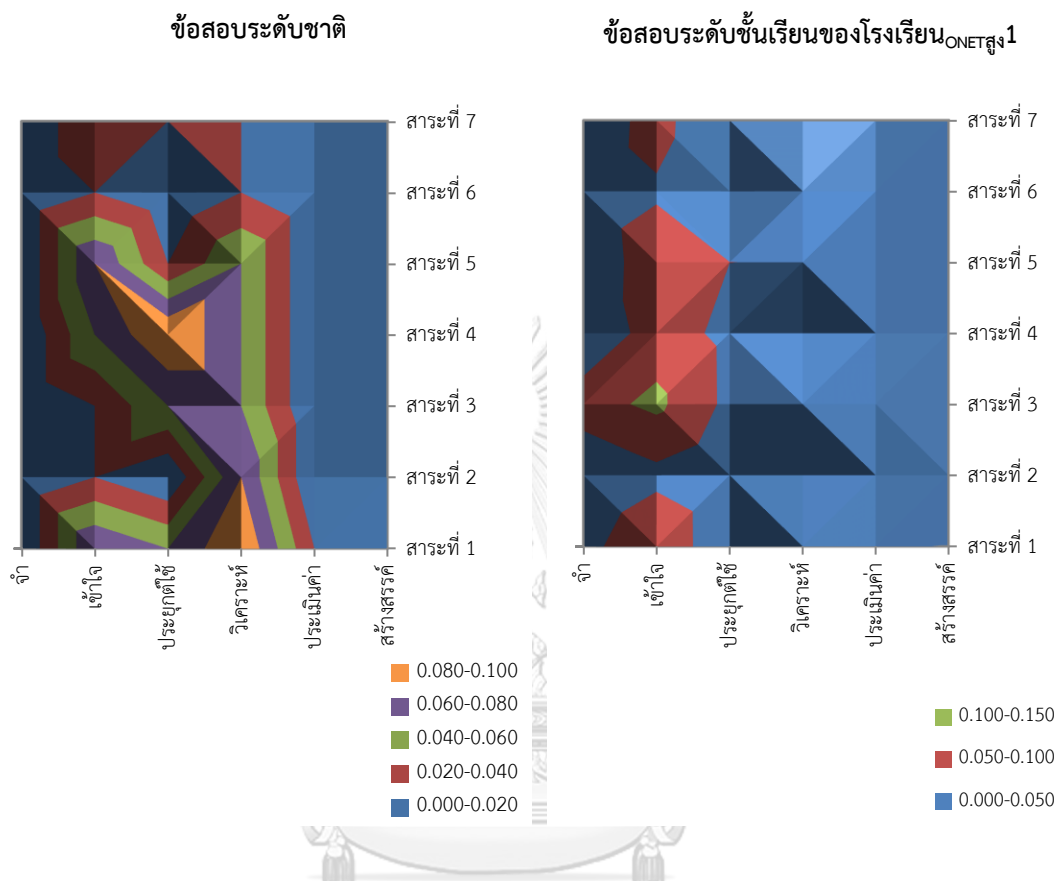
จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
0.032	0.016	0.056	0.059	0.018	0.000
0.000	0.016	0.003	0.076	0.001	0.000
0.080	0.091	0.023	0.018	0.003	0.000
0.004	0.016	0.063	0.050	0.002	0.000
0.020	0.006	0.030	0.011	0.002	0.000
0.024	0.025	0.002	0.004	0.002	0.000
0.022	0.026	0.015	0.015	0.000	0.000

alignment index = 0.584

ภาพ 22 เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 1

เมื่อนำเมทริกซ์สัดส่วน X เมทริกซ์สัดส่วน Y ของโรงเรียน<sub>ONETสูง1</sub> มาสร้างแผนที่เนื้อหา (content map) เพื่อให้เห็นภาพที่ชัดเจนขึ้น โดยมีการจำแนกตามสาระและระดับความซับซ้อนทางปัญญาของเนื้อหาข้อสอบที่นำมาเปรียบเทียบกัน พบว่า ข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONETสูง1</sub> มีการวัดเนื้อหาใน 7 สาระ โดยข้อสอบในการประเมินระดับชาติเน้นการวัดระดับวิเคราะห์มากที่สุด และไม่มีข้อสอบที่วัดระดับจำ ซึ่งแตกต่างจากข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONETสูง1</sub> ที่เน้นการวัดระดับเข้าใจ อย่างไรก็ตาม

ทั้งข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน ONET สูง 1 ไม่มี การวัดระดับสร้างสรรค์ แสดงดังภาพ 23



ภาพ 23 แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน ONET สูง 1

(3) ผลการวิเคราะห์ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 2

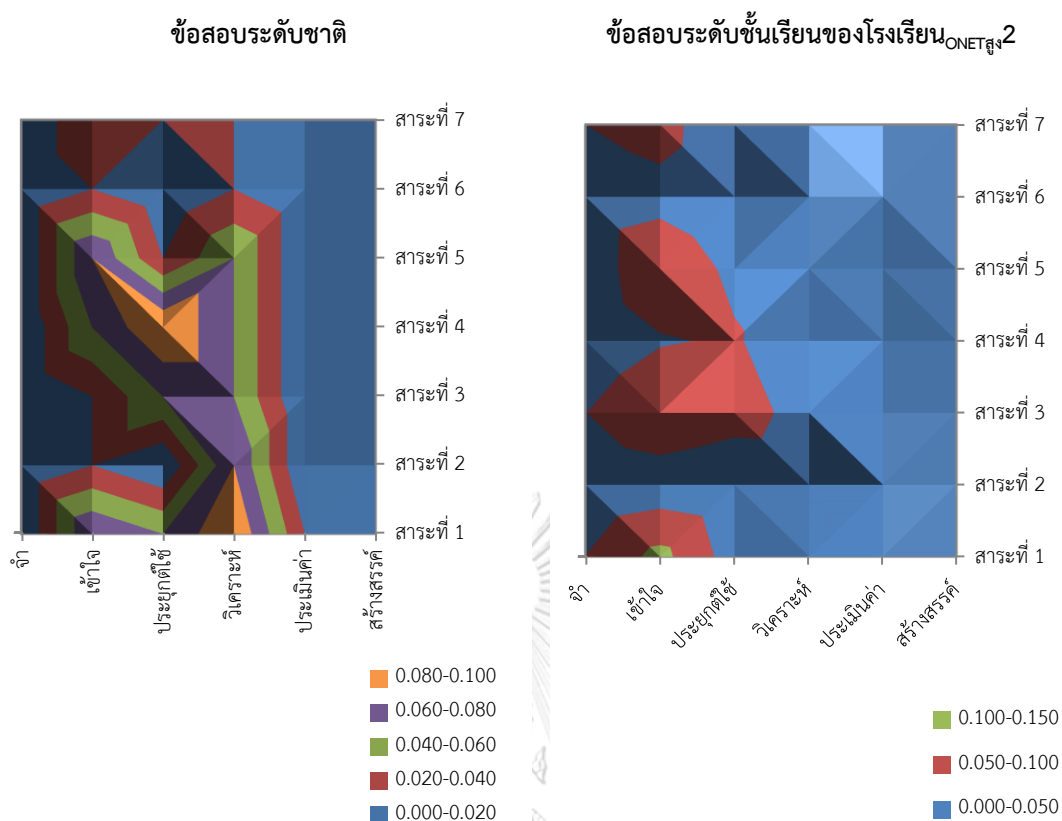
ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 2 (โรงเรียน ONET สูง 2) มีค่าเท่ากับ 0.588 รายละเอียดของเมทริกซ์สัดส่วนของข้อสอบในการประเมินระดับชาติ ระดับชาติ (เมทริกซ์ X) เมทริกซ์ในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 2 (เมทริกซ์ Y) และเมทริกซ์ผลต่างระหว่าง 2 เมทริกซ์ มีรายละเอียดดังภาพ 24

เมทริกซ์สัดส่วน X (ข้อสอบระดับชาติ)							เมทริกซ์สัดส่วน Y (ข้อสอบระดับชั้นเรียน ของโรงเรียน <sub>ONETสูง2</sub> )					
สาระ	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
1	0.000	0.080	0.060	0.100	0.020	0.000	0.051	0.116	0.025	0.037	0.007	0.000
2	0.000	0.020	0.000	0.080	0.000	0.000	0.000	0.017	0.000	0.005	0.005	0.000
3	0.000	0.020	0.060	0.060	0.000	0.000	0.049	0.097	0.076	0.027	0.002	0.000
4	0.000	0.060	0.100	0.060	0.000	0.000	0.009	0.045	0.056	0.008	0.000	0.000
5	0.000	0.080	0.020	0.060	0.000	0.000	0.020	0.089	0.038	0.012	0.002	0.000
6	0.000	0.020	0.000	0.020	0.000	0.000	0.018	0.033	0.002	0.012	0.006	0.000
7	0.000	0.040	0.020	0.020	0.000	0.000	0.050	0.070	0.007	0.009	0.000	0.000

เมทริกซ์ผลต่าง IX- YI						
จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	
0.051	0.036	0.035	0.063	0.013	0.000	
0.000	0.003	0.000	0.075	0.005	0.000	
0.049	0.077	0.016	0.033	0.002	0.000	
0.009	0.015	0.044	0.052	0.000	0.000	
0.020	0.009	0.018	0.048	0.002	0.000	
0.018	0.013	0.002	0.008	0.006	0.000	
0.050	0.030	0.013	0.011	0.000	0.000	alignment index = 0.588

ภาพ 24 เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง 2

เมื่อนำเมทริกซ์สัดส่วน X เมทริกซ์สัดส่วน Y ของโรงเรียน<sub>ONETสูง2</sub> มาสร้างแผนที่เนื้อหา (content map) เพื่อให้เห็นภาพที่ชัดเจนขึ้น โดยมีการจำแนกตามสาระและระดับความซับซ้อนทางปัญญาของเนื้อหาข้อสอบที่นำมาเปรียบเทียบกัน พบว่า ข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONETสูง2</sub> มีการวัดเนื้อหาใน 7 สาระ โดยข้อสอบในการประเมินระดับชาติเน้นการวัดระดับวิเคราะห์มากที่สุด และไม่มีข้อสอบที่วัดระดับจำ ซึ่งแตกต่างจากข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONETสูง2</sub> ที่เน้นการวัดระดับเข้าใจและระดับประยุกต์ใช้ อย่างไรก็ตาม ทั้งข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONETสูง2</sub> ไม่มีการวัดระดับสร้างสรรค์ แสดงดังภาพ 25



ภาพ 25 แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน<sub>ONETสูง2</sub>

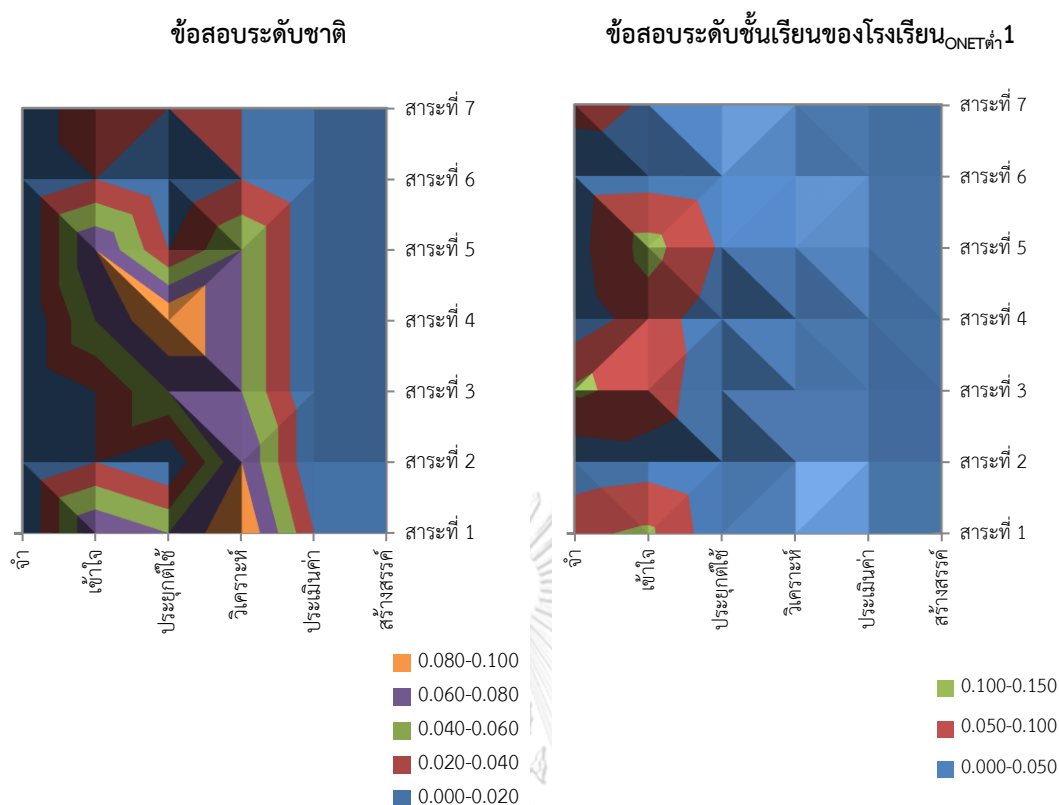
(4) ผลการวิเคราะห์ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 1 ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 1 (โรงเรียน<sub>ONETต่ำ1</sub>) มีค่าเท่ากับ 0.436 รายละเอียดของเมทริกซ์สัดส่วนของข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) เมทริกซ์ในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 1 (เมทริกซ์ Y) และเมทริกซ์ผลต่างระหว่าง 2 เมทริกซ์ มีรายละเอียดดังภาพ 26

เมทริกซ์สัดส่วน X (ข้อสอบระดับชาติ)							เมทริกซ์สัดส่วน Y (ข้อสอบระดับชั้นเรียน ของโรงเรียน <sub>ONET</sub> ต่ำ 1)						
สาระ	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	
1	0.000	0.080	0.060	0.100	0.020	0.000	0.089	0.110	0.013	0.015	0.003	0.000	
2	0.000	0.020	0.000	0.080	0.000	0.000	0.016	0.027	0.000	0.004	0.000	0.000	
3	0.000	0.020	0.060	0.060	0.000	0.000	0.108	0.082	0.009	0.004	0.000	0.000	
4	0.000	0.060	0.100	0.060	0.000	0.000	0.020	0.074	0.020	0.006	0.001	0.000	
5	0.000	0.080	0.020	0.060	0.000	0.000	0.032	0.119	0.043	0.012	0.002	0.000	
6	0.000	0.020	0.000	0.020	0.000	0.000	0.025	0.029	0.016	0.003	0.001	0.000	
7	0.000	0.040	0.020	0.020	0.000	0.000	0.062	0.047	0.005	0.003	0.000	0.000	

เมทริกซ์ผลต่าง IX- YI						
จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	
0.089	0.030	0.047	0.085	0.017	0.000	
0.016	0.007	0.000	0.076	0.000	0.000	
0.108	0.062	0.051	0.056	0.000	0.000	
0.020	0.014	0.080	0.054	0.001	0.000	
0.032	0.039	0.023	0.048	0.002	0.000	
0.025	0.009	0.016	0.017	0.001	0.000	
0.062	0.007	0.015	0.017	0.000	0.000	alignment index = 0.436

ภาพ 26 เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 1

เมื่อนำเมทริกซ์สัดส่วน X เมทริกซ์สัดส่วน Y ของโรงเรียน<sub>ONET</sub>ต่ำ 1 มาสร้างแผนที่เนื้อหา (content map) เพื่อให้เห็นภาพที่ชัดเจนขึ้น โดยมีการจำแนกตามสาระและระดับความซับซ้อนทางปัญญาของเนื้อหาข้อสอบที่นำมาเปรียบเทียบกัน พบว่า ข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONET</sub>ต่ำ 1 มีการวัดเนื้อหาใน 7 สาระ โดยข้อสอบในการประเมินระดับชาตินั้นการวัดระดับวิเคราะห์มากที่สุด และไม่มีข้อสอบที่วัดระดับจำ ซึ่งแตกต่างจากข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONET</sub>ต่ำ 1 ที่เน้นการวัดระดับเข้าใจและระดับจำ อย่างไรก็ตาม ทั้งข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONET</sub>ต่ำ 1 ไม่มีการวัดระดับสร้างสรรค์ แสดงดังภาพ 27



ภาพ 27 แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน ONET ต่ำ 1

(5) ผลการวิเคราะห์ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 2

ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 2 (โรงเรียน ONET ต่ำ 2) มีค่าเท่ากับ 0.565 รายละเอียดของเมทริกซ์สัดส่วนของข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) เมทริกซ์ในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 2 (เมทริกซ์ Y) และเมทริกซ์ผลต่างระหว่าง 2 เมทริกซ์ มีรายละเอียดดังภาพ 28

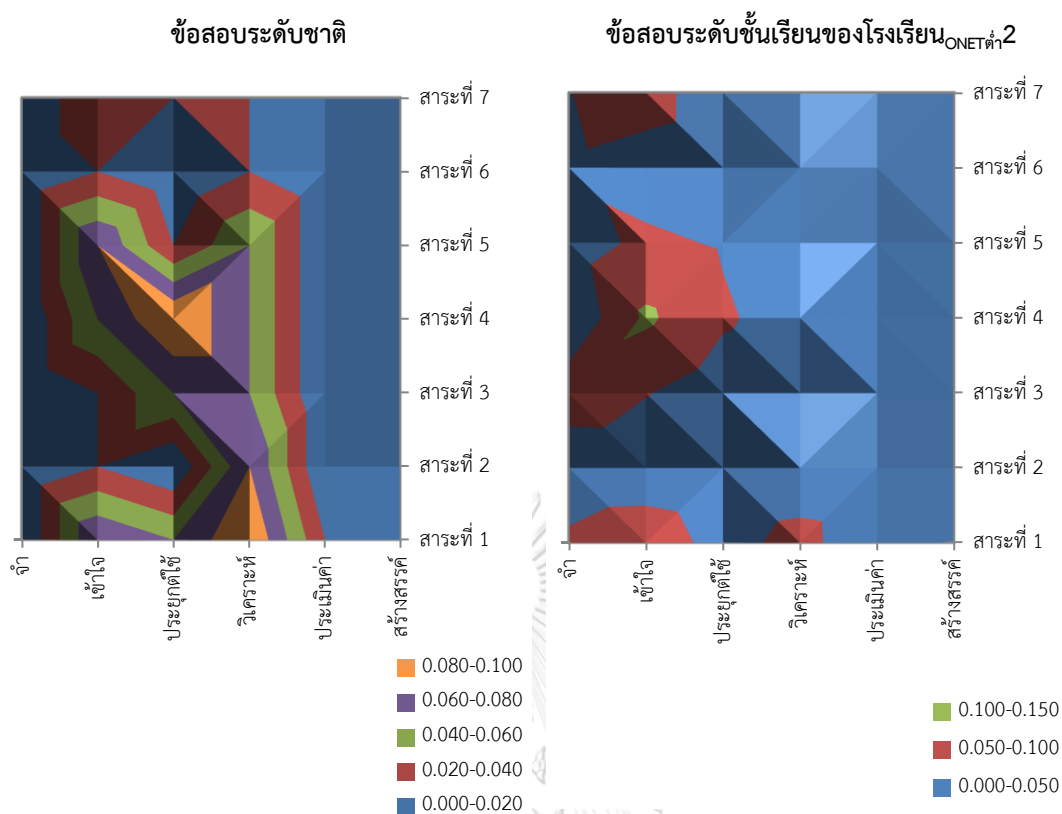


เมทริกซ์สัดส่วน X (ข้อสอบระดับชาติ)							เมทริกซ์สัดส่วน Y (ข้อสอบระดับชั้นเรียน ของโรงเรียน <sub>ONET</sub> ต่ำ 2)					
สาระ	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
1	0.000	0.080	0.060	0.100	0.020	0.000	0.061	0.085	0.028	0.070	0.002	0.000
2	0.000	0.020	0.000	0.080	0.000	0.000	0.012	0.015	0.001	0.011	0.000	0.000
3	0.000	0.020	0.060	0.060	0.000	0.000	0.080	0.053	0.024	0.005	0.000	0.000
4	0.000	0.060	0.100	0.060	0.000	0.000	0.007	0.108	0.057	0.023	0.000	0.000
5	0.000	0.080	0.020	0.060	0.000	0.000	0.026	0.061	0.045	0.010	0.001	0.000
6	0.000	0.020	0.000	0.020	0.000	0.000	0.040	0.020	0.002	0.011	0.003	0.000
7	0.000	0.040	0.020	0.020	0.000	0.000	0.048	0.080	0.003	0.005	0.001	0.000

เมทริกซ์ผลต่าง IX- YI						
จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	
0.061	0.005	0.032	0.030	0.018	0.000	
0.012	0.005	0.001	0.069	0.000	0.000	
0.080	0.033	0.036	0.055	0.000	0.000	
0.007	0.048	0.043	0.037	0.000	0.000	
0.026	0.019	0.025	0.050	0.001	0.000	
0.040	0.000	0.002	0.009	0.003	0.000	
0.048	0.040	0.017	0.015	0.001	0.000	alignment index = 0.565

ภาพ 28 เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ 2

เมื่อนำเมทริกซ์สัดส่วน X เมทริกซ์สัดส่วน Y ของโรงเรียน<sub>ONET</sub>ต่ำ 2 มาสร้างแผนที่เนื้อหา (content map) เพื่อให้เห็นภาพที่ชัดเจนขึ้น โดยมีการจำแนกตามสาระและระดับความซับซ้อนทางปัญญาของเนื้อหาข้อสอบที่นำมาเปรียบเทียบกัน พบว่า ข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONET</sub>ต่ำ 1 มีการวัดเนื้อหาใน 7 สาระ โดยข้อสอบในการประเมินระดับชาติเน้นการวัดระดับวิเคราะห์มากที่สุด และไม่มีข้อสอบที่วัดระดับจำ ซึ่งแตกต่างจากข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONET</sub>ต่ำ 1 ที่เน้นการวัดระดับเข้าใจและระดับจำ อย่างไรก็ตาม ทั้งข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียน<sub>ONET</sub>ต่ำ 2 ไม่มีการวัดระดับสร้างสรรค์ แสดงดังภาพ 29



ภาพ 29 แผนที่เนื้อหาข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียน ONET ต่ำ 2

4.3.2 ผลการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน

ผลการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน แบ่งการนำเสนอเป็น 2 ส่วนคือ 1) การเปรียบเทียบค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน และ 2) การเปรียบเทียบสัดส่วนระดับความซับซ้อนทางปัญญาระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน รายละเอียดมีดังนี้

### 1) การเปรียบเทียบค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนทั้ง 4 โรงเรียน พบว่า โดยภาพรวมมีค่าดัชนีความสอดคล้องเท่ากับ 0.560 เมื่อพิจารณาดัชนีความสอดคล้อง จำแนกตามผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียน พบว่า ค่าดัชนีความสอดคล้องของทั้ง 4 โรงเรียน มีค่าใกล้เคียงกัน โดยโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง มีค่าดัชนีความสอดคล้องสูงกว่าโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำเล็กน้อย โดยโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง มีค่าดัชนีความสอดคล้องเท่ากับ 0.584 และ 0.588 ในขณะที่โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ มีค่าดัชนีความสอดคล้องเท่ากับ 0.436 และ 0.555 รายละเอียดดังตาราง 31

ตาราง 31 ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

เมทริกซ์ X	เมทริกซ์ Y	ดัชนีความสอดคล้อง
ข้อสอบระดับชาติ	ข้อสอบระดับชั้นเรียน โรงเรียน ONET สูง 1	0.584
ข้อสอบระดับชาติ	ข้อสอบระดับชั้นเรียน โรงเรียน ONET สูง 2	0.588
ข้อสอบระดับชาติ	ข้อสอบระดับชั้นเรียน โรงเรียน ONET ต่ำ 1	0.436
ข้อสอบระดับชาติ	ข้อสอบระดับชั้นเรียน โรงเรียน ONET ต่ำ 2	0.565
ข้อสอบระดับชาติ	ข้อสอบระดับชั้นเรียน โดยภาพรวม	0.560

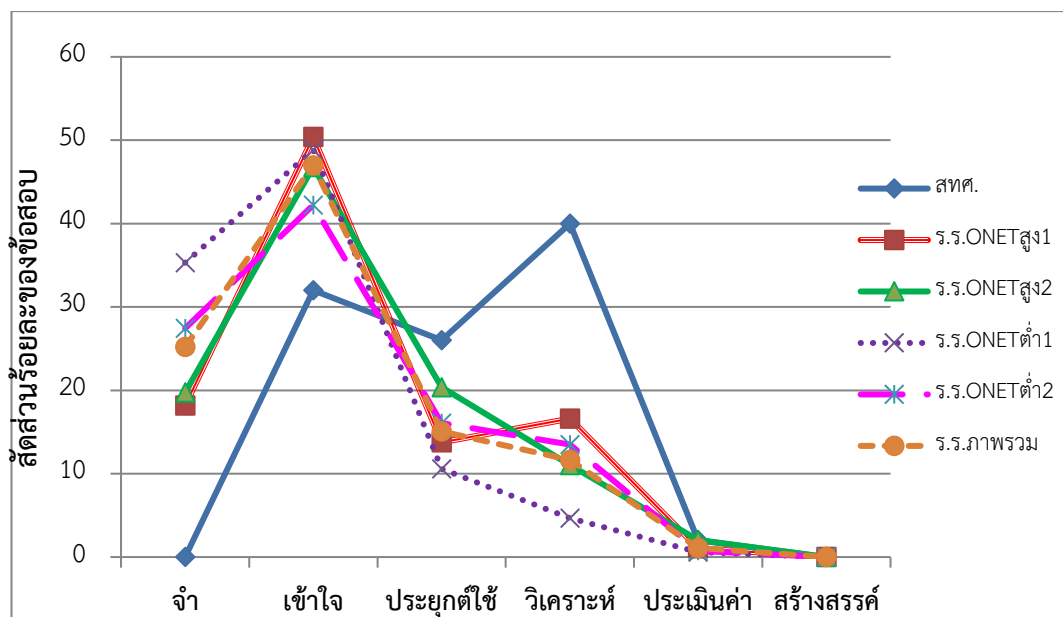
### 2) การเปรียบเทียบร้อยละสัดส่วนระดับความซับซ้อนทางปัญญาระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

ผลการเปรียบเทียบร้อยละสัดส่วนระดับความซับซ้อนทางปัญญาระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน พบว่า ข้อสอบในการประเมินระดับชาติมีร้อยละสัดส่วนของข้อสอบที่วัดความซับซ้อนทางปัญญาระดับวิเคราะห์มากที่สุด รองลงมาคือ ระดับเข้าใจ ระดับประยุกต์ใช้ และระดับประเมินค่า ตามลำดับ โดยไม่มีข้อสอบที่วัดระดับจำและระดับสร้างสรรค์ ในขณะที่ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนทั้ง 4 โรงเรียน โดยภาพรวมพบว่า มีร้อยละสัดส่วนของข้อสอบที่วัดระดับความเข้าใจมากที่สุด ร้อยละ 46.99 รองลงมาคือ ระดับจำ ร้อยละ 25.23 ระดับประยุกต์ใช้ ร้อยละ 15.04 ระดับวิเคราะห์ ร้อยละ 11.65 ระดับประเมินค่า ร้อยละ 1.09 และระดับสร้างสรรค์ ร้อยละ 0.00 ตามลำดับ รายละเอียดดังตาราง 32

เมื่อพิจารณาจำแนกตามระดับผลสัมฤทธิ์ทางการเรียนของโรงเรียน ในตาราง 32 และภาพ 25 พบว่า ทุกโรงเรียนมีร้อยละสัดส่วนระดับความซับซ้อนทางปัญญาของข้อสอบที่วัดในระดับเข้าใจมากที่สุด โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง (โรงเรียน<sub>ONETสูง1</sub> และโรงเรียน<sub>ONEสูง2</sub>) มีร้อยละสัดส่วนระดับความซับซ้อนทางปัญญาของข้อสอบที่วัดในระดับจำ ระดับประยุกต์ใช้ และระดับวิเคราะห์ที่ใกล้เคียงกัน คืออยู่ระหว่างร้อยละ 11.64 – 20.36 ในขณะที่โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ (โรงเรียน<sub>ONETต่ำ1</sub> และโรงเรียน<sub>ONEต่ำ2</sub>) จะมีร้อยละสัดส่วนระดับความซับซ้อนทางปัญญาในระดับจำ เท่ากับ 35.32 และ 27.45 ซึ่งค่อนข้างแตกต่างจากระดับประยุกต์ใช้และระดับวิเคราะห์ โดยมีร้อยละสัดส่วนระดับความซับซ้อนทางปัญญาในระดับประยุกต์ใช้ เท่ากับ 10.56 และ 16.06 ตามลำดับ และระดับวิเคราะห์ ร้อยละ 4.68 และ 0.75 ตามลำดับ นอกจากนี้ ทุกโรงเรียนมีร้อยละสัดส่วนระดับความซับซ้อนทางปัญญาของข้อสอบที่วัดในระดับประเมินค่าเล็กน้อยซึ่งสอดคล้องกับข้อสอบระดับชาติ โดยมีค่าร้อยละสัดส่วนเท่ากับ 1.06 2.06 0.06 และ 0.75 ตามลำดับ และไม่มีโรงเรียนที่มีข้อสอบที่วัดในระดับสร้างสรรค์ เช่นเดียวกับข้อสอบระดับชาติ

ตาราง 32 สรุปร้อยละของสัดส่วนระดับความซับซ้อนทางปัญญา ของข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน

ข้อสอบ	ร้อยละของสัดส่วนระดับความซับซ้อนทางปัญญา					
	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
ข้อสอบระดับชาติ	0.00	32.00	26.00	40.00	2.00	0.00
ข้อสอบโรงเรียน <sub>ONETสูง1</sub>	18.18	50.39	13.75	16.62	1.06	0.00
ข้อสอบโรงเรียน <sub>ONETสูง2</sub>	19.75	46.80	20.36	11.02	2.06	0.00
ข้อสอบโรงเรียน <sub>ONETต่ำ1</sub>	35.32	48.83	10.56	4.68	0.61	0.00
ข้อสอบโรงเรียน <sub>ONETต่ำ2</sub>	27.45	42.23	16.06	13.51	0.75	0.00
ข้อสอบโรงเรียนโดยภาพรวม	25.21	46.99	15.07	11.64	1.09	0.00



ภาพ 30 สรุปร้อยละของสัดส่วนระดับความซับซ้อนทางปัญญา ของข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน

ตอนที่ 5 ผลการประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปร่างอิงของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน

ผู้วิจัยนำข้อมูลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ของข้อสอบจำนวน 40 ข้อ (ชุดย่อย A) ที่ถูกประเมินด้วยผู้ประเมินจำนวน 20 คน มาวิเคราะห์เพื่อประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปร่างอิงสำหรับการตัดสินใจ (G-coefficient) ตามทฤษฎีการสรุปร่างอิงความน่าเชื่อถือของผลการวัด (G-theory) ผลการวิเคราะห์มีรายละเอียดดังนี้

5.1 ผลการประมาณค่าความแปรปรวนขององค์ประกอบที่มีผลต่อสัมประสิทธิ์การสรุปร่างอิงสำหรับการตัดสินใจ (G-coefficient) ของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน

ผลการวิเคราะห์ความแปรปรวนขององค์ประกอบที่มีผลต่อสัมประสิทธิ์การสรุปร่างอิงสำหรับการตัดสินใจ (G-coefficient) ของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ซึ่งได้ออกแบบการวัดเป็นแบบ one-facet design โดยมีรูปแบบการออกแบบการประเมินแบบผู้ประเมินแต่ละคนประเมินข้อสอบทุก

ข้อ (i x r design) และผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อ (r : i design) เมื่อ i แทน ข้อสอบ และ r แทน ผู้ประเมิน วิเคราะห์ความแปรปรวนแบบสองทาง (two-way ANOVA) ของแต่ละรูปแบบการออกแบบการประเมิน ทั้งการประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องในแนวเดียวกันๆ ด้วยมาตรประมาณค่า 5 ผลการวิเคราะห์ พบว่า ทั้งการประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ มีขนาดความแปรปรวนขององค์ประกอบเป็นไปในทิศทางเดียวกัน กล่าวคือ ความแปรปรวนจากแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนทั้งหมด มีขนาดองค์ประกอบความแปรปรวนเรียงลำดับจากมากไปหาน้อย ดังนี้ ความแปรปรวนของคะแนนความสอดคล้องในแนวเดียวกันของข้อสอบ ( $\sigma_i^2$ ) มีขนาดใหญ่ที่สุด รองลงมาคือ ความแปรปรวนของความคลาดเคลื่อนหรือส่วนเหลือ ( $\sigma_{ir,e}^2$ ) ซึ่งเป็นความแปรปรวนอันเนื่องมาจากความแตกต่างของความสอดคล้องในแนวเดียวกันของข้อสอบแต่ละข้อที่ได้รับการประเมินด้วยผู้ประเมินที่แตกต่างกัน และ/หรือเป็นแหล่งความแปรปรวนที่ไม่สามารถวัดได้ และความแปรปรวนอันเกิดจากผู้ประเมิน ( $\sigma_r^2$ ) รายละเอียดการศึกษารูปอ้างอิง (G-study) แสดงดังตาราง 33 และเมื่อพิจารณารายละเอียดจำแนกตามการประเมิน 2 ส่วน คือ การประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ เป็นดังนี้

การประเมินระดับความซับซ้อนทางปัญญา พบว่า ความแปรปรวนรวมทั้งหมด เท่ากับ 0.881 โดยความแปรปรวนของคะแนนความสอดคล้องในแนวเดียวกันของข้อสอบ ( $\sigma_i^2$ ) มีค่าสูงสุด เท่ากับ 0.482 คิดเป็นร้อยละ 54.72 ของความแปรปรวนทั้งหมด รองลงมาคือ ความแปรปรวนของความคลาดเคลื่อนหรือส่วนเหลือ ( $\sigma_{ir,e}^2$ ) ซึ่งเป็นความแปรปรวนอันเนื่องมาจากความแตกต่างของความสอดคล้องในแนวเดียวกันของข้อสอบแต่ละข้อที่ได้รับการประเมินด้วยผู้ประเมินที่แตกต่างกัน และ/หรือเป็นแหล่งความแปรปรวนที่ไม่สามารถวัดได้ มีค่าเท่ากับ 0.374 คิดเป็นร้อยละ 42.44 ของความแปรปรวนทั้งหมด และความแปรปรวนอันเกิดจากผู้ประเมิน ( $\sigma_r^2$ ) มีค่าเท่ากับ 0.025 คิดเป็นร้อยละ 2.84 ของความแปรปรวนทั้งหมด

การประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ พบว่า ความแปรปรวนรวมทั้งหมด เท่ากับ 0.487 โดยความแปรปรวนของคะแนนความสอดคล้องในแนวเดียวกันของข้อสอบ ( $\sigma_i^2$ ) มีค่าสูงสุด เท่ากับ 0.249 คิดเป็นร้อยละ 51.12 ของความแปรปรวนทั้งหมด รองลงมาคือ ความแปรปรวนของความคลาดเคลื่อนหรือส่วนเหลือ ( $\sigma_{ir,e}^2$ ) ซึ่งเป็นความแปรปรวนอันเนื่องมาจากความแตกต่างของความสอดคล้องในแนวเดียวกันของข้อสอบแต่ละข้อที่ได้รับการประเมินด้วยผู้ประเมินที่แตกต่างกัน และ/หรือเป็นแหล่งความแปรปรวนที่ไม่สามารถวัดได้ มีค่าเท่ากับ 0.177 คิดเป็นร้อยละ 36.36 ของความแปรปรวนทั้งหมด และ

ความแปรปรวนอันเกิดจากผู้ประเมิน ( $\sigma_r^2$ ) มีค่าเท่ากับ 0.061 คิดเป็นร้อยละ 12.52 ของความแปรปรวนทั้งหมด

**ตาราง 33 G-Study สำหรับประมาณค่าองค์ประกอบความแปรปรวนของการวัดในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบฯ**

Source of variance	Sum squares	df	Mean squares	Estimated variance component	% of total variance
<b>การประเมินระดับความซับซ้อนทางปัญญา</b>					
Items (i)	390.639	39	10.016	0.482	54.72
Rater (r)	26.364	19	1.388	0.025	2.84
Residual (ir,e)	276.986	741	0.374	0.374	42.44
<b>Total</b>	<b>693.989</b>	<b>799</b>		<b>0.881</b>	<b>100.00</b>
<b>การประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรฐานค่า 5 ระดับ</b>					
Items (i)	200.899	39	5.151	0.249	51.12
Rater (r)	49.524	19	2.607	0.061	12.52
Residual (ir,e)	131.226	741	0.177	0.177	36.36
<b>Total</b>	<b>381.649</b>	<b>799</b>		<b>0.487</b>	<b>100.00</b>

## 5.2 ผลการศึกษาเพื่อการตัดสินใจการสรุปอ้างอิง (D-study) ของการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน

การศึกษาเพื่อการตัดสินใจการสรุปอ้างอิง (D-study) ทำโดยการนำข้อมูลจาก G-study มาสรุปอ้างอิงคุณภาพของการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมิน 1 – 20 คน และรูปแบบการออกแบบการประเมินแบบ  $i \times r$  design และแบบ  $r : i$  design แสดงดังตาราง 34 และ 35 พบว่า ค่าความเที่ยงที่พิจารณาจากสัมประสิทธิ์การสรุปอ้างอิงสำหรับตัดสินใจของผลการประเมินความสอดคล้องในแนวเดียวกันๆ มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ประเมินเพิ่มขึ้นในทุกรูปแบบการออกแบบการประเมิน ทั้งการประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรฐานค่า 5 ระดับ และเมื่อพิจารณาผลการวิเคราะห์จำแนกตามการประเมิน 2 ส่วน คือ การประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรฐานค่า 5 พบว่า ทั้งรูปแบบการออกแบบการประเมินแบบ  $i \times r$  design และแบบ  $r : i$  design เมื่อจำนวนผู้ประเมินเท่ากัน จะมีค่าเที่ยงที่ใกล้เคียงกัน รายละเอียดผลการวิเคราะห์มีดังนี้

เมื่อพิจารณาการประเมินระดับความซับซ้อนทางปัญญา สำหรับรูปแบบการออกแบบการประเมินแบบ  $i \times r$  design ที่ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ ในตาราง 34 พบว่าผู้ประเมินจำนวน 1 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับปานกลาง โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (relative coefficient:  $\rho^2_{Rel}$ ) เท่ากับ 0.56 และมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (absolute coefficient:  $\rho^2_{Abs}$ ) เท่ากับ 0.55 ผู้ประเมินจำนวน 2 - 6 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับสูง โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho^2_{Rel}$ ) อยู่ระหว่าง 0.72 - 0.89 และมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) เท่ากับ 0.71 - 0.89 และผู้ประเมินจำนวนตั้งแต่ 7 - 20 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับสูงมาก โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho^2_{Rel}$ ) และมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) เท่ากัน อยู่ระหว่าง 0.91 - 0.96

เมื่อพิจารณาการประเมินระดับความซับซ้อนทางปัญญา สำหรับรูปแบบการออกแบบการประเมินแบบ  $r : i$  design ที่ผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อ พบว่า ให้ค่าความเที่ยงใกล้เคียงกับรูปแบบการออกแบบการประเมินแบบ  $i \times r$  design เมื่อจำนวนผู้ประเมินเท่ากัน โดยผู้ประเมินจำนวน 1 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับปานกลาง โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho^2_{Rel}$ ) และค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) เท่ากัน เท่ากับ 0.55 ผู้ประเมินจำนวน 2 - 7 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับสูง โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho^2_{Rel}$ ) และค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) เท่ากัน อยู่ระหว่าง 0.71 - 0.89 ตามลำดับ และผู้ประเมินจำนวนตั้งแต่ 8 - 20 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับสูงมาก โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho^2_{Rel}$ ) และค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) เท่ากัน อยู่ระหว่าง 0.91 - 0.96

เมื่อพิจารณาจำนวนผู้ประเมินที่ทำให้ผลการประเมินมีความน่าเชื่อถืออยู่ในระดับที่ยอมรับได้ โดยการพิจารณาค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) ที่มีค่าตั้งแต่ 0.70 ขึ้นไป (Nunnally, 1978) ในการประเมินระดับความซับซ้อนทางปัญญา พบว่า ผู้ประเมินจำนวนตั้งแต่ 2 คนขึ้นไป สามารถให้ผลการประเมินที่มีความน่าเชื่อถือในระดับที่ยอมรับได้ ในทุกรูปแบบการออกแบบการประเมิน



ตาราง 34 D-Study ของการประเมินระดับความซับซ้อนทางปัญญา

EFFEC <sup>c</sup>	$n_r =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
<b>i x r design</b>																					
item (i)	$\sigma_{i=}^2 =$	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
rater (r)	$\sigma_{r=}^2 =$	0.03	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
residual (i,r,e)	$\sigma_{i,r,e=}^2 =$	0.37	0.19	0.12	0.09	0.07	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
	$\sigma_{Rel=}^2 =$	0.37	0.19	0.12	0.09	0.07	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
	$\sigma_{Abs=}^2 =$	0.40	0.20	0.13	0.10	0.08	0.06	0.05	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
	$\rho_{Rel}^2 =$	0.56	0.72	0.80	0.84	0.87	0.89	0.91	0.91	0.92	0.92	0.94	0.94	0.94	0.94	0.96	0.96	0.96	0.96	0.96	0.96
	$\rho_{Abs}^2 =$	0.55	0.71	0.79	0.83	0.86	0.89	0.91	0.91	0.92	0.92	0.94	0.94	0.94	0.94	0.96	0.96	0.96	0.96	0.96	0.96
<b>r : i design</b>																					
item (i)	$\sigma_{i=}^2 =$	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
rater :	$\sigma_{i,r,e=}^2 =$	0.40	0.20	0.13	0.10	0.08	0.07	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02
item (r : i)																					
	$\sigma_{Rel=}^2 =$	0.40	0.20	0.13	0.10	0.08	0.07	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
	$\sigma_{Abs=}^2 =$	0.40	0.20	0.13	0.10	0.08	0.07	0.06	0.05	0.04	0.04	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02
	$\rho_{Rel}^2 =$	0.55	0.71	0.79	0.83	0.86	0.87	0.89	0.91	0.92	0.92	0.92	0.94	0.94	0.94	0.94	0.96	0.96	0.96	0.96	0.96
	$\rho_{Abs}^2 =$	0.55	0.71	0.79	0.83	0.86	0.87	0.89	0.91	0.92	0.92	0.92	0.94	0.94	0.94	0.94	0.96	0.96	0.96	0.96	0.96

เมื่อพิจารณาการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ สำหรับรูปแบบการออกแบบการประเมินแบบ i x r design ที่ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ รายละเอียดดังตาราง 35 พบว่า ผู้ประเมินจำนวน 1 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับปานกลาง โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho_{Rel}^2$ ) เท่ากับ 0.58 และมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho_{Abs}^2$ ) เท่ากับ 0.51 ผู้ประเมินจำนวน 2 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับปานกลาง - สูง โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho_{Rel}^2$ ) เท่ากับ 0.74 และมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho_{Abs}^2$ ) เท่ากับ 0.68 ผู้ประเมินจำนวน 3 - 7 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับสูง โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho_{Rel}^2$ ) เท่ากับ อยู่ระหว่าง 0.81 - 0.89 และมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho_{Abs}^2$ ) เท่ากับ 0.76 - 0.86 ผู้ประเมินจำนวน 8 - 11 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับสูง - สูงมาก โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho_{Rel}^2$ ) 0.93 และมีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho_{Abs}^2$ ) เท่ากับ 0.89 ตามลำดับ และผู้ประเมินจำนวนตั้งแต่ 12 - 20 คน มี

ค่าความเที่ยงในการประเมินอยู่ในระดับสูงมาก โดยมีค่าสัมประสิทธิ์การสรุปร่างสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho^2_{Rel}$ ) เท่ากับ 0.96 และมีค่าสัมประสิทธิ์การสรุปร่างสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) อยู่ระหว่าง 0.93 – 0.96

เมื่อพิจารณาการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ สำหรับรูปแบบการออกแบบการประเมินแบบ r : i design ที่ผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อ พบว่า ผู้ประเมินจำนวน 1 - 2 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับปานกลาง โดยมีค่าสัมประสิทธิ์การสรุปร่างสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho^2_{Rel}$ ) และค่าสัมประสิทธิ์การสรุปร่างสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) เท่ากัน เท่ากับ 0.51 และ 0.68 ตามลำดับ ผู้ประเมินจำนวน 3 - 8 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับสูง โดยมีค่าสัมประสิทธิ์การสรุปร่างสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho^2_{Rel}$ ) และมีค่าสัมประสิทธิ์การสรุปร่างสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) เท่ากัน อยู่ระหว่าง 0.76 – 0.89 และผู้ประเมินจำนวนตั้งแต่ 10 – 20 คน มีค่าความเที่ยงในการประเมินอยู่ในระดับสูงมาก โดยมีค่าสัมประสิทธิ์การสรุปร่างสำหรับการตัดสินใจเชิงสัมพัทธ์ ( $\rho^2_{Rel}$ ) และมีค่าสัมประสิทธิ์การสรุปร่างสำหรับการตัดสินใจเชิงสัมบูรณ์ ( $\rho^2_{Abs}$ ) เท่ากัน อยู่ระหว่าง 0.93 – 0.96

เมื่อพิจารณาจำนวนผู้ประเมินที่ทำให้ผลการประเมินมีความน่าเชื่อถืออยู่ในระดับที่ยอมรับได้ โดยการพิจารณาค่าสัมประสิทธิ์การสรุปร่างสำหรับการตัดสินใจเชิงสัมบูรณ์ที่มีค่าตั้งแต่ 0.70 ขึ้นไป (Nunnally, 1978) ในการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ พบว่า ผู้ประเมินจำนวนตั้งแต่ 3 คนขึ้นไป สามารถให้ผลการประเมินที่มีความน่าเชื่อถือในระดับที่ยอมรับได้ ในทุกรูปแบบการออกแบบการประเมิน

ตาราง 35 D-Study ของการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วย  
 มาตรฐานค่า 5 ระดับ

EFFEC	$n_r =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
<b>i x r design</b>																					
item (i)	$\sigma_i^2 =$	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
rater (r)	$\sigma_r^2 =$	0.06	0.03	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
residual (ir,e)	$\sigma_{ir,e}^2 =$	0.18	0.09	0.06	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	$\sigma_{Rel}^2 =$	0.18	0.09	0.06	0.04	0.04	0.03	0.03	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	$\sigma_{Abs}^2 =$	0.24	0.12	0.08	0.06	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	$\rho_{Rel}^2 =$	0.58	0.74	0.81	0.86	0.86	0.89	0.89	0.93	0.93	0.93	0.93	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	$\rho_{Abs}^2 =$	0.51	0.68	0.76	0.81	0.83	0.86	0.86	0.89	0.89	0.89	0.89	0.93	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
<b>r : i design</b>																					
item (i)	$\sigma_i^2 =$	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
rater :	$\sigma_{i,r,e}^2 =$	0.24	0.12	0.08	0.06	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01
item (r : i)																					
	$\sigma_{Rel}^2 =$	0.24	0.12	0.08	0.06	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01
	$\sigma_{Abs}^2 =$	0.24	0.12	0.08	0.06	0.05	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01
	$\rho_{Rel}^2 =$	0.51	0.68	0.76	0.81	0.83	0.86	0.89	0.89	0.89	0.93	0.93	0.93	0.93	0.93	0.93	0.96	0.96	0.96	0.96	0.96
	$\rho_{Abs}^2 =$	0.51	0.68	0.76	0.81	0.83	0.86	0.89	0.89	0.89	0.93	0.93	0.93	0.93	0.93	0.93	0.96	0.96	0.96	0.96	0.96

ตอนที่ 6 ผลการตรวจสอบคุณภาพของคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกัน  
 ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

ผลการตรวจสอบคุณภาพของคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน โดยใช้ผลการประเมินจากผู้ทรงคุณวุฒิมีความรู้และประสบการณ์ในการทำงานวิจัยด้านการศึกษาความสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบทางการศึกษา และมีความรู้ความสามารถด้านการวัดและประเมินผลการศึกษา จำนวน 5 ท่าน เพื่อตรวจสอบด้านตรงตามเนื้อหา ด้วยการวิเคราะห์ค่าดัชนี IOC และตรวจสอบความเหมาะสมของคู่มือฯ ตามมาตรฐานการประเมิน 4 ด้าน ได้แก่ ด้านอรรถประโยชน์ (utility) ด้านความเป็นไปได้ (feasibility) ด้านความเหมาะสม (propriety) และด้านความถูกต้อง (accuracy) มีรายละเอียดดังนี้

### 6.1 ผลการตรวจสอบคุณภาพด้านตรงตามเนื้อหาของคู่มือแนวทางการดำเนินการศึกษา ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมิน ระดับชั้นเรียน

ผลการตรวจสอบคุณภาพด้านตรงตามเนื้อหา โดยการตรวจสอบความสอดคล้องระหว่างคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กับวัตถุประสงค์ของคู่มือฯ พบว่า มีค่าดัชนี IOC อยู่ระหว่าง 0.80 – 1.00 ซึ่งผ่านเกณฑ์ทุกข้อ รายละเอียดดังตาราง 36 นอกจากนี้มีผู้ทรงคุณวุฒิจำนวน 1 ท่าน ที่ให้ข้อเสนอแนะในเรื่องคำจำกัดความที่ควรกระชับและเน้นแต่คำที่สำคัญเท่านั้น อย่างไรก็ตามผู้วิจัยไม่ได้ปรับแก้ตามข้อเสนอแนะดังกล่าว เนื่องจากผู้ทรงคุณวุฒิท่านอื่นเห็นว่าเหมาะสมแล้ว นอกจากนี้ ผู้ทรงคุณวุฒิได้ให้ข้อเสนอแนะว่า ควรเพิ่มรายละเอียดในตอนที่ 2 แนวคิดและทฤษฎีที่เกี่ยวข้อง เรื่องแนวคิดเกี่ยวกับ MFRM และควรมีการยกตัวอย่างของค่าเฉลี่ยของข้อสอบที่คัดเลือกในการศึกษา ซึ่งผู้วิจัยได้เพิ่มเติมในประเด็นดังกล่าว และจัดทำเป็นคู่มือฉบับสมบูรณ์

**ตาราง 36 ผลการตรวจสอบคุณภาพด้านตรงตามเนื้อหาของคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน**

รายการ	IOC	ความหมาย	ข้อเสนอแนะ
1. หลักการและวัตถุประสงค์ของคู่มือ			
1.1 หลักการและแนวคิด	1.00	สอดคล้อง	
1.2 วัตถุประสงค์ของคู่มือ	1.00	สอดคล้อง	
1.3 ประโยชน์ที่คาดว่าจะได้รับจากการใช้คู่มือ	1.00	สอดคล้อง	
1.4 คำจำกัดความที่ใช้ในคู่มือ	0.80	สอดคล้อง	ควรกระชับความ โดยการเน้นแต่คำสำคัญของแต่ละคำ
2. แนวคิดและทฤษฎีที่เกี่ยวข้อง	1.00	สอดคล้อง	ควรเพิ่มรายละเอียดของแนวคิด 2.2 (เรื่อง MFRM)
3. ขั้นตอนในการศึกษาความสอดคล้องในแนวเดียวกันฯ			
ขั้นตอนที่ 1 ขั้นเตรียมการ	1.00	สอดคล้อง	
ขั้นตอนที่ 2 ขั้นดำเนินการ	1.00	สอดคล้อง	ควรมีตัวอย่างของค่าเฉลี่ยของข้อสอบที่คัดเลือก
ขั้นตอนที่ 3 ขั้นสรุปผล	1.00	สอดคล้อง	

## 6.2 ผลการตรวจสอบคุณภาพด้านความเหมาะสมของคู่มือแนวทางการดำเนินการศึกษา ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมิน ระดับชั้นเรียน ตามมาตรฐานการประเมิน

ผลการตรวจสอบคุณภาพด้านความเหมาะสมของคู่มือฯ ตามมาตรฐานการประเมิน 4 ด้าน โดยผู้ทรงคุณวุฒิจำนวน 5 ท่าน มีข้อรายการประเมินคุณภาพในการประเมินคู่มือฯ ประกอบด้วย 18 ข้อรายการ ที่ครอบคลุมประเด็นใน 4 ด้าน ได้แก่ ด้านอรรถประโยชน์ (utility) ด้านความเป็นไปได้ (feasibility) ด้านความเหมาะสม (property) และด้านความถูกต้อง (accuracy) พบว่า ในภาพรวม คู่มือฯ มีระดับคุณภาพในระดับมากที่สุด (M=4.69, SD=0.49) โดยมีผลการประเมินระดับคุณภาพในระดับมากที่สุดทุกด้าน ด้านที่มีผลคะแนนเฉลี่ยมากที่สุด ในระดับมากที่สุด คือ ด้านความเหมาะสม (property) (M=4.95, SD=0.22) รองลงมาคือ ด้านความถูกต้อง (accuracy) (M=4.73, SD=0.46) ด้านความเป็นไปได้ (feasibility) ในระดับมากที่สุด (M=4.65, SD=0.49) และด้านอรรถประโยชน์ (utility) ในระดับมากที่สุด (M=4.53, SD=0.56) ตามลำดับ รายละเอียดดังตาราง 37 โดยเกณฑ์ในการแปลความหมายของระดับคุณภาพคู่มือฯ (นพวรรณ ศรีเกตุ, 2557) เป็นดังนี้

คะแนนเฉลี่ย 4.51 – 5.00 หมายถึง คู่มือฯ มีความเหมาะสมในระดับมากที่สุด

คะแนนเฉลี่ย 3.51 – 4.50 หมายถึง คู่มือฯ มีความเหมาะสมในระดับมาก

คะแนนเฉลี่ย 2.51 – 3.50 หมายถึง คู่มือฯ มีความเหมาะสมในระดับปานกลาง

คะแนนเฉลี่ย 1.51 – 2.50 หมายถึง คู่มือฯ มีความเหมาะสมในระดับน้อย

คะแนนเฉลี่ย 1.00 – 1.50 หมายถึง คู่มือฯ มีความเหมาะสมในระดับน้อยที่สุด

ตาราง 37 ผลการประเมินคุณภาพของคู่มือฯ ตามมาตรฐานการประเมิน

รายการประเมินคู่มือฯ	ผลการประเมิน		
	M	SD	ระดับคุณภาพ
<b>1. ด้านอรรถประโยชน์ (utility)</b>	<b>4.53</b>	<b>0.56</b>	<b>มากที่สุด</b>
1.1 คู่มือฯ สามารถเป็นแนวทางในการศึกษาความสอดคล้องระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน	4.60	0.55	มากที่สุด
1.2 คู่มือฯ สามารถเป็นแนวทางในการศึกษาความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน	4.60	0.55	มากที่สุด
1.3 คู่มือฯ สามารถเป็นแนวทางในการลดความคลาดเคลื่อนอันเกิดจากอิทธิพลของผู้ประเมินได้	4.40	0.89	มากที่สุด

ตาราง 37 (ต่อ)

รายการประเมินคู่มือฯ	ผลการประเมิน		
	M	SD	ระดับคุณภาพ
1.4 ผลการดำเนินงานตามคู่มือฯ สามารถช่วยให้เกิดการพัฒนา/ ปรับปรุงการวัดและประเมินผลของโรงเรียนให้สอดคล้องกับ มาตรฐานและตัวชี้วัดได้	4.40	0.55	มาก
1.5 ผลการดำเนินงานตามคู่มือฯ สามารถช่วยให้เกิดการพัฒนา/ ปรับปรุงการวัดและประเมินผลของโรงเรียนให้สอดคล้องการวัด และประเมินระดับชาติได้	4.25	0.50	มาก
1.6 สารสนเทศจากการดำเนินการตามคู่มือฯ เป็นประโยชน์ต่อ การพัฒนา/ปรับปรุงการออกข้อสอบในการประเมินระดับชั้นเรียน ให้ความครอบคลุมระดับพฤติกรรมกรวัดได้	4.60	0.55	มากที่สุด
1.7 สารสนเทศจากการดำเนินการตามคู่มือฯ เป็นประโยชน์ต่อพัฒนา/ ปรับปรุงการวัดและประเมินระดับชั้นเรียน	4.80	0.45	มากที่สุด
<b>2. ด้านความเป็นไปได้ (feasibility)</b>	<b>4.65</b>	<b>0.49</b>	<b>มากที่สุด</b>
2.1 คู่มือฯ ให้แนวทางในการสร้างเครื่องมือในการศึกษา ความสอดคล้องระหว่างองค์ประกอบทางการศึกษา	4.80	0.45	มากที่สุด
2.2 แนวทางการดำเนินงานตามคู่มือฯ สามารถนำไปปฏิบัติจริงได้	4.80	0.45	มากที่สุด
2.3 แนวทางการดำเนินงานตามคู่มือฯ สามารถนำไปประยุกต์ใช้ใน การศึกษาความสอดคล้องระดับองค์ประกอบทางการศึกษาได้	4.60	0.55	มากที่สุด
2.4 ประโยชน์ที่ได้จากการดำเนินงานตามคู่มือฯ มีความคุ้มค่า เมื่อเทียบกับเวลาและค่าใช้จ่าย	4.40	0.55	มาก
<b>3. ด้านความเหมาะสม (property)</b>	<b>4.95</b>	<b>0.22</b>	<b>มากที่สุด</b>
3.1 คู่มือฯ อยู่บนพื้นฐานของแนวคิดและทฤษฎีที่เหมาะสม	5.00	0.00	มากที่สุด
3.2 แนวทางการดำเนินงานตามคู่มือฯ มีความเหมาะสมตามหลักการ วัดและประเมินผล	5.00	0.00	มากที่สุด
3.3 เกณฑ์การตัดสินผลการประเมินความสอดคล้องในแนวเดียวกันมี ความเหมาะสมและยอมรับได้	5.00	0.00	มากที่สุด
3.4 ผู้ประเมินตามที่ระบุในคู่มือฯ มีความเหมาะสมและน่าเชื่อถือ	4.80	0.46	มากที่สุด
<b>4. ด้านความถูกต้อง (accuracy)</b>	<b>4.73</b>	<b>0.46</b>	<b>มากที่สุด</b>
4.1 ขั้นตอนในการดำเนินงานตามคู่มือฯ มีต่อเนื่องและชัดเจน	4.80	0.45	มากที่สุด
4.2 ผลจากการดำเนินงานตามคู่มือฯ สามารถให้ผลที่ชัดเจน	4.80	0.45	มากที่สุด
4.3 การรายงานผลตามคู่มือฯ มีความถูกต้องน่าเชื่อถือ	4.60	0.55	มากที่สุด
<b>รวม</b>	<b>4.69</b>	<b>0.49</b>	<b>มากที่สุด</b>

### 6.3 ข้อเสนอแนะเพิ่มเติมจากผู้ทรงคุณวุฒิในการพัฒนาคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

ผู้ทรงคุณวุฒิมีข้อเสนอแนะเพิ่มเติมในการพัฒนาคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ดังนี้

“เป็นคู่มือที่มีประโยชน์ ควรนำคู่มือนี้ให้นักวิชาการที่มีบทบาทเกี่ยวข้องกับ  
การประเมินผลระดับชาติ ระดับสถานศึกษา ระดับเขตพื้นที่การศึกษา ได้ใช้ประเมิน  
ความสอดคล้องของข้อสอบใช้เพื่อให้เกิดเป็นนวัตกรรมทางการวัดประเมินผลที่แพร่หลาย  
และเกิดการยอมรับและนิยมใช้ในวงการศึกษา”



## บทที่ 5

### สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) ตรวจสอบอิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น 2) เปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน 3) ศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน และ 4) ประเมินค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปร่างของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน

กลุ่มตัวอย่างในการวิจัย ได้แก่ (1) ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น (มัธยมศึกษาปีที่ 1 – 3) ของโรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ในกรุงเทพมหานคร ประกอบด้วยข้อสอบที่ครูสร้างขึ้น (teacher-made test) เพื่อประเมินผลการเรียนรู้ของผู้เรียนในระดับชั้นเรียน ทั้งข้อสอบระหว่างภาค และข้อสอบประจำภาค ปีการศึกษา 2559 จำนวน 1,089 ข้อ ที่ได้มาด้วยวิธีการสุ่มแบบหลายขั้นตอน (multi-stage random sampling) (2) ข้อสอบในการประเมินระดับชาติ คือ ข้อสอบในการทดสอบทางการศึกษาระดับชาติด้านขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 จำนวน 45 ข้อ และ (3) ผู้เชี่ยวชาญในการประเมินความสอดคล้องในแนวเดียวกัน ประกอบด้วย ครูสอนวิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น และ/หรือนักวิชาการ และ/หรืออาจารย์ระดับอุดมศึกษาที่สอนในรายวิชาวิทยาศาสตร์ และ/หรือสาขาวิชาที่เกี่ยวข้องกับการวัดและประเมินผลการศึกษา จำนวน 20 ท่าน ที่ได้มาจากการเลือกแบบเจาะจง (purposive sampling)

เครื่องมือในการวิจัยครั้งนี้ ได้แก่ (1) แบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (test-item specification) ในการทดสอบทางการศึกษาระดับขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 และ (2) แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์



ระดับชั้นมัธยมศึกษาตอนต้น ที่ผู้เชี่ยวชาญใช้ในการประเมินระดับความซับซ้อนทางปัญญาตามแนวคิด revised Bloom's taxonomy และประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัดที่มีลักษณะเป็นมาตรฐานค่า 5 ระดับ (0 - 4)

การเก็บรวบรวมข้อมูลในการวิจัยครั้งนี้แบ่งเป็น 2 ส่วน คือ (1) การเก็บรวบรวมข้อมูลลักษณะจำเพาะของข้อสอบในการประเมินผลระดับชาติ และ (2) การเก็บรวบรวมข้อมูลจากผู้เชี่ยวชาญในการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน

การวิเคราะห์ข้อมูลโดยการประยุกต์ใช้ MFRM ด้วยโปรแกรม FACETS ในการวิเคราะห์ความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เพื่อตรวจสอบอิทธิพลของผู้ประเมิน ได้แก่ อิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ตลอดจนการตรวจสอบความแม่นยำในการให้คะแนนของผู้ประเมิน นอกจากนี้ MFRM ใช้ในการควบคุมอิทธิพลการกดหรือปล่อยคะแนน และมีการเปรียบเทียบผลการประเมินความสอดคล้องฯ ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ผู้วิจัยทำการคัดเลือกข้อสอบในการประเมินระดับชั้นเรียนที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัด หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน เพื่อใช้ในการสร้างเมตริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมตริกซ์ Y) จากนั้นทำการวิเคราะห์ค่าดัชนีความสอดคล้อง (alignment index) ระหว่างเมตริกซ์ข้อสอบในการประเมินระดับชาติ (เมตริกซ์ X) กับเมตริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมตริกซ์ Y) โดยวิเคราะห์และเปรียบเทียบค่าดัชนีความสอดคล้องระหว่างโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน นอกจากนี้ ได้มีการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด เพื่อประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน ที่ทำผลการประเมินมีค่าน่าเชื่อถือในระดับที่ยอมรับได้

## สรุปผลการวิจัย

สรุปผลการวิจัย แบ่งการนำเสนอตามวัตถุประสงค์ของการวิจัย ดังนี้

**1. ผลการตรวจสอบอิทธิพลของผู้ประเมินในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน**

1.1 ผลการตรวจสอบอิทธิพลการกดหรือปล่อยคะแนน พบว่า เกิดอิทธิพลการกดหรือปล่อยคะแนนที่สามารถตรวจสอบได้จากการทดสอบสถิติไคสแควร์ (fixed chi-square test) ที่ได้

จากโปรแกรม FACETS พบว่า ค่าไคสแควร์มีนัยสำคัญที่ระดับ .01 ( $\chi^2 = 975.70$ , d.f. = 19,  $p = .00$ ) แสดงว่ามีผู้ประเมินอย่างน้อย 2 คนที่มีระดับการกดคะแนนแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และเมื่อพิจารณาคะแนนโลจิทที่แสดงถึงค่าการกดหรือปล่อยคะแนนของผู้ประเมินเป็นรายบุคคล พบว่า โดยภาพรวม ผู้ประเมินมีแนวโน้มที่จะมีลักษณะของการกดคะแนนมากกว่าปล่อยคะแนน คือ มีคะแนนโลจิทไปทางบวกมากกว่าทางลบ ทั้งนี้ ผู้ประเมินคนที่ 4 มีลักษณะกดคะแนนมากที่สุด (คะแนนโลจิท = 1.83) รองลงมาคือ ผู้ประเมินคนที่ 20 19 และ 16 ตามลำดับ (คะแนนโลจิท = 1.03, 0.93 และ 0.92 ตามลำดับ) ส่วนผู้ประเมินคนที่ 13 มีลักษณะการปล่อยคะแนนมากที่สุด (คะแนนโลจิท = -3.24) รองลงมาคือ ผู้ประเมินคนที่ 12 15 และ 14 ตามลำดับ (คะแนนโลจิท = -1.83, -1.47 และ -1.15 ตามลำดับ)

1.2 ผลการ อธิธิพลแนวโน้มการให้คะแนนค่ากลาง อธิธิพลการจำกัดช่วง และตรวจสอบความแม่นยำในการให้คะแนนโดยพิจารณาค่าสถิติ infit MNSQ และ outfit MNSQ ของผู้ประเมินพบว่า ผู้ประเมินส่วนใหญ่แสดงรูปแบบของความแม่นยำในการให้คะแนน (accurate) คือ มีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.70 – 1.30 จำนวน 16 คน คิดเป็นร้อยละ 80.00 ทั้งนี้ไม่มีผู้ประเมินที่แสดงอิทธิพลแนวโน้มการให้คะแนนค่ากลาง อธิธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน สำหรับผู้ประเมินอีกจำนวน 4 คน ได้แก่ ผู้ประเมินคนที่ 12 13 15 และ 18 คิดเป็นร้อยละ 20.00 ไม่ถูกจัดให้อยู่ในรูปแบบของผู้ประเมินใดๆ

## 2. ผลการเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

2.1 ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยพิจารณาข้อสอบที่มีค่า Obs. Avge ตั้งแต่ 3.00 คะแนนขึ้นไป พบว่า ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน จากข้อสอบทั้งหมด 1,089 ข้อ มีข้อสอบที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัดที่ระบุจำนวน 1,012 ข้อ คิดเป็นร้อยละ 92.93

2.2 ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยพิจารณาข้อสอบที่มีค่า Fair-M Avge ตั้งแต่ 3.00 คะแนนขึ้นไป และมีคะแนนเชิงประจักษ์สอดคล้องกับโมเดล คือ มีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.50 – 1.50 พบว่า หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน จากข้อสอบทั้งหมด 1,089 ข้อ มีข้อสอบที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัดที่ระบุจำนวน 902 ข้อ คิดเป็นร้อยละ 82.83 ซึ่งข้อสอบทั้ง 902 ข้อ จะถูกนำไปใช้ใน

การสร้างเมทริกซ์ Y และวิเคราะห์ดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียนตามแนวคิดของ Porter ต่อไป

2.3 ผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนการประเมินความสอดคล้องในแนวเดียวกัน ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Obs. Avge) และหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Fair-M Avge) ด้วยการทดสอบสถิติที่ (paired-samples t-test) พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ( $t = 17.044$ ,  $p = .00$ ) โดยขนาดอิทธิพล (effect size) ที่คำนวณจากสูตรของ Cohen มีค่าเท่ากับ 0.52 ซึ่งอยู่ในระดับปานกลาง (Cohen, 1988)

2.4 ผลการศึกษาการเปลี่ยนแปลงผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน พบว่า มีข้อสอบ จำนวน 21 ข้อ คิดเป็นร้อยละ 1.93 ที่มีการเปลี่ยนแปลงผลการประเมินไปจากก่อนมีการควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยมีข้อสอบจำนวน 11 ข้อ ที่มีการเปลี่ยนแปลงผลการประเมินจาก “สอดคล้องกับตัวชี้วัดที่ระบุ” คือมีค่า Fair-M Avge ตั้งแต่ 3.00 คะแนนขึ้นไป เปลี่ยนเป็น “ไม่สอดคล้องกับตัวชี้วัดที่ระบุ” คือมีค่า Fair-M Avge ต่ำกว่า 3.00 คะแนน และมีข้อสอบจำนวน 10 ข้อ ที่เปลี่ยนแปลงผลการประเมินจาก “ไม่สอดคล้องกับตัวชี้วัดที่ระบุ” คือมีค่า Fair-M Avge ต่ำกว่า 3.00 คะแนน เปลี่ยนเป็น “สอดคล้องกับตัวชี้วัดที่ระบุ” คือมีค่า Fair-M Avge ตั้งแต่ 3.00 คะแนนขึ้นไป

**3. ผลการศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียนของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน**

3.1 ผลการเปรียบเทียบค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนทั้ง 4 โรงเรียน โดยภาพรวมมีค่าดัชนีความสอดคล้องเท่ากับ 0.560 เมื่อพิจารณาดัชนีความสอดคล้องจำแนกตามผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียน พบว่า ค่าดัชนีความสอดคล้องของทั้ง 4 โรงเรียน มีค่าใกล้เคียงกัน โดยโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูงมีค่าดัชนีความสอดคล้องสูงกว่าโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำเล็กน้อย โดยโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง มีค่าดัชนีความสอดคล้องเท่ากับ 0.584 และ 0.588 ในขณะที่โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ มีค่าดัชนีความสอดคล้องเท่ากับ 0.436 และ 0.565

3.2 ผลการเปรียบเทียบร้อยละสัดส่วนระดับความซับซ้อนทางปัญญาระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน พบว่า ข้อสอบในการประเมินระดับชาติมีร้อยละสัดส่วนของข้อสอบที่วัดความซับซ้อนทางปัญญาระดับวิเคราะห์มากที่สุด ร้อยละ

40.00 รองลงมาคือ ระดับเข้าใจ ร้อยละ 36.00 ระดับประยุกต์ใช้ ร้อยละ 26.00 ระดับประเมินค่า ร้อยละ 2.00 และไม่มีข้อสอบที่วัดระดับจำ และระดับสร้างสรรค์ ในขณะที่ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนทั้ง 4 โรงเรียน โดยภาพรวม พบว่า มีร้อยละสัดส่วนของข้อสอบที่วัดระดับความเข้าใจมากที่สุด ร้อยละ 46.99 รองลงมาคือ ระดับจำ ร้อยละ 25.21 ระดับประยุกต์ใช้ ร้อยละ 15.07 ระดับวิเคราะห์ ร้อยละ 11.64 ระดับประเมินค่า ร้อยละ 1.09 และไม่มีข้อสอบที่วัดระดับสร้างสรรค์ เช่นเดียวกับข้อสอบในการประเมินระดับชาติ

#### 4. ผลการประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปร่างอิงสำหรับการตัดสินใจของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน

4.1 ผลการวิเคราะห์ขนาดขององค์ประกอบความแปรปรวนของคะแนนการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน พบว่า ทั้งการประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ มีผลเป็นไปในทิศทางเดียวกัน กล่าวคือ ความแปรปรวนจากแหล่งต่างๆ เมื่อเทียบกับความแปรปรวนทั้งหมดมีค่าเรียงลำดับจากมากไปหาน้อย ดังนี้ ความแปรปรวนของคะแนนความสอดคล้องในแนวเดียวกันของข้อสอบ (ร้อยละ 54.72 และ 51.12 ตามลำดับ) รองลงมาคือ ความแปรปรวนของความคลาดเคลื่อนหรือส่วนเหลือ ซึ่งเป็นความแปรปรวนอันเนื่องมาจากความแตกต่างของความสอดคล้องในแนวเดียวกันของข้อสอบ แต่ละข้อที่ได้รับการประเมินด้วยผู้ประเมินที่แตกต่างกัน และ/หรือเป็นแหล่งความแปรปรวนที่ไม่สามารถวัดได้ (ร้อยละ 42.44 และ 36.36 ตามลำดับ) และความแปรปรวนอันเกิดจากผู้ประเมิน (ร้อยละ 2.84 และ 12.52 ตามลำดับ) ตามลำดับ

4.2 ผลการเปรียบเทียบค่าสัมประสิทธิ์การสรุปร่างอิงของคะแนนการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน พบว่า ค่าสัมประสิทธิ์การสรุปร่างอิงสำหรับตัดสินใจเชิงสัมบูรณ์ของคะแนนการประเมินความสอดคล้องในแนวเดียวกันๆ มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ประเมินเพิ่มขึ้น ในทุกรูปแบบการออกแบบการประเมิน ทั้งการประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ

4.3 ผลการพิจารณาจำนวนผู้ประเมินที่ทำให้เกิดความน่าเชื่อถือของการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน ในระดับที่ยอมรับได้ โดยการพิจารณาค่าสัมประสิทธิ์การสรุปร่างอิงสำหรับตัดสินใจเชิงสัมบูรณ์ ตั้งแต่ 0.70 ขึ้นไป (Nunnally, 1987) ในการประเมินระดับความซับซ้อนทางปัญญา พบว่า ผู้ประเมิน

จำนวน 2 คน สามารถให้ผลการประเมินที่มีความน่าเชื่อถือในระดับที่ยอมรับได้ ในทุกรูปแบบการออกแบบการประเมิน โดยให้ค่าสัมประสิทธิ์การสรุปร่างอ้างอิงสำหรับตัดสินใจเชิงสัมบูรณ์ ตั้งแต่ 0.71 ขึ้นไป และพบว่า การประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรฐานค่า 5 ระดับ ต้องการผู้ประเมินจำนวน ตั้งแต่ 3 คนขึ้นไป ให้ผลการประเมินที่มีความน่าเชื่อถือในระดับที่ยอมรับได้ ในทุกรูปแบบการออกแบบการประเมิน โดยให้ค่าสัมประสิทธิ์การสรุปร่างอ้างอิงสำหรับตัดสินใจเชิงสัมบูรณ์ ตั้งแต่ 0.76 ขึ้นไป นอกจากนี้ เมื่อพิจารณาจำนวนผู้ประเมินที่เท่ากันทั้ง 2 รูปแบบการออกแบบการประเมิน คือ รูปแบบที่ 1 ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ และรูปแบบที่ 2 ผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อให้ จะให้ค่าสัมประสิทธิ์การสรุปร่างอ้างอิงที่ใกล้เคียงกัน

#### 5. ผลการตรวจสอบคุณภาพคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

ผลการตรวจสอบคุณภาพด้านความตรงตามเนื้อหาของคู่มือฯ โดยการตรวจสอบความสอดคล้องระหว่างคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กับวัตถุประสงค์ของคู่มือฯ พบว่า มีค่าดัชนี IOC อยู่ระหว่าง 0.80 – 1.00 ซึ่งผ่านเกณฑ์ทุกข้อ

ผลการตรวจสอบคุณภาพด้านความเหมาะสมของคู่มือฯ ตามมาตรฐานการประเมิน 4 ด้าน โดยผู้ทรงคุณวุฒิจำนวน 5 ท่าน มีข้อรายการประเมินคุณภาพในการประเมินคู่มือฯ พบว่า ในภาพรวมคู่มือฯ มีระดับคุณภาพในระดับมากที่สุด ( $M=4.69$ ,  $SD=0.49$ ) โดยมีผลการประเมินระดับคุณภาพในระดับมากที่สุดทุกด้าน ด้านที่มีผลคะแนนเฉลี่ยมากที่สุด ในระดับมากที่สุด คือ ด้านความเหมาะสม (property) ( $M=4.95$ ,  $SD=0.22$ ) รองลงมาคือ ด้านความถูกต้อง (accuracy) ( $M=4.73$ ,  $SD=0.46$ ) ด้านความเป็นไปได้ (feasibility) ในระดับมากที่สุด ( $M=4.65$ ,  $SD=0.49$ ) และด้านอรรถประโยชน์ (utility) ในระดับมากที่สุด ( $M=4.53$ ,  $SD=0.56$ ) ตามลำดับ

#### อภิปรายผลการวิจัย

##### 1. การตรวจสอบอิทธิพลการกดหรือปล่อยคะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน

1.1 จากผลการวิจัยที่พบว่า เกิดอิทธิพลการกดหรือปล่อยคะแนนในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ที่ตรวจสอบได้ทั้งในระดับกลุ่มและระดับบุคคล โดยผู้ประเมินมีแนวโน้มที่จะมีลักษณะของการกด

คะแนนมากกว่าปล่อยคะแนน ผู้ประเมินคนที่ 4 มีลักษณะกตคะแนนมากที่สุด (คะแนนโลจิท = 1.83) รองลงมาคือ ผู้ประเมินคนที่ 20 19 และ 16 ตามลำดับ (คะแนนโลจิท = 1.03, 0.93 และ 0.92 ตามลำดับ) ส่วนผู้ประเมินคนที่ 13 มีลักษณะการปล่อยคะแนนมากที่สุด (คะแนนโลจิท = -3.24) รองลงมาคือ ผู้ประเมินคนที่ 12 15 และ 14 ตามลำดับ (คะแนนโลจิท = -1.83, -1.47 และ -1.15 ตามลำดับ) สอดคล้องกับงานวิจัยของ Anderson, Irvin, Alonzo, และ Tindal (2015) ที่พบว่ามีความแตกต่างอย่างมีนัยสำคัญทางสถิติของผลการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับมาตรฐานคณิตศาสตร์ เมื่อมีการใช้สถิติในการควบคุมอิทธิพลการกตหรือปล่อยคะแนนของผู้ประเมิน ที่เป็นเช่นนี้อาจเป็นเพราะผู้ประเมินมีประสบการณ์ในการประเมินความสอดคล้องของข้อสอบแตกต่างกัน ซึ่งสอดคล้องกับงานวิจัยของอุษณีย์ บัวศิริพันธ์ (2543) และบุษวรรษ์ แสนปลื้ม และองอาจ นัยพัฒน์ (2556) ที่พบว่า ผู้ประเมินให้คะแนนที่มีประสบการณ์ต่างกันส่งผลให้คะแนนประเมินมีค่าแตกต่างกัน และถึงแม้ว่าในงานวิจัยนี้จะมีการดำเนินการเพื่อป้องกันหรือลดอิทธิพลของผู้ประเมินในวิธีการต่างๆ ตามหลักการของ Myford และ Wolfe (2003) แล้วก็ตาม ได้แก่ การคัดเลือกผู้ประเมิน การอบรมฝึกปฏิบัติให้กับผู้ประเมินเพื่อให้ผู้ประเมินมีความรู้ความเข้าใจในการดำเนินงานและเครื่องมือในการประเมิน การสร้างมาตรฐานประเมินที่มีหลายระดับ การใช้ผู้ประเมินหลายคนในการประเมิน เพื่อหาค่าเฉลี่ยของผลประเมิน แต่ยังคงพบว่าเกิดอิทธิพลการกตและปล่อยคะแนนของผู้ประเมิน อย่างไรก็ตามอิทธิพลของการกตหรือปล่อยคะแนนดังกล่าวได้ถูกควบคุมโดยประยุกต์ใช้ MFRM ซึ่งเป็นวิธีการทางสถิติในการปรับคะแนนความสอดคล้องในแนวเดียวกันๆ ของข้อสอบที่แตกต่างกันจากการกตหรือปล่อยคะแนนของผู้ประเมิน ทำให้ผลการประเมินมีความถูกต้องและปราศจากความเป็นไม่เป็นปรนัยของผู้ประเมิน (Turner, 2003)

1.2 จากผลการตรวจสอบอิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วงอิทธิพลของความไม่มีแบบแผน และความแม่นยำในการให้คะแนน โดยพิจารณาค่าสถิติ infit MNSQ และ outfit MNSQ ของผู้ประเมิน ที่พบว่า ผู้ประเมินส่วนใหญ่แสดงรูปแบบของความแม่นยำในการให้คะแนน จำนวน 16 คน คิดเป็นร้อยละ 80.00 ไม่มีผู้ประเมินที่แสดงอิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน สำหรับผู้ประเมินอีกจำนวน 4 คน ได้แก่ ผู้ประเมินคนที่ 12 13 15 และ 18 คิดเป็นร้อยละ 20.00 ไม่ถูกจัดให้อยู่ในรูปแบบของผู้ประเมินใดๆ Wolfe (2004) กล่าวว่าสาเหตุของการที่ผู้ประเมินมีความแม่นยำในการให้คะแนนนั้น อาจเนื่องมาจากหลายปัจจัย ได้แก่ ประสบการณ์ของผู้ประเมิน (เช่น การได้รับการฝึกอบรม ประวัติการศึกษา ประสบการณ์ในการทำงาน) ปัจจัยทางปัญญา (เช่น รูปแบบการคิด ความสามารถในการเรียนรู้) ลักษณะของเกณฑ์ในการประเมิน (ระดับความคล้ายคลึงของเกณฑ์ และความเข้าใจในประเด็นที่มีปัญหา) และสิ่งแวดล้อมในการประเมิน (เช่น ความเป็นอิสระจากสิ่งรบกวน รูปแบบของปฏิสัมพันธ์ทางสังคมที่เกิดขึ้นในการการประเมิน) ซึ่งในงานวิจัยนี้ผู้วิจัยตระหนักความสำคัญของ

ปัจจัยดังกล่าว และได้มีการดำเนินการเพื่อให้เกิดความแม่นยำในการให้คะแนน สรุปได้ 3 ประการ ดังนี้ **ประการแรก** การคัดเลือกผู้เชี่ยวชาญที่ทำหน้าที่เป็นผู้ประเมินที่มีคุณสมบัติที่เหมาะสม ซึ่งสอดคล้องกับ Davis-Becker และ Buckendahl (2013) La Marca, Redfield, และ Winter (2000) ที่กล่าวว่าผู้เชี่ยวชาญจะต้องมีความรู้และคุ้นเคยในเนื้อหาของวิชาที่ต้องการประเมิน ลักษณะของผู้สอบ หลักสูตร ตลอดจนมาตรฐานการเรียนรู้ที่ต้องการประเมิน โดยผู้วิจัยมีการคัดเลือกผู้เชี่ยวชาญแบบเฉพาะเจาะจง (purposive sampling) กำหนดเกณฑ์ในการคัดเลือก คือ ต้องเป็นผู้ที่มีความรู้ความสามารถด้านมาตรฐานและตัวชี้วัดกลุ่มสาระการเรียนรู้วิทยาศาสตร์ชั้นมัธยมศึกษาตอนต้น และด้านการวัดและประเมินทางการศึกษา โดยผู้ประเมินทุกคนสำเร็จการศึกษาอย่างน้อยในระดับปริญญาตรีทางการสอนวิทยาศาสตร์ และสำเร็จการศึกษาอย่างน้อยในระดับปริญญาโท สาขาวิชาการวัดและประเมินทางการศึกษาหรือสาขาวิชาที่เกี่ยวข้อง และ/หรือมีประสบการณ์สอนกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ไม่น้อยกว่า 3 ปี ทั้งนี้ ผู้ประเมินส่วนใหญ่เป็นครูโรงเรียนมัธยมศึกษา (ร้อยละ 70.00) รองลงมาคือ อาจารย์มหาวิทยาลัย (ร้อยละ 20.00) และนักวิชาการ (ร้อยละ 10.00) โดยผู้เชี่ยวชาญส่วนใหญ่ (ร้อยละ 85) มีประสบการณ์ด้านการสอนวิทยาศาสตร์ มากกว่า 6 ปี **ประการที่สอง** การจัดประชุมและฝึกปฏิบัติผู้เชี่ยวชาญก่อนทำการประเมินจริง โดยมีกิจกรรมเพื่อสร้างความเข้าใจที่ตรงกันระหว่างผู้เชี่ยวชาญในเรื่องต่างๆ เช่น การทบทวนนิยามของระดับความซับซ้อนทางปัญญา ตามแนวคิด revised Bloom's taxonomy การชี้แจงแนวทางในการดำเนินงานของผู้เชี่ยวชาญ การอธิบายและทำความเข้าใจถึงลักษณะของเครื่องมือและเกณฑ์การประเมินต่างๆ การให้ผู้เชี่ยวชาญทดลองฝึกปฏิบัติกระบวนการฉันทามติ และฝึกประเมินความสอดคล้องระหว่างข้อสอบกับตัวชี้วัดเป็นรายบุคคล ตลอดจนเปิดโอกาสให้ผู้ประเมินได้อภิปรายในภายในกลุ่มเพื่อให้ได้ผลการพิจารณาที่ถูกต้องและเข้าใจตรงกัน ซึ่งสอดคล้องกับ Webb (1999) ที่เน้นย้ำว่าการฝึกปฏิบัติก่อนดำเนินการประเมินความสอดคล้องในแนวเดียวกันๆ มีความสำคัญต่อการศึกษาความสอดคล้องในแนวเดียวกัน เนื่องจากการฝึกปฏิบัติจะช่วยให้ผู้ประเมินทุกคนมีความเข้าใจถึงกระบวนการประเมินความสอดคล้องในแนวเดียวกัน เข้าใจถึงหลักการจำแนก ระดับความซับซ้อนทางปัญญา ตลอดจนวิธีการในการพิจารณาตัดสินความสอดคล้องของผู้ประเมินในการประเมินจริง เพื่อให้สามารถทำหน้าที่ประเมินอย่างอิสระเป็นรายบุคคล นอกจากนี้ หลังจากการประชุมและฝึกปฏิบัติแล้วแล้ว ผู้ประเมินแต่ละคนจะทำหน้าที่ประเมินความสอดคล้องระหว่างตัวชี้วัดกับข้อสอบได้อย่างอิสระ โดยผู้วิจัยได้จัดทำเอกสารคู่มือประกอบการดำเนินงานประเมินความสอดคล้องในแนวเดียวกันๆ สำหรับผู้ประเมิน ซึ่งสาระสำคัญของคู่มือฯ ประกอบด้วย (1) คำชี้แจงในการดำเนินงานของผู้เชี่ยวชาญ (2) นิยามศัพท์ คำอธิบาย และตัวอย่างของการจำแนกพฤติกรรมการเรียนรู้ด้านพุทธิพิสัยตามแนวคิด revised Bloom's taxonomy และ (3) รายละเอียดของตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษา

ตอนต้น ตามหลักสูตรแกนกลางขั้นพื้นฐาน พุทธศักราช 2551 และ**ประกาศที่สาม** ผู้ประเมินมีความเป็นอิสระในการประเมิน เนื่องจากหลังการประชุมและฝึกปฏิบัติแล้ว ผู้ประเมินแต่ละคนจะทำหน้าที่ในการประเมินข้อสอบแบบอิสระเป็นรายบุคคลโดยปราศจากอิทธิพลของกลุ่ม ซึ่งผู้ประเมินสามารถเลือกดำเนินการประเมินโดยใช้แบบประเมินความสอดคล้องๆ รูปแบบใดรูปแบบหนึ่งจาก 2 รูปแบบ ได้แก่ รูปแบบเอกสาร หรือรูปแบบออนไลน์ ทั้งนี้ ขึ้นอยู่กับความสะดวกของผู้ประเมินแต่ละคน การประเมินแบบอิสระเป็นรายบุคคล นอกจากจะเป็นการช่วยประหยัดเวลาและค่าใช้จ่าย และมีความยืดหยุ่นในการดำเนินการมากกว่าการประชุมเพื่อประเมินโดยใช้ฉันทามติร่วมกันแล้ว ยังเป็นการป้องกันผลกระทบที่อาจเกิดขึ้นจากอิทธิพลของกลุ่มผู้ประเมินได้ เช่น การปรับคะแนนประเมินตามคนในกลุ่ม หรือการถูกโน้มน้าวให้มีความคล้อยตามความความคิดเห็นของผู้ประเมินส่วนใหญ่หรือผู้ประเมินบางคนในกลุ่ม (Bond, 2005; Kocher & Sutter, 2005)

อย่างไรก็ตาม ถึงแม้ในงานวิจัยนี้จะมีการดำเนินการต่างๆ ดังที่กล่าวข้างต้น เพื่อให้เกิดความแม่นยำในการให้คะแนน แต่ผลการวิจัยระบุว่า ยังมีผู้ประเมินจำนวน 4 คน ได้แก่ ผู้ประเมินคนที่ 12 13 15 และ 18 คิดเป็นร้อยละ 20.00 ที่ข้อมูลคะแนนเชิงประจักษ์ไม่สอดคล้องกับโมเดลคือมีค่า infit MNSQ และ outfit MNSQ ไม่อยู่ในช่วง 0.70 – 1.30 และไม่ถูกจัดให้อยู่ในรูปแบบของผู้ประเมินใดๆ ที่มีการศึกษาในงานวิจัยนี้ อันได้แก่ ความแม่นยำในการให้คะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลการจำกัดช่วง และอิทธิพลของความไม่มีแบบแผน สอดคล้องกับงานวิจัยของ Myford and Wolfe (2004) ที่ทำการศึกษาแหล่งความแปรปรวนในการประเมินการพูดภาษาอังกฤษ (test of spoken English: TSE) ที่มาจากลักษณะและอิทธิพลของผู้ประเมินประเภทต่างๆ ได้แก่ ความแม่นยำในการให้คะแนน อิทธิพลแนวโน้มการให้คะแนนค่ากลาง อิทธิพลฮาโล และอิทธิพลการให้คะแนนสุดโต่ง พบว่า มีผู้ประเมินร้อยละ 27 ที่บ่งชี้ว่าไม่ถูกจัดให้อยู่ในรูปแบบใดๆ ในการศึกษา ที่เป็นเช่นนี้อาจเป็นไปได้ว่าผู้ประเมินมีการให้คะแนนที่เบี่ยงเบนไปจากรูปแบบการให้คะแนนต่างๆ ที่ยากต่อการตรวจสอบ ซึ่งเป็นประเด็นที่น่าสนใจในการศึกษาวิจัยต่อไป (Myford & Wolfe: 2004) อย่างไรก็ตาม เมื่อพิจารณาผู้ประเมินคนที่ 12 13 และ 15 ยังพบว่ามีลักษณะการปล่อยคะแนน ซึ่งอาจเป็นสาเหตุหนึ่งที่ทำให้เกิดความไม่แม่นยำในการให้คะแนน

## **2. การเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน**

2.1 จากผลการวิจัยที่พบว่า ผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน มีความแตกต่างกันระหว่างก่อนและหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ซึ่งเป็นไปตามสมมติฐานการวิจัยข้อที่ 1 ที่กล่าวว่า “หากเกิดอิทธิพลการกดหรือปล่อยคะแนน ในการประเมินความสอดคล้องในแนวเดียวกันระหว่าง



มาตรฐานและตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้นแล้ว น่าจะมีการเปลี่ยนแปลงของค่าเฉลี่ยคะแนน ระหว่างคะแนนผลการประเมินความสอดคล้องในแนวเดียวกันๆ ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน ซึ่งเป็นค่าเฉลี่ยคะแนนดิบของข้อสอบ (*item raw mean*) กับคะแนนผลการประเมินความสอดคล้องในแนวเดียวกันๆ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน ซึ่งเป็นค่าเฉลี่ยคะแนนที่ได้จากโมเดล ที่มีการปรับแก้อิทธิพลการกดหรือปล่อยคะแนนแล้ว (*item adjusted mean*)” โดยพบว่า ภายหลังควบคุมอิทธิพลการกดปล่อยคะแนน มีจำนวนข้อสอบที่มีความสอดคล้องในแนวเดียวกันกับมาตรฐานและตัวชี้วัด แตกต่างจากจำนวนข้อสอบก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน และพบความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ระหว่างค่าเฉลี่ยคะแนนการประเมินความสอดคล้องในแนวเดียวกัน ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Obs. Avge) และหลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน (Fair-M Avge) จากการทดสอบด้วยสถิติที (paired-samples t-test) ( $t = 17.044$   $p = .00$ ) โดยมีขนาดอิทธิพล (effect size) เท่ากับ 0.52 แสดงให้เห็นว่า อิทธิพลการกดหรือปล่อยคะแนนมีอิทธิพลต่อคะแนนการประเมินความสอดคล้องในแนวเดียวกันๆ ในระดับปานกลาง (Cohen, 1988) และจากผลการศึกษาการเปลี่ยนแปลงผลการประเมินความสอดคล้องในแนวเดียวกันของข้อสอบ หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน ที่พบว่า หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน มีข้อสอบจำนวน 21 ข้อ คิดเป็นร้อยละ 1.93 ที่มีการเปลี่ยนแปลงผลการประเมิน โดยมีข้อสอบจำนวน 11 ข้อ ที่มีการเปลี่ยนแปลงผลการประเมินจาก “สอดคล้องกับตัวชี้วัดที่ระบุ” เปลี่ยนเป็น “ไม่สอดคล้องกับตัวชี้วัดที่ระบุ” และมีข้อสอบจำนวน 10 ข้อ ที่เปลี่ยนแปลงผลการประเมินจาก “ไม่สอดคล้องกับตัวชี้วัดที่ระบุ” เปลี่ยนเป็น “สอดคล้องกับตัวชี้วัดที่ระบุ” ซึ่งสอดคล้องกับงานวิจัยของ Anderson, Irvin, Alonzo, และ Tindal (2015) ที่พบว่า ภายหลังมีการควบคุมอิทธิพลการกดหรือปล่อยคะแนน มีข้อสอบร้อยละ 5.50 ที่มีการเปลี่ยนแปลงผลการประเมิน ที่เป็นเช่นนี้เนื่องจากการวิเคราะห์ด้วย MFRM สามารถควบคุมอิทธิพลการกดหรือปล่อยคะแนนได้ โดยแต่ละองค์ประกอบจะถูกวิเคราะห์พร้อมกัน แต่มีความเป็นอิสระจากกัน และจะถูกปรับ (calibrate) ให้อยู่ในมาตรเดียวกัน (logit scale) การปรับแต่ละองค์ประกอบทำให้สามารถวัดการกดคะแนนของผู้ประเมินที่อยู่บนมาตรเดียวกับความสอดคล้องในแนวเดียวกันระหว่างตัวชี้วัดกับข้อสอบได้ MFRM สามารถปรับแก้คะแนนที่เกิดจากความแตกต่างของการกดหรือปล่อยคะแนนของผู้ประเมิน ทำให้คะแนนการประเมินความสอดคล้องในแนวเดียวกันๆ ที่ได้มีความน่าเชื่อถือ เทียบตรงและยุติธรรมมากขึ้น (Iramaneerat & Yudkowsky, 2007; Linacre, 1994) ซึ่งนำไปสู่การตัดสินผลการประเมินที่มีความถูกต้องมากกว่าการพิจารณาตัดสินผลการประเมินจากค่าเฉลี่ยของคะแนนดิบที่ไม่มีการปรับแก้หรือควบคุมอิทธิพลการกดหรือปล่อยคะแนน

2.2 จากผลการวิจัยที่พบว่าภายหลังการควบคุมอิทธิพลการกดหรือปล่อยคะแนน มีข้อสอบที่มีความสอดคล้องในแนวเดียวกันกับมาตรฐานและตัวชี้วัด และมีความสอดคล้องกับโมเดล จำนวน 902 ข้อ คิดเป็นร้อยละ 82.83 ซึ่งค่อนข้างสูง สอดคล้องกับรายงานการวิจัยการติดตามและตรวจสอบคุณภาพของข้อสอบที่ใช้ในการวัดและประเมินผลในชั้นเรียนของสถานศึกษา (สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2559) ที่พบว่า มีข้อสอบที่ใช้ในการวัดและประเมินผลในชั้นเรียนของสถานศึกษาที่ตรงตามมาตรฐานและตัวชี้วัดกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ร้อยละ 98.54 ที่เป็นเช่นนี้อาจเป็นเพราะโรงเรียนได้รับการสนับสนุนด้านการวัดและประเมินผลการเรียนรู้จากสำนักงานเขตพื้นที่การศึกษาหรือหน่วยงานต้นสังกัด เช่น การสร้างความรู้ความเข้าใจให้กับบุคลากรในสถานศึกษาที่เกี่ยวกับการวัดและประเมินผลการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน การส่งเสริมและสนับสนุนการเรียนรู้ในกลุ่มสาระการเรียนรู้ต่างๆ การส่งเสริมให้ครูบุคลากรในสถานศึกษามีความรู้ความเข้าใจในแนวปฏิบัติการวัดและประเมินผล ตลอดจนความเข้าใจในเทคนิควิธีการวัดและประเมินผลรูปแบบต่างๆ โดยเน้นการประเมินตามสภาพจริง การส่งเสริมสนับสนุนให้สถานศึกษาพัฒนาเครื่องมือและการบริหารจัดการด้านการวัดและประเมินผลการเรียนรู้ ตลอดจนการให้คำปรึกษา แนะนำเกี่ยวกับการวัดและประเมินผลการเรียนรู้ การจัดทำเอกสารหลักฐานการศึกษา (กระทรวงศึกษาธิการ, สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2551) จึงทำให้โรงเรียนสามารถจัดการศึกษา และทำการวัดและประเมินผลได้สอดคล้องกับมาตรฐานและตัวชี้วัดได้ ซึ่งเป็นไปตามนโยบายของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ที่เป็นหน่วยงานที่รับผิดชอบการศึกษาขั้นพื้นฐานของประเทศ ที่ตระหนักถึงความสำคัญของการพัฒนาคุณภาพผู้เรียนตามระบบการศึกษาแบบอิงมาตรฐาน โดยได้จัดทำประกาศของสำนักงานคณะกรรมการ การศึกษาขั้นพื้นฐาน เรื่อง แนวปฏิบัติในการยกระดับคุณภาพนักเรียนระดับการศึกษาขั้นพื้นฐาน ที่มุ่งปฏิรูปการเรียนรู้ทั้งระบบให้สัมพันธ์เชื่อมโยงกันระหว่างหลักสูตรการเรียนการสอน การเรียนการสอน และการวัดและประเมินผล ซึ่งให้ดำเนินการอย่างจริงจังตั้งแต่ปีการศึกษา 2557 เป็นต้นมา (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2557: เว็บไซต์)

**3. การศึกษาและเปรียบเทียบผลการประเมินความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียนกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ของโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกัน**

3.1 ผลการเปรียบเทียบค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนทั้ง 4 โรงเรียน โดยภาพรวมมีค่าดัชนีความสอดคล้องเท่ากับ 0.559 เมื่อพิจารณาดัชนีความสอดคล้องจำแนกตามผลสัมฤทธิ์ทางวิทยาศาสตร์ของโรงเรียน

พบว่า ค่าดัชนีความสอดคล้องของทั้ง 4 โรงเรียน มีค่าใกล้เคียงกัน ซึ่งเป็นค่าที่อยู่ในระดับปานกลาง โดยโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูงมีค่าดัชนีความสอดคล้องสูงกว่าโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำเล็กน้อย โดยโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง มีค่าดัชนีความสอดคล้องเท่ากับ 0.584 และ 0.588 ในขณะที่โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ มีค่าดัชนีความสอดคล้องเท่ากับ 0.436 และ 0.565 ซึ่งไม่เป็นไปตามสมมติฐานการวิจัยข้อที่ 2 ที่กล่าวว่า “โรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์แตกต่างกันน่าจะมีค่าดัชนีความสอดคล้อง (alignment index) ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น แตกต่างกัน โดยโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์สูง น่าจะมีค่าดัชนีความสอดคล้องสูงกว่าโรงเรียนที่มีผลสัมฤทธิ์ทางวิทยาศาสตร์ต่ำ ” ที่เป็นเช่นนี้อาจเป็นเพราะครูในโรงเรียนต่างๆ มีความสามารถในการออกข้อสอบให้สอดคล้องกับมาตรฐานและสอดคล้องข้อสอบ O-NET ใกล้เคียงกัน โดยจากการเก็บรวบรวมข้อมูลการออกข้อสอบข้อครูทั้ง 4 โรงเรียน พบว่า ทุกโรงเรียนมีกระบวนการออกข้อสอบที่มีประจวบพร้อมกันระหว่างครูผู้สอนเพื่อออกข้อสอบ (กรณีที่มีผู้สอนมากกว่า 1 คน) มีการวิเคราะห์ระบุเนื้อหา ตัวชี้วัดที่ และระดับพฤติกรรมการเรียนรู้ที่ต้องการวัดของข้อสอบแต่ละข้อ ตลอดจนมีการพิจารณา/ตรวจสอบความถูกต้องและความเหมาะสมของแบบทดสอบ โดยผู้อื่นที่ไม่ใช่ครูผู้ออกข้อสอบก่อนนำไปใช้จริง นอกจากนี้มีโรงเรียนจำนวน 3 ใน 4 โรงเรียนที่มีการจัดทำแผนผังในการออกข้อสอบ (test blueprint) มีการวิเคราะห์ข้อสอบหลังจากนำข้อสอบไปใช้จริง เช่น ค่าความยาก ค่าอำนาจจำแนก ประสิทธิภาพ ตัวลวง และนำผลการวิเคราะห์ข้อสอบไปใช้ในการปรับปรุงการออกข้อสอบในครั้งถัดไป นอกจากนี้โรงเรียนอาจได้รับการส่งเสริมสนับสนุนให้ความรู้และพัฒนาสมรรถนะครู ด้านการสร้างข้อสอบที่สอดคล้องกับตัวชี้วัด พฤติกรรมที่ต้องการวัดทั้งในระดับต้นสังกัดเขตพื้นที่ หรือกลุ่มโรงเรียน สอดคล้องกับผลการวิจัยการติดตามและตรวจสอบคุณภาพของข้อสอบที่ใช้ในการวัดและประเมินผลในชั้นเรียนของสถานศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2559) ที่พบว่า ทั้งกลุ่มโรงเรียนที่มีคะแนน O-NET สูงและต่ำ มีค่าเฉลี่ยร้อยละของข้อสอบที่ใช้ในการวัดและประเมินผลในชั้นเรียนของสถานศึกษาที่ตรงตามมาตรฐานและตัวชี้วัด ระดับชั้นมัธยมศึกษาปีที่ 3 กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ที่ไม่แตกต่างกัน ในระดับสูง เท่ากับร้อยละ 98.58 และ 95.89 ตามลำดับ และยังพบว่า ทั้งกลุ่มโรงเรียนที่มีคะแนน O-NET สูงและต่ำ มีความสัมพันธ์เชิงบวกระหว่างคะแนนวัดผลในชั้นเรียนกับคะแนนทดสอบ O-NET กลุ่มสาระการเรียนรู้วิทยาศาสตร์ โดยมีค่าเฉลี่ยร้อยละของข้อสอบที่ใช้ในการวัดและประเมินผลในชั้นเรียนที่ตรงตามมาตรฐานและตัวชี้วัด โดยภาพรวมสูงกว่า ร้อยละ 90 ทั้ง 2 กลุ่ม ซึ่งการที่ครูผู้สอนมีการวัดและประเมินผลโดยใช้ข้อสอบให้สอดคล้องหรือคล้ายคลึงกับข้อสอบ O-NET เป็นปัจจัยหนึ่งที่สามารถส่งเสริมการเพิ่มผลสัมฤทธิ์ของการสอบ O-NET ของผู้เรียนได้ (บริษัทอักษร

เจริญทัศน์ จำกัด, 2559) อย่างไรก็ตาม ปัจจัยที่มีความสัมพันธ์กับคะแนน O-NET ของโรงเรียน ยังมีอีกหลายปัจจัยที่มีความสัมพันธ์กับการเพิ่มผลสัมฤทธิ์ของการทดสอบ O-NET ได้แก่ ปัจจัยด้านตัวนักเรียน เช่น การตระหนักถึงความสำคัญของการสอบ O-NET พื้นฐานความรู้ของนักเรียน สภาพแวดล้อมทางครอบครัว สภาพแวดล้อมของโรงเรียน ปัจจัยด้านครูผู้สอน เช่น พฤติกรรมการสอนของครู ภาระงานของครู คุณวุฒิของครู การเข้าถึงข้อมูลของครู ปัจจัยด้านผู้บริหาร ปัจจัยด้านข้อสอบ O-NET เช่น ระดับความยากของข้อสอบ ช่วงเวลาในการสอบและการแจ้งผลการสอบ (เอื้อมพร หลินเจริญ สิริศักดิ์ อาจวิชัย และ ภริภา จันทร์อินทร์, 2552) ซึ่งปัจจัยดังกล่าวอาจเป็นปัจจัยหลักที่ส่งผลต่อผลสัมฤทธิ์ของการทดสอบ O-NET มากกว่าปัจจัยด้านการวัดและประเมินผลระดับชั้นเรียนเพียงปัจจัยเดียว

3.2 จากผลการศึกษาที่พบว่า ค่าดัชนีความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนทั้ง 4 โรงเรียน มีค่าอยู่ในระดับปานกลาง คือมีค่าดัชนีความสอดคล้องอยู่ระหว่าง 0.436 – 0.588 แสดงเห็นว่าข้อสอบในการประเมินระดับชั้นเรียนยังมีส่วนที่ค่อนข้างแตกต่างกับข้อสอบในการประเมินระดับชาติ ซึ่งความแตกต่างดังกล่าวอาจเนื่องมาจากข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน มีความแตกต่างกันในด้านความซับซ้อนของข้อสอบ (item complexity) ซึ่งเป็นความแตกต่างของจำนวนทักษะย่อย (subs-skills) ที่ผู้สอบจำเป็นต้องใช้ในการทำข้อสอบ โดยข้อสอบในการประเมินระดับชาติอาจมีความซับซ้อนของข้อสอบที่สูงกว่าข้อสอบในการประเมินระดับชั้นเรียน

3.3 จากผลการเปรียบเทียบร้อยละสัดส่วนระดับความซับซ้อนทางปัญญาระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน พบว่า ข้อสอบในการประเมินระดับชาติมีร้อยละสัดส่วนของข้อสอบที่วัดความซับซ้อนทางปัญญาระดับวิเคราะห์มากที่สุด รองลงมาคือ ระดับเข้าใจ ระดับประยุกต์ใช้ ระดับประเมินค่า และไม่มีข้อสอบที่วัดระดับจำ และระดับสร้างสรรค์ ในขณะที่ข้อสอบในการประเมินระดับชั้นเรียน ของโรงเรียนทั้ง 4 โรงเรียน โดยภาพรวม พบว่ามีร้อยละสัดส่วนของข้อสอบที่วัดระดับเข้าใจมากที่สุด รองลงมาคือ ระดับจำ ระดับประยุกต์ใช้ ระดับวิเคราะห์ ระดับประเมินค่า และไม่มีข้อสอบที่วัดระดับสร้างสรรค์ จะเห็นว่าข้อสอบในการประเมินระดับชาติจะเน้นการวัดความซับซ้อนทางปัญญาในระดับสูง คือระดับวิเคราะห์มากที่สุด ในขณะที่ข้อสอบในการประเมินระดับโรงเรียนจะเน้นการวัดพฤติกรรมการเรียนรู้ในระดับเข้าใจมากที่สุด รองลงมาคือ ระดับจำ ซึ่งทั้ง 2 ระดับ เป็นการวัดความซับซ้อนทางปัญญาสองระดับต่ำสุดตามแนวคิด revised Bloom's taxonomy ตลอดจนโรงเรียนมีสัดส่วนของข้อสอบที่วัดระดับวิเคราะห์ในสัดส่วนที่ต่ำเมื่อเทียบกับระดับเข้าใจและระดับจำ ซึ่งสอดคล้องกับรายงานการวิจัยการติดตามและตรวจสอบคุณภาพของข้อสอบที่ใช้ในการวัดและประเมินผลในชั้นเรียนของสถานศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษา

ขั้นพื้นฐาน, 2559) ที่พบว่า ข้อสอบที่ใช้ในการวัดผลในชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ส่วนใหญ่วัดพฤติกรรมระดับเข้าใจมากที่สุด รองลงมา คือ ระดับจำ ระดับประยุกต์ใช้ และระดับวิเคราะห์ ตามลำดับ ทั้งนี้อาจเนื่องมาจากการรับรู้และพฤติกรรมการออกข้อสอบของครูนั้น ข้อสอบที่วัดพฤติกรรมระดับความจำและความเข้าใจ เป็นข้อสอบที่สร้างขึ้นง่ายและสะดวก สอดคล้องกับงานวิจัยของพินดา วราสุนันท์ (2554) ที่พบว่าข้อสอบ O-NET วิทยาศาสตร์ มุ่งวัดระดับวิเคราะห์และเข้าใจ คำถามมีการยกสถานการณ์ที่ใหม่โดยเน้นให้นักเรียนได้อ่าน ในขณะที่ข้อสอบในการประเมินระดับชั้นเรียนวัดระดับความรู้ ตามแนวคิด Bloom taxonomy โดยเป็นข้อสอบที่ครูใช้ในกระบวนการเรียนการสอน และการวัดและประเมินผลในชั้นเรียน ส่วนใหญ่วัดระดับความรู้ นอกจากนี้ ยังพบว่าครูส่วนใหญ่รับรู้ว่ามีข้อสอบที่ใช้วัดระดับความรู้อยู่ในระดับปานกลาง และข้อสอบตามสภาพจริงที่ครูสร้างขึ้นในภาพรวมวัดระดับความรู้ถึงร้อยละ 58.64 และสอดคล้องกับงานวิจัยของ ตู จงรักษ์ (2543) ที่พบว่าข้อสอบในการประเมินระดับชั้นเรียนที่ครูสร้างขึ้นส่วนใหญ่มุ่งวัดพฤติกรรมระดับความรู้และความจำ ตามแนวคิด Bloom's taxonomy ความแตกต่างของระดับความซับซ้อนทางปัญญาระหว่างข้อสอบในการประเมินระดับชาติกับข้อสอบในการประเมินระดับชั้นเรียนดังกล่าว อาจเป็นสาเหตุที่ทำให้โดยภาพรวมโรงเรียนมีดัชนีความสอดคล้องอยู่ในระดับปานกลาง

3.4 จากผลการวิจัยที่พบว่า ข้อสอบในการศึกษาครั้งนี้ ไม่มีข้อสอบที่วัดระดับสร้างสรรค์ที่เป็นเช่นนี้อาจเป็นเพราะทั้งข้อสอบในการประเมินระดับชาติและข้อสอบในการประเมินระดับชั้นเรียน มีลักษณะเป็นข้อสอบปรนัย โดยข้อสอบในการประเมินระดับชาติทุกข้อเป็นข้อสอบแบบเลือกตอบหลายตัวเลือก และข้อสอบในการประเมินระดับชั้นเรียนเกือบทั้งหมดเป็นข้อสอบแบบเลือกตอบหลายตัวเลือก มีส่วนน้อยที่เป็นข้อสอบแบบเติมค่าและตอบสั้น ซึ่งข้อสอบปรนัยดังกล่าวมีข้อจำกัดในการวัดความรู้ความคิดระดับสูงในระดับสร้างสรรค์ ข้อสอบที่เหมาะสมสำหรับการวัดระดับสร้างสรรค์ จึงต้องเป็นข้อสอบแบบอัตนัย เนื่องจากการวัดความสามารถในการสร้างสิ่งใหม่จากสิ่งที่เคยเรียนรู้ หรือสิ่งที่พบเห็นในบริบทต่างๆ สามารถสร้างสรรค์งาน แผนงาน หรือผลิตภัณฑ์ หรือชิ้นงานที่แปลกใหม่ (Anderson et al., 2001)

**4. การประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน เมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน**

4.1 จากผลการวิจัยที่พบว่า ความแปรปรวนของคะแนนความสอดคล้องในแนวเดียวกันของข้อสอบมีค่ามากที่สุด เมื่อเทียบกับความแปรปรวนทั้งหมด ทั้งการประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรฐานประมาณค่า 5 ระดับ แสดงว่าความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัดของ

ข้อสอบแต่ละข้อมีความแตกต่างกัน อาจเนื่องมาจากข้อสอบแต่ละข้อได้รับการสุ่มจากโรงเรียนที่แตกต่างกัน ซึ่งปัจจัยพื้นฐานในการออกข้อสอบ หรือการบริหารจัดการด้านการวัดและประเมินผลที่แตกต่างกัน ส่วนความแปรปรวนอันเกิดจากผู้ประเมินมีค่าน้อยที่สุด เมื่อเทียบกับความแปรปรวนทั้งหมด แสดงให้เห็นว่าผู้ประเมินแต่ละคนมีการประเมินที่สอดคล้องกัน หรือมีความแตกต่างกันในการประเมินค่อนข้างน้อย ซึ่งสอดคล้องกับงานวิจัยของกนกวรรณ เอี่ยมชัย (2539) ที่พบว่า ผู้ตรวจเป็นแหล่งความแปรปรวนในการตรวจให้คะแนนแบบทดสอบความเรียงประยุกต์น้อยที่สุด สอดคล้องกับงานวิจัยของสมคิด เทียรพิสุทธิ์ (2550) ที่พบว่า ความแปรปรวนของผู้ตรวจมีค่าน้อยมาก และสอดคล้องกับงานวิจัยของจิรายุ เถาว์โท อนุ เจริญวงศ์ระยัย และปณณวิชญ์ ไบกุลหาลาบ (2559) ซึ่งพบว่า ความแปรปรวนของผู้ตรวจให้คะแนนแบบทดสอบอัตนัยมีค่าน้อยที่สุด ทั้งนี้อาจเนื่องมาจากงานวิจัยนี้ได้มีการคัดเลือกผู้เชี่ยวชาญที่ทำหน้าที่ประเมินที่มีคุณสมบัติเหมาะสม มีความรู้ความสามารถในด้านตัวชี้วัดและการสอนวิทยาศาสตร์ระดับชั้นมัธยมศึกษาตอนต้น อีกทั้งเป็นผู้ที่สำเร็จการศึกษาอย่างน้อยในระดับปริญญาตรีวิชาเอกด้านการสอนวิทยาศาสตร์ และสำเร็จการศึกษาอย่างน้อยในระดับปริญญาโททางด้านการวัดและประเมินทางการศึกษา และผู้เชี่ยวชาญส่วนใหญ่มีประสบการณ์ด้านการสอนวิทยาศาสตร์ไม่ต่ำกว่า 6 ปี นอกจากนี้ ยังมีการจัดประชุมและฝึกปฏิบัติก่อนทำหน้าที่ประเมินจริง เพื่อชี้แจงแนวทางการดำเนินงาน และทำความเข้าใจร่วมกับเกี่ยวกับรายละเอียดของเครื่องมือในการประเมินความสอดคล้องๆ ตลอดจนมีคู่มือสำหรับการดำเนินงานสำหรับผู้ประเมินที่ช่วยเพิ่มความชัดเจนของแนวทางในการดำเนินงาน ซึ่งสอดคล้องกับ Webb (1999) ที่ให้ความสำคัญกับการฝึกปฏิบัติของผู้เชี่ยวชาญก่อนดำเนินการประเมินความสอดคล้องในแนวเดียวกันๆ เพื่อให้ผู้เชี่ยวชาญทุกคนมีความเข้าใจถึงกระบวนการประเมินความสอดคล้องในแนวเดียวกัน หลักการจำแนกระดับความซับซ้อนทางปัญญา ตลอดจนการใช้ดุลพินิจในการตัดสินความสอดคล้องๆ ได้อย่างถูกต้องและปฏิบัติไปในทิศทางเดียวกัน และสอดคล้องกับ Patz, Junker, Johnson, and Mariano (2002) และ Welk, Schaben, and Morrow (2004) ที่กล่าวว่า ความคลาดเคลื่อนของการประเมินที่มาจากผู้ประเมินจะมีค่าน้อยมาก เมื่อมีเครื่องมือการประเมินที่กำหนดและอธิบายรายละเอียดของรายการประเมิน และมีเกณฑ์การประเมินที่ชัดเจน

4.2 จากผลการเปรียบเทียบค่าสัมประสิทธิ์การสรุปร่างของคะแนนการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนเมื่อจำนวนผู้ประเมินและรูปแบบการออกแบบการประเมินต่างกัน ที่พบว่าค่าสัมประสิทธิ์การสรุปร่างอ้างอิงสำหรับตัดสินใจเชิงสัมบูรณ์ของคะแนนการประเมินความสอดคล้องในแนวเดียวกันๆ มีค่าเพิ่มขึ้นเมื่อจำนวนผู้ประเมินเพิ่มขึ้น ในทุกรูปแบบการออกแบบการประเมิน ทั้งการประเมินระดับความซับซ้อนทางปัญญา และการประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ โดยค่าสัมประสิทธิ์การสรุปร่างอ้างอิงของคะแนนการประเมิน

ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด เมื่อมีจำนวนผู้ประเมินที่มากกว่า จะมีค่าความเที่ยงสูงกว่าจำนวนผู้เชี่ยวชาญที่น้อยกว่า ในทุกรูปแบบการออกแบบการประเมิน ซึ่งเป็นไปตามสมมติฐานการวิจัยข้อที่ 3 ที่กล่าวว่า “ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจของคะแนนผลการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ที่มีจำนวนผู้เชี่ยวชาญที่มากกว่า น่าจะมีค่าสูงกว่าจำนวนผู้ประเมินที่น้อยกว่า ในทุกรูปแบบการออกแบบการประเมิน ” สอดคล้องกับแนวคิดของ Herman, Webb, and Zuniga (2007) Lombardi, Sebuen, Conley, and Snow (2010) Porter and Smithson (2001) และ Webb (2007) ที่กล่าวว่าจำนวนผู้เชี่ยวชาญในการประเมินความสอดคล้องในแนวเดียวกันยิ่งมาก จะทำให้ผลการประเมินยิ่งมีความเที่ยงเพิ่มมากขึ้น และสอดคล้องกับงานวิจัยที่เป็นการศึกษาความเที่ยงในการตรวจประเมินที่พบว่า การเพิ่มจำนวนผู้ตรวจประเมินจะทำค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัดเพิ่มขึ้น เช่น งานวิจัยของวัชรียา ศรีวลีรัตน์ (2558) ที่พบว่า หากเพิ่มจำนวนผู้ตรวจข้อสอบเรียงความ ค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัดจะเพิ่มขึ้น และค่าความแปรปรวนของความคลาดเคลื่อนลดลงอย่างมีนัยสำคัญทางสถิติ เช่นเดียวกับงานวิจัยของดวงใจ สีเขียว (2549) ที่พบว่า คะแนนการประเมิน ผลสรุปรวม/สอบสอนของนิสิต/นักศึกษาฝึกประสบการณ์วิชาชีพครูที่มีจำนวนผู้ประเมินมากกว่าจะมีค่าความเที่ยงสูงกว่ากรณีที่มีผู้ประเมินน้อยกว่า ในทุกสถานการณ์การประเมิน และงานวิจัยของอังคณา กุลนภาดล (2558) ที่พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัดของคะแนนผังมโนทัศน์วิชาการวิจัยทางการศึกษา ที่มีจำนวนผู้ตรวจมากกว่า จะมีค่าสูงกว่าจำนวนผู้ตรวจน้อยกว่า ในทุกรูปแบบการตรวจ

นอกจากนี้ จากผลการวิจัยที่พบว่า รูปแบบการออกแบบการประเมิน 2 รูปแบบ คือ รูปแบบที่ 1 ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ และรูปแบบที่ 2 ผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อ ให้ค่าสัมประสิทธิ์การสรุปอ้างอิงที่ใกล้เคียงกัน เมื่อจำนวนผู้ประเมินเท่ากัน แสดงให้เห็นว่าในการประเมินความสอดคล้องในแนวเดียวกัน ผู้ประเมินแต่ละคนไม่จำเป็นต้องประเมินข้อสอบทุกข้อก็สามารถได้ผลการประเมินที่มีความน่าเชื่อถือใกล้เคียงกับการศึกษาที่ให้ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ ซึ่งเหมาะกับบริบทการศึกษาความสอดคล้องในแนวเดียวกันในกรณีที่มีข้อสอบในการศึกษาจำนวนมาก ที่อาจต้องใช้ผู้เชี่ยวชาญในการประเมินจำนวนมาก การออกแบบการประเมินโดยให้ผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อจะช่วยประหยัดเวลา แรงงาน และงบประมาณในการศึกษา ตลอดจนมีความเป็นไปได้ในการปฏิบัติ และให้ผลการศึกษาที่มีความน่าเชื่อถือเทียบเท่ากับการให้ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ

4.3 จากผลการพิจารณาจำนวนผู้ประเมินที่ทำให้เกิดความน่าเชื่อถือของการประเมินความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัดฯ ในระดับที่ยอมรับได้ โดยการพิจารณาค่าสัมประสิทธิ์การสรุปร่างอ้างอิงสำหรับตัดสินใจเชิงสัมบูรณ์ของสถานการณ์ในการประเมิน ตั้งแต่ 0.70 ขึ้นไป (Nunnally, 1987) พบว่า การประเมินระดับความซับซ้อนทางปัญญา จำนวนผู้ประเมิน 2 คนสามารถให้ผลการตัดสินที่มีความน่าเชื่อถืออยู่ในระดับที่ยอมรับได้ ในทุกรูปแบบการออกแบบการประเมิน ขณะที่การประเมินระดับความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับตัวชี้วัด ที่มีลักษณะเป็นมาตรฐานค่า 5 ระดับ ต้องการผู้ประเมินจำนวน ตั้งแต่ 3 คนขึ้นไป จึงจะให้ผลการประเมินที่มีความน่าเชื่อถืออยู่ในระดับที่ยอมรับได้ ในทุกรูปแบบการออกแบบการประเมิน ความแตกต่างของจำนวนผู้ประเมินดังกล่าวอาจเกิดจากคำอธิบายที่ใช้ในการประเมิน โดยในการประเมินระดับความซับซ้อนทางปัญญาตามแนวคิดของ revised Bloom's taxonomy ที่ประกอบด้วย 6 ระดับ ผู้ประเมินสามารถเข้าใจคำอธิบายและตัวอย่างของพฤติกรรมการเรียนรู้ในแต่ละระดับ เพื่อพิจารณาตัดสินระดับความซับซ้อนทางปัญญาของข้อสอบแต่ละข้อได้ดีกว่า การประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัดที่เป็นมาตรฐานค่า 5 ระดับ ที่ผู้ประเมินมีโอกาสที่ต้องใช้การตีความ และมีความเป็นปรนัยในการประเมินน้อยกว่าการประเมินตัดสินแต่ละระดับของความซับซ้อนทางปัญญา ซึ่งสอดคล้องกับงานวิจัยของ Lombardi, Sebuon, Conley, & Snow (2010) ที่พบว่า ในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบกับมาตรฐานวิชาคณิตศาสตร์และภาษาอังกฤษ ในการประเมินระดับความซับซ้อนทางปัญญาจะมีค่าความเที่ยงมากกว่าการประเมินด้วยมาตรฐานค่า 3 ระดับ เพื่อประเมินระดับความรู้ที่ผู้สอบต้องใช้ในการทำข้อสอบแต่ละข้อ ที่เรียกว่า rigor scale (โดยที่ 1 = ต่ำกว่าระดับความรู้ที่ผู้สอบควรมีในการสอบเข้ามหาวิทยาลัย, 2 = ระดับความรู้ที่ผู้สอบควรมีในการสอบเข้ามหาวิทยาลัย, 3 = สูงกว่าระดับความรู้ที่ผู้สอบควรมีในการสอบเข้ามหาวิทยาลัย) ทั้งนี้ ในภาพรวมพบว่า มาตรฐานประเมินความซับซ้อนทางปัญญามีคำอธิบายแต่ละระดับที่ชัดเจนมากกว่า จึงทำให้ง่ายต่อการประเมินและมีความเป็นปรนัยในการประเมิน

## ข้อเสนอแนะ

### ข้อเสนอแนะในการนำไปใช้

1. โรงเรียนที่ออกข้อสอบควรคัดเลือกข้อสอบที่มีความสอดคล้องในแนวเดียวกันกับมาตรฐานตัวชี้วัด จัดเก็บไว้ใช้ในการวัดและประเมินผลระดับชั้นเรียน ตลอดจนสามารถนำไปพัฒนาเป็นฐานข้อมูลหรือคลังข้อสอบต่อไป



2. องค์กรหรือหน่วยงาน เช่น สพฐ. สทศ. สำนักงานเขตพื้นที่การศึกษา ฯลฯ นักวิชาการ/ครู/ผู้บริหาร หรือผู้ที่สนใจศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานกับตัวชี้วัด กับข้อสอบในการประเมินชั้นเรียน ควรออกแบบการประเมินโดยใช้รูปแบบการออกแบบการประเมินที่ผู้ประเมินแต่ละคนประเมินข้อสอบบางข้อ โดยเฉพาะในกรณีที่มีการศึกษามีข้อสอบจำนวนมาก ซึ่งสามารถช่วยประหยัดเวลาและงบประมาณ ช่วยลดความซับซ้อนในการบริหารจัดการ และมีความเป็นไปได้ในการปฏิบัติ แต่ยังสามารถให้ค่าความเที่ยงที่ใกล้เคียงกับ การประเมินรูปแบบการออกแบบการประเมินที่ผู้ประเมินแต่ละคนประเมินข้อสอบทุกข้อ

3. องค์กรหรือหน่วยงาน เช่น สพฐ. สทศ. สำนักงานเขตพื้นที่การศึกษา ฯลฯ นักวิชาการ/ครู/ผู้บริหาร หรือผู้ที่สนใจศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานกับตัวชี้วัด กับข้อสอบในการประเมินชั้นเรียน ควรออกแบบการประเมินโดยใช้ผู้ประเมินข้อสอบในแต่ละข้ออย่างน้อย 2 คน ในการประเมินระดับความซับซ้อนทางปัญญา และใช้ผู้ประเมิน อย่างน้อย 3 คน ในการประเมินระดับความสอดคล้องฯ โดยใช้มาตรประมาณค่า ซึ่งจะให้ค่าความเที่ยงในการประเมินตั้งแต่ 0.70 ขึ้นไป อย่างไรก็ตาม หากเป็นการศึกษาความสอดคล้องในแนวเดียวกันของข้อสอบในระดับที่สูงกว่าข้อสอบที่ประเมินระดับชั้นเรียน ผู้วิจัยสามารถเลือกใช้เกณฑ์ความน่าเชื่อถือของการวัดที่มีความเหมาะสมกับงานวิจัย เช่น ข้อสอบประเมินระดับเขตพื้นที่ หรือระดับชาติ อาจกำหนดเกณฑ์ในการตัดสินความน่าเชื่อถือของผลการวัดที่มีความเข้มงวดมากขึ้น เช่น เกณฑ์ค่าความเที่ยง 0.80 (Sattler, 1998) ซึ่งจากผลการวิจัยต้องใช้ผู้ประเมินข้อสอบในแต่ละข้อ อย่างน้อย 5 คน หรือเกณฑ์ค่าความเที่ยง 0.90 (Gay, 1992) ซึ่งจากผลการวิจัยต้องใช้ผู้ประเมิน อย่างน้อย 8 คน

4. องค์กรหรือหน่วยงาน เช่น สพฐ. สทศ. สำนักงานเขตพื้นที่การศึกษา ฯลฯ นักวิชาการ/ครู/ผู้บริหาร หรือผู้ที่สนใจศึกษาความสอดคล้องในแนวเดียวกันข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ควรนำคู่มือแนวทางการประเมินความสอดคล้องในแนวเดียวกัน ข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียนตามแนวคิดของ Porter ที่ผู้วิจัยสร้างขึ้นจากผลการศึกษาของงานวิจัยนี้ ไปประยุกต์ใช้ในการออกแบบการประเมินความสอดคล้องในแนวเดียวกันทั้งระหว่างมาตรฐานตัวชี้วัดกับข้อสอบในกลุ่มสามารถการเรียนรู้และในระดับต่างๆ หรือระหว่างข้อสอบในการประเมินระดับชาติกับข้อสอบในการประเมินระดับชั้นเรียน

5. ในการประเมินความสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบทางการศึกษา ควรมีการอบรมและฝึกปฏิบัติผู้ประเมินให้ความเข้าใจที่ตรงกันถึงแนวทางในการดำเนินงาน เพื่อช่วยลดปัญหาอันเกิดจากอิทธิพลของผู้ประเมิน

6. โรงเรียนควรปรับปรุงการออกข้อสอบที่วัดความซับซ้อนทางปัญญาในระดับวิเคราะห์ และประยุกต์ใช้ให้มากขึ้น เพื่อให้ผู้เรียนฝึกการคิดในระดับสูง และเป็นการสร้างพื้นฐานที่เกิดความคุ้นเคย

และเตรียมตัวเพื่อรับการประเมินด้วยข้อสอบระดับชาติ ตลอดจนการประเมินในการศึกษาระดับที่สูงขึ้นต่อไป

### ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. เนื่องจากงานวิจัยนี้เป็นการศึกษาแบบย้อนหลัง (retrospective study) ในการวิจัยครั้งต่อไปจึงควรมีการประยุกต์ใช้ MFRM ในการวิเคราะห์ความสอดคล้องระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ที่เป็นการศึกษาแบบไปข้างหน้า (prospective study) โดยให้การศึกษาความสอดคล้องมาตรฐานและตัวชี้วัดกับข้อสอบ เป็นขั้นตอนหนึ่งในกระบวนการสร้างแบบสอบ ซึ่งสารสนเทศที่ได้จากการวิเคราะห์ความสอดคล้องฯ จะเป็นประโยชน์ต่อคุณภาพการสร้างแบบสอบในด้านความตรงมากยิ่งขึ้น

2. เนื่องจากการวิเคราะห์ด้วย MFRM ในงานวิจัยนี้ อยู่บนพื้นฐานของ Andrich's rating scale model (Andrich, 1978) ซึ่งมีข้อตกลงเบื้องต้นว่าผู้ประเมินทุกคนมีความสามารถในการจำแนกความสอดคล้องระหว่างข้อสอบกับมาตรฐานและตัวชี้วัดไม่แตกต่างกัน ในงานวิจัยครั้งต่อไปอาจเลือกใช้โมเดลอื่นในการศึกษา เช่น Masters's partial credit model (Masters, 1982) ซึ่งอนุญาตให้ความสามารถในการจำแนกระหว่างข้อสอบกับมาตรฐานและตัวชี้วัดแปรผันได้ตามผู้ประเมิน เช่นเดียวกับโมเดลทั่วไปที่กำหนดให้ผู้ประเมินเป็นองค์ประกอบอย่างสุ่ม (random facet) ที่แปรผันตามข้อสอบ (Muckle & Karabatsos, 2009)

3. ควรมีการประยุกต์ใช้ MFRM ในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบต่างๆ ในระบบการศึกษา เช่น ระหว่างมาตรฐานกับการจัดการเรียนการสอน หรือการประเมินผลกับการเรียนการสอน โดยอาจเป็นการพัฒนาวิธีการศึกษาความสอดคล้องในแนวเดียวกันขึ้นใหม่ หรือประยุกต์ใช้วิธีการที่มีอยู่ เช่น วิธีการของ Webb วิธีการของ Porter หรือวิธีการของบริษัท Achieve

4. ควรมีการศึกษาปัจจัยอื่นๆ ที่ก่อให้เกิดความไม่สอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติกับข้อสอบในการประเมินระดับชั้นเรียน เช่น ปัจจัยเกี่ยวกับนักเรียน ปัจจัยเกี่ยวกับครูผู้สอน ปัจจัยเกี่ยวกับผู้บริหาร ปัจจัยเกี่ยวกับการวัดและประเมินผล

5. การศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติกับข้อสอบในการประเมินระดับชั้นเรียนในครั้งต่อไป ควรมีการพิจารณาถึงความสอดคล้องระหว่างความซับซ้อนของข้อสอบ (item complexity) โดยเป็นการเปรียบเทียบจำนวนทักษะย่อย (subs-skills) ที่ใช้ในการออกข้อสอบ ซึ่งจะสามารถจำแนกความแตกต่างของข้อสอบได้อย่างละเอียดและชัดเจนกว่าการจำแนกตามระดับความซับซ้อนทางปัญญาเพียงอย่างเดียว

6. การศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียนในครั้งต่อไป อาจมีการศึกษาโดยใช้มาตรฐานประเมินความสอดคล้องๆ ที่ไม่จำเป็นต้องมี 5 ระดับ (ไม่สอดคล้อง ค่อนข้างไม่สอดคล้อง ไม่แน่ใจ ค่อนข้างสอดคล้อง และสอดคล้องโดยตรง) ตามที่ใช้ในงานวิจัยนี้ เช่น การใช้มาตรฐานประเมินที่แบ่งเป็น 3 ระดับ (ไม่สอดคล้อง ไม่แน่ใจ และสอดคล้อง) ซึ่งอาจทำให้การประเมินมีความเป็นปรนัยมากยิ่งขึ้น



## รายการอ้างอิง

### ภาษาไทย

- กนกวรรณ เอี่ยมชัย. (2539). การศึกษาค่าความเชื่อมั่นของแบบทดสอบ เอ็ม อี คิว วิชาแนวคิด  
พื้นฐานและหลักการพยาบาลที่ตรวจให้คะแนนต่างกัน (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต),  
สาขาวิชาการวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพมหานคร.
- กระทรวงศึกษาธิการ. (2544). หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544. กรุงเทพมหานคร:  
โรงพิมพ์คุรุสภาลาดพร้าว.
- กระทรวงศึกษาธิการ. (2552). หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551.  
กรุงเทพมหานคร: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย.
- กฤตวรรณ ทงมี. (2557). การประเมินความสอดคล้องระหว่างมาตรฐานคุณวุฒิระดับอุดมศึกษา  
แห่งชาติกับการจัดการเรียนการสอนของภาควิชาคัลยกรรม คณะแพทยศาสตร์ มหาวิทยาลัย  
นเรศวร. (วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต), สาขาวิชาการประเมินการศึกษา  
มหาวิทยาลัยสุโขทัยธรรมมาธิราช, นนทบุรี.
- ชาญวิทย์ จรัสสุทธิอิสร์. (2550). การแสดงหลักฐานความเที่ยงตรงเชิงโครงสร้างของคะแนนการ  
ประเมินทักษะกระบวนการทางวิทยาศาสตร์โดยใช้ทฤษฎีมาตรฐานเดิมและทฤษฎีการตอบ  
ข้อคำถามตามโมเดลของราล์ชแบบหลายองค์ประกอบ. (วิทยานิพนธ์ปริญญาการศึกษาดุขฎี  
บัณฑิต), บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพมหานคร.
- ดวงใจ สีเขียว. (2549). การพัฒนาระบบการประเมินนิสิต/นักศึกษาฝึกประสบการณ์วิชาชีพครู ตาม  
แนวคิดการประเมินแบบ 360 องศา โดยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิง. (วิทยานิพนธ์  
ปริญญาดุขฎีบัณฑิต), สาขาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์  
มหาวิทยาลัย, กรุงเทพมหานคร.
- ทรรศนีย์ ปั้นประเสริฐ. (2556). การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตร  
แกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กลุ่มสาระ การเรียนรู้ภาษาไทย ชั้น  
ประถมศึกษาปีที่ 6. (วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต), บัณฑิตวิทยาลัย  
มหาวิทยาลัยศิลปากร, นครปฐม.
- นราศักดิ์ ไชยเรือง. (2557). ผลกระทบของความไม่สอดคล้องระหว่างมาตรฐานการประเมินกับการ  
ประเมินของครูต่อผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 4

- ในจังหวัดเชียงใหม่. (วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต), สาขาวิชาการประเมิน การศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช, นนทบุรี.
- น้ำผึ้ง อินทเนตร. (2554). การศึกษาคุณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ เมื่อ จำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่างกัน โดยใช้โมเดลการสรุปอ้างอิงและ โมเดลหลายองค์ประกอบของราล์ซ. (วิทยานิพนธ์ปริญญาการศึกษาดุสิตบัณฑิต), บัณฑิต วิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพมหานคร. Retrieved from [http://thesis.swu.ac.th/swudis/Test\\_Mea/Nampueng\\_I.pdf](http://thesis.swu.ac.th/swudis/Test_Mea/Nampueng_I.pdf)
- บริษัทอักษรเจริญทัศน์ จำกัด. (2559). การเพิ่มผลสัมฤทธิ์ของการสอบ O-NET. Retrieved from [https://www.google.co.th/?gws\\_rd=cr,ssl&ei](https://www.google.co.th/?gws_rd=cr,ssl&ei)
- บุษยา ปักการะนิง. (2556). การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตร แกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กลุ่มสาระการเรียนรู้ภาษาไทย ชั้น มัธยมศึกษาปีที่ 4. (วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต), บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร, นครปฐม.
- บุษวรรรษ์ แสนปลื้ม และองอาจ นัยพัฒน์. (2556). การศึกษาความสอดคล้องของการระบุความ เข้มงวด/ใจดี และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ของผู้ตรวจให้ คะแนนความสามารถในการเขียนเรียงความของนักเรียนชั้นประถมศึกษาปีที่ 3. วารสาร ศิลปการศึกษาศาสตร์วิจัย, 5(2), 335-347.
- พดุม ไบระหมาน สังวรณั ังตกระโทก และ นลินี ฒ นคร. (2559). ผลของความสอดคล้องระหว่าง เครื่องมือประเมินกับกรอบมาตรฐานความสามารถทางภาษาของสหภาพยุโรปที่มีต่อ ผลสัมฤทธิ์ทางภาษาฝรั่งเศสของนักเรียนชั้นมัธยมศึกษาปีที่ 5. เอกสารประกอบการ ประชุมสัมมนา การวิจัย วัดผลสัมพัธ์แห่งประเทศไทย ครั้งที่ 24. มหาวิทยาลัยราชภัฏพิบูล สงคราม, พิษณุโลก.
- พินดา วราสุนันท์. (2554). การพัฒนาศักยภาพทางการประเมินในด้านการสร้างข้อสอบของครู ประถมศึกษาโดยใช้เครือข่ายมิตรวิพากษ์. (วิทยานิพนธ์ปริญญาครุศาสตรดุสิตบัณฑิต), สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพมหานคร.
- มนชิตา เรืองรัมย์. (2556). การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตรแกนกลาง การศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กลุ่มสาระการเรียนรู้ภาษาไทย ชั้นมัธยมศึกษาปีที่ 2. (วิทยานิพนธ์หลักสูตรปริญญาศึกษาศาสตรมหาบัณฑิต), บัณฑิตวิทยาลัย มหาวิทยาลัย ศิลปากร, นครปฐม.

- มาณี ปรีกมานนท์. (2556). *การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตรแกนกลาง การศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กลุ่มสาระการเรียนรู้ภาษาไทย ชั้นมัธยมศึกษาปีที่ 1*. (วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต), บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร, นครปฐม.
- รายุ เถาว์โท อนุ เจริญวงศ์ระยัย และปิ่นณวิชญ์ ไบกุลลาบ. (2559). การศึกษาค่าความเชื่อมั่นของ คะแนนแบบทดสอบอัตนัยวิชาคณิตศาสตร์ของ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ที่มีจำนวน ผู้ตรวจและรูปแบบการตรวจให้คะแนน ต่างกัน โดยใช้ทฤษฎีการสรุปอ้างอิง. *วารสาร หาดใหญ่วิชาการ*, 14(1), 1-14.
- วัชรียา ศรีวลีรัตน์. (2558). การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการ วัดของแบบทดสอบเรียงความวิชาภาษาไทย ชั้นมัธยมศึกษาปีที่ 6 ที่มีจำนวนข้อสอบและ จำนวนผู้ตรวจต่างกัน. *วารสารวิจัยและพัฒนาหลักสูตร*, 5(2), 14-19.
- วีรารวรรณ โฆษิตสุด. (2556). *การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตร แกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กลุ่มสาระการเรียนรู้ภาษาไทย ชั้น มัธยมศึกษาปีที่ 3*. (วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต), บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร, นครปฐม.
- ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่*. (พิมพ์ครั้งที่ 4). กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม*. (พิมพ์ครั้งที่ 7). กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศุภรัตน์ อิงชาติเจริญ. (2557). *การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชา ที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราล์ฟ*. (วิทยานิพนธ์ปริญญาครุศาสตรดุษฎีบัณฑิต), สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพมหานคร.
- สมคิด เทียรพิสุทธิ์. (2550). *การเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบความเรียงประยุกต์วิชา คณิตศาสตร์ที่มีวิธีการตรวจและจำนวนผู้ตรวจต่างกัน*. (วิทยานิพนธ์ปริญญาการมหาบัณฑิต), สาขาวิชาการวัดผลการศึกษา มหาวิทยาลัยทักษิณ, สงขลา.
- สังวรรณ รัตกระโทก. (2555). การวัดความสอดคล้องของมาตรฐานการเรียนรู้กับการจัดการเรียนการสอนและการประเมิน. In *เอกสารการสอนชุดวิชา การวัดและประเมินอิง มาตรฐานการเรียนรู้ (หน่วยที่ 6)*. นนทบุรี: สำนักพิมพ์มหาวิทยาลัยสุโขทัยธรรมมาธิราช.
- สำนักงานทดสอบแห่งชาติ. (2559). รูปแบบข้อสอบ O-NET ม. 3 ปีการศึกษา 2559. Retrieved from <http://www.niets.or.th/th/catalog/view/213>

- สำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2559). รายงานการติดตามและตรวจสอบคุณภาพของข้อสอบที่ใช้ในการวัดและประเมินผลในชั้นเรียนของสถานศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. Retrieved from <http://bet.obec.go.th/index/?p=4794>
- สำนักนายกรัฐมนตรี สำนักงานคณะกรรมการการศึกษาแห่งชาติ. (2542). พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542. กรุงเทพมหานคร: โรงพิมพ์องค์การรับส่งสินค้าและพัสดุภัณฑ์ (ร.ส.พ.).
- สุพร ลัทธิจิตต์. (2556). การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กลุ่มสาระการเรียนรู้ภาษาไทย ชั้นประถมศึกษาปีที่ 4. (วิทยานิพนธ์ปริญญาศึกษาศาสตรมหาบัณฑิต), บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร, นครปฐม.
- อังคณา กุลภาดล. (2557). การเปรียบเทียบค่าสัมประสิทธิ์การสรุปร่างของคะแนนแผนผังมโนทัศน์วิชาการวิจัยทางการศึกษา เมื่อรูปแบบการตรวจและจำนวนผู้ตรวจต่างกัน. *วารสารศึกษาศาสตร์*, 25(3), 39-50.
- อุษณีย์ บัวศิริพันธุ์. (2543). การเปรียบเทียบค่าสัมประสิทธิ์การสรุปร่างของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจจำนวนผู้ตรวจและประสบการณ์ของผู้ตรวจแตกต่างกัน. (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต), สาขาวิชาการวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ, กรุงเทพมหานคร.
- เอี่ยมพร หลินเจริญ สิริศักดิ์ อาจวิชัย และภริภา จันทร์อินทร์. (2552). ปัจจัยเชิงสาเหตุที่ทำให้คะแนนการทดสอบ O-NET ของนักเรียนชั้นประถมศึกษาปีที่ 6 และชั้นมัธยมศึกษาปีที่ 6 ต่ำ. Retrieved from <http://www.niets.or.th/th/content/download/279>

### ภาษาอังกฤษ

- Ananda, S. (2003). *Rethinking Issues of Alignment under "No Child Left Behind."* Knowledge Brief. San Francisco: WestEd.
- Anderson, D., Irvin, S., Alonzo, J., & Tindal, G. A. (2015). Gauging item alignment through online systems while controlling for rater effects. *Educational Measurement: Issues and Practice*, 34(1), 22-33.
- Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory into practice*, 41(4), 255-260.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., . . . Wittrock, M. C. (2001). *A taxonomy for learning, teaching,*

- and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete ed.). New York: Longman.
- Andrew, T. R., & Dawn, M. (2010). Aligning an early childhood assessment to state kindergarten content standards: application of a nationally recognized alignment framework. *Educational Measurement: Issues and Practice*, 29(1), 25-37.
- Andrich, D. A. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573. doi:10.1007/BF02293814
- Barrett, S. (2005). Raters and examination. In S. Alagumalai, D. D. Curtis, & N. Hungi (Eds.), *Applied Rasch measurement: A book of exemplars* (pp. 159-178). Netherlands: Springer.
- Basturk, R. (2008). Applying the many-facet Rasch model to evaluate PowerPoint presentation performance in higher education. *Assessment & Evaluation in Higher Education*, 33(4), 431–444.
- Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice*, 22(3), 21-29.
- Biggs, J. (2003). *Teaching for quality learning at university*. Glasgow: The Society for Research into Higher Education & Open University Press.
- Blank, R. K., Porter, A., & Smithson, J. (2001). *New tools for analyzing teaching, curriculum and standards in mathematics & science. Results from survey of enacted curriculum project. final report*. Washington, DC: Council of Chief State School Officers.
- Blank, R. K., & Smithson, J. L. (2010). *Alignment content analysis of NAEP 2009 science assessments with the methodology of surveys of enacted curriculum*. Washington, DC: Council of Chief State School Officers.
- Case, B., Jorgensen, M. A., & Zucker, S. (2004). *Alignment in educational assessment*. San Antonio: Harcourt Assessment.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* Hillsdale, NJ: Lawrence Earlbaum Associates.



- D'Agostino, V. J., Welsh, E. M., Cimetta, D. A., Falco, D. L., Smith, S., VanWinkle, H. W., & Powers, J. S. (2008). The rating and matching item-objective alignment methods. *Applied Measurement in Education*, 21, 1–21.  
doi:10.1080/08957340701580728
- Davis-Becker, S. L., & Buckendahl, C. W. (2013). A proposed framework for evaluating alignment studies. *Educational Measurement: Issues and Practice*, 32(1), 23–33.
- Eckes, T. (2005). Examining rater effects in TestDaF writing and speaking performance assessments: A many-facet Rasch analysis. *Language Assessment Quarterly*, 2(3), 197–221.
- Eckes, T. (2008). Rater types in writing performance assessments: a classification approach to rater variability. *Language Testing*, 25(2), 155-185.
- Eckhout, T. J., Plake, B. S., Smith, D. L., & Larsen, A. (2007). Aligning a state's alternative standards to regular core content standards in reading and mathematics: A case study. *Applied Measurement in Education*, 20(1), 79-100.
- Engelhard, G. J. (1992). The measurement of writing ability with a many-facet Rasch model. *Applied Measurement in Education*, 5(3), 171-191.
- Engelhard, G. J. (1994). Examining rater errors in the assessment of written composition with a many-facet Rasch model. *Journal of Educational Measurement*, 31(2), 93-112.
- Engelhard G. J. (1996). Evaluating rater accuracy in the performance assessments. *Journal of Educational Measurement*, 33(1), 56-70.
- Farenga, S. J., Joyce, B. A., & Ness, D. (2002). Reaching the zone of optimal learning: The alignment of curriculum, instruction, and assessment. In R. W. Bybees (Ed.), *Learning science and the science of learning* (pp. 51-62). Arlington: NSTA press.
- Farrokhi, F., & Esfandiari, R. (2011). A many-facet Rasch model to detect halo effect in three types of raters. *Theory and Practice in Language Studies*, 1(11), 1531-1540. doi:10.4304/tpls.1.11.1531-1540
- Farrokhi, F., Esfandiari, R., & Dalili, V. M. (2011). Applying the many-facet Rasch model to detect centrality in self-assessment, peer-assessment and teacher

- assessment. *World Applied Sciences Journal*, 15(Innovation and Pedagogy for Lifelong Learning), 70-77.
- Farrokhi, F., Esfandiari, R., & Schaefer, E. (2012). A many-facet Rasch measurement of differential rater severity/leniency in three types of assessment. *JALT Journal*, 34(1), 79-101.
- Fisher, W. J. (1992). Reliability, separation, strata statistics. *Rasch Measurement Transaction*, 6(3), 238.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378-382. doi:10.1037/h0031619
- Fleiss, J. L., Levin, B. A., & Paik, M. C. (2003). *Statistical methods for rates and proportions*. Hoboken, NJ: John Wiley & Sons, Inc.
- Fonthal, G. (2004). *Alignment of state assessments and higher education expectations: Definition and utilization of an alignment index*. University of California, Irvine.
- Fulmer, G. W., & Polikoff, M. S. (2014). Tests of alignment among assessment, standards, and instruction using generalized linear model regression. *Educational Assessment, Evaluation and Accountability*, 26(3), 225-240.
- Gay, L. R. (1992). *Educational research competencies for analysis and application*. New York: Macmillan Publishing.
- Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, 9(3), 330-338.
- Herman, J. L., Webb, N. M., & Zuniga, S. A. (2007). Measurement issues in the alignment of standards and assessments. *Applied Measurement in Education*, 20(1), 101-126. doi:10.1080/08957340709336732
- Hung, S.-P., Chen, P.-H., & Chen, H.-C. (2012). Improving creativity performance assessment: A rater effect examination with many facet Rasch model. *Creativity Research Journal*, 24(4), 345-357.
- Iramaneerat, C., & Yudkowsky, R. (2007). Rater errors in a clinical skills assessment of medical students. *Evaluation & the Health Professions*, 30(3), 266-283.

- Iramaneerat, C., Yudkowsky, R., Myford, C. M., & Downing, S. M. (2008). Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. *Advances in Health Sciences Education, 13*(4), 479-493.
- Karakaya, İ. (2015). Comparison of self, peer and instructor assessments in the portfolio assessment by using many facet RASCH model. *Journal of Education and Human Development, 4*(2), 182-192.
- Kassim, N. L. A. (2011). Judging behaviour and rater errors: an application of the many-facet Rasch model. *GEMA Online Journal of Language Studies, 11*(3), 179-197.
- Kocher, M. G., & Sutter, M. (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *The Economic Journal, 115*(500), 200-223.
- La Marca, P. M. (2001). Alignment of standards and assessments as an accountability criterion. In: ERIC Clearinghouse on Assessment and Evaluation.
- La Marca, P. M., Redfield, D., & Winter, P. C. (2000). *State standards and state assessment systems: a guide to alignment*. Washington, DC: Council of Chief State School Officers.
- Linacre, J. M. (1994). *Many-facet Rasch measurement*. Chicago: MESA Press.
- Linacre, J. M. (2017). *Facets computer program for man-facet Rasch measurement (Version 3.80.3)*. Beaverton, OR: Winsteps.com.
- Linacre, J. M. (2018). *A user's guide to FACETS Rasch-model computer programs. Program Manual 3.80.4*. Beaverton, Oregon: Winsteps.com.
- Liu, J., & Xie, L. (2014). Examining rater effects in a WDCT pragmatics test. *Iranian Journal of Language Testing, 4*(1), 50-65.
- Liu, X., & Fulmer, G. (2008). Alignment between the science curriculum and assessment in selected NY state regents exams. *Journal of Science Education and Technology, 17*, 373-383.
- Lombardi, A., Seburn, M., Conley, D., & Snow, E. (2010). *A generalizability investigation of cognitive demand and rigor ratings of items and standards in an alignment study*. Paper presented at the Annual Meeting of the American Educational Research Association, Denver, CO.

- Lunz, M. E., Wright, B. D., & Linacre, J. M. (1990). Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3(4), 331-345.
- Marzano, R. J., & Kendall, J. S. (2001). *The new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174.
- Muckle, T. J., & Karabatsos, G. (2009). Hierarchical generalized linear models for the analysis of judge ratings. *Journal of Educational Measurement*, 46(2), 198-219.
- Myford, C. M., & Wolfe, E. W. (2000). *Monitoring sources of variability within the test of spoken english assessment system (TOEFL Research Report No. 65)*. Retrieved from Princeton, NJ: <http://www.ets.org/Media/Research/pdf/RR-00-06-Myford.pdf>
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of applied measurement*, 4(4), 386-422.
- Myford, C. M., & Wolfe, E. W. (2004). Detecting and measuring rater effects using many-facet Rasch measurement: Part II. *Journal of applied measurement*, 5(2), 189-227.
- Myford, C. M., & Wolfe, E. W. (2009). Monitoring rater performance over time: A framework for detecting differential accuracy and differential scale category use. *Journal of Educational Measurement*, 46(4), 371-389.
- Nakamura, Y. (2002). Effectiveness of paired rating in the assessment of English compositions. *The Japan Language Testing Association Journal*, 5, 61-71.
- Näsström, G., & Henriksson, W. (2008). Alignment of standards and assessment: A theoretical and empirical study of methods for alignment. *Electronic Journal of Research in Educational Psychology*, 6(3), 667-690.
- Nunnally, J. C. (1987). *Psychometric theory*. New York, NY: McGraw-Hill.
- Park, T. (2004). An investigation of an ESL placement test of writing using many-facet Rasch measurement. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 4(1), 1-21. doi:10.7916/D8T15G49

- Patz, R. J., Junker, B. W., Johnson, M. S., & Mariano, L. T. (2002). The hierarchical rater model for rated test items and its application to large-scale educational assessment data. *Journal of Educational and Behavioral Statistics, 27*(4), 341-384.
- Polikoff, M. S. (2015). How well aligned are textbooks to the common core standards in mathematics? *American Educational Research Journal, 52*(6), 1185-1211.
- Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis, 36*(4), 399-416. doi:10.3102/0162373714531851
- Polikoff, M. S., Zhou, N., & Campbell, S. E. (2015). Methodological choices in the content analysis of textbooks for measuring alignment with standards. *Educational Measurement: Issues and Practice, 34*(3), 10-17.
- Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, Va: Association for Supervision and Curriculum Development.
- Porter, A. C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher, 31*(7), 3-14
- Porter, A. C., Smithson, J., Blank, R., & Zeidner, T. (2007). Alignment as a teacher variable. *Applied Measurement in Education, 20*(1), 27-51. doi:10.1207/s15324818ame2001\_3
- Porter, A. C., & Smithson, J. L. (2001). *Defining, developing, and using curriculum indicators*. CPRE research report series RR-048. Retrieved from Philadelphia, PA: University of Pennsylvania Graduate School of Education:
- Porter, A. C., & Smithson, J. L. (2002). *Alignment of assessments, standards and instruction using curriculum indicator data*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.
- Portney, L. G., & Watkins, M. P. (2015). *Foundations of clinical research: Applications to practice*. Philadelphia, PA: F. A. Davis Company.
- Rothman, R., Slattery, J. B., Vranek, J. L., & Resnick, L. B. (2002). *Benchmarking and alignment of standards and testing*. CSE technical report (CSE-TR-566). Retrieved from

<http://ezproxy.car.chula.ac.th/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED466642&site=eds-live>

- Sattler, J. M. (1988). *Assessment of children*. San Diego, CA: J.M. Sattler.
- Scullen, S. E., Mount, M. K., & Goff, M. (2000). Understanding the latent structure of job performance ratings. *Journal of Applied Psychology, 85*(6), 956-970.
- Shankararaman, V., & Gottipati, S. (2014). *Aligning assessments with competencies using keyphrase extraction*. Paper presented at the 2014 International Conference on Teaching, Assessment and Learning (TALE) Wellington, New Zealand.
- Shelton, A. R., & Brown, R. S. (2008). *Measuring the alignment of high school and community college math assessments*. Paper presented at the annual meeting of the American Educational Research Association, New York City.
- Smith, J. E. V., & Kulikowich, J. M. (2004). An application of generalizability theory and many-facet Rasch measurement using a complex problem-solving skills assessment. *Educational and Psychological Measurement, 64*(4), 617-639.
- Smith, M. S., & O'Day, J. (1990). Systemic school reform. In S. H. Fuhman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-267). London: The Falmer Press.
- Song, T., & Wolfe, E. W. (2015). *Distinguishing several rater effects with the Rasch model*. Paper presented at the National Council of Measurement in Education Annual Meeting, Chicago, IL.
- Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2005). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing, 9*(3), 239-261.
- The Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards : how to assess evaluations of educational programs*. Thousand Oaks, California: Sage.
- Toffoli, S. F. L., de Andrade, D. F., & Bornia, A. C. (2016). Evaluation of open items using the many-facet Rasch model. *Journal of Applied Statistics, 43*(2), 299-316. doi:10.1080/02664763.2015.1049938

- Turner, J. (2003). *Examining an art portfolio assessment using a many-facet Rasch measurement model*. (doctoral dissertation), Boston College.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. (Research Monograph No. 8). Washington, DC: Council of Chief State School Officers.
- Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states*. (Research Monograph No. 18). Washington, DC: Council of Chief State School Officers.
- Webb, N. L. (2002). *An analysis of the alignment between mathematics standards and assessments for three states*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20(1), 7-25.
- Welk, G. J., Schaben, J. A., & Morrow, J. J. (2004). Reliability of accelerometry-based activity monitors: A generalizability study. *Medicine and science in sports and exercise*, 36(9), 1637-1645.
- Wesolowski, B. C., Wind, S. A., & Engelhard Jr, G. (2015). Rater fairness in music performance assessment: evaluating model-data fit and differential rater functioning. *Musicae Scientiae*, 19(2), 147-170.
- Wiseman, C. S. (2012). A comparison of the performance of analytic vs. holistic scoring rubrics to assess L2 writing. *Iranian Journal of Language Testing*, 2(1), 59-92.
- Wolfe, E. W. (2004). Identifying rater effects using latent trait models. *Psychology Science*, 46(1), 35-51.
- Wolfe, E. W., & McVay, A. (2012). Application of latent trait Models to identifying substantively interesting raters. *Educational Measurement: Issues & Practice*, 31(3), 31-37. doi:10.1111/j.1745-3992.2012.00241.x
- Wright, B. D., & Master, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago, IL: MESA Press.

- Wu, S., & Dou, T. (2015). Validation of an oral english test based on many-faceted Rasch model. *Journal of Language Teaching and Research*, 6(4), 866-872. doi:10.17507/jltr.0604.21
- Zhu, W., Ennis, C. D., & Chen, A. (1998). Many-faceted Rasch modeling expert judgment in test development. *Measurement in Physical Education and Exercise Science*, 2(1), 21-39.







ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**



ภาคผนวก ก  
รายนามผู้เชี่ยวชาญในการวิจัย

จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

### รายนามผู้เชี่ยวชาญในการวิจัย

การตรวจสอบและทดลองใช้แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น

- |                                    |   |
|------------------------------------|---|
| 1. อาจารย์ ดร.ศศิธร เขียวกอ        | ครูโรงเรียนพญาไท  |
| 2. อาจารย์ ดร.สุรศักดิ์ เก้าเอี้ยน | อาจารย์ประจำคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย                                      |
| 3. อาจารย์ ดร.เมษา นวลศรี          | อาจารย์ประจำคณะครุศาสตร์<br>มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์         |
| 4. อาจารย์ ดร.นิติกร อ่อนโยน       | อาจารย์ประจำคณะครุศาสตร์<br>มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์         |
| 5. อาจารย์ธัญปนา จ้อยเจริญ         | อาจารย์ประจำคณะครุศาสตร์<br>มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์         |
| 6. อาจารย์ ดร.อนันดา สันฐิตินิชย์  | อาจารย์ประจำมหาวิทยาลัยสุโขทัยธรรมมาธิราช   |
| 7. ดร.จุฑาภรณ์ มาสันเทียะ          | คณาจารย์บัณฑิต สาขาการวัดและประเมินผลการศึกษา<br>คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย |

การประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น

- |  |  |
|--|--|
| 1. อาจารย์ ดร.เพียรกิจ นิมิตรดี                          | ครูโรงเรียนแม่ใจวิทยาคม                              |
| 2. อาจารย์ ดร.ขวัญฤทัย<br>(เที่ยงจันทร์ทาทิพย์) วงศ์ใหญ่ | ครูโรงเรียนเฉลิมพระเกียรติสมเด็จพระศรีนครินทร์ พะเยา |
| 3. อาจารย์ ดร.กุลรดี เอกสุวรรณ                           | ครูโรงเรียนบ้านหมีวิทยา                              |
| 4. อาจารย์มัณฑนา แยมสูงเนิน                              | ครูโรงเรียนทวีธาภิเศก                                |
| 5. อาจารย์สมฤทัย คงงาม                                   | ครูโรงเรียนสตรีสมุทรปราการ                           |
| 6. อาจารย์นริรักษ์ ทองสะอาด                              | ครูโรงเรียนวัดสังเวช                                 |
| 7. อาจารย์ภรณ์วัน สินฉิม                                 | ครูโรงเรียนวัดบูรณาวาส                               |
| 8. อาจารย์ณพัชร์ เสงสมบูรณ์                              | ครูโรงเรียนสนามชัยเขต                                |
| 9. อาจารย์จารุวรรณ ฤทธิเพชร                              | ครูโรงเรียนนนทรีวิทยา                                |

- |                                   |  |
|-----------------------------------|--|
| 10. อาจารย์พรพิมล ค่อมสิงห์       | ครูโรงเรียนสตรีศรีสุริโยทัย  |
| 11. ดร.นริศรา เสือคล้าย           | อาจารย์ประจำมหาวิทยาลัยพะเยา   |
| 12. อาจารย์ ดร.นิติกร อ่อนโยน     | อาจารย์ประจำคณะครุศาสตร์<br>มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์      |
| 13. อาจารย์ธำปนา จ้อยเจริญ        | อาจารย์ประจำคณะครุศาสตร์<br>มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์      |
| 14. อาจารย์ ดร.เมษา นวลศรี        | อาจารย์ประจำคณะครุศาสตร์<br>มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์      |
| 15. ดร.จุฑาภรณ์ มาสันเทียะ        | คณบดีบัณฑิต สาขาการวัดและประเมินผลการศึกษา<br>คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย |
| 16. อาจารย์รัฐภิกษา ตั้งพุทธพิงศ์ | อาจารย์ประจำโรงเรียนสาธิตจุฬาลงกรณ์มหาวิทยาลัย<br>ฝ่ายมัธยม                      |
| 17. อาจารย์น้ำฝน นาสวาสดี         | อาจารย์ประจำโรงเรียนสาธิตจุฬาลงกรณ์มหาวิทยาลัย<br>ฝ่ายมัธยม                      |
| 18. อาจารย์สุนทร ภูรีปริชาเลิศ    | อาจารย์ประจำโรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒ<br>ปทุมวัน                   |
| 19. อาจารย์ทวิช มณีพนา            | นักวิชาการ และครูโรงเรียนกวดวิชา   |
| 20. นางสาวพนิดา พาณิชวัฒน์        | นักวิชาการ และครูโรงเรียนกวดวิชา   |

จุฬาลงกรณ์มหาวิทยาลัย

Chulalongkorn University

การตรวจสอบด้านตรงตามเนื้อหาและความเหมาะสมของคู่มือแนวทางการดำเนินการศึกษา  
ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบใน  
การประเมินระดับชั้นเรียน

- |                               |  |
|-------------------------------|--|
| 1. อาจารย์พฤตม ไบระหมาน       | ครูโรงเรียนวรนาธิเฉลิม จังหวัดสงขลา                            |
| 2. อาจารย์ นราศักดิ์ ไชยเรือง | ครูโรงเรียนมารีย์อุปถัมภ์                                      |
| 3. อาจารย์ธัญญา แก้วศรีหาวงษ์ | ครูโรงเรียนบ้านหมี่วิทยา                                       |
| 4. อาจารย์เบญจวรรณ รัตนะ      | ครูโรงเรียนชะรัดชนูปถัมภ์                                      |
| 5. นางกฤตวรรณ ทองมี           | นักวิชาการศึกษา ศูนย์แพทยศาสตรศึกษาชั้นคลินิก<br>โรงพยาบาลแพร์ |



ภาคผนวก ข  
เครื่องมือในการวิจัย

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

แบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (test-item specification)  
ในการทดสอบทางการศึกษาระดับขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์  
ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559

(ตัวอย่างข้อรายการจำนวน 6 ข้อ จากทั้งหมด 45 ข้อ)

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

**แบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (Test-Item Specification)**  
 ในการทดสอบทางการศึกษาระดับขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559

สิ่งที่ส่งมาด้วย

แบบบันทึกนี้จัดทำขึ้นโดยมีวัตถุประสงค์เพื่อใช้ในการบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (Test-item Specification) ในการทดสอบทางการศึกษาระดับขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 ของ สทศ. ข้อมูลที่ได้จะถูกนำไปใช้ในการสร้างแมทริกซ์ข้อสอบในการประเมินระดับชาติ และใช้ในการคำนวณค่าความสอดคล้องในแนวเดียวกัน (alignment index) กับข้อสอบในการประเมินระดับชั้นเรียนต่อไป

**คำชี้แจง:** ขอความกรุณาระบุข้อมูลลักษณะเฉพาะของข้อสอบแต่ละข้อดังนี้

1. ระบุมาตรฐานของกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ที่ สทศ. ใช้ในการออกข้อสอบแต่ละข้อลงในคอลัมน์ที่ 3 "มาตรฐาน"
2. ระบุระดับชั้นและตัวชี้วัดของกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ที่ สทศ. ใช้ในการออกข้อสอบแต่ละข้อลงในคอลัมน์ที่ 4 "ตัวชี้วัด" โดยระบุเป็น "ม...../ระดับชั้น...../.....(ตัวชี้วัดที่.....)"
3. ทำเครื่องหมาย ✓ ในช่อง  ที่ตรงกับระดับการวัดด้านพุทธิปัญญาที่จำแนกตามแนวคิด Revised Bloom's Taxonomy ของข้อสอบแต่ละข้อในคอลัมน์ที่ 5 "ระดับการวัดด้านพุทธิปัญญา"

**ตัวอย่างการกรอกข้อมูล**

ข้อสอบข้อที่ 0 มีรายละเอียดของมาตรฐาน ระดับชั้นและตัวชี้วัด และระดับการวัดด้านพุทธิปัญญา ที่ สทศ. ใช้ในการออกข้อสอบดังนี้

- (1) มาตรฐาน ว 1.1 เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้าง และหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต ให้ระบุในคอลัมน์ที่ 3 ว่า "ว.1.1"
- (2) ระดับชั้นมัธยมศึกษาปีที่ 2 "ตัวชี้วัดที่ 3 สังเกตและอธิบายพฤติกรรมของมนุษย์และสัตว์ที่ตอบสนองต่อสิ่งแวดล้อมและภายใน" ให้ระบุในคอลัมน์ที่ 4 ว่า "ม.2/3"
- (3) ระดับการวัดด้านพุทธิปัญญา คือ ระดับวิเคราะห์ ให้ทำเครื่องหมาย ✓ ในคอลัมน์ที่ 5 ที่ตรงกับ "ช่อง วิเคราะห์"

ข้อที่	ข้อความ-ตัวเลือก	มาตรฐาน	ตัวชี้วัด	ระดับการวัดด้านพุทธิปัญญา					คะแนนเต็ม รายชื่อ (คะแนน)	
				จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า		สร้างสรรค์
0	.....	ว 1.1	ม.2/3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
		(1)	(2)	(3)						

ขอขอบพระคุณเป็นอย่างสูงในความกรุณาให้ข้อมูลอันเป็นประโยชน์ต่อการศึกษาวิจัยในครั้งนี้

นางสาวบุษยา หัตถ์จันทร์ (นักวิจัย)

1/44

Tel: 085-098-8055 e-mail: bjanpraseri@gmail.com

**นิยามศัพท์ที่เกี่ยวข้อง**

1. ระดับการวัดด้านพุทธิปัญญา หมายถึง ลำดับขั้นของกระบวนการทางสมองหรือความลึกของความรู้ในตัวผู้เรียนได้มาตรฐานต่างๆ และข้อสอบในการประเมินผลการเรียนรู้ ซึ่งวัดได้จาก การพิจารณาของผู้เชี่ยวชาญ ซึ่งความซับซ้อนทางปัญญาจำแนกตามแนวคิดการจำแนกพฤติกรรมการเรียนรู้ด้านพุทธิปัญญาของ Bloom ที่ได้ปรับปรุงใหม่ (revised Bloom's taxonomy) (Anderson et al., 2001) โดยแบ่งลำดับขั้นของกระบวนการทางปัญญา (cognitive process dimension) ได้เป็น 6 ชั้น เรียงลำดับจากความรู้ระดับต่ำไปย้ความรู้อันสูง ดังนี้
    - (1) จำ (remember) เป็นความสามารถในการระลึกหรือจำความรู้หรือสารสนเทศที่เก็บไว้ในสมอง ซึ่งเป็นความรู้จำระยะยาว
    - (2) เข้าใจ (understand) เป็นความสามารถในการสร้างความหมายหรือความรู้จากสื่อหรือเครื่องมือทางการศึกษาด้วยตนเอง เช่น จากการ์ตูน การอธิบายของครูที่ขบถของของ ความสามารถในขั้นนี้ ได้แก่ การแปลความหมาย (interpreting) การให้ตัวอย่าง (exemplifying) การจัดจำแนก (classifying) การสรุป (summarizing) การสรุปอ้างอิง (inferring) การเปรียบเทียบ (comparing) และการอธิบาย (explaining)
    - (3) ประยุกต์ใช้ (apply) เป็นความสามารถในการใช้กระบวนการที่ได้เรียนรู้มาในสถานการณ์ใหม่หรือสถานการณ์ที่คล้ายคลึงกัน
    - (4) วิเคราะห์ (analyze) เป็นความสามารถในการแยกความรู้ออกเป็นส่วนใหญ่ๆ โดยสามารถให้เหตุผลว่า ความรู้ส่วนย่อยที่แยกแต่ละส่วนมีความเกี่ยวข้องกับโครงสร้างของความรู้ทั้งหมด อย่างไร สามารถจำแนกความแตกต่าง จัดระบบความรู้ และบอกที่มาของความรู้หรือองค์ประกอบแต่ละส่วนได้
    - (5) ประเมินค่า (evaluate) เป็นความสามารถในการตรวจสอบ วิจารณ์ และตัดสิน
    - (6) สร้างสรรค์ (create) เป็นความสามารถสร้างสรรค์สิ่งใหม่จากสิ่งที่เคยเรียนรู้ หรือสิ่งที่พบเห็นในบริบทต่างๆ สามารถสร้างสรรค์งาน แผนงาน หรือผลิตภัณฑ์ หรือชิ้นงานที่แปลกใหม่
  2. มาตรฐาน หมายถึง คุณภาพที่ต้องการให้เกิดขึ้นในตัวผู้เรียน เป็นสิ่งที่ผู้เรียนพึงรู้และปฏิบัติได้ในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ตามหลักสูตรแกนกลาง การศึกษาระดับขั้นพื้นฐาน พุทธศักราช 2551
  3. ตัวชี้วัด หมายถึง สิ่งที่ยู่เรียนพึงรู้และปฏิบัติได้ในกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ซึ่งสะท้อนถึงมาตรฐานการเรียนรู้ มีลักษณะเฉพาะเจาะจงและมีความเป็นรูปธรรมในการนำไปใช้ในการกำหนดเนื้อหา การเรียนการสอน และเป็นเกณฑ์สำคัญ สำหรับการวัดและประเมินผลเพื่อตรวจสอบ คุณภาพผู้เรียน
- รายละเอียดของมาตรฐานและตัวชี้วัดกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จำแนกตามระดับชั้น แสดงดังตารางต่อไปนี้

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
สาระที่ 1 สิ่งมีชีวิตกับ กระบวนการ ดำรงชีวิต	มาตรฐาน ว 1.1 เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้าง และหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต	1. สังเกตและอธิบายรูปร่าง ลักษณะและองค์ประกอบของสิ่งมีชีวิตเซลล์เดียวและของสิ่งมีชีวิตหลายเซลล์ 2. สังเกตและเปรียบเทียบส่วนประกอบสำคัญของเซลล์พืชและเซลล์สัตว์ 3. ทดลองและอธิบายหน้าที่ของส่วนประกอบที่สำคัญของเซลล์พืชและเซลล์สัตว์ 4. ทดลองและอธิบายกระบวนการผ่านเซลล์และการแพร่และออสโมซิส 5. ทดลองอธิบายกระบวนการที่จำเป็นต่อการมีละวัชพืชของพืชและอธิบายความง คอเลอเลียม แก๊สที่พบในโกล-	1. อธิบายโครงสร้างและการทำงานของระบบอาหาร ระบบหมุนเวียนเลือด ระบบหายใจ ระบบสืบพันธุ์ ของมนุษย์และสัตว์ รวมทั้งระบบประสาทของมนุษย์ 2. อธิบายความสัมพันธ์ของระบบต่างๆ ของ มนุษย์และสัตว์ ความรู้ทั่วไปเกี่ยวกับ 3. สังเกตและอธิบายพฤติกรรมของมนุษย์และสัตว์ที่ตอบสนองต่อสิ่งเร้าภายนอกและภายใน 4. อธิบายหลักการและผลของการให้แก๊สไนโตรเจนในการขยายพันธุ์ ปรับปรุงพันธุ์ และเพิ่มผลผลิตของสัตว์และนำความรู้ไปใช้ประโยชน์	

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
		<p>ข้อใด นำ น้ำแข็งจืด ที่จำเป็น ต้องใช้ในการสังเคราะห์ด้วยแสง</p> <p>6. ทดลองและอธิบายผลที่ได้จากการสังเคราะห์ด้วยแสงของพืช</p> <p>7. อธิบายความสัมพันธ์ของกระบวนการสังเคราะห์ด้วยแสงของพืช สัตว์มีวิวัฒนาการและสิ่งแวดล้อม</p> <p>8. ทดลองและอธิบายคุณสมบัติที่เกี่ยวกับความสัมพันธ์ของพืช</p> <p>9. สังเกตและอธิบายโครงสร้างที่เกี่ยวกับระบบลำเลียงน้ำและอาหารของพืช</p> <p>10. ทดลองและอธิบายโครงสร้างของดอกที่เกี่ยวข้องกับการสืบพันธุ์ของพืช</p> <p>11. อธิบายกระบวนการสืบพันธุ์แบบอาศัยเพศของพืชดอกและการมีเพศสัมพันธ์แบบไม่อาศัยเพศของพืช โดยใช้ตัวอย่างของพืชเพื่อช่วยในการขยายพันธุ์</p> <p>12. ทดลองและอธิบายการตอบสนองของพืชต่อแสง น้ำ และการมีเมล็ด</p> <p>13. อธิบายหลักการและสองการใช้เทคโนโลยีชีวภาพในการขยายพันธุ์ วัชพืชพันธุ์ดี เพิ่มผลผลิตของพืชและนำความรู้ไปใช้ประโยชน์</p>	<p>5. ทดลอง วิเคราะห์ และอธิบาย สาเหตุอาหารในอาหารมีปริมาณพลังงานและสัดส่วนที่เหมาะสมกับประเภทและวัย</p> <p>6. อภิปรายผลของสารพันธุกรรมที่ถ่ายทอดกันมาของพ่อแม่ และแนวทางในการป้องกันตนเองจากสารเสพติด</p>	
	<p>มาตรฐาน ว 1.2 เข้าใจ กระบวนการและความสำคัญของการถ่ายทอดลักษณะทางพันธุกรรม วิวัฒนาการของสิ่งมีชีวิต ความหลากหลายทางชีวภาพ การใช้เทคโนโลยีชีวภาพที่มีผลกระทบต่อมนุษย์และสิ่งแวดล้อม มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์</p>			<p>1. สังเกตและอธิบายลักษณะของโครงสร้างที่มีหน่วยการรวมหรืออินในนิวเคลียส</p> <p>2. อธิบายความสัมพันธ์ของสารพันธุกรรมหรือดีเอ็นเอ และกระบวนการถ่ายทอดลักษณะทางพันธุกรรม</p> <p>3. อภิปรายวิวัฒนาการของสิ่งมีชีวิตที่เกิดจากความผิดปกติของยีนและโครโมโซมและนำความรู้ไปใช้ประโยชน์</p> <p>4. สังเกตและอธิบายความหลากหลายทางชีวภาพในท้องถิ่นที่หาได้ซึ่งมีขีดจำกัดหรืออยู่ได้อย่างสมดุล</p> <p>5. อธิบายผลของความหลากหลายทางชีวภาพที่มีต่อมนุษย์ สัตว์ พืช และสิ่งแวดล้อม</p> <p>6. อภิปรายผลของเทคโนโลยีชีวภาพต่อการดำรงชีวิตของมนุษย์และสิ่งแวดล้อม</p>

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
<p>สาระที่ 2 ชีวิตกับสิ่งแวดล้อม</p>	<p>มาตรฐาน ว 2.1 เข้าใจ สิ่งแวดล้อมในท้องถิ่น ความสัมพันธ์ระหว่างสิ่งแวดล้อมกับสิ่งมีชีวิต ความสัมพันธ์ระหว่างสิ่งมีชีวิตต่างๆ ในระบบนิเวศ มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์</p> <p>มาตรฐาน ว 2.2 เข้าใจ ความสำคัญของทรัพยากรธรรมชาติ การใช้ทรัพยากรธรรมชาติในท้องถิ่น ประเทศ และโลก นำความรู้ไปใช้ในการจัดการทรัพยากรธรรมชาติและสิ่งแวดล้อมในท้องถิ่นอย่างยั่งยืน</p>			<p>1. สสำรวจระบบนิเวศต่างๆในท้องถิ่นและอธิบาย ความสัมพันธ์ขององค์ประกอบภายในระบบนิเวศ</p> <p>2. วิเคราะห์และอธิบายความสัมพันธ์ของการถ่ายทอดพลังงานของสิ่งมีชีวิตในรูปของอาหารและสายใยอาหาร</p> <p>3. อธิบายปัจจัยที่มีอิทธิพลต่อระบบนิเวศ</p> <p>4. อธิบายปัจจัยที่มีผลต่อการเปลี่ยนแปลงขนาดของประชากรในระบบนิเวศ</p>
<p>สาระที่ 3 สารและสมบัติของสาร</p>	<p>มาตรฐาน ว 3.1 เข้าใจสมบัติของสาร ความสัมพันธ์ระหว่างสมบัติของสารกับโครงสร้างและแรงยึดเหนี่ยวระหว่างอนุภาค มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์</p> <p>มาตรฐาน ว 3.2 เข้าใจ หลักการและธรรมชาติของสารเปลี่ยนแปลงสถานะของสาร การเกิดสารละลาย การเกิดปฏิกิริยา มีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ และนำความรู้ไปใช้ประโยชน์</p>	<p>1. ทดลองและจำแนกสารเป็นกลุ่มโดยใช้เนื้อสารหรือขนาดอนุภาคเป็นเกณฑ์ และอธิบายสมบัติของสารในแต่ละกลุ่ม</p> <p>2. อธิบายสมบัติและการเปลี่ยนแปลงของสาร โดยใช้สมบัติของสาร การจัดเรียงอนุภาคของสาร</p> <p>3. ทดลองและอธิบายสมบัติความเป็นกรด เบส ของสารละลาย</p> <p>4. ตรวจสอบค่า pH ของสารละลายและนำความรู้ไปใช้ประโยชน์</p>	<p>1. จำลองและอธิบายองค์ประกอบ สมบัติของธาตุและสารประกอบ</p> <p>2. สืบค้นข้อมูลและเปรียบเทียบสมบัติของธาตุโลหะ ธาตุกึ่งโลหะและธาตุที่มีครึ่งสังกะสีและนำความรู้ไปใช้ประโยชน์</p> <p>3. ทดลองและอธิบายหลักการแยกสารด้วยวิธีการกรอง การตกตะกอน การกลั่น และโครมาโทกราฟี และนำความรู้ไปใช้ประโยชน์</p>	
		<p>1. ทดลองและอธิบายวิธีเตรียมสารละลายที่มีความเข้มข้นเป็นร้อยละ และอภิปรายการนำความรู้เกี่ยวกับสารละลายไปใช้ประโยชน์</p> <p>2. ทดลองและอธิบายการเปลี่ยนแปลงสมบัติมวลและพลังงานของสาร เมื่อสารเปลี่ยนแปลงสถานะของสาร</p> <p>3. ทดลองและอธิบายปัจจัยที่มีผลต่อการเปลี่ยนแปลง และอัตราการขยายของสาร</p>	<p>1. ทดลองและอธิบายการเปลี่ยนแปลงสมบัติ มวล และพลังงานเมื่อสารเกิดปฏิกิริยาเคมี รวมทั้งอธิบายปัจจัยที่มีผลต่อการเกิดปฏิกิริยาเคมี</p> <p>2. ทดลอง อธิบายและเขียนสมการเคมีของปฏิกิริยาของสารต่างๆ และนำความรู้ไปใช้ประโยชน์</p> <p>3. สืบค้นข้อมูลและอภิปรายผลของสารเคมี ปฏิกิริยาเคมีของสิ่งมีชีวิตและสิ่งแวดล้อม</p> <p>4. สืบค้นข้อมูลและอธิบายการใช้สารเคมีอย่างถูกต้อง ปลอดภัย หรือป้องกันและแก้ไขอันตรายที่เกิดจากการใช้สารเคมี</p>	



สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
สาระที่ 4 แรงและการเคลื่อนที่	มาตรฐาน ว 4.1 เข้าใจธรรมชาติของแรงมีขนาด ไฟฟ้าแรงโน้มถ่วง และแรงนิวเคลียร์ มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์อย่างถูกต้องและมีคุณธรรม	1. อธิบายข้อสังเกต และอธิบายปริมาณ สเกลาร์ ปริมาณเวกเตอร์ 2. ทดลองและอธิบายของเหลว การกระจัด อัตราเร็วและ ความเร็ว ในการเคลื่อนที่ของวัตถุ	1. ทดลองและอธิบายการหาผลตั้งของแรงหลายแรงในระบบ เบี่ยงกันที่กระทำกับวัตถุ 2. อธิบายแรงลัพธ์ที่กระทำต่อวัตถุที่หยุดนิ่งหรือวัตถุเคลื่อนที่ด้วยความเร็วคงตัว	1. อธิบายความเร่งและผลของแรงลัพธ์ที่กระทำต่อวัตถุ 2. ทดลองและอธิบายแรงยกและแรงปฏิปักษ์ในระบบระหว่างวัตถุ และนำความรู้ไปใช้ประโยชน์ 3. ทดลองและอธิบายแรงของเหลวของเหลวที่กระทำต่อวัตถุ
	มาตรฐาน ว 4.2 เข้าใจลักษณะการเคลื่อนที่แบบต่างๆ ของวัตถุในธรรมชาติ มีกระบวนการสืบเสาะหาความรู้ และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์			1. ทดลองและอธิบายความแตกต่างระหว่างแรงเสียดทานสถิตกับแรงเสียดทานจลน์ และนำความรู้ไปใช้ประโยชน์ 2. ทดลองและวิเคราะห์มุมบังคับของแรง และนำความรู้ไปใช้ประโยชน์ 3. สังเกตและอธิบายการเคลื่อนที่ของวัตถุที่เป็นวงกลม และแนววงโค้ง
สาระที่ 5 พลังงาน	มาตรฐาน ว 5.1 เข้าใจความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การเปลี่ยนรูปพลังงาน ปฏิสัมพันธ์ระหว่างสารและพลังงาน ผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์	1. ทดลองและอธิบายอุณหภูมิและการวัดอุณหภูมิ 2. สังเกตและอธิบายการถ่ายโอนความร้อน และนำความรู้ไปใช้ประโยชน์ 3. อธิบายการดูดกลืน การคายความร้อน โดยการแผ่รังสี และนำความรู้ไปใช้ประโยชน์ 4. อธิบายสมมูลความร้อนและผลของความร้อนต่อการขยายตัวของสาร และนำความรู้ไปใช้ในชีวิตประจำวัน	1. ทดลองและอธิบายการสะท้อนของแสง การหักเหของแสง และนำความรู้ไปใช้ประโยชน์ 2. อธิบายผลของความสว่างที่มีต่อมนุษย์และสิ่งมีชีวิตอื่นๆ 3. ทดลองและอธิบายการดูดกลืนและแผ่ การมองเห็นสีของวัตถุ และนำความรู้ไปใช้ประโยชน์	1. อธิบายงาน พลังงานจลน์ พลังงานศักย์โน้มถ่วง กฎการอนุรักษ์พลังงาน และความสัมพันธ์ระหว่างปริมาณเหล่านี้ รวมทั้งนำความรู้ไปใช้ประโยชน์ 2. ทดลองและอธิบายความสัมพันธ์ระหว่างความดันกับปริมาตร ความดันอากาศ และนำความรู้ไปใช้ประโยชน์ 3. ทำบอลลูนพลังงานไฟฟ้าของเครื่องใช้ไฟฟ้า และนำความรู้ไปใช้ประโยชน์ 4. สังเกตและอธิบายการตั้งวงโคจรของดาวเทียมในแนวนอนอย่างถูกต้องปลอดภัย และประหยัด 5. อธิบายตัวนำทาง ไทเทเนียม ทรานซิสเตอร์ และหลอดฟลูออโรสเซียมที่พบในชีวิตประจำวัน
	มาตรฐาน ว 5.1 เข้าใจกระบวนการต่างๆ ที่เกิดขึ้นบนผิวโลกและภายในโลก ความสัมพันธ์ของกระบวนการต่างๆ ที่มีผลต่อการเปลี่ยนแปลงภูมิอากาศ ภูมิประเพณี และสิ่งแวดล้อม มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์	1. อธิบายและอธิบายองค์ประกอบและการแบ่งชั้นบรรยากาศที่ปกคลุมผิวโลก 2. ทดลองและอธิบายความสัมพันธ์ระหว่าง อุณหภูมิ ความชื้นและความกดอากาศที่มีผลต่อปรากฏการณ์ทางลมฟ้าอากาศ 3. สังเกต วิเคราะห์และ อธิบายการเกิดปรากฏการณ์ทางลมฟ้าอากาศที่มีผลต่อมนุษย์ 4. อธิบาย วิเคราะห์ และแปลความหมายข้อมูลจากการพยากรณ์อากาศ 5. อธิบาย วิเคราะห์ และอธิบายผลของลมฟ้าอากาศต่อการดำรงชีวิตของสิ่งมีชีวิต และสิ่งแวดล้อม	1. อธิบาย ทดลองและอธิบายลักษณะของชั้นบรรยากาศชั้นโทรโพสเฟียร์ และกระบวนการเกิดเมฆ 2. อธิบาย วิเคราะห์และอธิบายการเปลี่ยนแปลงสภาพภูมิอากาศ 3. ทดลองสังเกตแบบเพื่ออธิบายกระบวนการเกิด และลักษณะของพายุฝนฟ้าคะนอง 4. ทดลอง และสังเกตรูปแบบของเมฆชนิดต่างๆ และนำความรู้ไปใช้ประโยชน์ 5. ตรวจสอบและอธิบายลักษณะทางกายภาพของเมฆ และการนำไปใช้ประโยชน์	


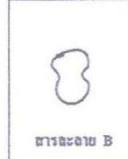
5/44

สาระ	มาตรฐาน	ตัวชี้วัด		
		ม. 1	ม. 2	ม. 3
สาระที่ 7 ดาราศาสตร์และอวกาศ	มาตรฐาน ว 7.1 เข้าใจวิวัฒนาการของระบบสุริยะ กาแล็กซีและเอกภพ การปฏิสัมพันธ์ภายในระบบสุริยะ และผลต่อสิ่งมีชีวิตบนโลก มีกระบวนการสืบเสาะหาความรู้ และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์	6. อธิบาย วิเคราะห์ และอธิบายปัจจัยทางธรรมชาติและการกระทำของมนุษย์ที่มีผลต่อการเปลี่ยนแปลงอุณหภูมิของโลก ภูมิทวีปอากาศ และฝนกรด 7. อธิบาย วิเคราะห์และอธิบายผลกระทบของ รังสีคอสมิก และฝนกรด ที่มีต่อสิ่งมีชีวิตและสิ่งแวดล้อม	6 อธิบายและอธิบายกระบวนการเกิดชีวิตและระบบนิเวศของพืช สัตว์และมนุษย์บนโลก และนำความรู้ไปใช้ประโยชน์ 7. อธิบายและอธิบายลักษณะของชั้นบรรยากาศ การเกิดเมฆ การเกิดฟ้าผ่า การเกิดสายฟ้า และผลของ ระบบการเกิดฟ้าผ่า 10. อธิบาย ลักษณะของระบบสุริยะ และอธิบายโครงสร้างและองค์ประกอบของสุริยะ	1. อธิบายและอธิบายความสัมพันธ์ระหว่างเอกภพกับชีวิต ชีวิตบนโลก และผลต่อสิ่งมีชีวิตบนโลก 2. อธิบายและอธิบายองค์ประกอบของเอกภพ กาแล็กซี และระบบสุริยะ 3. ระบุตำแหน่งของอวกาศ และนำความรู้ไปใช้ประโยชน์
	มาตรฐาน ว 7.2 เข้าใจความสัมพันธ์ของพลังงานอวกาศที่นำมาใช้ในการสำรวจอวกาศและการสื่อสาร การสื่อสาร การสืบเสาะหาความรู้ และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์อย่างมีคุณธรรมต่อชีวิตและสิ่งแวดล้อม			1. อธิบายและอธิบายความก้าวหน้าของเทคโนโลยีอวกาศที่ใช้สำรวจอวกาศ วิทยุสื่อสาร สatelites อวกาศ หรือการรวมระบบการสื่อสาร และการสื่อสาร
สาระที่ 8 ธรรมชาติของวิทยาศาสตร์และเทคโนโลยี	มาตรฐาน ว 8.1 ใช้กระบวนการ ทดลองหรือศึกษาค้นคว้าเพื่อสืบเสาะหาความรู้ และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์อย่างมีคุณธรรมต่อชีวิตและสิ่งแวดล้อม	1. ตั้งคำถามที่กำกวมประเด็นหรือสัมพันธ์กับสิ่งที่ศึกษาหรือสงสัยได้ 2. สร้างสมมติฐานที่สามารถตรวจสอบได้และวางแผนการสำรวจตรวจสอบอย่างยุติธรรม 3. เลือกเทคนิควิธีการสำรวจตรวจสอบทั้งเชิงปริมาณและเชิง	1. ตั้งคำถามที่กำกวมประเด็นหรือสัมพันธ์กับสิ่งที่ศึกษาหรือสงสัยได้ 2. สร้างสมมติฐานที่สามารถตรวจสอบได้และวางแผนการสำรวจตรวจสอบอย่างยุติธรรม 3. เลือกเทคนิควิธีการสำรวจตรวจสอบทั้งเชิงปริมาณและเชิง	1. ตั้งคำถามที่กำกวมประเด็นหรือสัมพันธ์กับสิ่งที่ศึกษาหรือสงสัยได้ 2. สร้างสมมติฐานที่สามารถตรวจสอบได้และวางแผนการสำรวจตรวจสอบอย่างยุติธรรม 3. เลือกเทคนิควิธีการสำรวจตรวจสอบทั้งเชิงปริมาณและเชิง

6/44

สาระ	มาตรฐาน	ตัวชี้วัด		
		ว. 1	ว. 2	ว. 3
<p>นักเรียนสามารถอธิบายและตรวจสอบได้ ภายใต้ข้อมูลและเครื่องมือที่มีอยู่ในช่วงเวลาสั้นๆ เข้าใจว่า วิทยาศาสตร์ยุคโมเดิร์น สังคม และสิ่งแวดล้อม มีความเกี่ยวข้องสัมพันธ์กัน</p>	<p>คุณภาพที่ได้ผลเพียงตรงและปลอดภัย โดยไม่ใช้วัตถุและเครื่องมือที่เฉพาะสม</p> <p>4. รวบรวมข้อมูล จัดกระทำข้อมูลเชิงปริมาณและคุณภาพ</p> <p>5. วิเคราะห์และประเมินความสอดคล้องของประจักษ์พยานกับข้อสรุป ที่ที่สนับสนุนหรือขัดแย้งกับสมมติฐาน และความคิดปกติของข้อมูลจากการสำรวจตรวจสอบ. สร้างแบบจำลอง หรือรูปแบบ ที่อธิบายผลหรือผลของการสำรวจตรวจสอบ</p> <p>7. สร้างคำถามที่นำไปสู่การสำรวจตรวจสอบ ในเรื่องที่เกี่ยวข้อง และนำความรู้ที่ได้ไปใช้ในสถานการณ์ใหม่หรืออธิบายเกี่ยวกับแนวคิดกระบวนการ และผลของโครงการหรือชิ้นงานให้ผู้อื่นเข้าใจ</p> <p>8. บันทึกและอธิบายผลการสังเกต การสำรวจ ตรวจสอบ ค้นคว้าเพิ่มเติมจากแหล่งความรู้ต่างๆ ไม่ได้ข้อมูลที่เชื่อถือได้ และยอมรับการเปลี่ยนแปลงความรู้ ที่ค้นพบเมื่อมีข้อมูลและประจักษ์พยานใหม่เพิ่มขึ้นหรือได้มาจากเดิม</p> <p>9. จัดแสดงงาน เขียนรายงาน และ/หรืออธิบายเกี่ยวกับแนวคิด กระบวนการ และผลของโครงการหรือชิ้นงานให้ผู้อื่นเข้าใจ</p>	<p>คุณภาพที่ได้ผลที่ตรงและปลอดภัย โดยไม่ใช้วัตถุและเครื่องมือที่เฉพาะสม</p> <p>4. รวบรวมข้อมูล จัดกระทำข้อมูลเชิงปริมาณและคุณภาพ</p> <p>5. วิเคราะห์และประเมินความสอดคล้องของประจักษ์พยานกับข้อสรุป ที่ที่สนับสนุนหรือขัดแย้งกับสมมติฐาน และความคิดปกติของข้อมูลจากการสำรวจตรวจสอบ. สร้างแบบจำลอง หรือรูปแบบ ที่อธิบายผลหรือผลของการสำรวจตรวจสอบ</p> <p>7. สร้างคำถามที่นำไปสู่การสำรวจตรวจสอบ ในเรื่องที่เกี่ยวข้อง และนำความรู้ที่ได้ไปใช้ในสถานการณ์ใหม่หรืออธิบายเกี่ยวกับแนวคิดกระบวนการ และผลของโครงการหรือชิ้นงานให้ผู้อื่นเข้าใจ</p> <p>8. บันทึกและอธิบายผลการสังเกต การสำรวจ ตรวจสอบ ค้นคว้าเพิ่มเติมจากแหล่งความรู้ต่างๆ ไม่ได้ข้อมูลที่เชื่อถือได้ และยอมรับการเปลี่ยนแปลงความรู้ ที่ค้นพบเมื่อมีข้อมูลและประจักษ์พยานใหม่เพิ่มขึ้นหรือได้มาจากเดิม</p> <p>9. จัดแสดงงาน เขียนรายงาน และ/หรืออธิบายเกี่ยวกับแนวคิด กระบวนการ และผลของโครงการหรือชิ้นงานให้ผู้อื่นเข้าใจ</p>	<p>คุณภาพที่ได้ผลที่ตรงและปลอดภัย โดยไม่ใช้วัตถุและเครื่องมือที่เฉพาะสม</p> <p>4. รวบรวมข้อมูล จัดกระทำข้อมูลเชิงปริมาณและคุณภาพ</p> <p>5. วิเคราะห์และประเมินความสอดคล้องของประจักษ์พยานกับข้อสรุป ที่ที่สนับสนุนหรือขัดแย้งกับสมมติฐาน และความคิดปกติของข้อมูลจากการสำรวจตรวจสอบ. สร้างแบบจำลอง หรือรูปแบบ ที่อธิบายผลหรือผลของการสำรวจตรวจสอบ</p> <p>7. สร้างคำถามที่นำไปสู่การสำรวจตรวจสอบ ในเรื่องที่เกี่ยวข้อง และนำความรู้ที่ได้ไปใช้ในสถานการณ์ใหม่หรืออธิบายเกี่ยวกับแนวคิดกระบวนการ และผลของโครงการหรือชิ้นงานให้ผู้อื่นเข้าใจ</p> <p>8. บันทึกและอธิบายผลการสังเกต การสำรวจ ตรวจสอบ ค้นคว้าเพิ่มเติมจากแหล่งความรู้ต่างๆ ไม่ได้ข้อมูลที่เชื่อถือได้ และยอมรับการเปลี่ยนแปลงความรู้ ที่ค้นพบเมื่อมีข้อมูลและประจักษ์พยานใหม่เพิ่มขึ้นหรือได้มาจากเดิม</p> <p>9. จัดแสดงงาน เขียนรายงาน และ/หรืออธิบายเกี่ยวกับแนวคิด กระบวนการ และผลของโครงการหรือชิ้นงานให้ผู้อื่นเข้าใจ</p>	

แบบบันทึกข้อมูลลักษณะเฉพาะของข้อสอบ (Test-item Specification)  
 ในการทดสอบทางการศึกษาระดับพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559

ข้อที่	ข้อความ-ตัวเลือก	มาตรฐาน	ตัวชี้วัด	ระดับการวัดด้านพุทธิปัญญา						คะแนนเต็ม รายข้อ (คะแนน)
				จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	
1	<p>นำขนาดชนิดเดียวกัน ขนาดเท่ากัน ไปป้อนฮารอะฉาย A B และ C ที่มีความเข้มข้นแตกต่างกันเป็นเวลานานเท่าๆ กัน ได้ผลดังภาพ</p> <div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;">  <p>ฮารอะฉาย A</p> </div> <div style="text-align: center;">  <p>ฮารอะฉาย B</p> </div> <div style="text-align: center;">  <p>ฮารอะฉาย C</p> </div> </div> <p>จงเรียงลำดับความเข้มข้นของฮารอะฉาย A B และ C จากความเข้มข้นน้อยที่สุดไปมากที่สุด</p> <ol style="list-style-type: none"> <li>1. A B C</li> <li>2. A C B</li> <li>3. B A C</li> <li>4. C A B</li> </ol>	ว.1.1	ว.1/4	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2

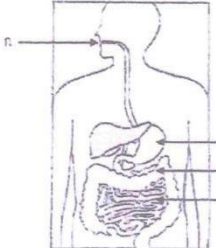
ข้อที่	ข้อความ-ตัวเลือก	มาตรฐาน	ตัวชี้วัด	ระดับการวัดด้านพุทธิปัญญา						คะแนนเต็ม รายข้อ (คะแนน)												
				จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์													
2	<p>การทดลองเพื่อศึกษาการแพร่ของค่างทับทิม 3 ชุดทดลอง ได้ผลการดังต่อไปนี้</p> <table border="1"> <thead> <tr> <th>ชุดทดลองที่</th> <th>ตัวกลาง</th> <th>ผลการสังเกต</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>A</td> <td>มีวงของค่างทับทิมกระจายไปทุกทิศทุกทางอย่างรวดเร็ว</td> </tr> <tr> <td>2</td> <td>B</td> <td>มีวงของค่างทับทิมค่อยๆ กระจายไปทุกทิศทุกทาง อย่างช้าๆ</td> </tr> <tr> <td>3</td> <td>ไม่มีสาร ตัวกลาง</td> <td>อนุภาคของค่างทับทิมไม่มีการเคลื่อนที่</td> </tr> </tbody> </table> <p>ตัวแปรต้นของการทดลองนี้คืออะไร</p> <ol style="list-style-type: none"> <li>ปริมาณของค่างทับทิม</li> <li>ขนาดอนุภาคของค่างทับทิม</li> <li>ชนิดของตัวกลาง</li> <li>อัตราเร็วในการแพร่ของค่างทับทิม</li> </ol>	ชุดทดลองที่	ตัวกลาง	ผลการสังเกต	1	A	มีวงของค่างทับทิมกระจายไปทุกทิศทุกทางอย่างรวดเร็ว	2	B	มีวงของค่างทับทิมค่อยๆ กระจายไปทุกทิศทุกทาง อย่างช้าๆ	3	ไม่มีสาร ตัวกลาง	อนุภาคของค่างทับทิมไม่มีการเคลื่อนที่	3.1.1	3.1.1/4	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
ชุดทดลองที่	ตัวกลาง	ผลการสังเกต																				
1	A	มีวงของค่างทับทิมกระจายไปทุกทิศทุกทางอย่างรวดเร็ว																				
2	B	มีวงของค่างทับทิมค่อยๆ กระจายไปทุกทิศทุกทาง อย่างช้าๆ																				
3	ไม่มีสาร ตัวกลาง	อนุภาคของค่างทับทิมไม่มีการเคลื่อนที่																				

9/44

ข้อที่	ข้อความ-ตัวเลือก	มาตรฐาน	ตัวชี้วัด	ระดับการวัดด้านพุทธิปัญญา						คะแนนเต็ม รายข้อ (คะแนน)																				
				จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์																					
3	<p>พิจารณาตารางแสดงปัจจัยการสังเคราะห์ด้วยแสงของพืชชนิดเดียวกัน มีจำนวนใบเท่ากัน จำนวน 4 ต้น แล้วตอบคำถาม</p> <table border="1"> <thead> <tr> <th>พืชต้นที่</th> <th>ระยะเวลาที่ได้รับแสง (ชั่วโมง)</th> <th>ปริมาณน้ำที่ได้รับ (ลูกบาศก์เซนติเมตร)</th> <th>ปริมาณแก๊ส คาร์บอนไดออกไซด์ที่ได้รับ (ลูกบาศก์เซนติเมตร)</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0</td> <td>10</td> <td>2</td> </tr> <tr> <td>2</td> <td>6</td> <td>0</td> <td>3</td> </tr> <tr> <td>3</td> <td>8</td> <td>0.5</td> <td>0</td> </tr> <tr> <td>4</td> <td>12</td> <td>0.5</td> <td>3</td> </tr> </tbody> </table> <p>พืชต้นใดสังเคราะห์ด้วยแสงได้ดีที่สุด</p> <ol style="list-style-type: none"> <li>ต้นที่ 1</li> <li>ต้นที่ 2</li> <li>ต้นที่ 3</li> <li>ต้นที่ 4</li> </ol>	พืชต้นที่	ระยะเวลาที่ได้รับแสง (ชั่วโมง)	ปริมาณน้ำที่ได้รับ (ลูกบาศก์เซนติเมตร)	ปริมาณแก๊ส คาร์บอนไดออกไซด์ที่ได้รับ (ลูกบาศก์เซนติเมตร)	1	0	10	2	2	6	0	3	3	8	0.5	0	4	12	0.5	3	3.1.1	3.1.1/5	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
พืชต้นที่	ระยะเวลาที่ได้รับแสง (ชั่วโมง)	ปริมาณน้ำที่ได้รับ (ลูกบาศก์เซนติเมตร)	ปริมาณแก๊ส คาร์บอนไดออกไซด์ที่ได้รับ (ลูกบาศก์เซนติเมตร)																											
1	0	10	2																											
2	6	0	3																											
3	8	0.5	0																											
4	12	0.5	3																											

10/44

ข้อที่	ข้อความ-ตัวเลือก	มาตรฐาน	ตัวชี้วัด	ระดับการวัดด้านพุทธิปัญญา						คะแนนเต็ม รายชื่อ (คะแนน)															
				จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์																
4	พิจารณาสมการการสังเคราะห์ด้วยแสง แล้วตอบคำถาม  <div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 10px auto;"> <math display="block">\text{น้ำ} + \text{แก๊สชนิด A} \xrightarrow{\text{แสง}} \text{แก๊สชนิด B} + \text{น้ำตาล} + \text{น้ำ}</math> <p style="text-align: center;">คลอโรพลาสต์</p> </div> <p>ข้อใดกล่าวถึงแก๊สชนิด A และ B ได้ถูกต้อง</p> <table border="1" style="width: 100%; border-collapse: collapse; text-align: center;"> <thead> <tr> <th>ข้อ</th> <th>แก๊สชนิด A</th> <th>แก๊สชนิด B</th> </tr> </thead> <tbody> <tr> <td>1.</td> <td>ทำให้ไฟฟัด</td> <td>ทำให้หน้าปุ่นใสขุ่น</td> </tr> <tr> <td>2.</td> <td>ทำให้หน้าปุ่นใสขุ่น</td> <td>ได้จากกรหายใจของสัตว์</td> </tr> <tr> <td>3.</td> <td>ทำให้หน้าปุ่นใสขุ่น</td> <td>ใช้ในการหายใจของพืช</td> </tr> <tr> <td>4.</td> <td>ทำให้ไฟฟัด</td> <td>ใช้ในการหายใจของสัตว์</td> </tr> </tbody> </table>	ข้อ	แก๊สชนิด A	แก๊สชนิด B	1.	ทำให้ไฟฟัด	ทำให้หน้าปุ่นใสขุ่น	2.	ทำให้หน้าปุ่นใสขุ่น	ได้จากกรหายใจของสัตว์	3.	ทำให้หน้าปุ่นใสขุ่น	ใช้ในการหายใจของพืช	4.	ทำให้ไฟฟัด	ใช้ในการหายใจของสัตว์	ง.1.1	ผ.1/8	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2
ข้อ	แก๊สชนิด A	แก๊สชนิด B																							
1.	ทำให้ไฟฟัด	ทำให้หน้าปุ่นใสขุ่น																							
2.	ทำให้หน้าปุ่นใสขุ่น	ได้จากกรหายใจของสัตว์																							
3.	ทำให้หน้าปุ่นใสขุ่น	ใช้ในการหายใจของพืช																							
4.	ทำให้ไฟฟัด	ใช้ในการหายใจของสัตว์																							
5	เด็กหญิงเดือน เจนป่วยง่าย มีภูมิคุ้มกันโรคต่ำ ข้อใดอธิบายเกี่ยวกับเด็กของ เด็กหญิงเดือนได้ถูกต้อง <ol style="list-style-type: none"> <li>1. เม็ดเลือดขาวต่ำกว่าปกติ</li> <li>2. เม็ดเลือดขาวสูงกว่าปกติ</li> <li>3. เม็ดเลือดต่ำกว่าปกติ</li> <li>4. เม็ดเลือดสูงกว่าปกติ</li> </ol>	ง.1.1	ผ.2/2, ผ.2/3	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2															

ข้อที่	ข้อความ-ตัวเลือก	มาตรฐาน	ตัวชี้วัด	ระดับการวัดด้านพุทธิปัญญา						คะแนนเต็ม รายชื่อ (คะแนน)
				จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	
6	พิจารณาภาพทางเดินอาหารของมนุษย์ แล้วตอบคำถาม   <p>การย่อยแป้งด้วยเอนไซม์อะไมเลสแล้วได้น้ำตาลกลูโคส เกิดขึ้นที่ส่วนใด ของทางเดินอาหาร</p> <ol style="list-style-type: none"> <li>1. ก และ ข</li> <li>2. ก และ ง</li> <li>3. ข และ ค</li> <li>4. ก และ ง</li> </ol>	ง.1.1	ผ.2/2, ผ.2/3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	2

แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด  
กับข้อสอบในการประเมินระดับชั้นเรียน  
กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น

(ตัวอย่างข้อรายการในการประเมินจำนวน 10 ข้อ จากทั้งหมด 1,089 ข้อ)

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY



ID\_A.....

แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด  
กับข้อสอบในการประเมินระดับชั้นเรียน  
กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

# SCIENCE



**แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด  
กับข้อสอบในการประเมินระดับชั้นเรียน  
กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น**

แบบประเมินฉบับนี้จัดทำขึ้นโดยมีวัตถุประสงค์เพื่อใช้เป็นเครื่องมือสำหรับผู้เชี่ยวชาญในการประเมินความสอดคล้องระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียนกลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ข้อมูลการประเมินความสอดคล้องจากผู้เชี่ยวชาญจะให้สารสนเทศที่เป็นข้อมูลป้อนกลับไปยังผู้มีส่วนได้ส่วนเสียในการประเมินระดับชั้นเรียนให้สอดคล้องกับมาตรฐานและตัวชี้วัด นอกจากนี้ ข้อมูลดังกล่าวจะถูกนำไปใช้ในการวิเคราะห์ความสอดคล้องในแนวเดียวกันระหว่างการประเมินระดับชั้นเรียนกับการประเมินระดับชาติเพื่อใช้เป็นแนวทางในการพัฒนาการออกแบบการประเมิน รวมถึงการเรียนการสอนให้สอดคล้องกับมาตรฐานที่กำหนด ให้มุ่งไปสู่เป้าหมายเดียวกัน อันเป็นประโยชน์ต่อการสร้างโอกาสในการเรียนรู้ อย่าง เท่าเทียมกันให้กับผู้เรียนทุกคน

**★★ คำชี้แจงสำหรับผู้เชี่ยวชาญ ★★**

ขอความกรุณาท่านพิจารณาข้อสอบเป็นรายข้อ และโปรดดำเนินการดังนี้

1. ระบุระดับความซับซ้อนทางปัญญาที่ข้อสอบแต่ละข้อ ที่ผู้สอบจำเป็นต้องใช้ในการทำข้อสอบได้ถูกต้อง โดยกรุณาทำเครื่องหมาย  ลงในช่อง  ที่ตรงกับระดับความซับซ้อนทางปัญญาของข้อสอบแต่ละข้อ เพียงระดับเดียว

2. ระบุระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัดที่กำหนด โดยกรุณาทำเครื่องหมาย  ลงในช่อง  ที่ตรงกับระดับความสอดคล้องระหว่างข้อสอบแต่ละข้อกับตัวชี้วัดที่ระบุ เพียงระดับเดียว

3. เติมข้อความที่เป็นข้อเสนอแนะเพิ่มเติมที่มีต่อข้อสอบ (ถ้ามี)

ทั้งนี้ท่านสามารถศึกษานิยามศัพท์และคำอธิบายที่เกี่ยวข้องใน **เอกสารประกอบการประเมินความสอดคล้องระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น**

ข้อที่ A1 1-1114.1-122	<p>นึ่งน้ำตาลทำการทดลองโดยตอกไข่ไก่ 1 ฟอง ใส่ลงในถ้วยแล้วใช้ส้อมตีไข่แดงและไข่ขาวปนกันเป็นเนื้อเดียว แล้วเติมสารละลายน้ำตาลกลูโคสลงไป 2 ช้อนโต๊ะ บรรจุลงในถุงกระดาษเซลโลเฟน (กระดาษแก้ว) ผูกปากถุงให้แน่น นำไปแช่ในน้ำกลั่น 1 ชั่วโมง นักเรียนคิดว่าสารใดสามารถผ่านกระดาษเซลโลเฟนออกมาได้</p> <ol style="list-style-type: none"> <li>1. ไข่ขาว</li> <li>2. เฉพาะไข่แดง</li> <li>3. เฉพาะน้ำ</li> <li>4. น้ำและน้ำตาลกลูโคส</li> </ol>
คะแนนรายข้อ	0.50 คะแนน
ระดับชั้น	มัธยมศึกษาปีที่ 1
มาตรฐาน	ว 1.1 เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้าง และหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต
ตัวชี้วัด	ม. 1/4 ทดลองและอธิบายกระบวนการเคลื่อนที่ของสารผ่านเซลล์โดยการแพร่และออสโมซิส

1. ระดับความซับซ้อนทางปัญญา (เลือกได้เพียงระดับเดียว)

- จำ       เข้าใจ       ประยุกต์ใช้       วิเคราะห์       ประเมินค่า       สร้างสรรค์

2. ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด (เลือกได้เพียงระดับเดียว)

- ไม่สอดคล้อง       ค่อนข้างไม่สอดคล้อง       ไม่แน่ใจ       ค่อนข้างสอดคล้อง       สอดคล้องโดยตรง
- (0)                      (1)                      (2)                      (3)                      (4)

3. ข้อเสนอแนะเพิ่มเติม (ถ้ามี)

.....

.....

.....



ข้อที่ A2 1-1115.5-211	<p>ในกระบวนการสังเคราะห์ด้วยแสง ข้อความใดถูกต้อง</p> <ol style="list-style-type: none"> <li>กระบวนการสังเคราะห์ด้วยแสงเกิดเฉพาะที่ใบเท่านั้น</li> <li>แก๊สออกซิเจนและน้ำเป็นวัตถุดิบที่ต้องใช้ในกระบวนการสังเคราะห์ด้วยแสง</li> <li>แป้งเป็นผลิตภัณฑ์ที่เกิดจากกระบวนการสังเคราะห์ด้วยแสง สามารถทดสอบได้ด้วยสารละลายไอโอดีน</li> <li>แก๊สออกซิเจนและแก๊สคาร์บอนไดออกไซด์เกิดจากกระบวนการสังเคราะห์ด้วยแสง</li> </ol>
คะแนนรายข้อ	0.50 คะแนน
ระดับชั้น	มัธยมศึกษาปีที่ 1
มาตรฐาน	ว 1.1 เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้าง และหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต
ตัวชี้วัด	ม. 1/5 ทดลองหาปัจจัยบางประการที่จำเป็นต่อการสังเคราะห์ด้วยแสงของพืชและอธิบายว่าแสง คลอโรฟิลล์ แก๊สคาร์บอนไดออกไซด์ น้ำ เป็นปัจจัย ที่จำเป็นต้องใช้ในการสังเคราะห์ด้วยแสง

1. ระดับความซับซ้อนทางปัญญา (เลือกได้เพียงระดับเดียว)

- จำ       เข้าใจ       ประยุกต์ใช้       วิเคราะห์       ประเมินค่า       สร้างสรรค์

2. ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด (เลือกได้เพียงระดับเดียว)

- ไม่สอดคล้อง     ค่อนข้างไม่สอดคล้อง     ไม่แน่ใจ     ค่อนข้างสอดคล้อง     สอดคล้องโดยตรง
- (0)                      (1)                      (2)                      (3)                      (4)

3. ข้อเสนอแนะเพิ่มเติม (ถ้ามี)

.....

.....

.....

ข้อที่ A3 1-3111.34-312	สารละลายมีขนาดเส้นผ่านศูนย์กลางของอนุภาคเท่าไร 1. ขนาดอนุภาคน้อยกว่า $10^{-7}$ เซนติเมตร 2. ขนาดอนุภาคมากกว่า $10^{-4}$ เซนติเมตร 3. ขนาดอนุภาค ระหว่าง $10^{-4}$ และ $10^{-7}$ เซนติเมตร 4. ขนาดอนุภาคเท่ากับ $10^{-7}$ เซนติเมตร
คะแนนรายข้อ	0.50 คะแนน
ระดับชั้น	มัธยมศึกษาปีที่ 1
มาตรฐาน	ว 3.1 เข้าใจสมบัติของสาร ความสัมพันธ์ระหว่างสมบัติของสารกับโครงสร้างและแรงยึดเหนี่ยวระหว่างอนุภาค มีกระบวนการสืบเสาะ หาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ นำความรู้ไปใช้ประโยชน์
ตัวชี้วัด	ม. 1/1 ทดลองและจำแนกสารเป็นกลุ่มโดยใช้เนื้อสารหรือขนาดอนุภาคเป็นเกณฑ์ และอธิบายสมบัติของสารในแต่ละกลุ่ม

## 1. ระดับความซับซ้อนทางปัญญา (เลือกได้เพียงระดับเดียว)

จำ  เข้าใจ  ประยุกต์ใช้  วิเคราะห์  ประเมินค่า  สร้างสรรค์

## 2. ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด (เลือกได้เพียงระดับเดียว)

ไม่สอดคล้อง  ค่อนข้างไม่สอดคล้อง  ไม่แน่ใจ  ค่อนข้างสอดคล้อง  สอดคล้องโดยตรง  
(0) (1) (2) (3) (4)

## 3. ข้อเสนอแนะเพิ่มเติม (ถ้ามี)

.....

.....

.....

ข้อที่ A4 1-3112.25-411	สาร X และสาร Y ไม่ผสมเป็นเนื้อเดียวกัน สาร Y ลอยอยู่เหนือสาร X เมื่อเติมสาร Z ลงไป แล้วเขย่าสาร X และ Y ผสมเป็นเนื้อเดียวกับสาร Z คือข้อใด 1. คอลลอยด์ 2. สารละลาย 3. อิมัลชัน 4. อิมัลซิฟายเออร์
คะแนนรายข้อ	0.50 คะแนน
ระดับชั้น	มัธยมศึกษาปีที่ 1
มาตรฐาน	ว 3.1 เข้าใจสมบัติของสาร ความสัมพันธ์ระหว่างสมบัติของสารกับโครงสร้างและแรงยึดเหนี่ยวระหว่างอนุภาค มีกระบวนการสืบเสาะ หาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ นำความรู้ไปใช้ประโยชน์
ตัวชี้วัด	ม. 1/2 อธิบายสมบัติและการเปลี่ยนแปลงสถานะของสาร โดยใช้แบบจำลองการจัดเรียงอนุภาคของสาร

1. ระดับความซับซ้อนทางปัญญา (เลือกได้เพียงระดับเดียว)

- จำ     เข้าใจ     ประยุกต์ใช้     วิเคราะห์     ประเมินค่า     สร้างสรรค์

2. ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด (เลือกได้เพียงระดับเดียว)

- ไม่สอดคล้อง     ค่อนข้างไม่สอดคล้อง     ไม่แน่ใจ     ค่อนข้างสอดคล้อง     สอดคล้องโดยตรง  
(0)                      (1)                      (2)                      (3)                      (4)

3. ข้อเสนอแนะเพิ่มเติม (ถ้ามี)

.....

.....

.....



ข้อที่ A6 1-3211.16-212	นำเอทานอล 10 ลูกบาศก์เซนติเมตร มาละลายน้ำปริมาตร 90 ลูกบาศก์เซนติเมตร สารละลายที่ได้มีความเข้มข้นเท่าใด 1. สารละลายเอทานอลเข้มข้นร้อยละ 10 โดยมวลต่อปริมาตร 2. สารละลายเอทานอลเข้มข้นร้อยละ 11 โดยมวลต่อปริมาตร 3. สารละลายเอทานอลเข้มข้นร้อยละ 10 โดยปริมาตรต่อปริมาตร 4. สารละลายเอทานอลเข้มข้นร้อยละ 11 โดยปริมาตรต่อปริมาตร
คะแนนรายข้อ	0.50 คะแนน
ระดับชั้น	มัธยมศึกษาปีที่ 1
มาตรฐาน	ว 3.2 เข้าใจหลักการและธรรมชาติของการเปลี่ยนแปลงสถานะของสาร การเกิดสารละลาย การเกิดปฏิกิริยา มีกระบวนการสืบเสาะ หาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ และนำความรู้ไปใช้ประโยชน์
ตัวชี้วัด	ม. 1/1 ทดลองและอธิบายวิธีเตรียมสารละลายที่มีความเข้มข้นเป็นร้อยละ และอธิบายการนำความรู้เกี่ยวกับสารละลายไปใช้ประโยชน์

1. ระดับความซับซ้อนทางปัญญา (เลือกได้เพียงระดับเดียว)

- จำ       เข้าใจ       ประยุกต์ใช้       วิเคราะห์       ประเมินค่า       สร้างสรรค์

2. ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด (เลือกได้เพียงระดับเดียว)

- ไม่สอดคล้อง       ค่อนข้างไม่สอดคล้อง       ไม่แน่ใจ       ค่อนข้างสอดคล้อง       สอดคล้องโดยตรง  
(0)                                      (1)                                      (2)                                      (3)                                      (4)

3. ข้อเสนอแนะเพิ่มเติม (ถ้ามี)

.....

.....

.....

ข้อที่ A7 1-3212.17-312	ข้อใดเป็นสารที่ได้จากการตกผลึก 1. ผงชูรส 2. น้ำตาลทราย 3. สารส้ม 4. ถูกทุกข้อ
คะแนนรายข้อ	0.50 คะแนน
ระดับชั้น	มัธยมศึกษาปีที่ 1
มาตรฐาน	ว 3.2 เข้าใจหลักการและธรรมชาติของการเปลี่ยนแปลงสถานะของสาร การเกิดสารละลาย การเกิดปฏิกิริยา มีกระบวนการสืบเสาะ หาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้ และนำความรู้ไปใช้ประโยชน์
ตัวชี้วัด	ม. 1/2 ทดลองและอธิบายการเปลี่ยนแปลงสมบัติ มวลและพลังงานของสาร เมื่อสารเปลี่ยนสถานะและเกิดการละลาย

1. ระดับความซับซ้อนทางปัญญา (เลือกได้เพียงระดับเดียว)

- จำ     เข้าใจ     ประยุกต์ใช้     วิเคราะห์     ประเมินค่า     สร้างสรรค์

2. ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด (เลือกได้เพียงระดับเดียว)

- ไม่สอดคล้อง     ค่อนข้างไม่สอดคล้อง     ไม่แน่ใจ     ค่อนข้างสอดคล้อง     สอดคล้องโดยตรง  
(0)                      (1)                      (2)                      (3)                      (4)

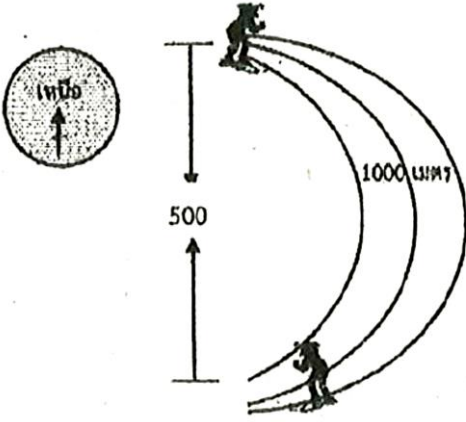
3. ข้อเสนอแนะเพิ่มเติม (ถ้ามี)

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

.....

.....

.....

<b>ข้อที่ A8</b> 1-4112.25-422	เจมส์บอลเล่นสกีจากยอดเขาลงมาถึงเนินเขาตามเส้นทางดังรูป  จากรูป ถ้าเจมส์บอลใช้เวลาในการเล่นสกีจากยอดเขาถึงเนินเขา 5 นาที เจมส์บอลเคลื่อนที่ด้วยอัตราเร็วเท่าใด 1. 1.67 เมตรวินาที/                      2. 2.33 เมตรวินาที/ 3. 3.33 เมตรวินาที                        4. 4.00 เมตรวินาที/
<b>คะแนนรายข้อ</b>	0.25 คะแนน
<b>ระดับชั้น</b>	มัธยมศึกษาปีที่ 1
<b>มาตรฐาน</b>	ว 4.1 เข้าใจธรรมชาติของแรงแม่เหล็กไฟฟ้า แรงแม่โน้มถ่วง และแรงนิวเคลียร์ มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์อย่างถูกต้องและมีคุณธรรม
<b>ตัวชี้วัด</b>	ม. 1/2 ทดลองและอธิบายระยะทาง การกระจัด อัตราเร็วและความเร็วในการเคลื่อนที่ของวัตถุ

**1. ระดับความซับซ้อนทางปัญญา (เลือกได้เพียงระดับเดียว)**

- 
- จำ
- 
- เข้าใจ
- 
- ประยุกต์ใช้
- 
- วิเคราะห์
- 
- ประเมินค่า
- 
- สร้างสรรค์

**2. ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด (เลือกได้เพียงระดับเดียว)**

- 
- ไม่สอดคล้อง
- 
- ค่อนข้างไม่สอดคล้อง
- 
- ไม่แน่ใจ
- 
- ค่อนข้างสอดคล้อง
- 
- สอดคล้องโดยตรง
- 
- (0)                      (1)                      (2)                      (3)                      (4)

**3. ข้อเสนอแนะเพิ่มเติม (ถ้ามี)**

.....

.....

.....

<b>ข้อที่ A9</b> <b>1-5112.12-312</b>	<p>จากภาพ ถ้านักเรียนจับที่หูของภาชนะที่ทำจากโลหะ เหตุใดจึงรู้สึกร้อน</p>  <ol style="list-style-type: none"> <li>1. มีการแผ่รังสีจากน้ำร้อนมายังมือ</li> <li>2. อนุภาคของอากาศนำความร้อนจากน้ำมายังมือ</li> <li>3. อนุภาคความร้อนเคลื่อนที่จากน้ำร้อนผ่านโลหะมาสู่มือ</li> <li>4. อนุภาคของน้ำเคลื่อนที่จากน้ำร้อนผ่านโลหะมาสู่มือ</li> </ol>
<b>คะแนนรายข้อ</b>	0.50 คะแนน
<b>ระดับชั้น</b>	มัธยมศึกษาปีที่ 1
<b>มาตรฐาน</b>	ว 5.1 เข้าใจความสัมพันธ์ระหว่างพลังงานกับการดำรงชีวิต การเปลี่ยนรูปพลังงาน ปฏิสัมพันธ์ระหว่างสารและพลังงาน ผลของการใช้พลังงานต่อชีวิตและสิ่งแวดล้อม มีกระบวนการสืบเสาะ หาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์
<b>ตัวชี้วัด</b>	ม. 1/2 สังเกตและอธิบายการถ่ายโอนความร้อน และนำความรู้ไปใช้ประโยชน์

1. ระดับความซับซ้อนทางปัญญา (เลือกได้เพียงระดับเดียว)

- จำ     
  เข้าใจ     
  ประยุกต์ใช้     
  วิเคราะห์     
  ประเมินค่า     
  สร้างสรรค์

2. ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด (เลือกได้เพียงระดับเดียว)

- ไม่สอดคล้อง     
  ค่อนข้างไม่สอดคล้อง     
  ไม่แน่ใจ     
  ค่อนข้างสอดคล้อง     
  สอดคล้องโดยตรง
- (0)                      (1)                      (2)                      (3)                      (4)

3. ข้อเสนอแนะเพิ่มเติม (ถ้ามี)

.....

.....

.....





แบบตรวจสอบความเหมาะสมของคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องใน  
แนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน



แบบตรวจสอบความเหมาะสมของคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องใน  
แนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

### ส่วนที่ 1 ข้อมูลส่วนตัวของผู้ทรงคุณวุฒิ

ชื่อ.....ตำแหน่ง.....  
สถานที่ทำงาน.....  
วุฒิการศึกษา.....

### ส่วนที่ 2 ความสอดคล้องระหว่างคู่มือฯ กับวัตถุประสงค์ของคู่มือ

คำชี้แจง โปรดทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับความคิดเห็นของท่านมากที่สุด

+1 หมายถึง มีความสอดคล้อง/เหมาะสม 0 หมายถึง ไม่แน่ใจ -1 หมายถึง ไม่สอดคล้อง/ไม่เหมาะสม

รายการ	ความคิดเห็น			ข้อเสนอแนะ
	+1	0	-1	
1. หลักการและวัตถุประสงค์ของคู่มือ				
1.1 หลักการและแนวคิด				
1.2 วัตถุประสงค์ของคู่มือ				
1.3 ประโยชน์ที่คาดว่าจะได้รับการใช้คู่มือ				
1.4 คำจำกัดความที่ใช้ในคู่มือ				
2. แนวคิดและทฤษฎีที่เกี่ยวข้อง				
3. ขั้นตอนในการศึกษาความสอดคล้องในแนวเดียวกันฯ				
ขั้นตอนที่ 1 ชั้นเตรียมการ				
ขั้นตอนที่ 2 ชั้นดำเนินการ				
ขั้นตอนที่ 3 ชั้นสรุปผล				

### ส่วนที่ 3 การประเมินคู่มือฯ ตามมาตรฐานด้านอรรถประโยชน์ ความเป็นไปได้ ความเหมาะสม

#### และความถูกต้อง

5 หมายถึง เห็นด้วยระดับมากที่สุด, 4 หมายถึง เห็นด้วยระดับมาก, 3 หมายถึง เห็นด้วยระดับปานกลาง,  
2 หมายถึง เห็นด้วยระดับน้อย, 1 หมายถึง เห็นด้วยระดับน้อยที่สุด

รายการ	ระดับความคิดเห็น				
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
<b>ด้านอรรถประโยชน์ (utility)</b>					
1. คู่มือฯ สามารถเป็นแนวทางในการศึกษาความ สอดคล้องระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบใน การประเมินระดับชั้นเรียน					

รายการ	ระดับความคิดเห็น				
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
2. คู่มือฯ สามารถเป็นแนวทางในการศึกษาความสอดคล้องระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน					
3. คู่มือฯ สามารถเป็นแนวทางในการลดความคลาดเคลื่อนอันเกิดจากอิทธิพลของผู้ประเมินได้					
4. ผลการดำเนินงานตามคู่มือฯ สามารถช่วยให้เกิดการพัฒนาปรับปรุงการวัดและประเมินผลของโรงเรียนให้/สอดคล้องกับมาตรฐานและตัวชี้วัดได้					
5. ผลการดำเนินงานตามคู่มือฯ สามารถช่วยให้เกิดการพัฒนาปรับปรุงการวัดและประเมินผลของโรงเรียนให้/สอดคล้องการวัดและประเมินระดับชาติได้					
6. สารสนเทศจากการดำเนินการตามคู่มือฯ เป็นประโยชน์ต่อการพัฒนา/ปรับปรุงการออกข้อสอบในการประเมินระดับชั้นเรียนให้ครอบคลุมระดับพฤติกรรมกรวัดได้					
7. สารสนเทศจากการดำเนินการตามคู่มือฯ เป็นประโยชน์ต่อพัฒนา/ปรับปรุงการวัดและประเมินระดับชั้นเรียน					
<b>ด้านความเป็นไปได้ (feasibility)</b>					
1. คู่มือฯ ให้แนวทางในการสร้างเครื่องมือในการศึกษาความสอดคล้องระหว่างองค์ประกอบทางการศึกษา					
2. แนวทางการดำเนินงานตามคู่มือฯ สามารถนำไปปฏิบัติจริงได้					
3. แนวทางการดำเนินงานตามคู่มือฯ สามารถนำไปประยุกต์ใช้ในการศึกษาความสอดคล้องระดับองค์ประกอบทางการศึกษาได้					
4. ประโยชน์ที่ได้จากการดำเนินงานตามคู่มือฯ มีความคุ้มค่า เมื่อเทียบกับเวลาและค่าใช้จ่าย					
<b>ด้านความเหมาะสม (propriety)</b>					
1. คู่มือฯ อยู่บนพื้นฐานของแนวคิดและทฤษฎีที่เหมาะสม					
2. แนวทางการดำเนินงานตามคู่มือฯ มีความเหมาะสมตามหลักการวัดและประเมินผล					
3. เกณฑ์การตัดสินผลการประเมินความสอดคล้องในแนวเดียวกันมีความเหมาะสมและยอมรับได้					

รายการ	ระดับความคิดเห็น				
	มากที่สุด	มาก	ปานกลาง	น้อย	น้อยที่สุด
4. ผู้ประเมินตามที่ระบุในคู่มือฯ มีความเหมาะสมและน่าเชื่อถือ					
<b>ด้านความถูกต้อง (accuracy)</b>					
1. ขั้นตอนในการดำเนินงานตามคู่มือฯ มีต่อเนื่องและชัดเจน					
2. ผลจากการดำเนินงานตามคู่มือฯ สามารถให้ผลที่ชัดเจน					
3. การรายงานผลตามคู่มือฯ มีความถูกต้องน่าเชื่อถือ					

### ข้อเสนอแนะ

.....

.....

.....

ขอกราบขอบพระคุณเป็นอย่างสูง  
นางสาวบุษยารัตน์ จันทร์ประเสริฐ  
นิสิตหลักสูตรปริญญาคุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา  
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY



ภาคผนวก ค  
ผลการวิเคราะห์ด้วยโปรแกรม FACETS

จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

## ผลการวิเคราะห์ด้วยโปรแกรม Facets

Facets (Many-Facet Rasch Measurement) Version No. 3.80.3 Copyright ©(c) 1987-2017, John M. Linacre.  
All rights reserved.27/5/2561 20:04:20

Alignment\_Science Item Tests 27/5/2561 20:04:20  
Table 1. Specifications from file "E:\BACKUP\\_Ph.D\\_Dissertation\thesis\analysis\Analyze\  
align final\alignmet\_final.txt".

Title = Alignment\_Science Item Tests 27/5/2561 20:04:20

```
;; Comments =
  Duplicate specification ignored: Positive = 1 ; item are measured with higher score=greater measure
```

```
Data file = (E:\BACKUP\_Ph.D\_Dissertation\thesis\analysis\Analyze\align final\alignmet_final.txt)
Output file = E:\BACKUP\_Ph.D\_Dissertation\thesis\analysis\Analyze\align final\alignmet_final.out.txt
Anchor output file = alignmenta.txt
Graph output file = alignmentg.txt
Residual output file = alignmentr.txt
```

```
; Data specification
Facets = 2
Delements = N
Non-centered = 1
Positive = 1
Labels =
  1,item ; (elements = 1089)
  2,rater ; (elements = 20)
Model = ?,?,ALIGN,1
Rating (or other) scale = ALIGN,R4,General,Ordinal
```

```
; Output description
Arrange tables in order = FN,mN
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
Pt-biserial = Yes
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 2
Scorefile = alignments.txt
Barchart = Yes
Total score for elements = Yes
T3onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 3
Ushort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y
```

```
; Convergence control
Convergence = .5, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5 ;(estimation, bias)
```

Alignment\_Science Item Tests 27/5/2561 20:04:20  
Table 2. Data Summary Report.

```
Assigning models to Data= "E:\BACKUP\_Ph.D\_Dissertation\thesis\analysis\Analyze\  
align final\alignmet_final.txt"
Total lines in data file = 1999
Total data lines = 1881
Responses matched to model: ?,?,ALIGN,1 = 6048
  Total non-blank responses found = 6048
  Number of blank lines = 117
Valid responses used for estimation = 6048
```

Alignment\_Science Item Tests 27/5/2561 20:04:20  
 Table 3. Iteration Report.

Iteration	Max. Score Elements	Residual %	Categories	Max. Logit Elements	Change Steps
PROX 1	Recount required			4.3694	
PROX 2				.9623	
JMLE 3	-96.2231	-31.7	-1123.0156	-.9920	1.5694
JMLE 4	33.9130	-11.1	392.5959	.5831	-.1827
JMLE 5	23.7949	-5.4	46.7468	.2572	.1435
JMLE 6	29.2865	-3.8	107.7744	.2119	.0693
JMLE 7	27.3676	-3.0	95.1763	.1569	.0743
JMLE 8	25.5070	-2.7	95.9236	.1296	.0652
JMLE 9	23.3421	-2.5	91.2642	.1108	.0600
JMLE 10	21.2861	-2.2	86.3386	.0973	.0551
JMLE 11	19.3916	-2.0	81.1016	.0868	.0509
JMLE 12	17.6822	-1.8	75.9800	.0792	.0476
JMLE 13	16.1367	-1.7	71.0513	.0727	.0445
JMLE 14	14.7441	-1.5	66.3801	.0671	.0417
JMLE 15	13.4911	-1.4	61.9939	.0622	.0392
JMLE 16	12.3633	-1.3	57.8936	.0578	.0368
JMLE 17	11.3467	-1.2	54.0691	.0539	.0346
JMLE 18	10.4289	-1.1	50.5059	.0503	.0326
JMLE 19	9.5982	1.0	47.1841	.0470	.0307
JMLE 20	8.8450	.9	44.0886	.0439	.0289
JMLE 21	8.1606	.9	41.2021	.0411	.0272
JMLE 22	7.5371	.8	38.5090	.0386	.0256
JMLE 23	6.9682	.8	35.9949	.0362	.0241
JMLE 24	6.4480	.7	33.6465	.0339	.0228
JMLE 25	5.9713	.7	31.4512	.0319	.0215
JMLE 26	5.5337	.6	29.3972	.0299	.0202
JMLE 27	5.1313	.6	27.4758	.0281	.0191
JMLE 28	4.7608	.5	25.6770	.0264	.0180
JMLE 29	4.4192	.5	23.9927	.0248	.0169
JMLE 30	4.1036	.5	22.4143	.0233	.0160
JMLE 31	3.8120	.4	20.9368	.0219	.0151
JMLE 32	3.5419	.4	19.5505	.0206	.0142
JMLE 33	3.2917	.4	18.2522	.0194	.0134
JMLE 34	3.0596	.3	17.0342	.0182	.0126
JMLE 35	2.8442	.3	15.8931	.0171	.0118
JMLE 36	2.6441	.3	14.8232	.0161	.0111
JMLE 37	2.4581	.3	13.8201	.0151	.0105
JMLE 38	2.2850	.3	12.8796	.0142	.0099
JMLE 39	2.1241	.2	11.9985	.0133	.0093
JMLE 40	1.9742	.2	11.1726	.0125	.0087



JMLE 41	1.8346	.2	10.3992	.0118	.0082
JMLE 42	1.7045	.2	9.6743	.0110	.0077
JMLE 43	1.5834	.2	8.9956	.0104	.0072
JMLE 44	1.4705	.2	8.3606	.0097	.0068
JMLE 45	1.3652	.2	7.7656	.0092	.0063
JMLE 46	1.2670	.1	7.2095	.0086	.0060
JMLE 47	1.1757	.1	6.6897	.0081	.0056
JMLE 48	1.0904	.1	6.2026	.0076	.0052
JMLE 49	1.0110	.1	5.7490	.0071	.0049
JMLE 50	.9370	.1	5.3242	.0067	.0046
JMLE 51	.8680	.1	4.9282	.0063	.0043
JMLE 52	.8038	.1	4.5583	.0059	.0040
JMLE 53	.7440	.1	4.2141	.0055	.0037
JMLE 54	.6884	.1	3.8926	.0051	.0035
JMLE 55	.6366	.1	3.5933	.0048	.0032
JMLE 56	.5884	.1	3.3149	.0045	.0030
JMLE 57	.5436	.1	3.0554	.0042	.0028
JMLE 58	.5021	.1	2.8152	.0039	.0026
JMLE 59	.4633	.1	2.5903	.0037	.0024

-----+  
Subset connection O.K.

Alignment\_Science Item Tests 27/5/2561 20:04:20  
Table 4. Unexpected Responses - appears after Table 8.

Alignment\_Science Item Tests 27/5/2561 20:04:20  
Table 5. Measurable Data Summary.

Cat	Score	Exp.	Resd	StRes	
3.56	3.56	3.56	.00	.02	Mean (Count: 5803)
.71	.71	.51	.50	.94	S.D. (Population)
.71	.71	.51	.50	.94	S.D. (Sample)

-----+  
Data log-likelihood chi-square = 7079.6519

Approximate degrees of freedom = 4741

Chi-square significance prob. = .0000

	Count	Mean	S.D.	Params
Responses used for estimation	= 5803	3.56	0.71	1062
Responses in one extreme score	= 245	4.00	0.00	49
All Responses	= 6048	3.58	0.70	1111
Count of measurable responses	= 6048			
Raw-score variance of observations	= 0.50	100.00%		
Variance explained by Rasch measures	= 0.25	49.55%		

Alignment\_Science Item Tests 27/5/2561 20:04:20

Table 6.0 All Facet Vertical "Rulers".

Vertical = (1\*,2A,S) Yardstick (columns lines low high extreme)= 0,4,-4,7,End

7	+	**.	+		+	(4)
6	+	.	+		+	
5	+	.	+		+	
		***.				
		*****.				
		***.				
4	+	***.	+		+	
		****.				
		*****.				
		**.				
3	+	**.	+		+	
		**				---
		.				
		*				
2	+	.	+		+	
		*		4		
		.				3
		.				
1	+	.	+	16	19	20
		.		18	2	7
		.		10	11	5
		.		1	8	
*	0	*	*	9		*
		.		17	6	
		.		3		2
		.				
-1	+	.	+			---
		.		14		
		.		15		1
		.		12		
-2	+	.	+			---
		.				
		.				
-3	+	.	+			
		.		13		
		.				
-4	+	.	+			(0)
-----						
Measr	*	= 22		-rater		ALIGN
-----						



Alignment\_Science Item Tests 27/5/2561 20:04:20  
 Table 7.1.1 item Measurement Report (arranged by mN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	+ Measure	Model S.E.	Infit MnSq ZStd	Outfit MnSq ZStd	Estim. Discrm	Corr. PtBis	Num item
20	5	4.00	3.97	( 6.25	1.90)	Maximum			.00	47 47
20	5	4.00	3.97	( 6.23	1.90)	Maximum			.00	242 242
20	5	4.00	3.97	( 6.23	1.90)	Maximum			.00	337 337
20	5	4.00	3.97	( 6.17	1.91)	Maximum			.00	106 106
20	5	4.00	3.97	( 6.17	1.91)	Maximum			.00	110 110
20	5	4.00	3.97	( 6.14	1.92)	Maximum			.00	1001 1001
20	5	4.00	3.97	( 6.14	1.92)	Maximum			.00	1016 1016
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	601 601
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	614 614
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	615 615
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	626 626
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	632 632
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	648 648
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	709 709
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	726 726
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	737 737
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	755 755
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	756 756
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	766 766
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	793 793
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	810 810
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	811 811
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	833 833
20	5	4.00	3.96	( 5.79	1.88)	Maximum			.00	130 130
20	5	4.00	3.96	( 5.79	1.88)	Maximum			.00	170 170
20	5	4.00	3.96	( 5.79	1.88)	Maximum			.00	196 196
20	5	4.00	3.96	( 5.76	1.89)	Maximum			.00	450 450
20	5	4.00	3.96	( 5.76	1.89)	Maximum			.00	456 456
20	5	4.00	3.95	( 5.71	1.87)	Maximum			.00	267 267
20	5	4.00	3.97	( 6.23	1.90)	Maximum			.00	337 337
20	5	4.00	3.95	( 5.64	1.88)	Maximum			.00	348 348
20	5	4.00	3.95	( 5.64	1.88)	Maximum			.00	349 349
20	5	4.00	3.95	( 5.64	1.88)	Maximum			.00	355 355
20	5	4.00	3.95	( 5.64	1.88)	Maximum			.00	359 359
20	5	4.00	3.95	( 5.64	1.88)	Maximum			.00	378 378
20	5	4.00	3.95	( 5.53	1.88)	Maximum			.00	388 388
20	5	4.00	3.95	( 5.53	1.88)	Maximum			.00	405 405
20	5	4.00	3.96	( 5.76	1.89)	Maximum			.00	450 450
20	5	4.00	3.96	( 5.76	1.89)	Maximum			.00	456 456
20	5	4.00	3.93	( 5.33	1.89)	Maximum			.00	505 505
20	5	4.00	3.93	( 5.26	1.90)	Maximum			.00	539 539
20	5	4.00	3.94	( 5.43	1.88)	Maximum			.00	552 552
20	5	4.00	3.94	( 5.43	1.88)	Maximum			.00	556 556
20	5	4.00	3.94	( 5.43	1.88)	Maximum			.00	560 560
20	5	4.00	3.94	( 5.43	1.88)	Maximum			.00	574 574
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	601 601
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	614 614
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	615 615
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	626 626
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	632 632
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	648 648
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	709 709
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	726 726
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	737 737
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	755 755
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	756 756
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	766 766
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	793 793
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	810 810
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	811 811
20	5	4.00	3.96	( 5.88	1.89)	Maximum			.00	833 833

20	5	4.00	3.96	(	5.79	1.88)	Maximum							.00	130	130
20	5	4.00	3.96	(	5.79	1.88)	Maximum							.00	170	170
20	5	4.00	3.96	(	5.79	1.88)	Maximum							.00	196	196
20	5	4.00	3.96	(	5.76	1.89)	Maximum							.00	450	450
20	5	4.00	3.96	(	5.76	1.89)	Maximum							.00	456	456
20	5	4.00	3.95	(	5.71	1.87)	Maximum							.00	267	267
20	5	4.00	3.95	(	5.67	1.90)	Maximum							.00	870	870
20	5	4.00	3.95	(	5.67	1.90)	Maximum							.00	879	879
20	5	4.00	3.95	(	5.67	1.90)	Maximum							.00	891	891
20	5	4.00	3.95	(	5.67	1.90)	Maximum							.00	902	902
20	5	4.00	3.95	(	5.64	1.88)	Maximum							.00	348	348
20	5	4.00	3.95	(	5.64	1.88)	Maximum							.00	349	349
20	5	4.00	3.95	(	5.64	1.88)	Maximum							.00	355	355
20	5	4.00	3.95	(	5.64	1.88)	Maximum							.00	359	359
20	5	4.00	3.95	(	5.64	1.88)	Maximum							.00	378	378
20	5	4.00	3.95	(	5.56	1.91)	Maximum							.00	915	915
20	5	4.00	3.95	(	5.56	1.91)	Maximum							.00	932	932
20	5	4.00	3.95	(	5.56	1.91)	Maximum							.00	947	947
20	5	4.00	3.95	(	5.53	1.88)	Maximum							.00	388	388
20	5	4.00	3.95	(	5.53	1.88)	Maximum							.00	405	405
20	5	4.00	3.94	(	5.43	1.88)	Maximum							.00	552	552
20	5	4.00	3.94	(	5.43	1.88)	Maximum							.00	556	556
20	5	4.00	3.94	(	5.43	1.88)	Maximum							.00	560	560
20	5	4.00	3.94	(	5.43	1.88)	Maximum							.00	574	574
20	5	4.00	3.93	(	5.33	1.89)	Maximum							.00	505	505
20	5	4.00	3.93	(	5.26	1.90)	Maximum							.00	539	539
79	20	3.95	3.97		6.03	1.03	.97	.2	.55	.3	1.04	.36	.21	.21	21	21
79	20	3.95	3.97		6.03	1.03	.77	.0	.26	.0	1.15	.56	.26	.26	26	26
79	20	3.95	3.97		6.03	1.03	.98	.2	.61	.4	1.03	.20	.34	.34	34	34
78	20	3.90	3.93		5.27	.76	1.00	.2	.70	.1	1.02	.21	.08	.08	8	8
78	20	3.90	3.93		5.27	.76	.84	.0	.69	.1	1.09	.36	.13	.13	13	13
78	20	3.90	3.93		5.27	.76	.88	.0	1.05	.4	1.03	.30	.15	.15	15	15
78	20	3.90	3.93		5.27	.76	.81	-.1	.58	.0	1.12	.44	.16	.16	16	16
78	20	3.90	3.93		5.27	.76	1.00	.2	.70	.1	1.02	.21	.17	.17	17	17
78	20	3.90	3.93		5.27	.76	.88	.0	1.05	.4	1.03	.30	.19	.19	19	19
78	20	3.90	3.93		5.27	.76	1.00	.2	.70	.1	1.02	.21	.33	.33	33	33
19	5	3.80	3.90		4.84	1.13	.50	-.4	.33	-.5	1.42	.91	.41	.41	41	41
19	5	3.80	3.90		4.84	1.13	.50	-.4	.33	-.5	1.42	.91	.42	.42	42	42
19	5	3.80	3.90		4.84	1.13	.50	-.4	.33	-.5	1.42	.91	.45	.45	45	45
19	5	3.80	3.90		4.84	1.13	.50	-.4	.33	-.5	1.42	.91	.50	.50	50	50
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.51	.51	51	51
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.58	.58	58	58
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.59	.59	59	59
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.60	.60	60	60
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.63	.63	63	63
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.65	.65	65	65
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.66	.66	66	66
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.81	.81	81	81
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.82	.82	82	82
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.83	.83	83	83
19	5	3.80	3.90		4.84	1.13	1.12	.4	.93	.2	.92	.09	.84	.84	84	84
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	219	219	219	219
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	223	223	223	223
19	5	3.80	3.90		4.83	1.13	1.17	.4	1.04	.3	.87	-.10	236	236	236	236
19	5	3.80	3.90		4.83	1.13	1.17	.4	1.04	.3	.87	-.10	237	237	237	237
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	239	239	239	239
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	241	241	241	241
19	5	3.80	3.90		4.83	1.13	1.17	.4	1.04	.3	.87	-.10	247	247	247	247
19	5	3.80	3.90		4.83	1.13	.49	-.4	.32	-.5	1.43	.95	304	304	304	304
19	5	3.80	3.90		4.83	1.13	1.17	.4	1.04	.3	.87	-.10	310	310	310	310
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	311	311	311	311
19	5	3.80	3.90		4.83	1.13	.49	-.4	.32	-.5	1.43	.95	330	330	330	330
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	331	331	331	331
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	332	332	332	332
19	5	3.80	3.90		4.83	1.13	1.17	.4	1.04	.3	.87	-.10	333	333	333	333
19	5	3.80	3.90		4.83	1.13	1.17	.4	1.04	.3	.87	-.10	334	334	334	334
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	338	338	338	338
19	5	3.80	3.90		4.83	1.13	1.17	.4	1.04	.3	.87	-.10	341	341	341	341
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	343	343	343	343
19	5	3.80	3.90		4.83	1.13	1.27	.5	1.33	.6	.74	-.08	344	344	344	344

77	20	3.85	3.89	4.79	.64	.82	-.2	.66	.0	1.12	.40	1 1
77	20	3.85	3.89	4.79	.64	1.13	.4	1.14	.4	.87	.02	6 6
77	20	3.85	3.89	4.79	.64	.91	.0	.95	.2	1.02	.28	9 9
77	20	3.85	3.89	4.79	.64	.83	-.2	.84	.1	1.09	.48	12 12
77	20	3.85	3.89	4.79	.64	.96	.0	1.16	.4	.95	.19	22 22
77	20	3.85	3.89	4.79	.64	.80	-.2	.64	.0	1.14	.43	27 27
77	20	3.85	3.89	4.79	.64	.76	-.3	.54	-.2	1.18	.59	29 29
77	20	3.85	3.89	4.79	.64	.88	-.1	.90	.2	1.05	.33	31 31
77	20	3.85	3.89	4.79	.64	.93	.0	1.10	.4	.98	.29	35 35
77	20	3.85	3.89	4.79	.64	.88	-.1	.90	.2	1.05	.33	37 37
77	20	3.85	3.89	4.79	.64	.74	-.4	.49	-.3	1.21	.64	38 38
77	20	3.85	3.89	4.79	.64	.92	.0	.98	.3	1.01	.29	39 39
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	90 90
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	91 91
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	93 93
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	94 94
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	103 103
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	104 104
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	109 109
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	111 111
19	5	3.80	3.89	4.75	1.14	.47	-.5	.30	-.5	1.44	.91	113 113
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	114 114
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	116 116
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	117 117
19	5	3.80	3.89	4.75	1.14	1.13	.4	.87	.2	.93	.16	126 126
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1009 1009
19	5	3.80	3.88	4.67	1.17	1.41	.7	1.18	1.0	.55	-.04	1012 1012
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1013 1013
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1014 1014
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1015 1015
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1020 1020
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1022 1022
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1023 1023
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1024 1024
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1026 1026
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1038 1038
19	5	3.80	3.88	4.67	1.17	1.09	.3	.67	.8	.95	.40	1039 1039
19	5	3.80	3.88	4.67	1.17	1.41	.7	1.18	1.0	.55	-.04	1041 1041
19	5	3.80	3.88	4.67	1.17	1.41	.7	1.18	1.0	.55	-.04	1042 1042
19	5	3.80	3.88	4.67	1.17	1.41	.7	1.18	1.0	.55	-.04	1043 1043
19	5	3.80	3.88	4.67	1.17	1.41	.7	1.18	1.0	.55	-.04	1044 1044
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	594 594
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	595 595
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	597 597
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	602 602
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	604 604
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	605 605
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	606 606
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	607 607
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	608 608
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	609 609
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	610 610
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	611 611
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	612 612
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	613 613
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	616 616
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	617 617
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	618 618
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	620 620
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	623 623
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	624 624
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	625 625
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	627 627
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	628 628
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	629 629
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	630 630
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	631 631
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	634 634
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	637 637
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	639 639
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	640 640
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	647 647





19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	783	783
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	784	784
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	785	785
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	789	789
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	795	795
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	797	797
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	798	798
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	801	801
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	805	805
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	808	808
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	812	812
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	817	817
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	818	818
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	820	820
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	821	821
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	822	822
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	828	828
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	830	830
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	832	832
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	834	834
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	835	835
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	836	836
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	837	837
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	838	838
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	841	841
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	842	842
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	843	843
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	844	844
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	845	845
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	849	849
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	850	850
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	852	852
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	854	854
19	5	3.80	3.86	4.51	1.12	.94	.1	.74	.1	1.07	.26	855	855
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	856	856
19	5	3.80	3.86	4.51	1.12	.83	.0	.60	.0	1.18	.43	857	857
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	863	863
19	5	3.80	3.86	4.51	1.12	.82	.0	.60	.0	1.18	.44	864	864
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	129	129
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	139	139
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	143	143
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	145	145
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	146	146
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	147	147
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	149	149
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	150	150
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	151	151
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	152	152
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	154	154
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	155	155
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	156	156
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	159	159
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	162	162
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	164	164
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	165	165
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	166	166
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	167	167
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	169	169
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	171	171
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	172	172
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	173	173
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	175	175
19	5	3.80	3.85	4.44	1.10	1.05	.3	1.08	.3	.92	-.28	177	177
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	179	179
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	181	181
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	183	183
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	184	184
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	185	185
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	189	189
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	190	190
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	193	193
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	194	194

19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	195	195
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	198	198
19	5	3.80	3.85	4.44	1.10	1.04	.3	1.07	.3	.93	-.25	199	199
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	202	202
19	5	3.80	3.85	4.44	1.10	.79	.0	.63	-.1	1.21	.83	203	203
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	204	204
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	205	205
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	206	206
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	208	208
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	211	211
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	212	212
19	5	3.80	3.85	4.44	1.10	.86	.0	.72	.0	1.14	.26	213	213
76	20	3.80	3.85	4.43	.57	.97	.0	.85	.0	1.02	.32	5	5
76	20	3.80	3.85	4.43	.57	.82	-.3	.77	.0	1.13	.50	24	24
76	20	3.80	3.85	4.43	.57	.80	-.3	.75	.0	1.14	.52	25	25
76	20	3.80	3.85	4.43	.57	.72	-.6	.52	-.4	1.26	.58	36	36
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	429	429
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	430	430
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	431	431
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	433	433
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	434	434
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	435	435
19	5	3.80	3.84	4.38	1.12	.69	-.1	.48	.0	1.30	.78	436	436
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	437	437
19	5	3.80	3.84	4.38	1.12	.69	-.1	.48	.0	1.30	.78	440	440
19	5	3.80	3.84	4.38	1.12	.69	-.1	.48	.0	1.30	.78	442	442
19	5	3.80	3.84	4.38	1.12	.69	-.1	.48	.0	1.30	.78	447	447
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	448	448
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	449	449
19	5	3.80	3.84	4.38	1.12	.99	.2	.81	.2	1.01	.09	452	452
19	5	3.80	3.84	4.38	1.12	.99	.2	.81	.2	1.01	.09	453	453
19	5	3.80	3.84	4.38	1.12	.69	-.1	.48	.0	1.30	.78	458	458
19	5	3.80	3.84	4.38	1.12	.99	.2	.81	.2	1.01	.09	459	459
19	5	3.80	3.84	4.38	1.12	.69	-.1	.48	.0	1.30	.78	460	460
19	5	3.80	3.84	4.38	1.12	.99	.2	.81	.2	1.01	.09	461	461
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	462	462
19	5	3.80	3.84	4.38	1.12	.98	.2	.79	.2	1.03	.10	463	463
19	5	3.80	3.84	4.38	1.12	.99	.2	.81	.2	1.01	.09	464	464
19	5	3.80	3.84	4.38	1.12	.99	.2	.81	.2	1.01	.09	465	465
19	5	3.80	3.84	4.38	1.12	.69	-.1	.48	.0	1.30	.78	468	468
19	5	3.80	3.84	4.38	1.12	.99	.2	.81	.2	1.01	.09	469	469
19	5	3.80	3.84	4.37	1.10	.93	.1	.86	.1	1.06	.58	260	260
19	5	3.80	3.84	4.37	1.10	.93	.1	.86	.1	1.06	.58	261	261
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	264	264
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	265	265
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	266	266
19	5	3.80	3.84	4.37	1.10	.93	.1	.86	.1	1.06	.58	268	268
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	269	269
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	270	270
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	271	271
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	276	276
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	277	277
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	279	279
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	280	280
19	5	3.80	3.84	4.37	1.10	.93	.1	.86	.1	1.06	.58	283	283
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	285	285
19	5	3.80	3.84	4.37	1.10	.80	.0	.67	-.1	1.19	.52	286	286
19	5	3.80	3.84	4.37	1.10	.93	.1	.86	.1	1.06	.58	287	287
19	5	3.80	3.84	4.37	1.10	.99	.2	.99	.2	.99	-.20	292	292
19	5	3.80	3.84	4.37	1.10	.99	.2	.99	.2	.99	-.20	295	295
19	5	3.80	3.84	4.37	1.10	.99	.2	.99	.2	.99	-.20	299	299
19	5	3.80	3.84	4.37	1.10	.99	.2	.99	.2	.99	-.20	300	300
19	5	3.80	3.84	4.37	1.10	.99	.2	.99	.2	.99	-.20	302	302
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	346	346
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	347	347
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	350	350
19	5	3.80	3.83	4.27	1.11	.82	.0	.61	.0	1.19	.34	353	353
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	354	354
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	358	358
19	5	3.80	3.83	4.27	1.11	1.16	.4	1.25	.6	.78	-.15	360	360
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	361	361



19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	362	362
19	5	3.80	3.83	4.27	1.11	1.16	.4	1.25	.6	.78	-.15	365	365
19	5	3.80	3.83	4.27	1.11	1.16	.4	1.25	.6	.78	-.15	366	366
19	5	3.80	3.83	4.27	1.11	1.16	.4	1.25	.6	.78	-.15	367	367
19	5	3.80	3.83	4.27	1.11	.82	.0	.61	.0	1.19	.34	369	369
19	5	3.80	3.83	4.27	1.11	1.16	.4	1.25	.6	.78	-.15	371	371
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	375	375
19	5	3.80	3.83	4.27	1.11	1.16	.4	1.25	.6	.78	-.15	376	376
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	377	377
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	379	379
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	380	380
19	5	3.80	3.83	4.27	1.11	1.16	.4	1.25	.6	.78	-.15	381	381
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	384	384
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	385	385
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	386	386
19	5	3.80	3.83	4.27	1.11	.90	.1	.71	.1	1.10	.36	387	387
19	5	3.80	3.83	4.25	1.15	1.56	.8	3.28	1.4	-.04	-.33	865	865
19	5	3.80	3.83	4.25	1.15	.70	-.1	.43	.4	1.31	.73	873	873
19	5	3.80	3.83	4.25	1.15	.70	-.1	.43	.4	1.31	.73	874	874
19	5	3.80	3.83	4.25	1.15	.70	-.1	.43	.4	1.31	.73	876	876
19	5	3.80	3.83	4.25	1.15	.70	-.1	.43	.4	1.31	.73	877	877
19	5	3.80	3.83	4.25	1.15	.70	-.1	.43	.4	1.31	.73	880	880
19	5	3.80	3.83	4.25	1.15	1.17	.4	.92	.7	.82	-.07	882	882
19	5	3.80	3.83	4.25	1.15	.70	-.1	.43	.4	1.31	.73	883	883
19	5	3.80	3.83	4.25	1.15	.70	-.1	.43	.4	1.31	.73	884	884
19	5	3.80	3.83	4.25	1.15	1.17	.4	.92	.7	.82	-.07	889	889
19	5	3.80	3.83	4.25	1.15	1.17	.4	.92	.7	.82	-.07	892	892
19	5	3.80	3.83	4.25	1.15	1.17	.4	.92	.7	.82	-.07	893	893
19	5	3.80	3.83	4.25	1.15	1.17	.4	.92	.7	.82	-.07	894	894
19	5	3.80	3.83	4.25	1.15	1.17	.4	.92	.7	.82	-.07	896	896
19	5	3.80	3.83	4.25	1.15	1.17	.4	.92	.7	.82	-.07	897	897
19	5	3.80	3.83	4.25	1.15	.70	-.1	.43	.4	1.31	.73	900	900
19	5	3.80	3.83	4.25	1.15	1.17	.4	.92	.7	.82	-.07	901	901
19	5	3.80	3.83	4.25	1.15	.70	-.1	.43	.4	1.31	.73	904	904
19	5	3.80	3.83	4.25	1.15	1.17	.4	.92	.7	.82	-.07	908	908
19	5	3.80	3.81	4.16	1.11	.98	.2	.84	.2	1.02	.05	390	390
19	5	3.80	3.81	4.16	1.11	.98	.2	.84	.2	1.02	.05	391	391
19	5	3.80	3.81	4.16	1.11	.83	.0	.64	.0	1.17	.46	392	392
19	5	3.80	3.81	4.16	1.11	.98	.2	.84	.2	1.02	.05	395	395
19	5	3.80	3.81	4.16	1.11	1.11	.3	1.12	.5	.85	-.08	397	397
19	5	3.80	3.81	4.16	1.11	1.11	.3	1.12	.5	.85	-.08	398	398
19	5	3.80	3.81	4.16	1.11	.98	.2	.84	.2	1.02	.05	399	399
19	5	3.80	3.81	4.16	1.11	.98	.2	.84	.2	1.02	.05	400	400
19	5	3.80	3.81	4.16	1.11	.83	.0	.64	.0	1.17	.46	403	403
19	5	3.80	3.81	4.16	1.11	.85	.0	.65	.0	1.16	.43	404	404
19	5	3.80	3.81	4.16	1.11	.98	.2	.84	.2	1.02	.05	406	406
19	5	3.80	3.81	4.16	1.11	.85	.0	.65	.0	1.16	.43	408	408
19	5	3.80	3.81	4.16	1.11	.85	.0	.65	.0	1.16	.43	410	410
19	5	3.80	3.81	4.16	1.11	.83	.0	.64	.0	1.17	.46	411	411
19	5	3.80	3.81	4.16	1.11	.85	.0	.65	.0	1.16	.43	413	413
19	5	3.80	3.81	4.16	1.11	.83	.0	.64	.0	1.17	.46	414	414
19	5	3.80	3.81	4.16	1.11	.85	.0	.65	.0	1.16	.43	415	415
19	5	3.80	3.81	4.16	1.11	.85	.0	.65	.0	1.16	.43	416	416
19	5	3.80	3.81	4.16	1.11	.83	.0	.64	.0	1.17	.46	418	418
19	5	3.80	3.81	4.16	1.11	.85	.0	.65	.0	1.16	.43	422	422
75	20	3.75	3.81	4.13	.52	1.16	.5	.82	.0	.99	.51	11	11
19	5	3.80	3.80	4.12	1.16	.65	-.2	.38	.3	1.34	.75	913	913
19	5	3.80	3.80	4.12	1.16	1.46	.7	1.55	.9	.44	-.19	918	918
19	5	3.80	3.80	4.12	1.16	.82	.0	.48	.4	1.20	.37	919	919
19	5	3.80	3.80	4.12	1.16	.65	-.2	.38	.3	1.34	.75	926	926
19	5	3.80	3.80	4.12	1.16	.65	-.2	.38	.3	1.34	.75	933	933
19	5	3.80	3.80	4.12	1.16	.82	.0	.48	.4	1.20	.37	937	937
19	5	3.80	3.80	4.12	1.16	.65	-.2	.38	.3	1.34	.75	938	938
19	5	3.80	3.80	4.12	1.16	.65	-.2	.38	.3	1.34	.75	939	939
19	5	3.80	3.80	4.12	1.16	.65	-.2	.38	.3	1.34	.75	940	940
19	5	3.80	3.80	4.12	1.16	.65	-.2	.38	.3	1.34	.75	945	945
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	553	553
19	5	3.80	3.80	4.06	1.12	.79	.0	.58	.0	1.21	.51	554	554
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	555	555
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	557	557
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	558	558
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	559	559

19	5	3.80	3.80	4.06	1.12	.79	.0	.58	.0	1.21	.51	562	562
19	5	3.80	3.80	4.06	1.12	.79	.0	.58	.0	1.21	.51	563	563
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	564	564
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	565	565
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	566	566
19	5	3.80	3.80	4.06	1.12	.79	.0	.58	.0	1.21	.51	567	567
19	5	3.80	3.80	4.06	1.12	.79	.0	.58	.0	1.21	.51	568	568
19	5	3.80	3.80	4.06	1.12	.79	.0	.58	.0	1.21	.51	569	569
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	570	570
19	5	3.80	3.80	4.06	1.12	1.09	.3	1.03	.4	.89	-.03	571	571
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	572	572
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	573	573
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	575	575
19	5	3.80	3.80	4.06	1.12	1.09	.3	1.03	.4	.89	-.03	578	578
19	5	3.80	3.80	4.06	1.12	1.09	.3	1.03	.4	.89	-.03	579	579
19	5	3.80	3.80	4.06	1.12	1.09	.3	1.03	.4	.89	-.03	580	580
19	5	3.80	3.80	4.06	1.12	.79	.0	.58	.0	1.21	.51	581	581
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	582	582
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	584	584
19	5	3.80	3.80	4.06	1.12	1.09	.3	1.03	.4	.89	-.03	587	587
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	591	591
19	5	3.80	3.80	4.06	1.12	.80	.0	.60	.0	1.20	.48	592	592
19	5	3.80	3.79	4.02	1.16	1.32	.6	1.18	.7	.66	-.16	1045	1045
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1049	1049
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1057	1057
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1062	1062
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1066	1066
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1070	1070
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1071	1071
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1072	1072
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1073	1073
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1074	1074
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1075	1075
19	5	3.80	3.79	4.02	1.16	.55	-.3	.33	.2	1.41	.83	1079	1079
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1083	1083
19	5	3.80	3.79	4.02	1.16	1.00	.2	.65	.5	1.04	.22	1084	1084
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	470	470
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	471	471
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	472	472
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	473	473
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	475	475
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	477	477
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	478	478
19	5	3.80	3.77	3.94	1.13	1.09	.3	.93	.3	.91	.01	482	482
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	483	483
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	484	484
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	485	485
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	487	487
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	488	488
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	489	489
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	491	491
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	492	492
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	493	493
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	496	496
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	497	497
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	498	498
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	500	500
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	501	501
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	502	502
19	5	3.80	3.77	3.94	1.13	.77	.0	.52	.0	1.24	.53	503	503
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	506	506
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	507	507
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	508	508
19	5	3.80	3.77	3.94	1.13	.78	.0	.53	.0	1.22	.50	509	509
74	20	3.70	3.76	3.88	.49	.93	.0	.85	.0	1.03	.24	2	2
74	20	3.70	3.76	3.88	.49	.81	-.4	.72	-.2	1.18	.38	10	10
18	5	3.60	3.76	3.86	.89	.69	-.3	.73	-.2	1.31	.44	46	46
18	5	3.60	3.76	3.86	.89	.46	-.8	.45	-.8	1.63	.82	48	48
18	5	3.60	3.76	3.86	.89	.46	-.8	.45	-.8	1.63	.82	49	49
18	5	3.60	3.76	3.86	.89	.69	-.3	.73	-.2	1.31	.44	52	52
18	5	3.60	3.76	3.86	.89	.69	-.3	.73	-.2	1.31	.44	53	53
18	5	3.60	3.76	3.86	.89	.63	-.4	.64	-.3	1.40	.63	64	64

18	5	3.60	3.76	3.86	.89	.69	-.3	.73	-.2	1.31	.44	68	68
18	5	3.60	3.76	3.86	.89	.46	-.8	.45	-.8	1.63	.82	69	69
18	5	3.60	3.76	3.86	.89	.63	-.4	.64	-.3	1.40	.63	70	70
18	5	3.60	3.76	3.86	.89	.63	-.4	.64	-.3	1.40	.63	71	71
18	5	3.60	3.76	3.86	.89	.69	-.3	.73	-.2	1.31	.44	72	72
18	5	3.60	3.76	3.86	.89	.46	-.8	.45	-.8	1.63	.82	73	73
18	5	3.60	3.76	3.86	.89	.69	-.3	.73	-.2	1.31	.44	76	76
18	5	3.60	3.76	3.86	.89	.63	-.4	.64	-.3	1.40	.63	77	77
18	5	3.60	3.76	3.85	.88	.50	-.7	.50	-.7	1.59	.70	218	218
18	5	3.60	3.76	3.85	.88	.50	-.7	.50	-.7	1.59	.70	221	221
18	5	3.60	3.76	3.85	.88	1.46	.8	1.12	.3	.81	.95	225	225
18	5	3.60	3.76	3.85	.88	1.46	.8	1.12	.3	.81	.95	226	226
18	5	3.60	3.76	3.85	.88	.50	-.7	.50	-.7	1.59	.70	227	227
18	5	3.60	3.76	3.85	.88	.50	-.7	.50	-.7	1.59	.70	228	228
18	5	3.60	3.76	3.85	.88	.50	-.7	.50	-.7	1.59	.70	233	233
18	5	3.60	3.76	3.85	.88	1.46	.8	1.12	.3	.81	.95	234	234
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	243	243
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	244	244
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	245	245
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	246	246
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	248	248
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	249	249
18	5	3.60	3.76	3.85	.88	.50	-.7	.50	-.7	1.59	.70	250	250
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	253	253
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	255	255
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	256	256
18	5	3.60	3.76	3.85	.88	.50	-.7	.50	-.7	1.59	.70	257	257
18	5	3.60	3.76	3.85	.88	.50	-.7	.50	-.7	1.59	.70	258	258
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	303	303
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	305	305
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	306	306
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	307	307
18	5	3.60	3.76	3.85	.88	.61	-.5	.63	-.4	1.43	.72	308	308
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	309	309
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	314	314
18	5	3.60	3.76	3.85	.88	1.46	.8	1.12	.3	.81	.95	315	315
18	5	3.60	3.76	3.85	.88	1.46	.8	1.12	.3	.81	.95	319	319
18	5	3.60	3.76	3.85	.88	1.46	.8	1.12	.3	.81	.95	320	320
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	321	321
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	323	323
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	324	324
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	325	325
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	326	326
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	327	327
18	5	3.60	3.76	3.85	.88	.67	-.3	.73	-.2	1.33	.51	336	336
18	5	3.60	3.76	3.85	.88	1.46	.8	1.12	.3	.81	.95	340	340
19	5	3.80	3.76	3.85	1.15	.79	.0	.49	.3	1.22	.51	511	511
19	5	3.80	3.76	3.85	1.15	.77	.0	.48	.3	1.24	.53	512	512
19	5	3.80	3.76	3.85	1.15	.79	.0	.49	.3	1.22	.51	513	513
19	5	3.80	3.76	3.85	1.15	.79	.0	.49	.3	1.22	.51	514	514
19	5	3.80	3.76	3.85	1.15	.79	.0	.49	.3	1.22	.51	515	515
19	5	3.80	3.76	3.85	1.15	.79	.0	.49	.3	1.22	.51	517	517
19	5	3.80	3.76	3.85	1.15	.79	.0	.49	.3	1.22	.51	519	519
19	5	3.80	3.76	3.85	1.15	.77	.0	.48	.3	1.24	.53	523	523
19	5	3.80	3.76	3.85	1.15	.77	.0	.48	.3	1.24	.53	524	524
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	525	525
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	526	526
19	5	3.80	3.76	3.85	1.15	.79	.0	.49	.3	1.22	.51	529	529
19	5	3.80	3.76	3.85	1.15	.77	.0	.48	.3	1.24	.53	530	530
19	5	3.80	3.76	3.85	1.15	.79	.0	.49	.3	1.22	.51	531	531
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	532	532
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	533	533
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	534	534
19	5	3.80	3.76	3.85	1.15	.77	.0	.48	.3	1.24	.53	537	537
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	542	542
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	543	543
19	5	3.80	3.76	3.85	1.15	.77	.0	.48	.3	1.24	.53	544	544
19	5	3.80	3.76	3.85	1.15	.77	.0	.48	.3	1.24	.53	545	545
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	546	546
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	547	547
19	5	3.80	3.76	3.85	1.15	.79	.0	.49	.3	1.22	.51	548	548
19	5	3.80	3.76	3.85	1.15	1.13	.4	.86	.6	.86	.11	551	551

18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	85 85
18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	86 86
18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	87 87
18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	88 88
18	5	3.60	3.74	3.75	.90	1.67	1.0	1.82	1.1	.10	-.54	89 89
18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	92 92
18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	96 96
18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	97 97
18	5	3.60	3.74	3.75	.90	.64	-.4	.65	-.3	1.36	.53	99 99
18	5	3.60	3.74	3.75	.90	.64	-.4	.65	-.3	1.36	.53	100 100
18	5	3.60	3.74	3.75	.90	.64	-.4	.65	-.3	1.36	.53	101 101
18	5	3.60	3.74	3.75	.90	.64	-.4	.65	-.3	1.36	.53	107 107
18	5	3.60	3.74	3.75	.90	.64	-.4	.65	-.3	1.36	.53	108 108
18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	112 112
18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	115 115
18	5	3.60	3.74	3.75	.90	.64	-.4	.65	-.3	1.36	.53	118 118
18	5	3.60	3.74	3.75	.90	.64	-.4	.65	-.3	1.36	.53	119 119
18	5	3.60	3.74	3.75	.90	1.43	.7	1.06	.3	.86	.91	123 123
18	5	3.60	3.74	3.75	.90	.64	-.4	.65	-.3	1.36	.53	124 124
18	5	3.60	3.74	3.75	.90	.80	.0	.99	.2	1.08	.40	125 125
73	20	3.65	3.71	3.65	.47	1.16	.5	.97	.1	.92	.41	3 3
73	20	3.65	3.71	3.65	.47	.89	-.2	1.20	.5	.97	.38	20 20
30	8	3.75	3.71	3.65	.86	1.23	.5	.85	.4	.81	.04	1089 1089
18	5	3.60	3.70	3.59	.95	.66	-.2	.51	.2	1.31	.53	1008 1008
18	5	3.60	3.70	3.59	.95	.66	-.2	.51	.2	1.31	.53	1017 1017
18	5	3.60	3.70	3.59	.95	.31	-1.0	.25	.0	1.67	.90	1028 1028
18	5	3.60	3.70	3.59	.95	.31	-1.0	.25	.0	1.67	.90	1034 1034
18	5	3.60	3.70	3.59	.95	.31	-1.0	.25	.0	1.67	.90	1036 1036
18	5	3.60	3.70	3.59	.95	.31	-1.0	.25	.0	1.67	.90	1040 1040
18	5	3.60	3.69	3.54	.89	.65	-.4	.55	-.4	1.45	.58	603 603
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	636 636
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	638 638
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	642 642
18	5	3.60	3.69	3.54	.89	.65	-.4	.55	-.4	1.45	.58	665 665
18	5	3.60	3.69	3.54	.89	.65	-.4	.55	-.4	1.45	.58	675 675
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	679 679
18	5	3.60	3.69	3.54	.89	.65	-.4	.55	-.4	1.45	.58	681 681
18	5	3.60	3.69	3.54	.89	.65	-.4	.55	-.4	1.45	.58	684 684
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	697 697
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	710 710
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	718 718
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	722 722
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	724 724
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	738 738
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	741 741
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	744 744
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	746 746
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	752 752
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	754 754
18	5	3.60	3.69	3.54	.89	1.16	.4	1.11	.3	.73	-.07	760 760
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	761 761
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	786 786
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	787 787
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	788 788
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	790 790
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	791 791
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	792 792
18	5	3.60	3.69	3.54	.89	1.05	.2	1.03	.2	.86	.07	794 794
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	796 796
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	799 799
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	800 800
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	802 802
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	803 803
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	804 804
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	807 807
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	813 813
18	5	3.60	3.69	3.54	.89	.65	-.4	.55	-.4	1.45	.58	815 815
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	829 829
18	5	3.60	3.69	3.54	.89	.54	-.6	.46	-.5	1.58	.72	831 831
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	839 839
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	846 846
18	5	3.60	3.69	3.54	.89	1.04	.2	1.02	.2	.87	.08	847 847



18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	853	853
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	859	859
18	5	3.60	3.69	3.54	.89	.65	-.4	.55	-.4	1.45	.58	860	860
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	861	861
18	5	3.60	3.69	3.54	.89	.66	-.4	.55	-.4	1.44	.57	862	862
18	5	3.60	3.68	3.52	.86	.84	-.1	.80	-.1	1.21	-.02	131	131
18	5	3.60	3.68	3.52	.86	.84	-.1	.80	-.1	1.21	-.02	132	132
18	5	3.60	3.68	3.52	.86	.78	-.2	.75	-.2	1.31	.45	133	133
18	5	3.60	3.68	3.52	.86	.59	-.6	.55	-.7	1.60	.90	134	134
18	5	3.60	3.68	3.52	.86	.78	-.2	.75	-.2	1.31	.45	135	135
18	5	3.60	3.68	3.52	.86	.59	-.6	.55	-.7	1.60	.90	137	137
18	5	3.60	3.68	3.52	.86	.77	-.2	.74	-.2	1.32	.47	140	140
18	5	3.60	3.68	3.52	.86	.78	-.2	.75	-.2	1.31	.45	141	141
18	5	3.60	3.68	3.52	.86	.59	-.6	.55	-.7	1.60	.90	153	153
18	5	3.60	3.68	3.52	.86	.59	-.6	.55	-.7	1.60	.90	157	157
18	5	3.60	3.68	3.52	.86	.78	-.2	.75	-.2	1.31	.45	158	158
18	5	3.60	3.68	3.52	.86	.59	-.6	.55	-.7	1.60	.90	160	160
18	5	3.60	3.68	3.52	.86	.59	-.6	.55	-.7	1.60	.90	161	161
18	5	3.60	3.68	3.52	.86	.78	-.2	.75	-.2	1.31	.45	163	163
18	5	3.60	3.68	3.52	.86	.78	-.2	.75	-.2	1.31	.45	168	168
18	5	3.60	3.68	3.52	.86	1.02	.2	.99	.1	.93	-.43	174	174
18	5	3.60	3.68	3.52	.86	.84	-.1	.80	-.1	1.21	-.02	176	176
18	5	3.60	3.68	3.52	.86	.78	-.2	.75	-.2	1.31	.45	178	178
18	5	3.60	3.68	3.52	.86	.78	-.2	.75	-.2	1.31	.45	191	191
18	5	3.60	3.68	3.52	.86	.78	-.2	.75	-.2	1.31	.45	192	192
18	5	3.60	3.68	3.52	.86	.77	-.2	.74	-.2	1.32	.47	197	197
18	5	3.60	3.68	3.52	.86	.77	-.2	.74	-.2	1.32	.47	200	200
18	5	3.60	3.68	3.52	.86	.77	-.2	.74	-.2	1.32	.47	201	201
18	5	3.60	3.68	3.52	.86	.77	-.2	.74	-.2	1.32	.47	207	207
18	5	3.60	3.68	3.52	.86	.77	-.2	.74	-.2	1.32	.47	209	209
18	5	3.60	3.68	3.52	.86	.77	-.2	.74	-.2	1.32	.47	210	210
18	5	3.60	3.68	3.52	.86	.84	-.1	.80	-.1	1.21	-.02	214	214
18	5	3.60	3.68	3.52	.86	.77	-.2	.74	-.2	1.32	.47	215	215
18	5	3.60	3.68	3.52	.86	2.01	1.5	1.74	1.2	.00	.82	216	216
18	5	3.60	3.67	3.46	.86	.68	-.4	.66	-.5	1.48	.92	262	262
18	5	3.60	3.67	3.46	.86	.68	-.4	.66	-.5	1.48	.92	263	263
18	5	3.60	3.67	3.46	.86	.94	.0	.92	.0	1.06	-.35	274	274
18	5	3.60	3.67	3.46	.86	.94	.0	.92	.0	1.06	-.35	275	275
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	278	278
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	281	281
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	284	284
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	288	288
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	289	289
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	290	290
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	291	291
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	294	294
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	296	296
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	297	297
18	5	3.60	3.67	3.46	.86	.75	-.3	.74	-.3	1.36	.23	298	298
18	5	3.60	3.67	3.46	.86	.74	-.3	.73	-.3	1.37	.27	301	301
18	5	3.60	3.66	3.41	.89	1.10	.3	.97	.2	.84	-.10	438	438
18	5	3.60	3.66	3.41	.89	.83	.0	.78	.0	1.17	.44	439	439
18	5	3.60	3.66	3.41	.89	.83	.0	.78	.0	1.17	.44	441	441
18	5	3.60	3.66	3.41	.89	.83	.0	.78	.0	1.17	.44	443	443
18	5	3.60	3.66	3.41	.89	.57	-.6	.49	-.4	1.55	.72	451	451
18	5	3.60	3.66	3.41	.89	.58	-.6	.50	-.4	1.53	.71	454	454
18	5	3.60	3.66	3.41	.89	.58	-.6	.50	-.4	1.53	.71	455	455
18	5	3.60	3.66	3.41	.89	.83	.0	.78	.0	1.17	.44	466	466
18	5	3.60	3.66	3.41	.89	.83	.0	.78	.0	1.17	.44	467	467
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	955	955
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	956	956
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	957	957
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	958	958
18	5	3.60	3.66	3.40	.96	1.15	.4	1.56	.8	.59	.32	966	966
18	5	3.60	3.66	3.40	.96	1.15	.4	1.56	.8	.59	.32	967	967
18	5	3.60	3.66	3.40	.96	1.15	.4	1.56	.8	.59	.32	969	969
18	5	3.60	3.66	3.40	.96	1.15	.4	1.56	.8	.59	.32	970	970
18	5	3.60	3.66	3.40	.96	1.15	.4	1.56	.8	.59	.32	971	971
18	5	3.60	3.66	3.40	.96	1.15	.4	1.56	.8	.59	.32	972	972
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	975	975
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	976	976
18	5	3.60	3.66	3.40	.96	1.48	.7	.89	.4	.83	.72	978	978

18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	979	979
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	980	980
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	981	981
18	5	3.60	3.66	3.40	.96	1.48	.7	.89	.4	.83	.72	982	982
18	5	3.60	3.66	3.40	.96	1.48	.7	.89	.4	.83	.72	983	983
18	5	3.60	3.66	3.40	.96	1.48	.7	.89	.4	.83	.72	984	984
18	5	3.60	3.66	3.40	.96	1.15	.4	1.56	.8	.59	.32	985	985
18	5	3.60	3.66	3.40	.96	1.48	.7	.89	.4	.83	.72	986	986
18	5	3.60	3.66	3.40	.96	1.48	.7	.89	.4	.83	.72	987	987
18	5	3.60	3.66	3.40	.96	1.48	.7	.89	.4	.83	.72	988	988
18	5	3.60	3.66	3.40	.96	1.15	.4	1.56	.8	.59	.32	989	989
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	990	990
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	991	991
18	5	3.60	3.66	3.40	.96	1.48	.7	.89	.4	.83	.72	992	992
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	996	996
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	997	997
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	998	998
18	5	3.60	3.66	3.40	.96	.96	.2	.88	.4	.94	.39	999	999
18	5	3.60	3.63	3.32	.88	.63	-.5	.54	-.4	1.50	.58	351	351
18	5	3.60	3.63	3.32	.88	.63	-.5	.54	-.4	1.50	.58	352	352
18	5	3.60	3.63	3.32	.88	.97	.1	.87	.1	1.01	.18	356	356
18	5	3.60	3.63	3.32	.88	.97	.1	.87	.1	1.01	.18	357	357
18	5	3.60	3.63	3.32	.88	.97	.1	.87	.1	1.01	.18	363	363
18	5	3.60	3.63	3.32	.88	.97	.1	.87	.1	1.01	.18	368	368
18	5	3.60	3.63	3.32	.88	.97	.1	.87	.1	1.01	.18	372	372
18	5	3.60	3.63	3.32	.88	.97	.1	.87	.1	1.01	.18	373	373
18	5	3.60	3.63	3.32	.88	.90	.0	.81	.0	1.10	.16	382	382
18	5	3.60	3.63	3.32	.88	.97	.1	.87	.1	1.01	.18	383	383
71	20	3.55	3.62	3.25	.43	.66	-1.0	.60	-.8	1.39	.51	14	14
18	5	3.60	3.61	3.22	.93	.69	-.2	.52	.0	1.34	.54	866	866
18	5	3.60	3.61	3.22	.93	.69	-.2	.52	.0	1.34	.54	869	869
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	881	881
18	5	3.60	3.61	3.22	.93	.69	-.2	.52	.0	1.34	.54	885	885
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	886	886
18	5	3.60	3.61	3.22	.93	.69	-.2	.52	.0	1.34	.54	888	888
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	890	890
18	5	3.60	3.61	3.22	.93	1.21	.5	1.45	.7	.51	.33	895	895
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	898	898
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	899	899
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	903	903
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	905	905
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	906	906
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	907	907
18	5	3.60	3.61	3.22	.93	.83	.0	.60	.1	1.21	.23	909	909
18	5	3.60	3.61	3.21	.88	.86	.0	.77	.0	1.17	.32	389	389
18	5	3.60	3.61	3.21	.88	.86	.0	.77	.0	1.17	.32	393	393
18	5	3.60	3.61	3.21	.88	.86	.0	.77	.0	1.17	.32	394	394
18	5	3.60	3.61	3.21	.88	.86	.0	.77	.0	1.17	.32	396	396
18	5	3.60	3.61	3.21	.88	1.00	.1	.88	.1	.98	-.02	401	401
18	5	3.60	3.61	3.21	.88	1.31	.6	2.64	1.7	.02	-.39	407	407
18	5	3.60	3.61	3.21	.88	.74	-.2	.63	-.2	1.36	.40	409	409
18	5	3.60	3.61	3.21	.88	.87	.0	.78	.0	1.15	.29	417	417
18	5	3.60	3.61	3.21	.88	.86	.0	.77	.0	1.17	.32	419	419
18	5	3.60	3.61	3.21	.88	1.00	.1	.88	.1	.98	-.02	420	420
18	5	3.60	3.61	3.21	.88	.87	.0	.78	.0	1.15	.29	421	421
18	5	3.60	3.61	3.21	.88	.87	.0	.78	.0	1.15	.29	423	423
18	5	3.60	3.61	3.21	.88	.87	.0	.78	.0	1.15	.29	424	424
18	5	3.60	3.61	3.21	.88	.86	.0	.77	.0	1.17	.32	425	425
18	5	3.60	3.61	3.21	.88	.87	.0	.78	.0	1.15	.29	426	426
18	5	3.60	3.61	3.21	.88	.87	.0	.78	.0	1.15	.29	427	427
17	5	3.40	3.59	3.16	.79	.40	-1.0	.43	-1.0	1.76	.71	43	43
17	5	3.40	3.59	3.16	.79	.47	-.8	.49	-.8	1.67	.52	54	54
17	5	3.40	3.59	3.16	.79	.47	-.8	.49	-.8	1.67	.52	55	55
17	5	3.40	3.59	3.16	.79	.99	.1	.89	.0	1.13	.96	56	56
17	5	3.40	3.59	3.16	.79	.47	-.8	.49	-.8	1.67	.52	61	61
17	5	3.40	3.59	3.16	.79	1.48	.8	1.42	.8	.53	.54	62	62
17	5	3.40	3.59	3.16	.78	1.02	.2	.93	.0	1.09	.91	217	217
17	5	3.40	3.59	3.16	.78	.50	-.8	.54	-.7	1.64	.44	222	222
17	5	3.40	3.59	3.16	.78	.44	-.9	.47	-.9	1.75	.64	224	224
17	5	3.40	3.59	3.16	.78	1.19	.5	1.10	.3	.82	.79	229	229
17	5	3.40	3.59	3.16	.78	.50	-.8	.54	-.7	1.64	.44	231	231
17	5	3.40	3.59	3.16	.78	1.13	.4	1.03	.2	.93	.92	235	235

17	5	3.40	3.59	3.16	.78	1.13	.4	1.03	.2	.93	.92	238	238
17	5	3.40	3.59	3.16	.78	1.19	.5	1.10	.3	.82	.79	240	240
17	5	3.40	3.59	3.16	.78	1.19	.5	1.10	.3	.82	.79	259	259
17	5	3.40	3.59	3.16	.78	.50	-.8	.54	-.7	1.64	.44	312	312
17	5	3.40	3.59	3.16	.78	.50	-.8	.54	-.7	1.64	.44	318	318
17	5	3.40	3.59	3.16	.78	.50	-.8	.54	-.7	1.64	.44	322	322
17	5	3.40	3.59	3.16	.78	.61	-.5	.64	-.5	1.48	.45	329	329
17	5	3.40	3.59	3.16	.78	.61	-.5	.64	-.5	1.48	.45	335	335
17	5	3.40	3.59	3.16	.78	1.13	.4	1.03	.2	.93	.92	342	342
17	5	3.40	3.59	3.16	.78	1.13	.4	1.03	.2	.93	.92	345	345
18	5	3.60	3.58	3.11	.88	.79	-.1	.69	-.1	1.27	.38	576	576
18	5	3.60	3.58	3.11	.88	.79	-.1	.69	-.1	1.27	.38	583	583
18	5	3.60	3.58	3.11	.88	.79	-.1	.69	-.1	1.27	.38	590	590
70	20	3.50	3.56	3.06	.42	.75	-.7	.68	-.6	1.28	.32	7	7
18	5	3.60	3.56	3.06	.93	1.14	.4	1.25	.6	.68	.37	921	921
18	5	3.60	3.56	3.06	.93	.91	.1	.75	.2	1.06	.46	922	922
18	5	3.60	3.56	3.06	.93	1.14	.4	1.25	.6	.68	.37	923	923
18	5	3.60	3.56	3.06	.93	1.14	.4	1.25	.6	.68	.37	924	924
18	5	3.60	3.56	3.06	.93	1.14	.4	1.25	.6	.68	.37	925	925
18	5	3.60	3.56	3.06	.93	1.05	.3	.83	.3	.93	.15	927	927
18	5	3.60	3.56	3.06	.93	1.05	.3	.83	.3	.93	.15	928	928
18	5	3.60	3.56	3.06	.93	.91	.1	.75	.2	1.06	.46	929	929
18	5	3.60	3.56	3.06	.93	.91	.1	.75	.2	1.06	.46	930	930
18	5	3.60	3.56	3.06	.93	.91	.1	.75	.2	1.06	.46	931	931
18	5	3.60	3.56	3.06	.93	1.05	.3	.83	.3	.93	.15	934	934
18	5	3.60	3.56	3.06	.93	.91	.1	.75	.2	1.06	.46	936	936
18	5	3.60	3.56	3.06	.93	1.05	.3	.83	.3	.93	.15	941	941
18	5	3.60	3.56	3.06	.93	.91	.1	.75	.2	1.06	.46	942	942
18	5	3.60	3.56	3.06	.93	.91	.1	.75	.2	1.06	.46	944	944
18	5	3.60	3.56	3.06	.93	.91	.1	.75	.2	1.06	.46	946	946
18	5	3.60	3.56	3.06	.93	1.05	.3	.83	.3	.93	.15	948	948
18	5	3.60	3.56	3.06	.93	.91	.1	.75	.2	1.06	.46	949	949
18	5	3.60	3.56	3.06	.93	1.05	.3	.83	.3	.93	.15	950	950
18	5	3.60	3.56	3.06	.93	1.05	.3	.83	.3	.93	.15	951	951
18	5	3.60	3.56	3.06	.93	1.14	.4	1.25	.6	.68	.37	952	952
18	5	3.60	3.56	3.06	.93	1.05	.3	.83	.3	.93	.15	953	953
18	5	3.60	3.56	3.06	.93	1.05	.3	.83	.3	.93	.15	954	954
17	5	3.40	3.56	3.04	.79	1.37	.7	1.30	.6	.69	.61	98	98
17	5	3.40	3.56	3.04	.79	1.37	.7	1.30	.6	.69	.61	102	102
17	5	3.40	3.56	3.04	.79	1.37	.7	1.30	.6	.69	.61	105	105
17	5	3.40	3.56	3.04	.79	.54	-.6	.61	-.5	1.51	.53	120	120
17	5	3.40	3.56	3.04	.79	.36	-1.1	.39	-1.1	1.80	.66	121	121
17	5	3.40	3.56	3.04	.79	.54	-.6	.61	-.5	1.51	.53	122	122
17	5	3.40	3.56	3.04	.79	.54	-.6	.61	-.5	1.51	.53	127	127
17	5	3.40	3.56	3.04	.79	.54	-.6	.61	-.5	1.51	.53	128	128
18	5	3.60	3.54	2.98	.92	1.01	.2	1.13	.5	.81	.46	1053	1053
18	5	3.60	3.54	2.98	.92	1.01	.2	1.13	.5	.81	.46	1054	1054
18	5	3.60	3.54	2.98	.92	1.01	.2	1.13	.5	.81	.46	1055	1055
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1056	1056
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1058	1058
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1059	1059
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1061	1061
18	5	3.60	3.54	2.98	.92	1.01	.2	1.13	.5	.81	.46	1063	1063
18	5	3.60	3.54	2.98	.92	.68	-.2	.56	.0	1.32	.55	1064	1064
18	5	3.60	3.54	2.98	.92	.68	-.2	.56	.0	1.32	.55	1065	1065
18	5	3.60	3.54	2.98	.92	.68	-.2	.56	.0	1.32	.55	1068	1068
18	5	3.60	3.54	2.98	.92	.68	-.2	.56	.0	1.32	.55	1069	1069
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1076	1076
18	5	3.60	3.54	2.98	.92	.68	-.2	.56	.0	1.32	.55	1077	1077
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1080	1080
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1081	1081
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1082	1082
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1086	1086
18	5	3.60	3.54	2.98	.92	1.06	.3	.78	.2	.96	.05	1087	1087
18	5	3.60	3.53	2.96	.90	1.12	.3	1.25	.5	.73	.18	474	474
18	5	3.60	3.53	2.96	.90	1.12	.3	1.25	.5	.73	.18	476	476
18	5	3.60	3.53	2.96	.90	1.41	.7	1.47	.7	.39	-.22	481	481
18	5	3.60	3.53	2.96	.90	1.12	.3	1.25	.5	.73	.18	490	490
18	5	3.60	3.53	2.96	.90	1.12	.3	1.25	.5	.73	.18	494	494
18	5	3.60	3.53	2.96	.90	.42	-.9	.36	-.8	1.68	.85	504	504
18	5	3.60	3.53	2.96	.90	1.41	.7	1.47	.7	.39	-.22	510	510
17	5	3.40	3.50	2.85	.78	.90	.0	.93	.0	1.05	-.02	148	148

17	5	3.40	3.49	2.83	.80	1.40	.7	1.21	.5	.70	.66	593	593
17	5	3.40	3.49	2.83	.80	1.40	.7	1.21	.5	.70	.66	619	619
17	5	3.40	3.49	2.83	.80	1.40	.7	1.21	.5	.70	.66	622	622
17	5	3.40	3.49	2.83	.80	1.50	.8	1.29	.6	.59	.57	633	633
17	5	3.40	3.49	2.83	.80	1.50	.8	1.29	.6	.59	.57	635	635
17	5	3.40	3.49	2.83	.80	1.50	.8	1.29	.6	.59	.57	662	662
17	5	3.40	3.49	2.83	.80	1.40	.7	1.21	.5	.70	.66	667	667
17	5	3.40	3.49	2.83	.80	1.90	1.3	1.65	1.0	.09	.26	670	670
17	5	3.40	3.49	2.83	.80	1.40	.7	1.22	.5	.69	.66	692	692
17	5	3.40	3.49	2.83	.80	1.50	.8	1.29	.6	.59	.57	694	694
17	5	3.40	3.49	2.83	.80	1.40	.7	1.22	.5	.69	.66	696	696
17	5	3.40	3.49	2.83	.80	1.40	.7	1.22	.5	.69	.66	714	714
17	5	3.40	3.49	2.83	.80	1.60	1.0	1.40	.7	.47	.48	723	723
17	5	3.40	3.49	2.83	.80	1.50	.8	1.29	.6	.59	.57	735	735
17	5	3.40	3.49	2.83	.80	.22	-1.6	.23	-1.4	1.90	.94	743	743
17	5	3.40	3.49	2.83	.80	.72	-.2	.67	-.3	1.29	.30	747	747
17	5	3.40	3.49	2.83	.80	1.40	.7	1.22	.5	.69	.66	757	757
17	5	3.40	3.49	2.83	.80	1.60	1.0	1.40	.7	.47	.48	768	768
17	5	3.40	3.49	2.83	.80	1.50	.8	1.29	.6	.59	.57	769	769
17	5	3.40	3.49	2.83	.80	1.40	.7	1.22	.5	.69	.66	774	774
17	5	3.40	3.49	2.83	.80	1.50	.8	1.29	.6	.59	.57	778	778
17	5	3.40	3.49	2.83	.80	1.60	1.0	1.40	.7	.47	.48	814	814
17	5	3.40	3.49	2.83	.80	1.40	.7	1.22	.5	.69	.66	819	819
17	5	3.40	3.49	2.83	.80	1.40	.7	1.22	.5	.69	.66	823	823
17	5	3.40	3.49	2.83	.80	1.60	1.0	1.40	.7	.47	.48	858	858
18	5	3.60	3.49	2.81	.93	1.34	.6	1.85	.9	.31	-.01	522	522
18	5	3.60	3.49	2.81	.93	1.34	.6	1.85	.9	.31	-.01	538	538
18	5	3.60	3.49	2.81	.93	1.34	.6	1.85	.9	.31	-.01	549	549
18	5	3.60	3.49	2.81	.93	1.34	.6	1.85	.9	.31	-.01	550	550
17	5	3.40	3.48	2.79	.84	1.26	.5	.97	.3	.87	.75	1002	1002
17	5	3.40	3.48	2.79	.84	1.26	.5	.97	.3	.87	.75	1010	1010
17	5	3.40	3.48	2.79	.84	1.26	.5	.97	.3	.87	.75	1011	1011
17	5	3.40	3.48	2.79	.84	1.26	.5	.97	.3	.87	.75	1019	1019
17	5	3.40	3.48	2.79	.84	1.26	.5	.97	.3	.87	.75	1021	1021
17	5	3.40	3.48	2.79	.84	1.26	.5	.97	.3	.87	.75	1025	1025
17	5	3.40	3.48	2.79	.84	.77	.0	.92	.3	1.03	.58	1030	1030
17	5	3.40	3.48	2.79	.84	.77	.0	.92	.3	1.03	.58	1031	1031
17	5	3.40	3.48	2.79	.84	.77	.0	.92	.3	1.03	.58	1032	1032
17	5	3.40	3.48	2.79	.84	.77	.0	.92	.3	1.03	.58	1035	1035
17	5	3.40	3.48	2.79	.84	.79	.0	.59	.0	1.25	.89	1037	1037
17	5	3.40	3.45	2.71	.80	1.57	.9	1.39	.7	.50	.49	432	432
17	5	3.40	3.45	2.71	.80	1.57	.9	1.39	.7	.50	.49	444	444
17	5	3.40	3.45	2.71	.80	1.30	.6	1.95	1.2	.26	-.17	457	457
17	5	3.40	3.42	2.60	.84	1.02	.2	.79	.1	1.10	.81	963	963
17	5	3.40	3.42	2.60	.84	1.02	.2	.79	.1	1.10	.81	964	964
17	5	3.40	3.42	2.60	.84	1.02	.2	.79	.1	1.10	.81	965	965
17	5	3.40	3.42	2.60	.84	1.02	.2	.79	.1	1.10	.81	968	968
17	5	3.40	3.42	2.60	.84	1.02	.2	.79	.1	1.10	.81	973	973
17	5	3.40	3.42	2.60	.84	1.02	.2	.79	.1	1.10	.81	974	974
17	5	3.40	3.42	2.60	.84	1.02	.2	.79	.1	1.10	.81	977	977
17	5	3.40	3.42	2.60	.84	1.02	.2	.79	.1	1.10	.81	993	993
17	5	3.40	3.42	2.60	.84	1.02	.2	.79	.1	1.10	.81	995	995
16	5	3.20	3.42	2.59	.73	.74	-.2	.73	-.2	1.30	.80	232	232
16	5	3.20	3.42	2.59	.73	1.27	.6	1.33	.6	.69	.39	44	44
16	5	3.20	3.42	2.59	.73	1.11	.3	1.13	.4	.89	.51	75	75
67	20	3.35	3.41	2.57	.39	1.32	.9	1.23	.6	.71	.43	23	23
17	5	3.40	3.37	2.45	.84	1.39	.7	1.04	.3	.74	.69	871	871
17	5	3.40	3.37	2.45	.84	1.10	.3	.85	.1	1.01	.91	875	875
17	5	3.40	3.37	2.45	.84	1.10	.3	.85	.1	1.01	.91	887	887
66	20	3.30	3.36	2.41	.39	1.19	.6	1.08	.3	.83	.66	40	40
17	5	3.40	3.35	2.41	.80	1.34	.6	1.19	.4	.74	.75	588	588
17	5	3.40	3.35	2.41	.80	1.34	.6	1.19	.4	.74	.75	589	589
16	5	3.20	3.31	2.29	.73	1.05	.2	1.03	.2	.97	.90	142	142
17	5	3.40	3.31	2.28	.83	1.49	.8	1.13	.4	.62	.65	920	920
17	5	3.40	3.31	2.28	.83	.24	-1.3	.24	-.8	1.73	.77	935	935
17	5	3.40	3.30	2.24	.80	.89	.0	.86	.0	1.07	.22	479	479
17	5	3.40	3.30	2.24	.80	1.20	.5	1.05	.2	.92	.77	499	499
16	5	3.20	3.29	2.23	.75	5.53	3.5	8.20	4.7	-3.64	-.70	599	599
16	5	3.20	3.29	2.23	.75	1.24	.5	1.11	.3	.78	.43	645	645
16	5	3.20	3.29	2.23	.75	1.63	1.0	1.65	1.0	.35	-.10	655	655
16	5	3.20	3.29	2.23	.75	.91	.0	1.06	.3	.92	-.28	703	703
16	5	3.20	3.29	2.23	.75	1.24	.5	1.11	.3	.78	.43	715	715



16	5	3.20	3.29	2.23	.75	1.77	1.1	1.61	.9	.27	.72	717	717
16	5	3.20	3.29	2.23	.75	.75	-.1	.72	-.2	1.30	.85	728	728
16	5	3.20	3.29	2.23	.75	.75	-.1	.72	-.2	1.30	.85	729	729
16	5	3.20	3.29	2.23	.75	1.77	1.1	1.61	.9	.27	.72	764	764
16	5	3.20	3.29	2.23	.75	.75	-.1	.72	-.2	1.30	.85	806	806
16	5	3.20	3.29	2.23	.75	1.73	1.0	1.71	1.1	.26	.01	824	824
16	5	3.20	3.29	2.23	.75	1.73	1.0	1.71	1.1	.26	.01	825	825
17	5	3.40	3.29	2.22	.83	.90	.0	.71	.0	1.19	.94	1046	1046
17	5	3.40	3.29	2.22	.83	.90	.0	.71	.0	1.19	.94	1048	1048
17	5	3.40	3.29	2.22	.83	.90	.0	.71	.0	1.19	.94	1052	1052
17	5	3.40	3.29	2.22	.83	1.23	.5	.95	.2	.86	.75	1060	1060
17	5	3.40	3.29	2.22	.83	.90	.0	.71	.0	1.19	.94	1078	1078
17	5	3.40	3.29	2.22	.83	1.23	.5	.95	.2	.86	.75	1085	1085
16	5	3.20	3.26	2.15	.77	.42	-.6	.43	-.4	1.49	.67	1000	1000
16	5	3.20	3.26	2.15	.77	.69	-.1	.56	-.2	1.34	.78	1003	1003
16	5	3.20	3.26	2.15	.77	.69	-.1	.56	-.2	1.34	.78	1004	1004
16	5	3.20	3.26	2.15	.77	.69	-.1	.56	-.2	1.34	.78	1006	1006
16	5	3.20	3.26	2.15	.77	.99	.2	.87	.1	1.07	.54	1018	1018
16	5	3.20	3.26	2.15	.77	.99	.2	.87	.1	1.07	.54	1027	1027
16	5	3.20	3.24	2.10	.75	.65	-.3	.94	.1	1.16	.22	446	446
15	5	3.00	3.24	2.08	.69	.25	-1.3	.23	-1.5	1.84	.00	74	74
17	5	3.40	3.22	2.03	.84	1.17	.4	.90	.1	.96	.79	518	518
17	5	3.40	3.22	2.03	.84	1.17	.4	.90	.1	.96	.79	527	527
17	5	3.40	3.22	2.03	.84	1.17	.4	.90	.1	.96	.79	528	528
17	5	3.40	3.22	2.03	.84	1.17	.4	.90	.1	.96	.79	536	536
16	5	3.20	3.21	2.02	.75	.64	-.3	.89	.0	1.18	.13	364	364
15	5	3.00	3.19	1.97	.69	1.07	.3	1.13	.4	.90	.09	95	95
16	5	3.20	3.19	1.96	.75	.53	-.4	.49	-.4	1.46	.60	994	994
16	5	3.20	3.18	1.92	.75	1.20	.5	1.09	.3	.84	.46	402	402
16	5	3.20	3.18	1.92	.75	.94	.1	.88	.0	1.10	.75	412	412
15	5	3.00	3.12	1.77	.70	.56	-.5	.56	-.5	1.42	.90	144	144
15	5	3.00	3.12	1.77	.70	.79	.0	.84	.0	1.20	.20	186	186
15	5	3.00	3.12	1.77	.70	.62	-.4	.63	-.4	1.37	.53	188	188
15	5	3.00	3.10	1.72	.70	.95	.1	.95	.1	1.02	.49	272	272
15	5	3.00	3.10	1.72	.70	1.08	.3	1.08	.3	.89	-.51	273	273
15	5	3.00	3.10	1.71	.70	.40	-.8	.39	-.9	1.62	.80	596	596
15	5	3.00	3.10	1.71	.70	.90	.0	.90	.0	1.06	.91	719	719
15	5	3.00	3.10	1.71	.70	.90	.0	.90	.0	1.06	.91	734	734
15	5	3.00	3.10	1.71	.70	.90	.0	.90	.0	1.06	.91	771	771
15	5	3.00	3.10	1.71	.70	.90	.0	.90	.0	1.06	.91	781	781
15	5	3.00	3.10	1.71	.70	.90	.0	.90	.0	1.06	.91	809	809
15	5	3.00	3.10	1.71	.70	.90	.0	.90	.0	1.06	.91	816	816
15	5	3.00	3.10	1.71	.70	.90	.0	.90	.0	1.06	.91	848	848
16	5	3.20	3.07	1.65	.76	1.18	.4	1.13	.4	.86	.40	943	943
14	5	2.80	3.07	1.64	.65	1.22	.5	1.38	.7	.58	.30	220	220
14	5	2.80	3.07	1.64	.65	1.12	.3	1.32	.6	.68	.29	230	230
14	5	2.80	3.07	1.64	.65	.59	-.4	.51	-.6	1.47	-.10	251	251
14	5	2.80	3.07	1.64	.65	.59	-.4	.51	-.6	1.47	-.10	254	254
14	5	2.80	3.07	1.64	.65	1.12	.3	1.32	.6	.68	.29	316	316
14	5	2.80	3.07	1.64	.65	.75	-.1	.72	-.2	1.30	-.32	317	317
14	5	2.80	3.07	1.64	.65	.75	-.1	.72	-.2	1.30	-.32	339	339
14	5	2.80	3.07	1.64	.65	.69	-.2	.81	.0	1.20	.56	57	57
14	5	2.80	3.07	1.64	.65	.69	-.2	.81	.0	1.20	.56	78	78
14	5	2.80	3.07	1.64	.65	.56	-.5	.47	-.7	1.51	.09	79	79
14	5	2.80	3.07	1.64	.65	.56	-.5	.47	-.7	1.51	.09	80	80
15	5	3.00	3.06	1.61	.70	.95	.1	.82	.0	1.10	.40	1033	1033
16	5	3.20	3.05	1.59	.76	2.53	1.7	2.04	1.2	-.36	.04	1067	1067
15	5	3.00	2.99	1.45	.69	.31	-1.0	.35	-.9	1.63	.94	960	960
15	5	3.00	2.97	1.39	.70	.64	-.3	.68	-.3	1.36	.52	428	428
58	20	2.90	2.96	1.36	.34	1.20	.6	1.52	1.3	.58	.30	28	28
14	5	2.80	2.93	1.31	.66	.91	.1	.96	.1	1.01	.31	180	180
14	5	2.80	2.93	1.31	.66	.91	.1	.96	.1	1.01	.31	182	182
14	5	2.80	2.93	1.31	.66	.91	.1	.96	.1	1.01	.31	187	187
15	5	3.00	2.93	1.29	.70	1.03	.2	1.01	.2	.92	.89	561	561
15	5	3.00	2.92	1.27	.70	1.33	.6	1.06	.3	.90	.90	867	867
15	5	3.00	2.92	1.27	.70	1.33	.6	1.06	.3	.90	.90	868	868
14	5	2.80	2.91	1.26	.65	.37	-.9	.38	-1.0	1.62	.44	621	621
14	5	2.80	2.91	1.26	.65	.38	-.9	.40	-.9	1.58	.91	641	641
14	5	2.80	2.91	1.26	.65	2.90	2.0	3.14	2.2	-.93	.09	666	666
14	5	2.80	2.91	1.26	.65	4.07	2.7	3.71	2.6	-1.19	.65	673	673
14	5	2.80	2.91	1.26	.65	.48	-.6	.47	-.7	1.53	.26	690	690
14	5	2.80	2.91	1.26	.65	.38	-.9	.38	-.9	1.62	.43	711	711

13	5	2.60	2.90	1.24	.62	1.02	.2	.91	.1	1.05	-.33	328	328
14	5	2.80	2.87	1.16	.65	2.04	1.3	1.79	1.1	.21	.24	1007	1007
15	5	3.00	2.86	1.12	.70	.25	-1.2	.23	-1.3	1.73	.89	916	916
15	5	3.00	2.83	1.06	.70	1.15	.4	.91	.1	1.02	.94	1047	1047
14	5	2.80	2.82	1.03	.65	.75	-.1	.78	-.1	1.22	.61	374	374
14	5	2.80	2.80	1.00	.64	4.93	3.3	5.40	3.5	-2.20	.04	962	962
12	5	2.40	2.75	.88	.58	3.13	2.3	3.14	2.2	-.25	-.09	252	252
13	5	2.60	2.74	.86	.61	3.03	2.1	2.70	1.9	-.52	.60	767	767
13	5	2.60	2.74	.86	.61	1.89	1.2	2.11	1.4	.17	-.35	779	779
15	5	3.00	2.74	.86	.70	.74	-.1	1.16	.4	.99	.65	516	516
15	5	3.00	2.74	.86	.70	.70	-.2	.68	-.2	1.27	.81	540	540
15	5	3.00	2.74	.86	.70	1.07	.3	1.00	.2	.79	.69	541	541
13	5	2.60	2.73	.84	.62	.65	-.3	.68	-.2	1.30	-.14	282	282
13	5	2.60	2.73	.84	.62	1.07	.3	1.15	.4	.75	.43	293	293
14	5	2.80	2.73	.84	.65	1.75	1.1	1.69	1.0	.27	.85	577	577
14	5	2.80	2.73	.84	.65	1.75	1.1	1.69	1.0	.27	.85	585	585
14	5	2.80	2.73	.84	.65	2.02	1.3	1.95	1.3	.03	.65	586	586
14	5	2.80	2.72	.82	.65	4.57	2.9	3.81	2.6	-2.00	.56	872	872
13	5	2.60	2.70	.77	.61	1.23	.5	1.35	.6	.59	.30	1029	1029
13	5	2.60	2.67	.72	.61	.51	-.6	.56	-.5	1.45	.44	445	445
13	5	2.60	2.62	.61	.61	4.41	3.2	4.79	3.4	-2.05	.16	959	959
11	5	2.20	2.59	.55	.56	1.44	.8	1.72	1.1	.42	-.04	313	313
12	5	2.40	2.57	.51	.57	2.19	1.5	3.12	2.3	-.65	-.60	644	644
13	5	2.60	2.45	.29	.61	.77	-.1	.83	.0	1.20	.18	480	480
13	5	2.60	2.45	.29	.61	1.03	.2	1.04	.2	.95	-.22	486	486
13	5	2.60	2.45	.29	.60	.94	.1	.99	.1	.96	.15	912	912
46	20	2.30	2.41	.23	.28	.61	-1.3	.63	-1.2	1.30	.66	18	18
46	20	2.30	2.41	.23	.28	1.18	.6	1.46	1.3	.59	.28	30	30
11	5	2.20	2.41	.22	.55	1.34	.6	1.28	.6	.59	-.01	136	136
11	5	2.20	2.41	.22	.55	1.21	.5	1.35	.6	.31	.87	138	138
13	5	2.60	2.41	.22	.61	.88	.0	1.12	.3	.90	.46	1050	1050
11	5	2.20	2.40	.20	.54	2.13	1.6	2.09	1.5	.01	.15	656	656
11	5	2.20	2.40	.20	.54	2.65	2.0	2.71	2.1	-.74	.37	705	705
11	5	2.20	2.40	.20	.54	1.84	1.3	2.70	2.1	-.48	-.50	733	733
11	5	2.20	2.32	.08	.57	2.42	1.9	2.06	1.5	-.18	.40	1005	1005
44	20	2.20	2.32	.08	.28	.67	-1.1	.65	-1.1	1.20	.72	4	4
43	20	2.15	2.27	.00	.27	.79	-.6	.79	-.6	1.05	.57	32	32
10	5	2.00	2.22	-.08	.52	2.96	2.5	3.75	3.0	-2.36	.11	598	598
10	5	2.00	2.22	-.08	.52	3.55	2.9	3.82	3.0	-1.84	.24	840	840
12	5	2.40	2.19	-.13	.58	.61	-.5	.63	-.5	1.27	.75	1051	1051
10	5	2.00	2.06	-.30	.52	.47	-.9	.47	-.9	1.65	.49	370	370
12	5	2.40	2.06	-.31	.57	3.86	3.0	5.37	4.0	-2.53	-.16	521	521
9	5	1.80	2.03	-.35	.51	2.41	2.0	2.58	2.1	-.83	-.40	826	826
8	5	1.60	1.83	-.61	.52	3.25	2.8	3.49	2.9	-2.66	-.35	646	646
11	5	2.20	1.83	-.61	.55	3.19	2.7	4.61	3.7	-2.17	.00	520	520
11	5	2.20	1.83	-.61	.55	3.78	3.1	5.05	3.9	-3.10	.00	535	535
6	5	1.20	1.64	-.84	.52	.14	-2.4	.16	-2.0	1.89	.95	67	67
7	5	1.40	1.60	-.88	.53	2.59	2.2	2.74	2.2	-1.46	-.23	643	643
7	5	1.40	1.60	-.88	.53	.59	-.6	.60	-.6	1.44	.66	689	689
9	5	1.80	1.55	-.93	.52	1.57	1.0	1.86	1.3	-.12	.69	495	495
9	5	1.80	1.49	-1.00	.56	5.27	3.8	4.67	3.5	-3.71	.28	917	917
6	5	1.20	1.35	-1.17	.54	3.03	2.4	3.48	2.7	-1.69	-.91	669	669
6	5	1.20	1.35	-1.17	.54	3.52	2.8	3.47	2.7	-1.65	-.32	827	827
6	5	1.20	1.35	-1.17	.54	1.92	1.4	1.82	1.2	.29	.06	851	851
8	5	1.60	1.18	-1.36	.56	.64	-.4	.59	-.5	1.88	.89	1088	1088
6	5	1.20	.79	-1.83	.63	6.30	3.8	7.04	3.8	-2.90	-.09	878	878
3	5	.60	.53	-2.25	.68	1.95	1.2	3.42	1.8	.19	-.05	600	600
5	5	1.00	.47	-2.37	.66	1.09	.3	3.04	1.5	.19	.50	961	961
5	5	1.00	.44	-2.42	.65	.95	.1	.92	.2	.92	.74	911	911
4	5	.80	.27	-2.87	.71	2.37	1.5	6.19	2.5	-.22	.22	910	910
3	5	.60	.15	-3.43	.79	.21	-.9	.17	-.4	1.29	.80	914	914
19.9	5.6	3.57	3.62	3.62	.99	.99	.1	.89	.2		.37	Mean (Count: 1089)	
10.6	2.8	.44	.45	1.30	.27	.51	.6	.64	.6		.30	S.D. (Population)	
10.6	2.8	.44	.45	1.30	.27	.51	.6	.64	.6		.30	S.D. (Sample)	

Table 7.2.1 rater Measurement Report (arranged by FN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	- Measure	Model S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Corr. PtBis	Exact Obs %	Agree. Exp %	Nu rater
1207	307	3.93	3.98	-3.24	.27	.98	.0	.18	-2.3	1.14	.67	64.3	62.6	13 13
1118	287	3.90	3.94	-1.83	.22	1.79	3.4	.51	-1.9	1.06	.42	69.7	69.0	12 12
1200	312	3.85	3.91	-1.47	.18	1.69	3.5	.55	-2.0	1.04	.68	66.6	66.0	15 15
1084	298	3.64	3.63	.22	.12	.77	-2.5	.70	-3.1	1.22	.41	60.8	62.5	8 8
933	304	3.07	3.09	1.83	.09	.71	-3.7	.73	-3.6	1.35	.39	33.6	43.5	4 4
1152	307	3.75	3.88	-1.15	.15	1.01	.1	.76	-1.4	1.06	.64	63.7	64.0	14 14
1132	306	3.70	3.78	-.47	.14	1.20	1.7	.79	-1.5	1.08	.56	60.4	63.4	3 3
1104	304	3.63	3.62	.24	.12	.93	-.6	.86	-1.3	1.08	.41	58.0	62.8	1 1
1054	287	3.67	3.71	-.09	.13	.98	-.1	.90	-.8	1.02	.36	62.6	65.8	9 9
1059	312	3.39	3.41	.93	.10	.98	-.2	.92	-.9	1.09	.52	53.7	58.1	19 19
1015	287	3.54	3.55	.49	.11	.99	-.1	.96	-.3	1.02	.30	54.3	61.6	10 10
1013	287	3.53	3.55	.51	.11	.99	.0	.98	-.1	.97	.26	57.8	61.4	11 11
1057	299	3.54	3.50	.66	.11	.99	.0	1.02	.2	.94	.20	55.6	59.9	7 7
997	307	3.25	3.38	1.03	.10	1.09	1.0	1.08	.9	.90	.52	47.2	52.2	20 20
1069	303	3.53	3.58	.40	.11	1.10	1.0	1.09	.9	.91	.40	51.8	59.3	5 5
1108	299	3.71	3.74	-.26	.13	1.10	.9	1.12	.9	.94	.36	52.4	63.5	6 6
1049	306	3.43	3.46	.80	.11	1.17	1.8	1.25	2.6	.79	.41	47.9	57.8	2 2
1060	312	3.40	3.42	.92	.10	1.27	2.8	1.12	1.4	.84	.56	55.5	58.2	16 16
1147	312	3.68	3.74	-.23	.13	1.28	2.2	.83	-1.3	1.01	.63	64.1	65.3	17 17
1079	312	3.46	3.49	.71	.11	1.31	3.1	1.30	3.1	.61	.35	45.9	60.2	18 18
1081.9	302.4	3.58	3.62	.00	.13	1.12	.7	.88	-.5		.45			Mean (Count: 20)
65.6	8.8	.21	.22	1.14	.04	.26	1.9	.26	1.8		.13			S.D. (Population)
67.3	9.0	.21	.22	1.17	.04	.26	1.9	.27	1.8		.14			S.D. (Sample)

Model, Populn: RMSE .14 Adj (True) S.D. 1.13 Separation 8.14 Strata 11.19 Reliability (not inter-rater) .99  
 Model, Sample: RMSE .14 Adj (True) S.D. 1.16 Separation 8.35 Strata 11.47 Reliability (not inter-rater) .99  
 Model, Fixed (all same) chi-square: 975.7 d.f.: 19 significance (probability): .00  
 Model, Random (normal) chi-square: 18.4 d.f.: 18 significance (probability): .43  
 Inter-Rater agreement opportunities: 18105 Exact agreements: 10188 = 56.3% Expected: 11013.9 = 60.8%

Table 8.1 Category Statistics.

Model = ?,?,ALIGN  
 Rating (or partial credit) scale = ALIGN,R4,G,0

Score	DATA				QUALITY CONTROL				RASCH-ANDRICH		EXPECTATION		MOST	RASCH-	Cat	Obsd-Expd	Response
	Category	Counts	Cum. %	Cum. %	Avg Meas	Exp Meas	OUTFIT MnSq	Thresholds Measure	S.E.	Category	at -0.5	from	PROBABLE	THURSTONE	PEAK	Diagnostic	Category
0	46	46	1%	1%	-1.40	-1.86	2.2			(-2.93)		low	low	100%			not aligned
1	59	59	1%	2%	-.25	-.55	1.5	-1.44	.20	-1.56	-2.29	-1.44	-1.93	32%			somewhat nonaligned
2	280	280	5%	7%	.85	1.00	.7	-1.38	.14	-.39	-.99	-1.38	-1.10	47%			vaguely aligned
3	1634	1634	28%	35%	2.74	2.78	.8	.16	.07	1.46	.39	.16	.26	63%			somewhat aligned
4	4029	3784	65%	100%	4.25	4.23	1.0	2.66	.03	(3.79)	2.85	2.66	2.73	100%			directly aligned
										(Mean)		(Modal)		(Median)			

Table 4.1 Unexpected Responses (37 residuals sorted by u).

Cat	Score	Exp.	Resd	StRes	Num	item	Nu	ra
1	1	3.7	-2.7	-5.9	599	599	15	15
2	2	.1	1.9	5.4	910	910	2	2
0	0	3.0	-3.0	-4.8	521	521	12	12
0	0	3.0	-3.0	-4.7	962	962	3	3
0	0	2.9	-2.9	-4.4	520	520	12	12
0	0	2.9	-2.9	-4.4	535	535	12	12
3	3	.4	2.6	4.3	878	878	2	2
0	0	2.8	-2.8	-4.3	959	959	3	3
3	3	3.9	-.9	-4.2	15	15	6	6
3	3	3.9	-.9	-4.2	19	19	6	6
2	2	.2	1.8	3.9	600	600	19	19
3	3	3.9	-.9	-3.9	865	865	14	14
4	4	.8	3.2	3.7	917	917	2	2
3	3	3.9	-.9	-3.6	35	35	3	3
1	1	3.2	-2.2	-3.6	644	644	15	15
0	0	2.6	-2.6	-3.5	252	252	5	5
1	1	.1	.9	3.5	961	961	4	4
3	3	3.9	-.9	-3.4	20	20	15	15
1	1	3.1	-2.1	-3.4	30	30	15	15
3	3	3.9	-.9	-3.3	6	6	6	6
3	3	3.9	-.9	-3.3	9	9	6	6
3	3	3.9	-.9	-3.3	12	12	6	6
3	3	3.9	-.9	-3.3	22	22	6	6
3	3	3.9	-.9	-3.3	31	31	6	6
3	3	3.9	-.9	-3.3	34	34	19	19
3	3	3.9	-.9	-3.3	37	37	6	6
3	3	3.9	-.9	-3.3	39	39	6	6
0	0	2.5	-2.5	-3.3	669	669	15	15
0	0	2.5	-2.5	-3.3	673	673	16	16
1	1	3.1	-2.1	-3.3	733	733	15	15
3	3	3.9	-.9	-3.2	13	13	1	1
3	3	3.9	-.9	-3.2	21	21	20	20
3	3	3.9	-.9	-3.2	407	407	12	12
1	1	3.0	-2.0	-3.2	666	666	17	17
1	1	3.0	-2.0	-3.1	598	598	15	15
0	0	2.4	-2.4	-3.1	840	840	17	17
1	1	3.0	-2.0	-3.1	878	878	13	13
Cat	Score	Exp.	Resd	StRes	Num	item	Nu	ra



ภาคผนวก ง

คู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่าง  
ข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

---

โดย  
นางสาวบุษยรัตน์ จันทน์ประเสริฐ  
ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง (อาจารย์ที่ปรึกษา)  
ผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ รัตกระโทก (อาจารย์ที่ปรึกษาร่วม)



เอกสารนี้เป็นส่วนหนึ่งของวิทยานิพนธ์ปริญญาโท  
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา  
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2560

## คำนำ

การจัดทำคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียนในครั้งนี้ เป็นส่วนหนึ่งของงานวิทยานิพนธ์เรื่อง “ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราส์ช และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด” ของนางสาวบุษยรัตน์ จันทร์ประเสริฐ นิสิตหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา จุฬาลงกรณ์มหาวิทยาลัย โดยมี ผู้ช่วยศาสตราจารย์ ดร. ณีฎฐภรณ์ หลาวทอง และผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ ังคระโทก เป็นอาจารย์ที่ปรึกษา ซึ่งมีเป้าหมายสำคัญในการจัดทำคู่มือ คือ เพื่อเสนอแนะแนวทางการปฏิบัติพร้อมทั้งเครื่องมือที่ใช้ในการตรวจสอบความสอดคล้องระหว่างข้อสอบในการประเมินระดับชั้นเรียนกับมาตรฐานและตัวชี้วัดตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน โดยมีการควบคุมความคลาดเคลื่อนอันเกิดจากอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน ตลอดจนการตรวจสอบความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

ในโอกาสนี้ผู้วิจัยขอขอบพระคุณท่านอาจารย์ที่ปรึกษา ท่านอาจารย์ผู้ทรงคุณวุฒิ สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ตลอดจนโรงเรียนกลุ่มตัวอย่าง ซึ่งได้ให้ความอนุเคราะห์ช่วยเหลือเป็นอย่างดีทั้งในขั้นตอนของการพัฒนาเครื่องมือ การเก็บรวบรวมข้อมูล และการพัฒนาเป็นคู่มือนี้

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

บุษยรัตน์ จันทร์ประเสริฐ  
ผู้จัดทำคู่มือ

## สารบัญ

	หน้า
บทที่ 1 หลักการและวัตถุประสงค์ของคู่มือ	...
1.1 หลักการและแนวคิด	...
1.2 วัตถุประสงค์ของคู่มือ	...
1.3 ประโยชน์ที่คาดว่าจะได้รับจากการใช้คู่มือ	...
1.4 คำจำกัดความที่ใช้ในคู่มือ	...
บทที่ 2 แนวคิดและทฤษฎีที่เกี่ยวข้อง	...
2.1 แนวคิดการจำแนกพฤติกรรมการเรียนรู้ด้านพุทธิพิสัยของ Bloom ที่ได้รับการปรับปรุงใหม่ (revised Bloom's taxonomy)	...
2.2 แนวคิดพื้นฐานของโมเดลหลายองค์ประกอบของราล์ฟ	...
2.3 แนวคิดของวิธีการศึกษาความสอดคล้องในแนวเดียวกันของ Porter	...
บทที่ 3 ขั้นตอนในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter	...
ขั้นตอนที่ 1 ขั้นเตรียมการ	...
ขั้นตอนที่ 2 ขั้นดำเนินการ	...
ขั้นตอนที่ 3 ขั้นสรุปผล	...
รายการอ้างอิง	...



## บทที่ 1

### หลักการและวัตถุประสงค์ของคู่มือ

#### 1.1 หลักการและแนวคิด

ระบบการศึกษาแบบอิงมาตรฐาน (standard-based educational system) ประกอบด้วย 3 องค์ประกอบหลัก ได้แก่ มาตรฐานการเรียนรู้ การเรียนการสอน และการประเมินผล ความสอดคล้องไปในทิศทางเดียวกันของทั้ง 3 องค์ประกอบหลักดังกล่าวจึงเป็นเงื่อนไขพื้นฐานที่สำคัญของระบบการจัดการศึกษาแบบอิงมาตรฐาน (Fuhrman, 2001; Smith & O’Day, 1990) การที่องค์ประกอบทางการศึกษามีความสอดคล้องสัมพันธ์กันจะช่วยในการส่งเสริมการเรียนรู้ของผู้เรียน (Anderson, 2002; La Marca, Redfield, Winter, Bailey & Hansche, 2000; Biggs, 2003) ความสอดคล้องกันดังกล่าว ใช้คำว่า “alignment” แปลว่า ความสอดคล้องในแนวเดียวกัน ซึ่งในทางการศึกษาหมายถึง ระดับความเชื่อมโยงสอดคล้องสัมพันธ์กันขององค์ประกอบต่างๆ ในระบบการศึกษา เช่น ผลการเรียนรู้ที่คาดหวัง (เช่น มาตรฐานการเรียนรู้ ตัวชี้วัด วัตถุประสงค์) การจัดการเรียนการสอน และการประเมินผล (เช่น ข้อสอบ) (Ananda, 2003; Biggs, 2003; Bholá, Impara, & Buckendahl, 2003; Case, Jorgensen, & Zucker, 2004; La Marca, 2001; Webb, 1997)

การทดสอบความรู้ทางการศึกษาในระดับชาติขั้นพื้นฐานในแต่ละปีการศึกษาที่ผ่านมาสะท้อนให้เห็นถึงความรู้ความสามารถของผู้เรียนที่อยู่ในระดับต่ำ จากงานวิจัยของ เอ็มพร หลินเจริญ, สิริศักดิ์ อางวิชัย, และภริภา จันทรอินทร์ (2552) สรุปสาเหตุที่ทำให้คะแนนการทดสอบ O-NET ต่ำมาจากหลายปัจจัย เช่น ปัจจัยด้านผู้เรียน ปัจจัยด้านนโยบายในการพัฒนาคุณภาพวิชาการของโรงเรียนหรือสภาพแวดล้อมของโรงเรียน ปัจจัยด้านครูผู้สอน ซึ่งในการเพิ่มผลสัมฤทธิ์ของการสอบ O-NET นั้น ปัจจัยด้านครูผู้สอนนั้น ถือว่ามีความสำคัญและต้องได้รับพัฒนาเป็นลำดับต้นๆ เนื่องจากครูผู้สอนเป็นผู้ที่มีบทบาทสำคัญที่สุดในการจัดการเรียนการสอนที่เน้นผู้เรียนให้มีคุณภาพตามมาตรฐานและตัวชี้วัด ตลอดจนทำการวัดและประเมินผลระดับชั้นเรียนเพื่อหาคำตอบว่าการจัดกิจกรรมการเรียนการสอนส่งผลให้ผู้เรียนมีความก้าวหน้าด้านความรู้ ทักษะกระบวนการ ด้านคุณธรรม จริยธรรม และค่านิยมอันพึงประสงค์ตามมาตรฐานและตัวชี้วัดมากน้อยเพียงใด ซึ่งบริษัทอักษรเจริญทัศน์ จำกัด (2559: เว็บไซต์) ได้จัดทำสารสนเทศเกี่ยวกับการเพิ่มผลสัมฤทธิ์ของการสอบ O-NET โดยได้ให้ข้อเสนอแนะว่าครูยังต้องมีบทบาทในเรื่องต่างๆ โดยสรุปคือ การสอนโดยยึดตามหลักสูตรและสอนให้ครอบคลุมหลักสูตร จัดกิจกรรมการเรียนการสอนที่ส่งเสริมการคิดวิเคราะห์ (critical thinking) เพื่อให้สอดคล้องกับข้อสอบ O-NET ของ สทศ. ที่เน้นการคิดวิเคราะห์ ศึกษา แนวข้อสอบ O-NET ย้อนหลัง และควรทำ Test Blueprint ของข้อสอบ O-NET จัดให้ผู้เรียนได้ฝึกทำแบบทดสอบ

เสมือนจริง โดยจำลองบรรยากาศการสอบ O-NET ทำการวัดและประเมินผลโดยใช้ข้อสอบให้สอดคล้องหรือคล้ายคลึงกับข้อสอบ O-NET โดยเฉพาะข้อสอบแบบปรนัยที่มีรูปแบบข้อสอบที่แตกต่างจากที่โรงเรียนใช้ ทำโครงการต่างๆ เพื่อเพิ่มผลสัมฤทธิ์ของการสอบ O-NET ตลอดจนสร้างความตระหนักให้ผู้เรียนเห็นความสำคัญของการสอบ O-NET

ดังนั้น การตรวจสอบว่าครูทำการวัดและประเมินผลโดยใช้ข้อสอบให้สอดคล้องหรือคล้ายคลึงกับข้อสอบ O-NET รวมถึงออกข้อสอบให้เน้นระดับความซับซ้อนทางปัญญา เช่น การคิดวิเคราะห์ ให้สอดคล้องกับข้อสอบ O-NET ที่กล่าวข้างต้นนั้นจึงเป็นเรื่องที่มีความสำคัญ ซึ่งสามารถตรวจสอบได้จากการประเมินความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ (ข้อสอบ O-NET) กับข้อสอบประเมินระดับชั้นเรียน เพื่อให้ได้สารสนเทศว่าโรงเรียนสามารถวัดและประเมินผลในระดับชั้นเรียนได้สอดคล้องกับการประเมินระดับชาติมากน้อยเพียงใด อันจะนำไปสู่การปรับปรุงการเรียนการสอน และพัฒนาการวัดและประเมินในชั้นเรียน เพื่อสร้างความมั่นใจด้านคุณภาพของผู้เรียน และสนับสนุนเพิ่มผลสัมฤทธิ์ของการทดสอบระดับชาติ ซึ่งอาจนำไปสู่การเพิ่มความรู้อาสาสมัครของผู้เรียน นอกจากนี้ การประเมินดังกล่าวสามารถตอบคำถามในเชิงประจักษ์เพื่อยืนยันได้อีกว่าการวัดและประเมินผลของครูผู้สอนในระดับชั้นเรียน เป็นปัจจัยสำคัญที่มีความสัมพันธ์กับคะแนนการสอบ O-NET หรือไม่ หรือเป็นผลจากปัจจัยหลักด้านอื่นๆ

ปัจจุบันวิธีการศึกษาความสอดคล้องในแนวเดียวกันที่พบในงานวิจัยส่วนใหญ่มี 3 วิธี ได้แก่ (1) วิธีการของ Webb (1997) (2) วิธีการของ Porter หรือเรียกว่าวิธีการสำรวจหลักสูตรที่นำไปใช้จริง (survey of enacted curriculum: SEC) (Porter & Smithson, 2001) และ (3) วิธีการของบริษัท Achieve (Rothman, Slattery, & Vranek, 2002) โดยวิธีการที่ (1) และ (3) เป็นวิธีการที่ใช้ในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างเนื้อหาของมาตรฐานกับเนื้อหาของการประเมินที่มีความครอบคลุม ในขณะที่วิธีการที่ (2) เป็นวิธีการที่สามารถใช้ในการศึกษาระดับความสอดคล้องเชิงปริมาณได้ทั้ง 3 องค์ประกอบของระบบการศึกษา ได้แก่ มาตรฐาน การประเมิน ตำราเรียน และแม้แต่การสอนของครูในห้องเรียน (Porter, Smithson, Blank, & Zeidner, 2007: 27) ดังนั้นวิธีการของ Porter จึงเป็นวิธีการเดียวที่มีความยืดหยุ่นและเหมาะสมที่สุดในการนำมาประยุกต์ใช้ในงานวิจัยนี้ ซึ่งเป็นการศึกษาความสอดคล้องในแนวเดียวกันระหว่าง 2 การประเมิน ได้แก่ การประเมินระดับชาติ (ข้อสอบ O-NET) และการประเมินระดับชั้นเรียน (ข้อสอบในการประเมินผลการเรียนรู้ที่ครูผู้สอนสร้างขึ้น) จากแนวคิดที่กล่าวมาในข้างต้น ผู้จัดทำจึงพัฒนาคู่มือแนวทางการดำเนินการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter และมีการควบคุม ความคลาดเคลื่อน อันเกิดจากอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน โดยนำแนวคิดมาจากงานวิทยานิพนธ์ของผู้จัดทำ เรื่อง “ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบใน

การประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราส์ช และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด” มาประยุกต์ใช้ในการจัดทำคู่มือนี้

## 1.2 วัตถุประสงค์ของคู่มือ

- 1) ให้แนวทางในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน โดยมีการควบคุมความคลาดเคลื่อนอันเกิดจากอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน
- 2) ให้แนวทางในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter

## 1.3 ประโยชน์ที่คาดว่าจะได้รับจากการใช้คู่มือ

- 1) ได้แนวทางและเครื่องมือที่ใช้สำหรับศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน โดยมีการควบคุมความคลาดเคลื่อนอันเกิดจากอิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมิน
- 2) ได้แนวทางในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter
- 3) สามารถใช้เป็นแนวทางการทำวิจัยด้านการศึกษาความสอดคล้องในแนวเดียวกันระหว่างองค์ประกอบระบบการศึกษาแบบอิงมาตรฐาน

## 1.4 คำจำกัดความที่ใช้ในคู่มือ

**สาระ** หมายถึง กลุ่มย่อยภายในกลุ่มสาระการเรียนรู้ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

**มาตรฐาน** หมายถึง คุณภาพที่ต้องการให้เกิดขึ้นในตัวผู้เรียน เป็นสิ่งที่ผู้เรียนพึงรู้และปฏิบัติได้ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

**ตัวชี้วัด** หมายถึง สิ่งที่ผู้เรียนพึงรู้และปฏิบัติได้ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ซึ่งสะท้อนถึงมาตรฐานการเรียนรู้ มีลักษณะเฉพาะเจาะจงและมีความเป็นรูปธรรมในการนำไปใช้ในการกำหนดเนื้อหา การเรียนการสอน และเป็นเกณฑ์สำคัญสำหรับการวัดและประเมินผลเพื่อตรวจสอบคุณภาพผู้เรียน

**ข้อสอบในการประเมินระดับชั้นเรียน** หมายถึง ข้อคำถามที่ใช้ในการประเมิน ผลการเรียนรู้ของผู้เรียน ของโรงเรียน ซึ่งเป็นข้อสอบที่ครูสร้างขึ้น (teacher-made test) เพื่อประเมินผลการเรียนรู้ของผู้เรียน

**ข้อสอบในการประเมินระดับชาติ** หมายถึง ข้อสอบในการทดสอบทางการศึกษาระดับชาตินำพื้นฐาน (ordinary national educational test: O-NET)

**ความสอดคล้องในแนวเดียวกัน** หมายถึง ระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างองค์ประกอบต่างๆ ในระบบการศึกษา ได้แก่ ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน และความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน

**ความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน** หมายถึง ระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างข้อสอบในการประเมินผลระดับชาติ ได้แก่ ข้อสอบในการทดสอบทางการศึกษาระดับชาตินำพื้นฐาน (ordinary national educational test: O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ประจำปีการศึกษา 2559 กับ การประเมินผลระดับชั้นเรียน ได้แก่ ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ในปีการศึกษา 2559 ซึ่งวัดความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน โดยการคำนวณค่าดัชนีความสอดคล้อง (alignment index) ตามแนวคิดของ Porter

**ความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน** หมายถึง ระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างผลการเรียนรู้ที่คาดหวัง ได้แก่ มาตรฐานและตัวชี้วัด กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กับ การประเมิน ได้แก่ ข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น ของโรงเรียนมัธยมศึกษา การวัดความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ทำโดยให้ผู้เชี่ยวชาญเป็นผู้ประเมิน ด้วยแบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาตอนต้น ที่ผู้วิจัยสร้างขึ้น

**วิธีการศึกษาความสอดคล้องในแนวเดียวกันของ Porter** หรือวิธีการสำรวจหลักสูตรที่นำไปใช้จริง (Survey of enacted curriculum (SEC) methodology) หมายถึง การวัดระดับความเชื่อมโยงและประสานความสัมพันธ์ระหว่างเนื้อหาของ 2 องค์ประกอบในการศึกษาที่นำมาประเมิน แล้วนำผลการประเมินของผู้เชี่ยวชาญมาสร้างเมทริกซ์ 2 เมทริกซ์ ได้แก่ เมทริกซ์ข้อสอบใน

การประเมินระดับชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) มาคำนวณค่าดัชนีความสอดคล้อง (alignment index)

**ความซับซ้อนทางปัญญา** หมายถึง ลำดับขั้นของกระบวนการทางสมองหรือความลึกของความรู้ในตัวชี้วัดภายใต้มาตรฐานต่างๆ และข้อสอบในการประเมินผลการเรียนรู้ ซึ่งวัดได้จากการพิจารณาของผู้เชี่ยวชาญ ความซับซ้อนทางปัญญานำมาตามแนวทางการจำแนกพฤติกรรม การเรียนรู้ด้านพุทธิพิสัยของ Bloom ที่ได้รับการปรับปรุงใหม่ (revised Bloom's taxonomy) (Anderson et al., 2001) โดยแบ่งลำดับขั้นของกระบวนการทางปัญญา (cognitive process dimension) ได้เป็น 6 ชั้น เรียงลำดับจากความรู้ระดับต่ำไปยังความรู้ระดับสูง ดังนี้

1. จำ (remember) เป็นความสามารถในการระลึกหรือจำความรู้หรือสารสนเทศที่เก็บไว้ในสมอง ซึ่งเป็นความจำระยะยาว
2. เข้าใจ (understand) เป็นความสามารถในการสร้างความหมายหรือความรู้จากสื่อหรือเครื่องมือทางการศึกษาด้วยตนเอง
3. ประยุกต์ใช้ (apply) เป็นความสามารถในการใช้กระบวนการที่ได้เรียนรู้มาในสถานการณ์ใหม่หรือสถานการณ์ที่คล้ายคลึงกัน
4. วิเคราะห์ (analyze) เป็นความสามารถในการแยกความรู้ออกเป็นส่วนๆ โดยสามารถให้เหตุผลว่า ความรู้ส่วนย่อยที่แยกแต่ละส่วนมีความเกี่ยวข้องกับโครงสร้างของความรู้ทั้งหมดอย่างไร สามารถจำแนกความแตกต่าง จัดระบบความรู้ และบอกที่มาของความรู้หรือองค์ประกอบแต่ละส่วนได้
5. ประเมินค่า (evaluate) เป็นความสามารถในการตรวจสอบ วิพากษ์ และตัดสิน
6. สร้างสรรค์ (create) เป็นความสามารถการสร้างสิ่งใหม่จากสิ่งที่เคยเรียนรู้ หรือสิ่งที่พบเห็นในบริบทต่างๆ สามารถสร้างสรรค์งาน แผนงาน หรือผลิตภัณฑ์ หรือชิ้นงานที่แปลกใหม่

**อิทธิพลการกดหรือปล่อยคะแนน** (severity/leniency effect) หมายถึง ความคลาดเคลื่อนที่เกิดจากการที่ผู้ประเมินหรือผู้เชี่ยวชาญในการประเมินมีแนวโน้มจะให้คะแนนผู้ถูกตรวจประเมิน หรือสิ่งที่ถูกประเมินสูงหรือต่ำกว่าความเป็นจริง โดยผู้ประเมินมีลักษณะกดคะแนน (severity) จะมีแนวโน้มให้คะแนนโดยเฉลี่ยต่ำกว่าผู้ประเมินคนอื่น ซึ่งเป็นการให้คะแนนต่ำกว่าความเป็นจริงตลอดช่วงความสามารถ ส่วนผู้ประเมินมีลักษณะปล่อยคะแนน (leniency) จะมีแนวโน้มให้คะแนนโดยเฉลี่ยสูงกว่าผู้ประเมินคนอื่นตลอดช่วงความสามารถ ซึ่งตรวจสอบอิทธิพลการกดหรือปล่อยคะแนน โดยใช้โมเดลหลายองค์ประกอบของราส์ช (many-facets Rasch measurement: MFRM) โดยคะแนนในหน่วยโลจิท (rater logit) ของฟาเซทผู้ประเมิน หากมีคะแนนต่ำ หมายถึง ผู้ประเมินมีลักษณะปล่อยคะแนน แต่หากมีคะแนนสูง หมายถึง ผู้ประเมินมีลักษณะกดคะแนน

## บทที่ 2

### แนวคิดและทฤษฎีที่เกี่ยวข้อง

#### 2.1 แนวคิดการจำแนกพฤติกรรมการเรียนรู้ด้านพุทธิพิสัยของ Bloom ที่ได้รับการปรับปรุงใหม่ (revised Bloom's taxonomy)

แนวคิดการจำแนกพฤติกรรมการเรียนรู้ด้านพุทธิพิสัยของ Bloom ที่ได้รับการปรับปรุงใหม่ (revised Bloom's taxonomy) โดยแบ่งลำดับขั้นของกระบวนการทางปัญญา (cognitive process dimension) ได้เป็น 6 ชั้น เรียงลำดับจากความรู้ระดับต่ำไปยังความรู้ระดับสูง ได้แก่ จำ เข้าใจ ประยุกต์ใช้ วิเคราะห์ ประเมินค่า และสร้างสรรค์

คำสำคัญและตัวอย่างพฤติกรรมในกระบวนการทางปัญญา ตามแนวคิดของ Bloom ที่ได้รับการปรับปรุงใหม่ (revised Bloom's taxonomy) สรุปได้ดังตาราง 1

**ตาราง 1** คำสำคัญและตัวอย่างพฤติกรรมในกระบวนการทางปัญญา ตามแนวคิด revised Bloom's taxonomy

กระบวนการทางปัญญา	คำสำคัญ	ความหมายและตัวอย่าง
<b>1. จำ (remember): ระลึกได้จากความจำระยะยาว</b>		
1.1 การระลึกได้ (recognizing)	- การระบุ (identifying)	การหาตำแหน่งองค์ความรู้ในความจำระยะยาวที่สอดคล้องกับเนื้อหาที่นำเสนอ เช่น - บอกวันที่สำคัญของเหตุการณ์ทางประวัติศาสตร์ได้ - บอกได้ว่ารูปหกเหลี่ยมประกอบด้วยด้านทั้งหมดกี่ด้าน - ระบุชนิดต่างๆ ในแผนผังของสัตว์ครึ่งบกครึ่งน้ำ
1.2 การจำได้ (recalling)	- การระลึก (retrieving)	การระลึกถึงความรู้ที่เกี่ยวข้องจากความจำระยะยาว เช่น - จดจำวันที่ของเหตุการณ์ที่สำคัญทางประวัติศาสตร์ได้ - ตอบได้ว่า 7 คูณ 8 เท่ากับเท่าไร - บอกได้ว่าใครเป็นผู้เขียนแต่งเรื่อง Charge of the Light Brigade
<b>2. เข้าใจ (understand): สร้างความหมายจากข้อความการเรียนการสอน ประกอบด้วย พุด เขียน และการสื่อสารกราฟิก</b>		
2.1 การแปลความหมาย (interpreting)	- การจัดประเภท (clarifying) - การถอดความ (paraphrasing) - การนำเสนอ (representing) - การแปลความ (translating)	การแปลงข้อเท็จจริงจากรูปแบบหนึ่ง (เช่น ตัวเลข) ไปยังอีกรูปแบบหนึ่ง (เช่น ภาษา) ได้ เช่น - ถอดความที่สำคัญของการกล่าวสุนทรพจน์และเอกสาร - แปลความหมายโจทย์ปัญหาเป็นสมการพีชคณิตได้ - วาดรูปบรรยายอาหารได้ - เรียบเรียงคำปราศรัยของประธานาธิบดี Lincoln ได้



## ตาราง 1 (ต่อ)

กระบวนการทางปัญญา	คำสำคัญ	ความหมายและตัวอย่าง
2.2 การให้ตัวอย่าง (exemplifying)	<ul style="list-style-type: none"> <li>- แสดงภาพ (illustrating)</li> <li>- ยกตัวอย่างประกอบ (instantiating)</li> </ul>	<p>การยกตัวอย่างหรือแสดงแนวคิดหรือหลักเกณฑ์ เช่น</p> <ul style="list-style-type: none"> <li>- ยกตัวอย่างของรูปแบบการวาดภาพศิลปะ</li> <li>- วาดรูปที่มีสมบัติคู่ขนานกันได้</li> <li>- บอกชื่อสัตว์เลี้ยงลูกด้วยนมที่อาศัยอยู่ในแถบบ้านเรา</li> </ul>
2.3 การจัดจำแนกหมวดหมู่ (classifying)	<ul style="list-style-type: none"> <li>- จัดประเภท (categorizing)</li> <li>- จัดหมวดหมู่ (instantiating)</li> </ul>	<p>การระบุหมวดหมู่ให้กับสิ่งต่างๆ เช่น</p> <ul style="list-style-type: none"> <li>- จัดจำแนกหมวดหมู่ให้กับผู้มีความผิดปกติทางจิต</li> <li>- ทำรายการประเภทของรัฐบาลที่พบในสหรัฐอเมริกา</li> <li>- จัดกลุ่มสัตว์พื้นเมืองตามสปีชีส์</li> </ul>
2.4 การสรุป (summarizing)	<ul style="list-style-type: none"> <li>- การย่อความ (abstracting)</li> <li>- การสรุปอ้างอิง (generalizing)</li> </ul>	<p>การสรุปประเด็นหลักหรือสาระที่สำคัญ เช่น</p> <ul style="list-style-type: none"> <li>- เขียนสรุปสั้นๆ จากเหตุการณ์ที่เกิดขึ้นในวิดีโอเทป</li> <li>- ตั้งชื่อเรื่องบทความสั้นๆ ได้</li> </ul>
2.5 การสรุปอ้างอิง (inferring)	<ul style="list-style-type: none"> <li>- การสรุป (concluding)</li> <li>- การคาดการณ์ (extrapolating)</li> <li>- การแก้ไขเพิ่มเติม (interpolating)</li> <li>- การทำนาย (predicting)</li> </ul>	<p>การเขียนข้อสรุปอย่างมีเหตุมีผลจากข้อมูลที่นำเสนอ เช่น</p> <ul style="list-style-type: none"> <li>- สรุปหลักไวยากรณ์จากตัวอย่างต่างๆ ในวิชาภาษาต่างประเทศ</li> <li>- อ่านเรื่องราว บทสนทนาระหว่างบุคคลสองบุคคลพร้อมทั้งสรุปความสัมพันธ์ในอดีตของคนที่ทั้งสอง</li> <li>- บอกความหมายของคำใหม่จากบริบทของเนื้อเรื่อง</li> <li>- ทำนายจำนวนในอนาคตได้</li> </ul>
2.6 การเปรียบเทียบ (comparing)	<ul style="list-style-type: none"> <li>- การเปรียบเทียบความแตกต่าง (contrasting)</li> <li>- การแสดงแผนผัง (mapping)</li> <li>- การจับคู่ (matching)</li> </ul>	<p>การตรวจสอบความสอดคล้องระหว่าง 2 แนวคิด หรือ 2 สิ่ง เช่น</p> <ul style="list-style-type: none"> <li>- เปรียบเทียบเหตุการณ์ในประวัติศาสตร์กับสถานการณ์ในปัจจุบัน</li> <li>- อธิบายว่าหัวใจมีความเหมือนกับปืมน้ำอย่างไร</li> <li>- ใช้แผนภาพเวนในการอธิบายความเหมือนและความแตกต่างของหนังสือสองเล่มที่แต่งโดย Charles Dickens</li> </ul>
2.7 การอธิบาย (explaining)	<ul style="list-style-type: none"> <li>- การสร้าง (constructing)</li> <li>- การสร้างแบบจำลอง (modeling)</li> </ul>	<p>การสร้างแบบจำลองเชิงสาเหตุและผลของระบบ เช่น</p> <ul style="list-style-type: none"> <li>- อธิบายสาเหตุของเหตุการณ์ที่สำคัญในศตวรรษที่ 18 ของประเทศฝรั่งเศส</li> <li>- วาดแผนผังอธิบายว่าความดันอากาศส่งผลอย่างไรต่ออากาศ</li> </ul>
<b>3. ประยุกต์ใช้ (apply): ดำเนินการตามขั้นตอนหรือใช้ขั้นตอนในสถานการณ์ที่กำหนด</b>		
3.1 การปฏิบัติ (executing)	<ul style="list-style-type: none"> <li>- การดำเนินการให้ สำเร็จ (carrying out)</li> </ul>	<p>การประยุกต์ใช้ขั้นตอนกับงานที่คุ้นเคย เช่น</p> <ul style="list-style-type: none"> <li>- ทหารจำนวนเต็มจำนวนหนึ่งด้วยจำนวนเต็มอีกจำนวนหนึ่ง โดย - ที่ทั้ง 2 จำนวนเป็นเลขหลายหลัก</li> <li>- บวกเลขสองหลักในแนวตั้งได้</li> <li>- อ่านออกเสียงเรื่องราวที่เป็นภาษาต่างประเทศได้</li> </ul>

ตาราง 1 (ต่อ)

กระบวนการทางปัญญา	คำสำคัญ	ความหมายและตัวอย่าง
3.2 การนำไปใช้/ การปฏิบัติ (implementing)	- การใช้ (using)	การประยุกต์ใช้ขั้นตอนกับงานที่ไม่คุ้นเคย เช่น - ใช้กฎของนิวตันข้อที่ 2 ในสถานการณ์ที่มีความเหมาะสม - ออกแบบการทดลองเกี่ยวกับการเจริญเติบโตของพืชในดินต่างชนิดกัน
<b>4. วิเคราะห์ (analyze): จำแนกเนื้อหาออกเป็นส่วนย่อยๆ และสามารถบอกได้ว่าแต่ละส่วนมีความสัมพันธ์กับส่วนอื่นๆ และกับโครงสร้างหรือวัตถุประสงค์โดยรวมอย่างไร</b>		
4.1 การหาความเหมือน- ความแตกต่าง (differentiating)	- การจำแนก (discriminating) - การจำแนกความแตกต่าง (distinguishing) - การเน้น (focusing) - การเลือก (selecting)	การจำแนกประเด็นในส่วนที่เกี่ยวข้องกับส่วนที่ไม่เกี่ยวข้อง หรือ จำแนกส่วนที่สำคัญจากส่วนที่ไม่สำคัญของเนื้อหาที่นำเสนอ เช่น - จำแนกระหว่างตัวเลขที่เกี่ยวข้องกับตัวเลขที่ไม่เกี่ยวข้อง ใน โจทย์ปัญหาคณิตศาสตร์ - เลือกข้อความหรือคำสำคัญจากโจทย์ปัญหาทางคณิตศาสตร์ และตัดข้อความที่ไม่เกี่ยวข้องออก - วาดแผนภาพแสดงประเด็นหลักและประเด็นรองของ บทประพันธ์
4.2 การจัดระบบ (organizing)	- การหาความเชื่อมโยง (finding coherence) - การบูรณาการ (integrating) - การร่างภาพ (outlining) - การแจกแจง (parsing) - การสร้างโครงสร้าง (structuring)	การระบุส่วนประกอบที่เหมาะสมหรือสัมพันธ์กับโครงสร้าง เช่น - จัดวางหนังสือในห้องสมุดให้ตรงตามหมวดหมู่ - ทำแผนผังแสดงความสัมพันธ์ระหว่างพืชกับสัตว์
4.3 การให้เหตุผล หรืออ้างเหตุผล (attributing)	- การสร้างโครงสร้างใหม่ (deconstructing)	การระบุทรศนะ อคติ คุณค่า หรือเจตนาที่แฝงอยู่ในเนื้อหาที่ นำเสนอ เช่น - ระบุทรศนะที่สำคัญตามมุมมองทางการเมืองของผู้เขียน - ระบุตัวละครหรือเรื่องราวที่เป็นจุดเด่นในบทประพันธ์หรือ เรื่องสั้น - ตั้งสมมติฐานเกี่ยวกับมุมมองของนักการเมืองที่มีต่อปัญหา ต่างๆ
<b>5. การประเมิน (evaluating): ตัดสินบนพื้นฐานของเกณฑ์และมาตรฐาน</b>		
5.1 การตรวจสอบ (checking)	- การประสาน (coordinating) - การตรวจสอบ (detecting) - การติดตาม (monitoring) - การทดสอบ (testing)	การตรวจสอบความไม่สอดคล้องหรือข้อผิดพลาดของ กระบวนการหรือผลผลิต หรือการตัดสินได้ว่ากระบวนการหรือ ผลผลิตมีความสอดคล้องภายในหรือไม่ หรือการตรวจสอบ ประสิทธิผลของขั้นตอนที่ใช้ดำเนินงาน เช่น - การตรวจสอบว่าข้อสรุปของนักวิทยาศาสตร์ได้มาจากข้อมูลที่ เก็บรวบรวมได้ - มีส่วนร่วมในการทำงานกลุ่ม ให้ข้อมูลป้อนกลับในการทำงาน หรือข้อโต้แย้งอย่างมีเหตุผล - ทบทวนแผนของโครงการว่าได้รวมขั้นตอนที่จำเป็นไว้ใน การดำเนินงานหรือยัง



ตาราง 1 (ต่อ)

กระบวนการทางปัญญา	คำสำคัญ	ความหมายและตัวอย่าง
5.2 การวิพากษ์วิจารณ์ (critiquing)	- การตัดสิน (judging)	การตรวจสอบความไม่สอดคล้องระหว่างผลผลิตกับเกณฑ์ภายนอก หรือการตัดสินได้ว่าผลผลิตมีความสอดคล้องภายนอกหรือไม่ หรือการตรวจสอบความเหมาะสมของขั้นตอนที่ใช้สำหรับปัญหาที่กำหนด เช่น - พิจารณาระหว่าง 2 วิธีการ วิธีการใดจะใช้เป็นแนวทางในการแก้ปัญหาที่ดีที่สุด - ตัดสินว่าควรดำเนินการอย่างไรให้โครงการประสบความสำเร็จตามเกณฑ์ - เลือกวิธีการแก้ปัญหาทางคณิตศาสตร์ที่เหมาะสมที่สุด
<b>6. สร้างสรรค์ (Creating): รวมองค์ประกอบย่อยๆ ที่มีลักษณะคล้ายๆ กันเข้าด้วยกัน เพื่อสร้างรูปแบบหรือโครงสร้างใหม่</b>		
6.1 การสร้างสิ่งใหม่ (generating)	- การตั้งสมมติฐาน (hypothesizing)	การสร้างสมมติฐานทางเลือกบนพื้นฐานของหลักเกณฑ์ เช่น สร้างสมมติฐานเพื่ออธิบายปรากฏการณ์ที่เกิดขึ้น - - สร้างเกณฑ์หรือทางเลือกสำหรับการพัฒนาความสัมพันธ์ระหว่างเชื้อชาติในโรงเรียน - เสนอทางเลือกในการลดการพึ่งพาเชื้อเพลิงจากซากพืชซากสัตว์เพื่อลดปัญหาด้านเศรษฐกิจและสิ่งแวดล้อม
6.2 การวางแผน (planning)	- การออกแบบ (designing)	การออกแบบขั้นตอนการดำเนินงานให้บรรลุความสำเร็จ เช่น วางแผนการดำเนินการวิจัยตามหัวข้อที่ได้รับ - - ทำบทบาทย่นตรในการนำเสนอผลงานเกี่ยวกับแมลง - ออกแบบการทดลองเพื่อตรวจสอบผลของดนตรีประเภทต่างๆ ที่ส่งผลต่อการผลิตไข่ของไก่
6.3 การผลิตผลงาน (producing)	- การสร้าง (constructing)	การสร้างผลิตภัณฑ์ เช่น - สร้างที่อยู่อาศัยตามจุดประสงค์เฉพาะ - แสดงละครเกี่ยวกับเรื่องราวบางตอนในบทประพันธ์ที่ได้จากการอ่าน

## 2.2 แนวคิดพื้นฐานของโมเดลหลายองค์ประกอบของราส์ช

โมเดลหลายองค์ประกอบของราส์ช (many-facets Rasch measurement: MFRM) เป็นโมเดลทางสถิติที่สามารถใช้สำหรับศึกษาและอธิบายอิทธิพลของผู้ประเมิน (Engelhard, 1994; Iramaneerat & Yudkowsky, 2007: 270; Myford & Wolfe, 2000) โดย Linacre ได้พัฒนาโมเดลหลายองค์ประกอบของราส์ชขึ้นในปี 1989 โดยได้ปรับขยายจากโมเดลของราส์ช ให้สามารถใช้กับข้อสอบแบบลำดับขั้น (order categories) หรือข้อสอบที่มีการตรวจให้คะแนนแบบบางส่วน (partial credit) ซึ่งจากเดิมโมเดลของราส์ชจะมุ่งเน้นไปที่ 2 องค์ประกอบคือ ข้อสอบและผู้สอบ แต่ MFRM สามารถเพิ่มองค์ประกอบอื่นๆ ที่สนใจเข้าไปในโมเดล เช่น ความยากของข้อสอบ การกดคะแนนของผู้ประเมิน หรือโครงสร้างของมาตรฐานค่า ซึ่งฟังก์ชันความสัมพันธ์ขององค์ประกอบทั้งหมดจะ

แสดงด้วยสมการทางคณิตศาสตร์ ทั้งนี้ MFRM สามารถบ่งชี้พฤติกรรมการตรวจให้คะแนนว่ามีารกดหรือปล่อยคะแนนมากเพียงใด การวิเคราะห์ข้อมูลโดยใช้ MFRM จึงเป็นการศึกษาอิทธิพลของแหล่งความคลาดเคลื่อนต่างๆ ที่มีต่อคุณภาพของการวัดประเมิน โดยผู้วิจัยสามารถใช้โครงสร้างทางสถิติในการวิเคราะห์ผลการประเมินเพื่อการสรุปรูปแบบ (pattern) ของผลหลักสำหรับผู้ตรวจหรือผู้ประเมิน (rater) ผู้สอบ (ratee) และคุณลักษณะ (trait) หรือข้อสอบ (item) และตัวแปรอื่นๆ ที่สนใจ นอกจากนี้ MFRM ยังสามารถแยกวิเคราะห์ในแต่ละองค์ประกอบหรือฟาเซท (facet) อย่างเป็นอิสระจากองค์ประกอบอื่น ตลอดจนสามารถประเมินระดับสมาชิก (individual – level effects) ในแต่ละฟาเซทว่ามีลักษณะใด เช่น ผู้ประเมินแต่ละคน ผู้สอบแต่ละคน คุณลักษณะหรือข้อสอบแต่ละข้อ (Myford, & Wolfe, 2003: 402, 404)

โมเดลทางคณิตศาสตร์ของ MFRM ที่สอดคล้องกับข้อมูลการให้คะแนนแบบหลายค่าสามารถเขียนเป็นสมการได้ดังนี้ (Linacre, 1989 อ้างถึงใน Myford & Wolfe, 2003: 404)

$$\ln \left[ \frac{P_{nijk}}{P_{nijk-1}} \right] = B_n - D_i - C_j - F_k$$

เมื่อ  $P_{nijk}$  คือ ความน่าจะเป็นที่ผู้ถูกประเมินคนที่  $n$  ได้ระดับคะแนน  $k$  ของข้อรายการที่  $i$  จากผู้ประเมินคนที่  $j$

$P_{nijk-1}$  คือ ความน่าจะเป็นที่ผู้ถูกประเมินคนที่  $n$  ได้ระดับคะแนน  $k-1$  ของข้อรายการที่  $i$  จากผู้ประเมินคนที่  $j$

$B_n$  คือ ระดับความสามารถของผู้ถูกประเมินคนที่  $n$

$D_i$  คือ ค่าความยากของคุณลักษณะ  $i$

$C_j$  คือ ค่าการกดคะแนนของผู้ประเมิน (severity of rater)  $j$

$F_k$  คือ ค่าความยากประจำระดับที่  $k$  สัมพันธ์กับขั้นที่  $k-1$

โดยที่  $F_k$  ถือว่าพารามิเตอร์ไม่นับว่าเป็นองค์ประกอบของโมเดล

จากสมการเป็นโมเดลการวิเคราะห์ที่มีให้คะแนนแบบมาตราประมาณค่า (rating scale) ที่ประกอบด้วย 3 องค์ประกอบ ได้แก่ ความสามารถของผู้ถูกประเมิน ความยากของข้อรายการประเมิน และการกดคะแนนของผู้ประเมิน โดยมีคะแนนโลจิท เป็นตัวแปรตาม องค์ประกอบต่างๆ (facets) อันได้แก่ ผู้ถูกประเมิน คุณลักษณะที่ประเมิน และผู้ประเมิน เป็นตัวแปรอิสระ ความน่าจะเป็นของผู้ถูกประเมิน  $n$  ได้รับการให้คะแนน  $x$  ในรายการประเมินที่  $i$  จากผู้ประเมินคนที่  $j$

อย่างไรก็ตาม ในการประยุกต์ใช้ MFRM ในการตรวจสอบและความคุมอิทธิพลการกดหรือปล่อยคะแนนในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน ได้มีการปรับพารามิเตอร์ในการศึกษาจาก MFRM ทั่วไป โมเดลใน

การศึกษาครั้งนี้ประกอบด้วย 3 พารามิเตอร์ ได้แก่ ข้อสอบ (item) ผู้ประเมิน (rater) และระดับความกดคะแนนของผู้ประเมิน (rater threshold) โดยที่ให้พารามิเตอร์ข้อสอบ (item) เป็นสิ่งที่ถูกประเมินแทนที่จะเป็นผู้สอบ (person) และพารามิเตอร์ผู้ประเมิน (raters) แทนข้อสอบ (item) ซึ่งอยู่บนพื้นฐานของโมเดลมาตรฐานค่าของ Andrich (Andrich's rating scale model) (1978) จึงมีข้อตกลงเบื้องต้น 2 ประการ คือ (1) ผู้ประเมินทุกคนมีความสามารถในการจำแนกระหว่างข้อสอบที่สอดคล้องและไม่สอดคล้องกับมาตรฐานและตัวชี้วัด และ (2) ระดับความสอดคล้องในแนวเดียวกันที่ผู้ประเมินจะเปลี่ยนจากลำดับขั้นหนึ่งไปยังอีกขั้นหนึ่งของมาตรฐานค่ามีค่าเท่ากัน (เช่น จาก “ค่อนข้างสอดคล้อง (3)” ไป “ค่อนข้างสอดคล้องโดยตรง (4)”) ทั้งนี้ ผู้ประเมินจะเป็นผู้ตัดสินคุณลักษณะแฝง (latent trait) ของข้อสอบ โดยที่ข้อสอบจะอยู่บนเงื่อนไขของการกดหรือปล่อยคะแนนของผู้ประเมินให้คะแนนแต่ละคน สมการในการวิเคราะห์เขียนได้ดังนี้

$$\ln\left(\frac{P_{nik}}{P_{nik-1}}\right) = B_n - D_i - F_k$$

เมื่อ  $P_{nik}$  คือ ความน่าจะเป็นที่ผู้ประเมิน  $i$  จะประเมินความสอดคล้องในแนวเดียวกันกับมาตรฐานและตัวชี้วัดของข้อสอบข้อที่  $n$  โดยให้ระดับคะแนน  $k$

$P_{nik-1}$  คือ ความน่าจะเป็นที่ผู้ประเมิน  $i$  จะประเมินความสอดคล้องในแนวเดียวกันกับมาตรฐานและตัวชี้วัดของข้อสอบข้อที่  $n$  โดยให้ระดับคะแนน  $k-1$

$B_n$  คือ ความสอดคล้องในแนวเดียวกันกับมาตรฐานและตัวชี้วัดของข้อสอบ (latent alignment) ข้อที่  $n$

$D_i$  คือ อิทธิพลการกดหรือปล่อยคะแนนของผู้ประเมินคนที่  $i$

$F_k$  คือ พารามิเตอร์ประจำขั้นที่  $k$  สัมพันธ์กับขั้นที่  $k-1$

โดยที่  $F_k$  ถือว่าพารามิเตอร์ไม่นับว่าเป็นองค์ประกอบของโมเดล

จากสมการข้างต้น เป็นโมเดล 2 องค์ประกอบ ซึ่งตัวแปรตาม คือ คะแนนโลจิก ตัวแปรอิสระ คือ องค์ประกอบ (facets) ได้แก่ ข้อสอบ และผู้ประเมิน

### 2.3 แนวคิดของวิธีการศึกษาความสอดคล้องในแนวเดียวกันของ Porter

Porter และ Smithson (2001) ได้พัฒนาวิธีการวัดความสอดคล้องขึ้นมา ถึงแม้ว่าการประเมินความสอดคล้องด้วยวิธีการของ Webb (1997) จะใช้กันอย่างแพร่หลาย แต่เป็นวิธีการที่ค่อนข้างซับซ้อนและมีหลายขั้นตอน วิธีศึกษาความสอดคล้องที่พัฒนาโดย Porter and Smithson (2001) เป็นวิธีการที่สามารถศึกษาความสอดคล้องเชิงปริมาณได้ทั้ง 3 องค์ประกอบของระบบการศึกษา ได้แก่ มาตรฐาน การสอน และการประเมินผล ทำให้สามารถช่วยให้ผู้ที่เกี่ยวข้องกับ

กระบวนการทางการศึกษาได้เห็นความเชื่อมโยงระหว่างสิ่งที่เป่าหมายกับสิ่งที่สอนในห้องเรียน และสิ่งที่ประเมิน และยังสามารเปรียบเทียบระดับความสอดคล้องได้ทั้งระดับโรงเรียนและระดับรัฐ การวัดความสอดคล้องด้วยวิธีการของ Porter นี้จะนำผลการประเมินของผู้เชี่ยวชาญมาสร้างเมทริกซ์ 2 เมทริกซ์ เช่น เมทริกซ์มาตรฐานหรือเนื้อหาตามมาตรฐาน (เมทริกซ์ X) และเมทริกซ์ของเครื่องมือประเมิน (เมทริกซ์ Y) แล้วคำนวณค่าดัชนีความสอดคล้อง (alignment index) ซึ่งเป็นดัชนีสำหรับใช้ ประเมินว่ามาตรฐานและเครื่องมือประเมินสอดคล้องกันระดับใด

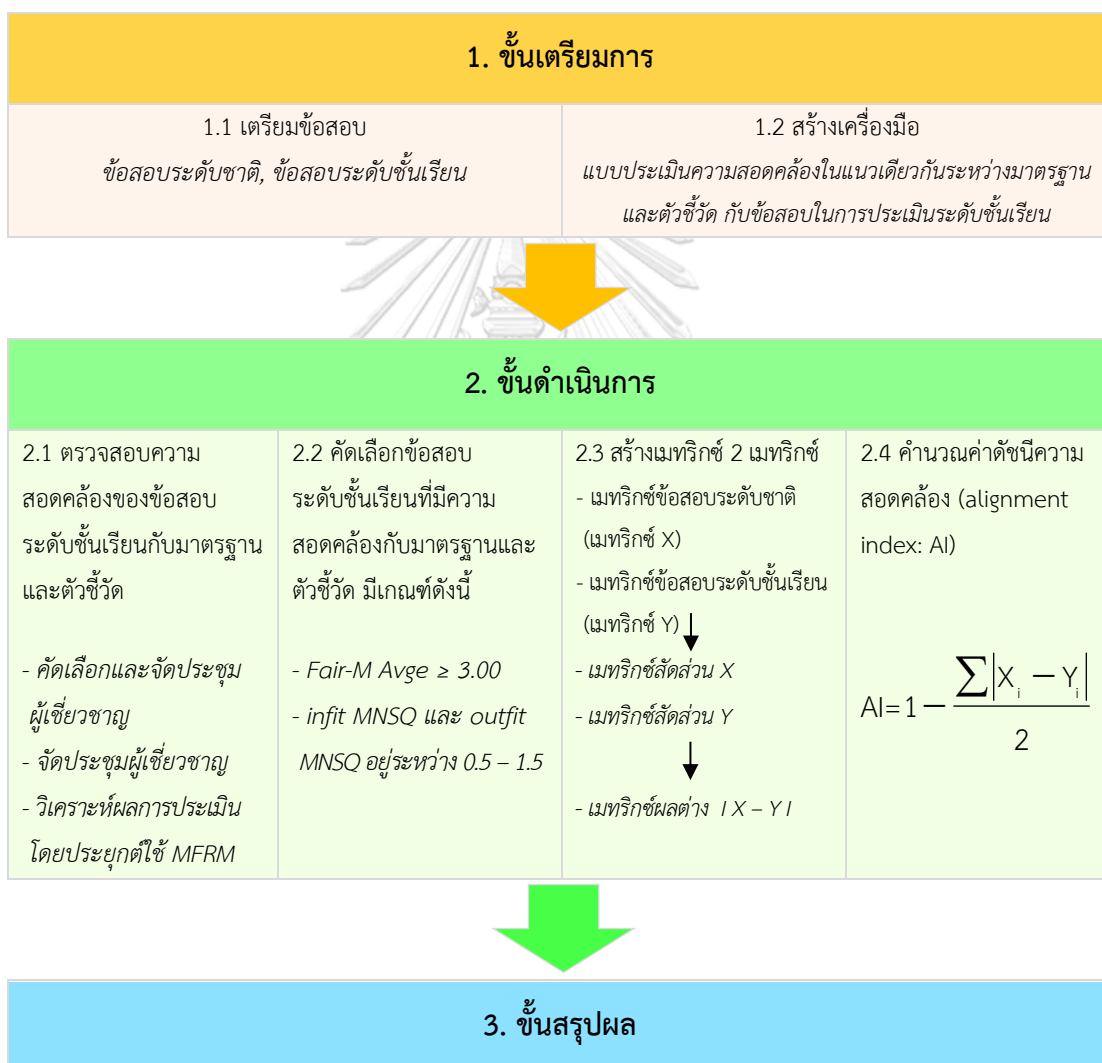
การเก็บรวบรวมข้อมูลในวิธีการของ Porter ขึ้นอยู่กับองค์ประกอบที่ต้องการวิเคราะห์ ความสอดคล้อง ในการอธิบายเนื้อหาของเครื่องมือประเมินและเนื้อหาของมาตรฐาน ใช้กระบวนการวิเคราะห์เนื้อหา โดยให้ผู้เชี่ยวชาญด้านเนื้อหาพิจารณาเนื้อหาของเครื่องมือประเมินและ เนื้อหามาตรฐาน และใช้กระบวนการที่เป็นระบบในการลงรหัส (code) ลงในตารางเนื้อหา 2 มิติ หรือเมทริกซ์ ซึ่งหัวข้อเรื่องถูกจัดอยู่ในแนวแถว และระดับความสามารถที่คาดหวังของผู้เรียน หรือ ความซับซ้อนทางปัญญจัดอยู่ในแนวคอลัมน์ สำหรับเครื่องมือในการประเมินหน่วยในการพิจารณา คือ ข้อคำถามในการประเมิน ส่วนมาตรฐานหรือกรอบหลักสูตร หน่วยในการพิจารณาจะถูกจัดระบบ ให้เป็นเนื้อหาย่อยที่เล็กที่สุด เช่น หัวข้อเรื่อง (topic) หรือหัวข้อเรื่องย่อย (subtopic) สำหรับเนื้อหา ของการสอนของครู ใช้การสำรวจหลักสูตรที่ครูนำไปใช้จริง ซึ่งครูจะรายงานเนื้อหาที่ครูสอนใน ห้องเรียน ซึ่งโดยทั่วไปครูจะรายงานข้อมูลนี้เมื่อสิ้นปีการศึกษา

ถัดมา ผลจากการเก็บข้อมูลเนื้อหาของหลักสูตรที่ถูกจัดให้อยู่ในรูปแบบเมทริกซ์ แต่ละช่อง หรือเซลล์ (cell) ของเมทริกซ์จะถูกคำนวณเป็นสัดส่วนของเนื้อหาในช่องนั้นต่อเนื้อหาทั้งหมดใน เมทริกซ์ ผลจากการเก็บข้อมูลเนื้อหาการสอนของครู อธิบายได้ด้วยสัดส่วนของเวลาที่ใช้ในการสอน เนื้อหาในช่องนั้นๆ ต่อเวลาที่ใช้สอนเนื้อหาทั้งหมด และผลจากการเก็บข้อมูลเนื้อหาของเครื่องมือ อธิบายได้ด้วยสัดส่วนของข้อคำถามในช่องนั้นๆ กับข้อคำถามทั้งหมดในเมทริกซ์

### บทที่ 3

## ขั้นตอนในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter

ขั้นตอนในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter สรุปได้ดังภาพ 1 ดังนี้



ภาพ 1 สรุปขั้นตอนในการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

รายละเอียดของการศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน ตามแนวคิดของ Porter มีดังต่อไปนี้


### ขั้นตอนที่ 1 ขั้นเตรียมการ

#### 1.1 การเตรียมข้อสอบที่ใช้ในการศึกษาความสอดคล้องเดียวกัน ประกอบด้วย ข้อสอบในการประเมินระดับชาติ และข้อสอบในการประเมินระดับชั้นเรียน

**ข้อสอบในการประเมินระดับชาติ** ได้แก่ ข้อสอบในการทดสอบทางการศึกษาระดับชาตินี้ขั้นพื้นฐาน (ordinary national educational test: O-NET) ในกลุ่มสาระการเรียนรู้และระดับชั้นที่ต้องการศึกษา การเตรียมข้อมูลข้อสอบในการประเมินระดับชาติทำโดยขอความอนุเคราะห์ไปยัง สทศ. เพื่อขอข้อมูลลักษณะเฉพาะของข้อสอบ (test-item specification) ซึ่งจะระบุว่าข้อสอบแต่ละข้อต้องการใช้มาตรฐานและตัวชี้วัดใด และวัดความซับซ้อนทางปัญญาระดับใด ยกตัวอย่างข้อมูลลักษณะเฉพาะของข้อสอบ O-NET วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559 ดังภาพ 2

(ตัวอย่าง) แบบบันทึกข้อมูลลักษณะจำเพาะของข้อสอบ (Test-item Specification)

ในการทดสอบทางการศึกษาขั้นพื้นฐาน (O-NET) กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2559

ข้อ ที่	ข้อคำถาม - ตัวเลือก	มาตรฐาน	ตัวชี้วัด	ระดับความซับซ้อนทางปัญญา						คะแนนเต็ม รายข้อ (คะแนน)
				จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์	
1	นำเซลล์ชนิดเดียวกัน ขนาดเท่ากัน ไปใส่ในสารละลาย A B และ C ที่มีความเข้มข้นแตกต่างกันเป็นเวลานานเท่าๆ กัน ได้ผลดังภาพ  จงเรียงลำดับความเข้มข้นของสารละลาย A B และ C จากความเข้มข้นน้อยที่สุดไปมากที่สุด 1. A B C 2. A C B 3. B A C 4. C A B  (ข้อสอบ O-NET วิทยาศาสตร์ ม.3 ปี 59)	ว.1.1	ว.1/4	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

ภาพ 2 ตัวอย่างแบบบันทึกลักษณะเฉพาะของข้อสอบ O-NET



จากภาพ 3 แสดงข้อรายการประเมินประกอบด้วย 2 ส่วนหลัก มีรายละเอียดดังนี้

ส่วนที่ 1 เป็นรายละเอียดของข้อสอบที่ครูสร้างขึ้นเพื่อใช้ในการประเมินผลการเรียนรู้ระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ระดับชั้นมัธยมศึกษาตอนต้น มีรายละเอียดดังนี้

1) ข้อสอบที่ต้องการประเมิน ซึ่งประกอบด้วย ข้อคำถามและตัวเลือก (กรณีเป็นข้อสอบหลายตัวเลือก)

2) คะแนนรายข้อ เป็นคะแนนเต็มรายข้อของข้อสอบที่ครูผู้ออกข้อสอบเป็นผู้กำหนด

3) ระดับชั้น เช่น ระดับชั้นมัธยมศึกษาปีที่ 1 – 3

4) มาตรฐานและตัวชี้วัด เป็นมาตรฐานการเรียนรู้และตัวชี้วัด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ที่ครูผู้สร้างข้อสอบระบุว่าต้องการวัดด้วยข้อสอบข้อนั้นๆ

ส่วนที่ 2 เป็นส่วนที่ให้ผู้เชี่ยวชาญพิจารณาประเมินหรือให้ข้อเสนอแนะต่อข้อสอบเป็นรายข้อใน 3 ประเด็น ได้แก่

1) ระดับความซับซ้อนทางปัญญา เป็นส่วนที่ให้ผู้เชี่ยวชาญพิจารณาระบุระดับความซับซ้อนทางปัญญาหรือความลึกของความรู้ในเนื้อหาข้อสอบ ที่ผู้เรียนต้องใช้เพื่อให้สามารถตอบข้อสอบได้ถูกต้อง ในตัวอย่างนี้จำแนกตามแนวคิด revised Bloom's taxonomy ทั้งนี้ ผู้เชี่ยวชาญระบุระดับความซับซ้อนทางปัญญาสูงสุดเพียง 1 ระดับ ที่ใช้ในการตอบข้อสอบข้อนั้นๆ ซึ่งข้อมูลในส่วนนี้จะนำไปวิเคราะห์ความสอดคล้องในแนวเดียวกันตามแนวคิดของ Porter เพื่อสร้างเมทริกซ์ข้อสอบ (Y) และนำไปคำนวณค่าดัชนีความสอดคล้องต่อไป

2) ระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด มีลักษณะเป็นมาตรประมาณค่า 5 ระดับ (0 - 4) ซึ่งให้ผู้เชี่ยวชาญพิจารณาตัดสินว่าเนื้อหาของข้อสอบกับเนื้อหาตัวชี้วัดที่กำหนด มีความสอดคล้องกันในระดับใด โดยรายละเอียดของแต่ละระดับมีดังนี้

0 คะแนน หมายถึง เนื้อหาของข้อสอบไม่สอดคล้องกับเนื้อหาของมาตรฐานและตัวชี้วัดที่กำหนด

1 คะแนน หมายถึง เนื้อหาของข้อสอบค่อนข้างไม่สอดคล้องกับเนื้อหาของมาตรฐานและตัวชี้วัดที่กำหนด

2 คะแนน หมายถึง ไม่แน่ใจว่าเนื้อหาของข้อสอบสอดคล้องกับเนื้อหาของมาตรฐานและตัวชี้วัดที่กำหนด

3 คะแนน หมายถึง เนื้อหาของข้อสอบค่อนข้างสอดคล้องกับเนื้อหาของมาตรฐานและตัวชี้วัดที่กำหนด

4 คะแนน หมายถึง เนื้อหาของข้อสอบสอดคล้องโดยตรงกับเนื้อหาของมาตรฐานและตัวชี้วัดที่กำหนด



ทั้งนี้ กำหนดเกณฑ์ที่ข้อสอบจะมีความสอดคล้องกับมาตรฐานและตัวชี้วัดก็ต่อเมื่อมีค่าเฉลี่ยคะแนนความสอดคล้องตั้งแต่ 3.00 ขึ้นไป

3) ข้อเสนอแนะเพิ่มเติม เป็นส่วนที่ให้ผู้เชี่ยวชาญเขียนแสดงความคิดเห็น หรือข้อสังเกต หรือข้อเสนอแนะอื่นๆ เพิ่มเติมเกี่ยวกับข้อสอบแต่ละข้อ

## ขั้นตอนที่ 2 ขั้นตอนดำเนินการ

### 2.1 ตรวจสอบความสอดคล้องของข้อสอบในการประเมินระดับชั้นเรียนกับมาตรฐานและตัวชี้วัด โดยมีการควบคุมอิทธิพลด้านการกดหรือปล่อยคะแนนของผู้ประเมิน

การดำเนินงานในขั้นนี้มีวัตถุประสงค์เพื่อตรวจสอบข้อสอบในการประเมินระดับชั้นเรียนที่มีความสอดคล้องกับมาตรฐานและข้อสอบตามที่ครูผู้สร้างข้อสอบระบุไว้หรือไม่ เพื่อคัดเลือกข้อสอบไปใช้ในการศึกษาความสอดคล้องในแนวเดียวกันกับข้อสอบในการประเมินระดับชาติต่อไป การดำเนินงานมีดังนี้

#### 1) การคัดเลือกและจัดประชุมผู้เชี่ยวชาญ มีรายละเอียดดังนี้

##### 1.1) การคัดเลือกผู้เชี่ยวชาญ

คุณสมบัติผู้เชี่ยวชาญ: ผู้เชี่ยวชาญต้องเป็นผู้ที่มีความรู้ความเชี่ยวชาญด้านมาตรฐานและตัวชี้วัดกลุ่มตามกลุ่มสาระการเรียนรู้และระดับชั้นที่ต้องการศึกษา และควรมีความรู้ในด้านการวัดและประเมินผลทางการศึกษา

จำนวนผู้เชี่ยวชาญ: ข้อสอบแต่ละข้อควรได้รับการประเมินจากผู้เชี่ยวชาญ อย่างน้อยจำนวน 2 - 3 คน ซึ่งจะให้ค่าความเที่ยงเท่ากับ 0.70 ขึ้นไป โดยการประเมินระดับความซับซ้อนทางปัญญาคควรใช้ผู้เชี่ยวชาญจำนวนอย่างน้อย 2 คน และการประเมินระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัดที่ระบุ ด้วยมาตรฐานประมาณค่า 5 ระดับ (0 - 4) ควรใช้ผู้เชี่ยวชาญจำนวนอย่างน้อย 3 คน

1.2) การจัดประชุมผู้เชี่ยวชาญ โดยมีจุดประสงค์เพื่อสร้างความเข้าใจที่ตรงกันระหว่างผู้เชี่ยวชาญถึงแนวทางในการดำเนินงาน ทดลองการฝึกปฏิบัติ และทำการปฏิบัติจริง ใช้ระยะเวลาประมาณครึ่งวัน เนื้อหาหรือกิจกรรมในประชุมมีดังนี้

1.2.1) ทบทวนนิยามของระดับความซับซ้อนทางปัญญาตามแนวคิด revised Bloom's taxonomy ทั้งนี้อาจมีเอกสารคู่มืออธิบายและยกตัวอย่างประกอบ

1.2.2) ให้ผู้เชี่ยวชาญทดลองฝึกปฏิบัติในกิจกรรมต่างๆ ได้แก่ การฝึกปฏิบัติกระบวนการค้นหามติเพื่อพิจารณาตัดสินการระบุระดับความซับซ้อนทางปัญญาของเนื้อหาข้อสอบ การฝึกปฏิบัติ การพิจารณาเพื่อประเมินความสอดคล้องระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบเป็นรายบุคคล และเปิดโอกาสให้ผู้เชี่ยวชาญได้อภิปรายภายในกลุ่มเพื่อให้ได้ผลการพิจารณาที่ถูกต้องและเข้าใจตรงกัน

1.2.3) ให้ผู้เชี่ยวชาญแต่ละคนทดลองทำหน้าที่จริงในการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบอย่างอิสระ เป็นรายบุคคล โดยยังไม่ปรึกษากัน ประมาณ 5 - 10 ข้อ

2) หลังจากการประชุมผู้เชี่ยวชาญแล้ว ผู้เชี่ยวชาญแต่ละคนจะทำหน้าที่ประเมินความสอดคล้องระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบฯ ได้อย่างอิสระเป็นรายบุคคล โดยใช้แบบประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบ และเอกสารคู่มือในการดำเนินงานสำหรับผู้เชี่ยวชาญ โดยข้อสอบแต่ละข้อควรได้รับการประเมินจากผู้เชี่ยวชาญข้อละ 2 - 3 คน ทั้งนี้ ในกรณีที่ข้อสอบมีจำนวนมากอาจมีการออกแบบให้ผู้เชี่ยวชาญแต่ละคนประเมินข้อสอบบางข้อ แทนการให้ผู้เชี่ยวชาญแต่ละคนประเมินข้อสอบทุกข้อ เพื่อประหยัดเวลา แรงงาน และงบประมาณในการดำเนินงาน ตลอดจนมีความเป็นไปได้ในการปฏิบัติ

3) นำผลการประเมินที่ผู้เชี่ยวชาญในส่วนของแต่ละระดับความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด ด้วยมาตรประมาณค่า 5 ระดับ (0 - 4) ของข้อสอบทุกข้อ มาวิเคราะห์ค่าเฉลี่ยความสอดคล้องระหว่างข้อสอบกับตัวชี้วัด เพื่อตัดสินระดับความสอดคล้องฯ โดยประยุกต์ใช้ MFRM ด้วยโปรแกรม FACETS เพื่อควบคุมความคลาดเคลื่อนของอิทธิพลการกดหรือปล่อยคะแนน ซึ่งจะให้ผลการประเมินที่ถูกต้องมากกว่าการใช้ค่าเฉลี่ยคะแนนดิบในการตัดสินความสอดคล้อง

โมเดลในการวิเคราะห์ประกอบด้วย 3 พารามิเตอร์ ได้แก่ ข้อสอบ (item) ผู้ประเมิน (rater) และระดับความกดคะแนนของผู้ประเมิน (rater threshold) โดยที่ให้พารามิเตอร์ข้อสอบ (item) เป็นสิ่งที่ถูกประเมินแทนที่จะเป็นผู้สอบ (person) และพารามิเตอร์ผู้ประเมิน (raters) แทนข้อสอบ (item)

## 2.2 คัดเลือกข้อสอบในการประเมินระดับชั้นเรียนที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัด

คัดเลือกข้อสอบในการประเมินระดับชั้นเรียนที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัด จากฟาเซทข้อสอบ (item facet) โดยพิจารณาค่า Fair-M Avgc ซึ่งเป็นค่าเฉลี่ยคะแนนที่ปรับใหม่ (adjusted mean) หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน โดยกำหนดเกณฑ์ความสอดคล้องระหว่างข้อสอบในการประเมินระดับชั้นเรียนกับมาตรฐานและตัวชี้วัดที่ระบุ ดังนี้

- 1) มีค่า Fair-M Avgc ตั้งแต่ 3.00 คะแนนขึ้นไป
- 2) มีข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล คือ มีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.50 - 1.50

ตัวอย่างผลการวิเคราะห์ความสอดคล้องระหว่างมาตรฐานและตัวชี้วัด กับข้อสอบในการประเมินระดับชั้นเรียน จากฟาเซทข้อสอบ จำนวน 10 ข้อ ที่วิเคราะห์จาก MFRM ด้วยโปรแกรม FACETS แสดงดังตาราง 2

ตาราง 2 ตัวอย่างผลการวิเคราะห์ฟาเซทข้อสอบ (item facet) ด้วย MFRM

ข้อที่	ชั้น	ตัวชี้วัด	Obs.Avg	FairM Avg	Measure	S.E.	Infit	Outfit
1	1	1114	3.85	3.89	4.79	0.64	0.82	0.66
2	2	4122	2.30	2.41	0.23	0.28	0.61	0.63
3	2	1124	3.60	3.74	3.75	0.90	1.67	1.82
4	3	2133	3.60	3.74	3.75	0.90	0.64	0.65
5	2	1125	3.00	3.12	1.77	0.70	0.62	0.63
6	2	1125	2.80	3.07	1.64	0.65	1.12	1.32
7	3	2131	3.60	3.76	3.85	0.88	0.67	0.73
8	1	3111	3.00	2.74	0.86	0.70	0.74	1.16
9	3	4131	3.20	3.29	2.23	0.75	1.24	1.11
10	1	4112	1.00	0.44	-2.42	0.65	0.95	0.92

หมายเหตุ Obs. Avg: ค่าเฉลี่ยคะแนนดิบ (raw mean) ก่อนควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Fair-M Avg: ค่าเฉลี่ยคะแนนที่ปรับใหม่ (adjusted mean) หลังควบคุมอิทธิพลการกดหรือปล่อยคะแนน

Measure: คะแนนในหน่วยลอจิทของฟาเซทข้อสอบ (item logit)

จากตาราง 2 พบว่า ภายหลังจากควบคุมอิทธิพลการกดหรือปล่อยคะแนน มีข้อสอบที่มีความสอดคล้องกับมาตรฐานและตัวชี้วัด จำนวน 6 ข้อ ได้แก่ ข้อ 1 4 5 6 7 และ 10 เนื่องจากมีค่า Fair-M Avg ตั้งแต่ 3.00 คะแนนขึ้นไป และมีข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล คือ มีค่า infit MNSQ และ outfit MNSQ อยู่ระหว่าง 0.50 – 1.50

### 2.3 สร้างเมทริกซ์ 2 เมทริกซ์ ได้แก่ เมทริกซ์ข้อสอบในการประเมินระดับชาติ และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน

การสร้างเมทริกซ์ 2 เมทริกซ์ ได้แก่ เมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y) โดยเมทริกซ์ X และ Y ในการประเมินความสอดคล้องในแนวเดียวกันแต่ละมาตรฐานนั้น จะกำหนดให้แถวของเมทริกซ์ (แนวนอน) เป็นตัวชี้วัดภายใต้มาตรฐาน และกำหนดให้คอลัมน์ (แนวตั้ง) เป็นระดับความซับซ้อนทางปัญญาตามลำดับขั้นของกระบวนการทางปัญญา ภายใต้แนวคิด revised Bloom's taxonomy ตัวอย่างเมทริกซ์ X และเมทริกซ์ Y แสดงดังภาพ 4 มีรายละเอียดดังนี้

1) การสร้างเมทริกซ์ X ทำโดยนำข้อมูลลักษณะเฉพาะของข้อสอบ (test-item specification) ของข้อสอบในการประเมินระดับชาติ มาจัดทำเมทริกซ์ X โดยแต่ละช่องในเมทริกซ์ X คือ ความถี่ที่ สทศ. กำหนดระบุว่าตัวชี้วัดนั้นๆ มีข้อสอบที่วัดระดับความซับซ้อนทางปัญญาในระดับใด เช่น สทศ.

ระบุว่าข้อสอบข้อที่ 1 มีคะแนนเต็มรายข้อเท่ากับ 2 คะแนน ข้อสอบข้อนี้ใช้วัดมาตรฐาน 1.1 ของระดับมัธยมศึกษาปีที่ 1 ตัวบ่งชี้ที่ 1 ในระดับวิเคราะห์ ผู้วิจัยจะใส่เลข 2 ในช่องตัวชี้วัดที่ 4 ที่ตัดกับระดับวิเคราะห์ และเติมหมายเลข 0 ลงในช่องที่ตัวชี้วัดที่ 1 ไม่ได้วัด ทำเช่นเดียวกันในข้อสอบข้ออื่นๆ

2) การสร้างเมทริกซ์ Y ทำโดยนำข้อมูลที่ผู้เชี่ยวชาญกำหนดระดับความซับซ้อนทางปัญญาของข้อสอบจากแบบประเมินความสอดคล้องระหว่างมาตรฐานและตัวชี้วัดกับข้อสอบในการประเมินระดับชั้นเรียน (ในส่วนของประเมินระดับความซับซ้อนทางปัญญา) โดยเป็นข้อสอบที่ถูกตรวจสอบและคัดเลือกแล้วว่ามี ความสอดคล้องกับมาตรฐานและตัวชี้วัดที่ระบุ ผู้เชี่ยวชาญแต่ละคนจะพิจารณาข้อสอบและทำการระบุว่าระดับความซับซ้อนทางปัญญาของเนื้อหาข้อสอบแต่ละข้อ โดยผู้เชี่ยวชาญจะเติมเลข 1 ลงในช่องที่ผู้เชี่ยวชาญระบุ ทั้งนี้ ผู้เชี่ยวชาญจะระบุระดับความซับซ้อนทางปัญญาของข้อสอบได้เพียง 1 ระดับ ซึ่งเป็นความซับซ้อนทางปัญญาที่ผู้เชี่ยวชาญเห็นว่าเป็นระดับสูงสุดที่ผู้เรียนต้องใช้ในการทำข้อสอบข้อนั้นๆ ได้อย่างถูกต้อง ซึ่งหลังจากเก็บข้อมูลจากผู้เชี่ยวชาญครบทุกคนแล้ว ผู้วิจัยจะทำการรวมคะแนนของข้อสอบทุกข้อในแต่ละตัวชี้วัด จำแนกตามระดับความซับซ้อนทางปัญญาแต่ละระดับ

3) นำผลรวมในเมทริกซ์ Y (ซึ่งเป็นตัวเลขในบรรทัดสุดท้ายของแต่ละเมทริกซ์ ในแต่ละสาระ) ของผู้เชี่ยวชาญแต่ละคนรวมกัน การหาผลรวมนี้เพื่อให้ได้ภาพรวมการประเมินของผู้เชี่ยวชาญแต่ละคน

4) ทำค่าในเมทริกซ์ผลรวม X และเมทริกซ์ผลรวม Y ให้เป็นเมทริกซ์สัดส่วน ซึ่งเป็นการทำให้ค่าในเมทริกซ์ผลรวม X และเมทริกซ์ผลรวม Y อยู่มาตรฐานเดียวกัน เพื่อให้สามารถเปรียบเทียบค่าระหว่าง 2 เมทริกซ์ได้ โดยนำค่าผลรวมของแต่ละเมทริกซ์ไปหารสมาชิกในเมทริกซ์เดียวกัน ผลของการหารสมาชิกทุกตัวของเมทริกซ์ด้วยผลรวม จะได้เป็นเมทริกซ์สัดส่วน และทำให้ผลรวมของค่าสัดส่วนในเมทริกซ์จะเท่ากับ 1.00 ดังนั้น จากกระบวนการนี้ จะทำให้ได้เมทริกซ์สัดส่วนจำนวน 2 เมทริกซ์ คือ เมทริกซ์สัดส่วน X และเมทริกซ์สัดส่วน Y

เมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X)						
มาตรฐาน/ตัวชี้วัด	ระดับความซับซ้อนทางปัญญา					
	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
สาระที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต 1						
มาตรฐาน ว 1.1 เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้าง และหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต						
ตัวชี้วัดที่ 4 ทดลองและอธิบายกระบวนการสารผ่านเซลล์ โดยการแพร่และออสโมซิส						
ตัวชี้วัดที่ 5 ทดลองหาปัจจัยบางประการที่จำเป็นต่อการสังเคราะห์ด้วยแสงของพืชและอธิบายว่าแสง คลอโรฟิลล์ แก๊สคาร์บอนไดออกไซด์ น้ำ เป็นปัจจัยที่จำเป็น ต้องใช้ในการสังเคราะห์ด้วยแสง						
.....						

เมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y)						
มาตรฐาน/ตัวชี้วัด	ข้อที่	ระดับความซับซ้อนทางปัญญา				
		จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า
สาระที่ 1 สิ่งมีชีวิตกับกระบวนการดำรงชีวิต						
มาตรฐาน ว 1.1 เข้าใจหน่วยพื้นฐานของสิ่งมีชีวิต ความสัมพันธ์ของโครงสร้าง และหน้าที่ของระบบต่างๆ ของสิ่งมีชีวิตที่ทำงานสัมพันธ์กัน มีกระบวนการสืบเสาะหาความรู้ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ในการดำรงชีวิตของตนเองและดูแลสิ่งมีชีวิต						
ตัวชี้วัดที่ 4 ทดลองและอธิบายกระบวนการสารผ่านเซลล์โดยการแพร่และออสโมซิส	1					
	2					
	...					
	...					
	n					
	รวม					
ตัวชี้วัดที่ 5 ทดลองหาปัจจัยบางประการที่จำเป็นต่อการสังเคราะห์ด้วยแสงของพืชและอธิบายว่าแสง คลอโรฟิลล์ แก๊สคาร์บอนไดออกไซด์ น้ำ เป็นปัจจัยที่จำเป็น ต้องใช้ในการสังเคราะห์ด้วยแสง	1					
	2					
	...					
	...					
	n					
	รวม					
.....						

ภาพ 4 ตัวอย่างเมทริกซ์ข้อสอบในการประเมินระดับชาติ (เมทริกซ์ X) และเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน (เมทริกซ์ Y)

## 2.4 คำนวณค่าดัชนีความสอดคล้อง (alignment index) ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

นำเมทริกซ์สัดส่วน  $X$  และเมทริกซ์สัดส่วน  $Y$  จากข้อ 2.3 (5) มาเปรียบเทียบกัน และคำนวณดัชนีความสอดคล้อง (alignment index) โดยการหาผลต่างของเมทริกซ์  $X$  และ  $Y$  หรือ  $X - Y$  ถ้าหาก 2 เมทริกซ์มีความสอดคล้องกัน ผลลัพธ์ของผลต่างควรมีค่าเป็นศูนย์หรือเข้าใกล้ศูนย์ สูตรการคำนวณความสอดคล้องของ Porter เป็นดังนี้

$$\text{Alignment Index} = 1 - \frac{\sum |X_i - Y_i|}{2}$$

เมื่อ  $X_i$  คือ ค่าสมาชิกของเมทริกซ์ข้อสอบในการประเมินระดับชาติ ตำแหน่งที่  $i$

$Y_i$  คือ ค่าสมาชิกของเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน ตำแหน่งที่  $i$

ทั้งนี้ ค่าดัชนีความสอดคล้อง (alignment index) มีค่าตั้งแต่ 0.0 ถึง 1.00 โดย 0.00 หมายถึง ไม่มีความสอดคล้องกัน และ 1.00 หมายถึง มีความสอดคล้องกันอย่างสมบูรณ์

ตัวอย่างการเมทริกซ์สัดส่วน  $X$  เมทริกซ์สัดส่วน  $Y$  และเมทริกซ์ผลต่าง  $IX - YI$  แสดงดังภาพ 5

### เมทริกซ์สัดส่วน $X$

#### (ข้อสอบในการประเมินระดับชาติ)

สาระที่	จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
1	0.000	0.080	0.060	0.100	0.020	0.000
2	0.000	0.020	0.000	0.080	0.000	0.000
3	0.000	0.020	0.060	0.060	0.000	0.000
4	0.000	0.060	0.100	0.060	0.000	0.000
5	0.000	0.080	0.020	0.060	0.000	0.000
6	0.000	0.020	0.000	0.020	0.000	0.000
7	0.000	0.040	0.020	0.020	0.000	0.000

### เมทริกซ์สัดส่วน $Y$

#### (ข้อสอบในการประเมินระดับชั้นเรียน)

จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
0.032	0.096	0.004	0.041	0.002	0.000
0.000	0.036	0.003	0.004	0.001	0.000
0.080	0.111	0.037	0.042	0.003	0.000
0.004	0.076	0.037	0.010	0.002	0.000
0.020	0.074	0.050	0.049	0.002	0.000
0.024	0.045	0.002	0.016	0.002	0.000
0.022	0.066	0.005	0.005	0.000	0.000

### เมทริกซ์ผลต่าง $IX - YI$

จำ	เข้าใจ	ประยุกต์ใช้	วิเคราะห์	ประเมินค่า	สร้างสรรค์
0.032	0.016	0.056	0.059	0.018	0.000
0.000	0.016	0.003	0.076	0.001	0.000
0.080	0.091	0.023	0.018	0.003	0.000
0.004	0.016	0.063	0.050	0.002	0.000
0.020	0.006	0.030	0.011	0.002	0.000
0.024	0.025	0.002	0.004	0.002	0.000
0.022	0.026	0.015	0.015	0.000	0.000

alignment index = 0.584

## ภาพ 5 เมทริกซ์ผลต่างระหว่างข้อสอบในการประเมินระดับชาติ กับเมทริกซ์ข้อสอบในการประเมินระดับชั้นเรียน

### ขั้นตอนที่ 3 ขั้นสรุปผล

#### 3.1 การสรุปผลการศึกษา

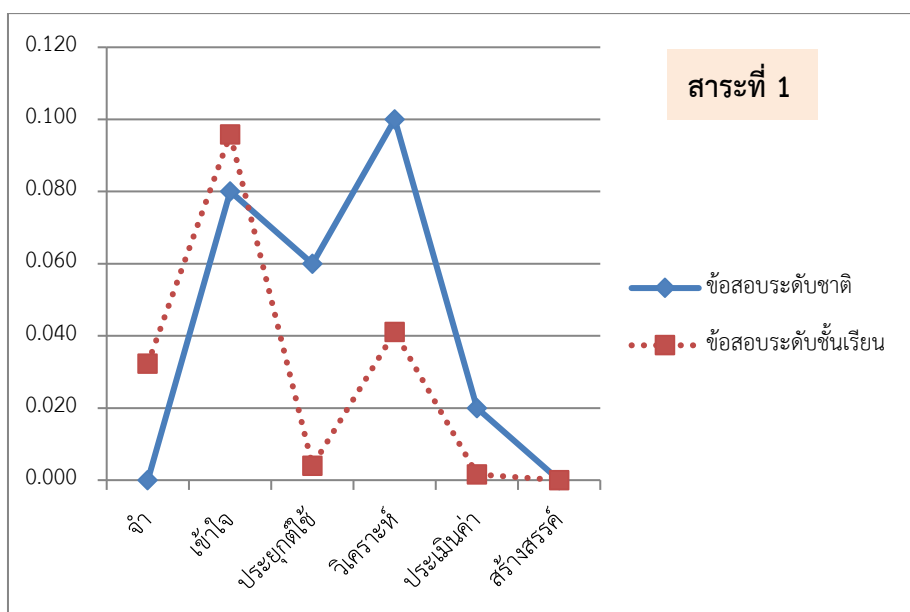
การสรุปผลการศึกษาด้วยการพิจารณาระดับความสอดคล้องในแนวเดียวกันสามารถดำเนินการได้ดังนี้

1. การพิจารณาความสอดคล้องในภาพรวมด้วยค่าดัชนีความสอดคล้อง (alignment index) ซึ่งหากยังมีค่าเข้าใกล้ 1.00 จะแสดงว่าสองเมทริกซ์ที่นำมาศึกษามีความสอดคล้องกันสูง

2. การพิจารณาความสอดคล้องในรายละเอียดของแต่ละช่อง (เซลล์) ของเมทริกซ์ โดยพิจารณาค่าในแต่ละเซลล์ของเมทริกซ์ผลต่าง  $|X - Y|$  หากค่าในแต่ละช่องมีค่าเข้าใกล้ศูนย์ แสดงว่าในเรื่องและระดับความซับซ้อนทางปัญญานั้นๆ มีความสอดคล้องกันสูง ทั้งนี้ หากพบว่าค่าในเมทริกซ์ช่องใดมีค่าที่สูง แสดงว่าเรื่องและระดับความซับซ้อนทางปัญญานั้นๆ มีความสอดคล้องกันต่ำ ซึ่งจะทำให้สารสนเทศอันเป็นประโยชน์ต่อผู้ที่เกี่ยวข้องกับการวัดและประเมินผลในการพัฒนา/ปรับปรุงการออกข้อสอบในการประเมินระดับชั้นเรียนต่อไป

3. สามารถพิจารณาเปรียบเทียบรายละเอียดของสัดส่วนข้อสอบที่จำแนกตามระดับความซับซ้อนทางปัญญาระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน โดยอาจนำผลจากเมทริกซ์ X และเมทริกซ์ Y มาสร้างแผนภูมิเส้น เพื่อให้เห็นภาพได้ชัดเจนมากขึ้น ยกตัวอย่างเช่น การนำค่าของสาระที่ 1 จากเมทริกซ์ X และเมทริกซ์ Y จากภาพ 5 มาแผนภูมิเส้นดังภาพ 6





ภาพ 6 เปรียบเทียบข้อสอบในสาระที่ 1 จำแนกตามระดับความซับซ้อนทางปัญญา ระหว่างข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน

จากภาพ 6 เมื่อพิจารณาความแตกต่างของสัดส่วนในการออกข้อสอบใช้สาระที่ 1 จำแนกตามระดับความซับซ้อนทางปัญญา จะเห็นว่าข้อสอบในการประเมินระดับชาติ กับข้อสอบในการประเมินระดับชั้นเรียน มีสัดส่วนของการออกข้อสอบในระดับจำ ประยุกต์ใช้ และวิเคราะห์ที่มีความแตกต่างกันมาก โดยข้อสอบในการประเมินระดับชาติ เน้นข้อสอบระดับวิเคราะห์มากที่สุด รองลงมาคือ ระดับเข้าใจ ระดับประยุกต์ใช้ และระดับประเมินค่า ในขณะที่ข้อสอบในการประเมินผลระดับชั้นเรียน เน้นข้อสอบระดับเข้าใจมากที่สุด รองลงมาคือ ระดับจำ ระดับวิเคราะห์ และมีข้อสอบที่วัดระดับประยุกต์ใช้ และระดับประเมินค่าจำนวนเล็กน้อย ดังนั้น จึงเป็นสารสนเทศเบื้องต้นว่าโรงเรียนควรปรับปรุงการออกข้อสอบให้วัดในระดับวิเคราะห์ และประยุกต์ใช้ให้มากขึ้น เพื่อให้ผู้เรียนฝึกการคิดในระดับสูง และเป็นการสร้างพื้นฐานที่เกิดความคุ้นเคยและเตรียมตัวเพื่อรับการประเมินด้วยข้อสอบระดับชาติ ตลอดจนการประเมินในการศึกษาระดับที่สูงขึ้นต่อไป

### 3.2 ข้อเสนอแนะเพิ่มเติม

1. ในส่วนของการประเมินระดับความซับซ้อนทางปัญญาของข้อสอบ ผู้ศึกษาสามารถจำแนกได้หลายแนวคิด เช่น แนวคิดการจำแนกลำดับชั้นของกระบวนการทางปัญญาของ Bloom ที่ได้รับการปรับปรุงใหม่ (revised Bloom's taxonomy) (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths, & Wittrock, 2001) จำแนกความซับซ้อนทางปัญญาได้เป็น 6 ชั้น ได้แก่ จำ เข้าใจ ประยุกต์ใช้ วิเคราะห์ ประเมินค่า และสร้างสรรค์ แนวคิดของ Webb (1997) จำแนกได้เป็น 4 ระดับ ได้แก่ ระลึกได้ ทักษะ/ความคิดรวบยอด การคิดเชิงกลยุทธ์ และการขยายความคิด



หรือแนวคิดของ Porter (Porter & Smith, 2001) จำแนกได้เป็น 5 ระดับ ได้แก่ จำ ปฏิบัติตาม ขั้นตอน สื่อสารความเข้าใจ แก้ปัญหาในสถานการณ์ใหม่ และคาดคะเน/สรุปอ้างอิง/พิสูจน์ แนวคิดของ Marzano (Marzano & Kendall, 2001) จำแนกได้เป็น 4 ระดับ ได้แก่ ดึงความรู้เดิมออกมาใช้ สร้างความคิดรวบยอด วิเคราะห์ความรู้ ใช้ประโยชน์จากความรู้ อภิปัญญา และจัดระบบแห่งตน

2. การศึกษาความสอดคล้องในแนวเดียวกันในคู่มือฯ เล่มนี้ อยู่บนพื้นฐานของงานวิจัยที่เป็นการศึกษาแบบย้อนหลัง (retrospective study) อย่างไรก็ตาม ผู้ศึกษาสามารถใช้แนวทางในคู่มือฯ นี้ในการศึกษาแบบไปข้างหน้า (prospective study) โดยเป็นขั้นตอนหนึ่งในกระบวนการสร้างแบบสอบ ซึ่งสารสนเทศที่ได้จากการวิเคราะห์ความสอดคล้องฯ จะเป็นประโยชน์ต่อคุณภาพการสร้างแบบสอบในด้านความตรงมากยิ่งขึ้น

3. ในคู่มือฯ นี้ระบุว่าการประเมินความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานกับตัวชี้วัดกับข้อสอบในการประเมินชั้นเรียน ใช้จำนวนผู้ประเมินความสอดคล้องฯ ของข้อสอบในแต่ละข้อ เพื่อการประเมินระดับความซับซ้อนทางปัญญา อย่างน้อย 2 คน และใช้จำนวนผู้ประเมิน อย่างน้อย 3 คน เพื่อการประเมินระดับความสอดคล้องฯ โดยใช้มาตรฐานประมาณค่า ซึ่งจะให้ค่าความเที่ยงในการประเมินมีค่าตั้งแต่ 0.70 ขึ้นไป อย่างไรก็ตาม การเพิ่มจำนวนผู้เชี่ยวชาญในการประเมินข้อสอบแต่ละข้อจะเป็นการเพิ่มความเที่ยงในการประเมิน เช่น หากต้องการให้การประเมินมีค่าความเที่ยงเท่ากับ 0.80 ต้องใช้จำนวนผู้ประเมินข้อสอบแต่ละข้อ อย่างน้อย 5 คน หรือต้องการให้การประเมินมีค่าความเที่ยง 0.90 ต้องใช้จำนวนผู้ประเมินข้อสอบแต่ละข้อ อย่างน้อย 8 คน ทั้งนี้ การออกแบบการประเมินดังกล่าวให้พิจารณาความเหมาะสมของเวลา ค่าใช้จ่าย แรงงาน ตลอดจนความเป็นไปได้ในทางปฏิบัติ

## รายการอ้างอิง

- บริษัทอักษรเจริญทัศน์ จำกัด. การเพิ่มผลสัมฤทธิ์ของการสอบ O-NET. สืบค้นข้อมูล เมื่อวันที่ 10 พฤศจิกายน พ.ศ. 2559 จาก [https://www.google.co.th/?gws\\_rd=cr,ssl&ei](https://www.google.co.th/?gws_rd=cr,ssl&ei).
- เอี่ยมพร หลินเจริญ, สิริศักดิ์ อาจวิชัย, ภริภา จันทรอินทร์ .(2552) .ปัจจัยเชิงสาเหตุที่ทำให้คะแนนการทดสอบ O-NET ของนักเรียนชั้นประถมศึกษาปีที่ 6 และชั้นมัธยมศึกษาปีที่ 6 สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) แขวงทุ่งพญาไท เขตราชเทวี กรุงเทพฯ.
- Ananda, S. (2003). *Rethinking Issues of Alignment under "No Child Left Behind."* Knowledge Brief. San Francisco: WestEd.
- Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory into practice*, 41(4), 255-260.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). New York: Longman.
- Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice*, 22(3), 21-29.
- Biggs, J. (2003). *Teaching for quality learning at university*. Glasgow: The Society for Research into Higher Education & Open University Press.
- Case, B., Jorgensen, M. A., & Zucker, S. (2004). *Alignment in educational assessment*. San Antonio: Harcourt Assessment.
- Engelhard, G., Jr. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Education Measurement*, 31(2), 93-112.
- Fuhrman, S. (Ed.) (2001). *From the capitol to the classroom: Standards-based reform in the states*. Chicago, IL: University of Chicago Press.
- Iramaneerat, C., & Yudkowsky, R. (2007). Rater errors in a clinical skills assessment of medical students. *Evaluation & the Health Professions*, 30(3), 266-283.
- La Marca, P. M. (2001). Alignment of standards and assessments as an accountability criterion. In: ERIC Clearinghouse on Assessment and Evaluation.

- La Marca, P. M., Redfield, D., & Winter, P. C. (2000). *State standards and state assessment systems: a guide to alignment*. Washington, DC: Council of Chief State School Officers.
- Marzano, R. J., & Kendall, J. S. (2001). *The new taxonomy of educational objectives*. Thousand Oaks, CA: Corwin.
- Myford, C. M., & Wolfe, E. W. (2000). *Monitoring sources of variability within the test of spoken english assessment system (TOEFL Research Report No. 65)*. Retrieved from Princeton, NJ: <http://www.ets.org/Media/Research/pdf/RR-00-06-Myford.pdf>
- Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many-facet Rasch measurement: Part I. *Journal of applied measurement*, 4(4), 386-422.
- Porter, A. C., & Smithson, J. L. (2001). *Defining, developing, and using curriculum indicators. CPRE research report series RR-048*. Retrieved from Philadelphia, PA: University of Pennsylvania Graduate School of Education.
- Porter, A. C., Smithson, J., Blank, R., & Zeidner, T. (2007). Alignment as a teacher variable. *Applied Measurement in Education*, 20(1), 27-51. doi:10.1207/s15324818ame2001\_3
- Rothman, R., Slattery, J. B., Vranek, J. L., & Resnick, L. B. (2002). *Benchmarking and alignment of standards and testing. CSE technical report (CSE-TR-566)*.
- Smith, M. S., & O'Day, J. (1990). Systemic school reform. In S. H. Fuhman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233-267). London: The Falmer Press.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. (Research Monograph No. 8). Washington, DC: Council of Chief State School Officers.



### ประวัติผู้เขียนวิทยานิพนธ์

นางสาวบุษยารัตน์ จันทร์ประเสริฐ เกิดเมื่อวันที่ 24 เมษายน พ.ศ. 2527 ที่จังหวัดฉะเชิงเทรา สำเร็จการศึกษาครุศาสตรบัณฑิต (เกียรตินิยมอันดับ 2) สาขาวิชามัธยมศึกษา-วิทยาศาสตร์ เอกชีววิทยาและวิทยาศาสตร์ทั่วไป คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2548 ครุศาสตรมหาบัณฑิต สาขาวิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2551 และศึกษาต่อในระดับปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2557 ปัจจุบันเป็นอาจารย์มหาวิทยาลัยเทคโนโลยีมหานคร

