

การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้
โดยใช้ปัญหาเป็นฐาน: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราสซ์



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 2557
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

DEVELOPMENT OF THE QUALITY RATING MODELS ACROSS GROUPS OF RATERS
IN PROBLEM BASED LEARNING: AN APPLICATION OF
MANY-FACET RASCH MEASUREMENT

Miss Sukolrat Ingchatcharoen



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 2014

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน: การ ประยุกต์ใช้โมเดลหลายองค์ประกอบของราสส์
โดย	นางสาวศุภกรรัตน์ อิงชาติเจริญ
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร. กมลวรรณ ตังธนากานนท์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	รองศาสตราจารย์ ดร. โชติกา ภาษีผล

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยเป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาศึกษาศาสตรบัณฑิต

..... คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร. บัญชา ชลาภิรมย์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร. กมลวรรณ ตังธนากานนท์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(รองศาสตราจารย์ ดร. โชติกา ภาษีผล)

..... กรรมการ
(รองศาสตราจารย์ ดร. วรรณิ แกมเกตุ)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. ณัฐภรณ์ หลาวทอง)

..... กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร. สุภมาส อังศุโชติ)

ศุภรัตน์ อิงชาติเจริญ : การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน: การประยุกต์ใช้โมเดลหลายองค์ประกอบของราสช์ (DEVELOPMENT OF THE QUALITY RATING MODELS ACROSS GROUPS OF RATERS IN PROBLEM BASED LEARNING: AN APPLICATION OF MANY-FACET RASCH MEASUREMENT) อ.ที่ปรึกษาวิทยานิพนธ์
 หลัก: ผศ. ดร. กมลวรรณ ตังธนกันนท์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: รศ. ดร. โชติกา ภาชีผล, 311 หน้า.

การวิจัยนี้มีวัตถุประสงค์ เพื่อ 1) วิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยใช้โมเดลหลายองค์ประกอบของราสช์ 2) พัฒนาและตรวจสอบโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน 3) ตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง และ 4) พัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน กลุ่มตัวอย่าง คือ อาจารย์และนักศึกษาในสถาบันอุดมศึกษาที่จัดการเรียนการสอนในรูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐาน เครื่องมือที่ใช้ในการวิจัย คือ แบบสอบถามคุณลักษณะของผู้ประเมิน และแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย การวิเคราะห์ข้อมูลใช้สถิติบรรยาย MFRM ตรวจสอบความตรงของโมเดล ตรวจสอบความไม่แปรเปลี่ยนของโมเดล ANCOVA และ MANCOVA ผลการวิจัยสรุปได้ดังนี้

1. ในภาพรวมผู้ประเมินทั้งสามกลุ่มมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนเพียงเล็กน้อย (Measure in Logits อยู่ระหว่าง -0.75 ถึง 0.47) โดยผู้ประเมินกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน ส่วนผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนเป็นอันดับสองรองจากผู้ประเมินกลุ่มตนเอง และทั้งหมดถือว่าการให้คะแนนที่มีความแม่นยำ (Infit MNSQ อยู่ระหว่าง 0.97 ถึง 1.05)

2. โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ประกอบด้วย 2 องค์ประกอบ คือ บริบทของผู้ประเมิน ซึ่งมีตัวบ่งชี้ 5 ตัว ได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้ และคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งมีตัวบ่งชี้ 3 ตัว ได้แก่ การรับรู้มาตรฐาน กระบวนการเปรียบเทียบ และความคลาดเคลื่อน โดยโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ ($\chi^2 = 22.444$, $df = 19$, $p = .263$, $GFI = 0.985$, $AGFI = 0.971$ และ $RMR = 0.035$)

3. โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน มีความไม่แปรเปลี่ยนของรูปแบบโมเดลระหว่างกลุ่มผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง แต่มีความแปรเปลี่ยนของค่าน้ำหนักองค์ประกอบของแต่ละตัวบ่งชี้

4. ภายหลังการทดลองใช้โมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในรูปแบบคู่มือ พบว่า กลุ่มทดลองมีความคลาดเคลื่อนน้อยกว่ากลุ่มควบคุมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ภาควิชา	วิจัยและจิตวิทยาการศึกษา	ลายมือชื่อนิติ
สาขาวิชา	การวัดและประเมินผลการศึกษา	ลายมือชื่อ อ.ที่ปรึกษาหลัก
ปีการศึกษา	2557	ลายมือชื่อ อ.ที่ปรึกษาร่วม

5484245027 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION

KEYWORDS: QUALITY RATING MODEL / PROBLEM-BASED LEARNING / MANY-FACET RASCH MEASUREMENT / MEASUREMENT INVARIANCE

SUKOLRAT INGCHATCHAROEN: DEVELOPMENT OF THE QUALITY RATING MODELS ACROSS GROUPS OF RATERS IN PROBLEM BASED LEARNING: AN APPLICATION OF MANY-FACET RASCH MEASUREMENT. ADVISOR: ASST. PROF. KAMONWAN TANGDHANAKANOND, Ph.D., CO-ADVISOR: ASSOC. PROF. SHOTIGA PASIPHOL, Ph.D., 311 pp.

The objectives of this research were 1) to analyze of rater errors in problem-based learning (PBL) using many-facet Rasch model, 2) to develop and validate the quality rating causal model in PBL, 3) to test the invariance of the quality rating causal model in PBL across tutor-, peer- and self-raters, and 4) to develop and test the quality rating model in PBL. The sample consisted of instructors and students of the institutes of higher education providing PBL. Research instruments were a set of rater characteristics questionnaires and a tutorial-based assessment form. Descriptive statistics, MFRM, SEM, MI, ANCOVA, and MANCOVA were employed to analyze the data. Research results were as follows:

1. In summary, three rater groups tend to have very little rater errors (Measure in Logits ranged from -0.75 to 0.47). Peer-raters tend to have the most lenient rater while tutor-raters tend to have the second severe raters next to self-raters. All ratings were considered accuracy (Infit MNSQ ranged from 0.97 to 1.05).

2. The quality rating causal model in PBL was comprised of the following two latent variable, such as 1) rater context measured from five observed variables consisting of rater goals, ability for rating, conscientiousness, rater's motivation and accountability, and 2) quality rating measured from three observed variables consisting of perception of PBL standards, comparison process, and rater error. The model was valid and fit to the empirical data (chi-square = 22.444, df = 19, p = .263, GFI = 0.985, AGFI = 0.971 and RMR = 0.035).

3. The quality rating causal model in PBL indicated invariance of model form among different rater groups, namely tutor-raters, peer-raters, and self-raters, but the model indicated variance of the factor loading of each indicators.

4. After a try out session of the quality rating model in PBL which was made available on manuals, it was found that the rater error of the experimental group was less than that of the control group at the statistically significant level .05.

Department: Educational Research and
Psychology

Field of Study: Educational Measurement and
Evaluation

Academic Year: 2014

Student's Signature

Advisor's Signature

Co-Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี ด้วยความเมตตาและกรุณาอย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ ดร.กมลวรรณ ดังธนกานนท์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำปรึกษาและคำแนะนำในการทำวิทยานิพนธ์ ตลอดจนสนับสนุนทั้งเวลาและกำลังใจให้แก่ผู้วิจัยทำให้สามารถพัฒนาผลงานอย่างถูกต้องสมบูรณ์มากยิ่งขึ้น รองศาสตราจารย์ ดร.โชติกา ภาซีผล อาจารย์ที่ปรึกษาร่วม ที่ได้กรุณาให้แนวคิด คำแนะนำและข้อเสนอแนะ ตลอดจนกำลังใจอย่างดียิ่ง เพื่อการแก้ไขปรับปรุงข้อบกพร่องในการทำวิทยานิพนธ์ตลอดมา ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ ประธานคณะกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.วรรณิ แกมเกตุ ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง และรองศาสตราจารย์ ดร.สุภมาส อังศุโชติ คณะกรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาให้คำแนะนำ ความรู้ และข้อเสนอแนะอันเป็นประโยชน์อย่างยิ่ง

ขอขอบพระคุณคณาจารย์ภาควิชาวิจัยและจิตวิทยาการศึกษาทุกท่านที่กรุณาประสิทธิ์ประสาทความรู้ทางวิชาการ ถ่ายทอดประสบการณ์ต่างๆ และเป็นต้นแบบที่ดีให้แก่ผู้วิจัย ทำให้ผู้วิจัยสามารถพัฒนาศักยภาพของตนทั้งทางด้านการวัดและประเมินผลการศึกษา การวิจัย และสถิติทางการศึกษา อันเป็นประโยชน์ต่อการทำวิทยานิพนธ์ การประกอบอาชีพ รวมถึงการเรียนรู้ต่อไปข้างหน้า

ขอขอบพระคุณบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย ที่กรุณาให้ทุนอุดหนุนการวิจัย “ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย” จากกองทุนรัชดาภิเษกสมโภช

ขอขอบพระคุณคณะพยาบาลศาสตร์ มหาวิทยาลัยมหิดลที่สนับสนุนเวลาในการศึกษาต่อและทำวิทยานิพนธ์ในครั้งนี้ ขอขอบคุณเพื่อนร่วมงาน เพื่อนร่วมชั้น รวมถึงรุ่นพี่ และรุ่นน้องของคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ให้ความช่วยเหลือและเป็นกำลังใจแก่ผู้วิจัยเป็นอย่างดี

ขอขอบพระคุณในความร่วมมืออย่างดียิ่งจากคณะผู้บริหาร คณาจารย์ บุคลากรสายสนับสนุน และนิสิต/นักศึกษา ที่ได้ให้ความอนุเคราะห์ในการเก็บรวบรวมข้อมูลในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน รวมถึงอีกหลายท่านที่ไม่ได้กล่าวในที่นี้ซึ่งได้ให้ความช่วยเหลือในการทำวิทยานิพนธ์ในครั้งนี้

ท้ายสุดนี้ขอขอบพระคุณครอบครัวของผู้วิจัยที่ให้กำลังใจ ความรัก ความห่วงใย และให้การสนับสนุนในทุกๆ ด้านตลอดมา ทำให้ผู้วิจัยสามารถทำวิทยานิพนธ์สำเร็จลุล่วงได้ด้วยดี คุณงามความดีที่ได้กล่าวมาทั้งหมดนี้ ขอจงบังเกิดเป็นโอกาสและสิ่งดีงามมอบให้แก่ทุกๆ ท่าน

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฎ
สารบัญภาพ	ฏ
บทที่ 1 บทนำ	1
ความเป็นมาและความสำคัญของปัญหา	1
คำถามการวิจัย	5
วัตถุประสงค์การวิจัย	6
ขอบเขตการวิจัย	6
คำจำกัดความที่ใช้ในการวิจัย.....	8
ประโยชน์ที่ได้รับจากการวิจัย	11
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง	13
ตอนที่ 1 การเรียนรู้โดยใช้ปัญหาเป็นฐานและการประเมินผลการเรียน	13
1.1 ความเป็นมาและประโยชน์ของการเรียนรู้โดยใช้ปัญหาเป็นฐาน	13
1.2 กระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐาน	15
1.3 การวัดและประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐาน	18
ตอนที่ 2 องค์ความรู้เกี่ยวกับคุณภาพการให้คะแนนการปฏิบัติ.....	26
2.1 ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนการปฏิบัติ.....	26
2.1.1 ความคลาดเคลื่อนของผู้ประเมิน (rater errors).....	28
2.1.2 ความแม่นยำในการให้คะแนน (rating accuracy)	32

2.1.3 ความสัมพันธ์ระหว่างความคลาดเคลื่อนของผู้ประเมินและความแม่นยำในการให้คะแนน.....	33
2.1.4 โมเดลหลายองค์ประกอบของราสช์ (Many-facet Rasch measurement: MFRM).....	35
2.1.5 ตัวอย่างงานวิจัยที่เกี่ยวข้องกับคุณภาพการให้คะแนนการปฏิบัติโดยใช้โมเดลหลายองค์ประกอบของราสช์.....	42
2.2 ปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ.....	53
2.2.1 แนวคิดทฤษฎีที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ.....	54
2.2.2 งานวิจัยที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ.....	62
ตอนที่ 3 การวิเคราะห์ความไม่แปรเปลี่ยนของโมเดล.....	82
3.1 หลักการวิเคราะห์โมเดลกลุ่มพหุ.....	82
3.2 ขั้นตอนการวิเคราะห์โมเดลกลุ่มพหุ.....	84
ตอนที่ 4 กรอบแนวคิดการวิจัย.....	90
บทที่ 3 วิธีดำเนินการวิจัย.....	91
ตอนที่ 1 การพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อนและตนเอง.....	91
1.1 ประชากรและตัวอย่าง.....	91
1.2 ตัวแปรที่ใช้ในการวิจัย.....	95
1.3 เครื่องมือที่ใช้ในการวิจัย.....	95
1.4 การเก็บรวบรวมข้อมูล.....	113
1.5 การวิเคราะห์ข้อมูล.....	115

ตอนที่ 2 การพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ ปัญหาเป็นฐาน.....	115
2.1 ประชากรและกลุ่มตัวอย่างในระยะที่ 2.....	116
2.2 เครื่องมือที่ใช้ในการวิจัยในระยะที่ 2.....	116
2.3 การดำเนินการทดลองและการเก็บรวบรวมข้อมูลในระยะที่ 2	118
2.4 การวิเคราะห์ข้อมูล	121
บทที่ 4 ผลการวิเคราะห์ข้อมูล.....	122
ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของกลุ่มตัวอย่าง.....	124
ตอนที่ 2 ผลการวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา	126
ตอนที่ 3 ผลการตรวจสอบโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้ โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา.....	130
ตอนที่ 4 ผลการตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนน ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาระหว่างกลุ่มผู้ประเมิน อาจารย์ เพื่อน และตนเอง.....	140
ตอนที่ 5 ผลการพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ ปัญหาเป็นฐานในระดับอุดมศึกษา.....	153
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	180
สรุปผลการวิจัย.....	181
ระยะที่ 1 การพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดย ใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ ประเมินอาจารย์ เพื่อน และตนเอง	181
ระยะที่ 2 การพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดย ใช้ปัญหาเป็นฐาน.....	184
อภิปรายผล.....	186

ระยะที่ 1 การพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง	186
ระยะที่ 2 การพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน.....	189
ข้อเสนอแนะ	191
ข้อเสนอแนะในการนำผลวิจัยไปใช้	191
ข้อเสนอแนะในการวิจัยต่อไป	192
รายการอ้างอิง	194
ภาคผนวก.....	201
ภาคผนวก ก รายนามผู้ทรงคุณวุฒิในการวิจัย.....	202
ภาคผนวก ข เครื่องมือที่ใช้ในการวิจัย	203
ภาคผนวก ค ใบรับรองโครงการวิจัย.....	209
ภาคผนวก ง คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา	211
ภาคผนวก จ ผลการวิเคราะห์โปรแกรม Facets.....	265
ภาคผนวก ฉ ตัวอย่างผลการวิเคราะห์โมเดลสมการโครงสร้างกลุ่มพหุด้วยโปรแกรม Lisrel	276
ประวัติผู้เขียนวิทยานิพนธ์	311

สารบัญตาราง

ตารางที่ 1	ลักษณะการวิเคราะห์ความคลาดเคลื่อนของผู้ประเมินด้วยวิธีโมเดล (Barrett, 2005: 164-5).....	42
ตารางที่ 2	การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับวิเคราะห์ความคลาดเคลื่อนของผู้ประเมิน.....	52
ตารางที่ 3	การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุที่ส่งผลต่อความแม่นยำในการให้คะแนน.....	80
ตารางที่ 4	การทดสอบชุดของสมมติฐานที่ซ้อนกันเป็นระดับลดหลั่นในการทดสอบกลุ่มพหุ.....	88
ตารางที่ 5	จำนวนตัวอย่าง อัตราการตอบกลับและอัตราความครบถ้วนของแบบสอบถามที่ได้รับจากกลุ่มตัวอย่าง.....	94
ตารางที่ 6	สาเหตุของความไม่สอดคล้องกันระหว่างคะแนนจากกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง.....	97
ตารางที่ 7	ผลการคำนวณค่าดัชนีความสอดคล้อง (IOC) และข้อเสนอแนะจากผู้ทรงคุณวุฒิ.....	98
ตารางที่ 8	แสดงการกำหนดจำนวนข้อคำถาม และพฤติกรรม/คุณลักษณะที่ต้องการวัด (table of specification).....	104
ตารางที่ 9	ค่าดัชนีของความสอดคล้องของแบบสอบถามคุณลักษณะของผู้ประเมิน.....	105
ตารางที่ 10	แสดงการกำหนดจำนวนข้อคำถาม และพฤติกรรม/คุณลักษณะที่ต้องการวัด (table of specification).....	109
ตารางที่ 11	ค่าดัชนีของความสอดคล้องของแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย.....	111
ตารางที่ 12	สรุปคุณลักษณะของเครื่องมือที่ใช้ในการวิจัย.....	112
ตารางที่ 13	ความถี่และร้อยละของกลุ่มตัวอย่างจำแนกตามตัวแปร เพศ คณะ วุฒิการศึกษา และชั้นปี.....	125
ตารางที่ 14	ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง จำแนกตามตัวแปร อายุ ระยะเวลาที่สอน GPAX และจำนวนวิชา PBL ที่เคยเรียน.....	126
ตารางที่ 15	แสดงการเปรียบเทียบระหว่างเพศ.....	128
ตารางที่ 16	ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน.....	129

ตารางที่ 17 ความถี่ของการให้คะแนนในแต่ละมาตรวัด	129
ตารางที่ 18 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่าสหสัมพันธ์ของตัวแปรสังเกตได้ในโมเดล (n= 360).....	130
ตารางที่ 19 ค่าประมาณพารามิเตอร์ และค่าสถิติที่เกี่ยวข้องในการตรวจสอบความตรงของ โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานตามกรอบ แนวคิด.....	132
ตารางที่ 20 ค่าสถิติผลการวิเคราะห์อิทธิพลของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนใน วิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานตามกรอบแนวคิด	134
ตารางที่ 21 ค่าพารามิเตอร์ และค่าสถิติที่เกี่ยวข้องในการตรวจสอบความตรงของโมเดลเชิง สาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (โมเดล ทางเลือก).....	138
ตารางที่ 22 ค่าสถิติผลการวิเคราะห์อิทธิพลของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนใน วิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (โมเดลทางเลือก).....	139
ตารางที่ 23 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่าสหสัมพันธ์ของตัวแปรในผู้ประเมินกลุ่ม อาจารย์ (n= 120).....	141
ตารางที่ 24 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่าสหสัมพันธ์ของตัวแปรในผู้ประเมินกลุ่ม เพื่อน (n= 120)	142
ตารางที่ 25 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่าสหสัมพันธ์ของตัวแปรในผู้ประเมินกลุ่ม ตนเอง (n= 120).....	143
ตารางที่ 26 ผลการทดสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมิน.....	144
ตารางที่ 27 ผลการวิเคราะห์ความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนน ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อนและตนเอง.....	151
ตารางที่ 28 แสดงความคิดเห็นเพิ่มเติมของผู้ทรงคุณวุฒิ (จำนวน 4 ท่าน)	155
ตารางที่ 29 จำนวนตัวอย่าง อัตราการตอบกลับและอัตราความครบถ้วนของแบบสอบถามที่ ได้รับจากกลุ่มตัวอย่าง (ในแต่ละกลุ่มประกอบด้วยอาจารย์ 1 คน และนักศึกษา 11-12 คน)	156

ตารางที่ 30 ผลการวิเคราะห์ความถี่และร้อยละของกลุ่มตัวอย่างจำแนกตามตัวแปร เพศ และ วุฒิการศึกษา.....	157
ตารางที่ 31 ผลการวิเคราะห์ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง จำแนกตาม ตัวแปร อายุ ระยะเวลาที่สอน และGPAX.....	158
ตารางที่ 32 ค่าสถิติจากการวิเคราะห์ข้อมูลพื้นฐานเปรียบเทียบระหว่างกลุ่มทดลองและกลุ่ม ควบคุม.....	159
ตารางที่ 33 แสดงการเปรียบเทียบระหว่างฟาเซท (กลุ่มที่ 1: กลุ่มทดลอง).....	163
ตารางที่ 34 ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน (กลุ่มที่ 1: กลุ่มทดลอง).....	164
ตารางที่ 35 ความถี่ของการให้คะแนนในแต่ละมาตรวัด (กลุ่มที่ 1: กลุ่มทดลอง).....	164
ตารางที่ 36 แสดงการเปรียบเทียบระหว่างฟาเซท (กลุ่มที่ 2: กลุ่มทดลอง).....	166
ตารางที่ 37 ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน (กลุ่มที่ 2: กลุ่มทดลอง).....	167
ตารางที่ 38 ความถี่ของการให้คะแนนในแต่ละมาตรวัด (กลุ่มที่ 2: กลุ่มทดลอง).....	167
ตารางที่ 39 แสดงการเปรียบเทียบระหว่างฟาเซท (กลุ่มที่ 3: กลุ่มควบคุม).....	169
ตารางที่ 40 ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน (กลุ่มที่ 3: กลุ่มควบคุม).....	170
ตารางที่ 41 ความถี่ของการให้คะแนนในแต่ละมาตรวัด (กลุ่มที่ 3: กลุ่มควบคุม).....	170
ตารางที่ 42 แสดงการเปรียบเทียบระหว่างฟาเซท (กลุ่มที่ 4: กลุ่มควบคุม).....	172
ตารางที่ 43 ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน (กลุ่มที่ 4: กลุ่มควบคุม).....	173
ตารางที่ 44 ความถี่ของการให้คะแนนในแต่ละมาตรวัด (กลุ่มที่ 4: กลุ่มควบคุม).....	173
ตารางที่ 45 ผลการวิเคราะห์ความแปรปรวนร่วมของคะแนนหลังทดลอง ระหว่างกลุ่มควบคุม และกลุ่มทดลองโดยมีคะแนนก่อนทดลองเป็นตัวแปรร่วม.....	174
ตารางที่ 46 ผลการทดสอบความแตกต่างของค่าเฉลี่ยคะแนนแรงจูงใจ.....	176
ตารางที่ 47 ผลการวิเคราะห์ความแปรปรวนตัวแปรร่วมพหุนามคะแนนหลังทดลองของตัวแปรการ รับรู้มาตรฐาน กระบวนการเปรียบเทียบ และความคลาดเคลื่อน ระหว่างกลุ่มควบคุมและกลุ่ม ทดลอง.....	178

สารบัญภาพ

ภาพที่ 1	วงจรกิจกรรมเรียนรู้โดยใช้ปัญหาเป็นฐาน (Hmelo-Silver, 2004: 237).....	16
ภาพที่ 2	ตัวอย่างการบันทึกข้อมูลบนกระดานหรือไวท์บอร์ด (Hmelo-Silver, 2004: 237)....	16
ภาพที่ 3	โมเดลองค์ประกอบในการให้คะแนนการปฏิบัติ (ที่มา: Landy & Farr, 1980: 73)....	55
ภาพที่ 4	โมเดลกระบวนการให้คะแนนการปฏิบัติ (ที่มา: Landy & Farr, 1980: 94).....	56
ภาพที่ 5	โมเดลสี่องค์ประกอบ (ที่มา: Murphy & Cleveland, 1995: 19).....	57
ภาพที่ 6	โมเดลสามขั้นตอน (ที่มา: Heidemeier & Moser, 2009: 355).....	59
ภาพที่ 7	โมเดลไนโปรแกรมลิสเรล (นงลักษณ์ วิรัชชัย, 2542: 26).....	86
ภาพที่ 8	กรอบแนวคิดเบื้องต้นในการวิจัย.....	90
ภาพที่ 9	ประชากรและกลุ่มตัวอย่างที่ใช้ในงานวิจัย.....	93
ภาพที่ 10	ขั้นตอนการเก็บรวบรวมข้อมูลในระยะที่ 1.....	114
ภาพที่ 11	ขั้นตอนการเก็บรวบรวมข้อมูลในระยะที่ 2.....	120
ภาพที่ 12	โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ในระดับอุดมศึกษา.....	135
ภาพที่ 13	โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ในระดับอุดมศึกษา (โมเดลทางเลือก).....	140
ภาพที่ 14	โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ใน ระดับอุดมศึกษาในผู้ประเมินกลุ่มอาจารย์.....	149
ภาพที่ 15	โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ใน ระดับอุดมศึกษาในผู้ประเมินกลุ่มเพื่อน.....	149
ภาพที่ 16	โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ใน ระดับอุดมศึกษาในผู้ประเมินกลุ่มตนเอง.....	150

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 แก้ไขเพิ่มเติมฉบับที่ 2 พ.ศ. 2545 และฉบับที่ 3 พ.ศ. 2553 กล่าวถึงความหมายของการศึกษาว่า เป็นกระบวนการเรียนรู้เพื่อความเจริญงอกงามของบุคคลและสังคม โดยมีหลักสำคัญ คือ ผู้เรียนทุกคนมีความสามารถเรียนรู้และพัฒนาตนเองได้ และผู้เรียนมีความสำคัญที่สุด ซึ่งกระบวนการจัดการศึกษาต้องส่งเสริมให้ผู้เรียนสามารถพัฒนาตามธรรมชาติและเต็มตามศักยภาพ ส่วนการประเมินผู้เรียนควรใช้วิธีการที่หลากหลาย เช่น ประเมินพัฒนาการของผู้เรียน สังเกตพฤติกรรมการเรียน การร่วมกิจกรรม และการทดสอบ โดยจัดให้มีควบคู่กันไปในกระบวนการเรียนการสอนตามความเหมาะสมของแต่ละระดับและรูปแบบการศึกษา นอกจากนี้ยังกำหนดให้มีระบบประกันคุณภาพการศึกษาเพื่อพัฒนาคุณภาพและมาตรฐานการศึกษาทุกระดับ (สำนักงานคณะกรรมการการศึกษา, 2542, 2545, 2553) จากแนวพระราชบัญญัติฯ นี้ สถาบันอุดมศึกษาจึงต้องมีพันธกิจหลัก คือ การผลิตบัณฑิต หรือการจัดกิจกรรมการเรียนการสอนเพื่อให้ผู้เรียนมีความรู้ในวิชาการและวิชาชีพ ตลอดจนมีคุณลักษณะตามที่หลักสูตรกำหนด ซึ่งการบริหารจัดการหลักสูตรและการจัดการเรียนการสอนจะต้องมีการวางระบบและกลไกในการควบคุมคุณภาพขององค์ประกอบต่างๆ ในการผลิตบัณฑิต ได้แก่ หลักสูตรการศึกษาในสาขาวิชาต่างๆ คณาจารย์และระบบการพัฒนาอาจารย์ สื่อการศึกษาและเทคนิคการสอน ห้องสมุดและแหล่งการเรียนรู้อื่น อุปกรณ์การศึกษา สภาพแวดล้อมในการเรียนรู้และบริการการศึกษา การวัดผลสัมฤทธิ์ทางการเรียนของนักศึกษา และองค์ประกอบอื่นตามที่แต่ละสถานศึกษาระดับอุดมศึกษาเห็นสมควรตามที่กฎกระทรวงว่าด้วยระบบ หลักเกณฑ์ และวิธีการประกันคุณภาพการศึกษา พ.ศ. 2553 ในขณะเดียวกันสังคมก็ต้องการให้ระบบอุดมศึกษามีความโปร่งใส (transparency) มีความรับผิดชอบที่ตรวจสอบได้ (accountability) และเปิดโอกาสให้ผู้มีส่วนได้ส่วนเสียได้มีส่วนร่วม (participation) ดังนั้นสถาบันอุดมศึกษาจึงต้องมีการให้ข้อมูลสาธารณะ (public information) ที่เป็นประโยชน์ต่อผู้มีส่วนได้ส่วนเสียทั้งนักศึกษา ผู้ปกครอง ผู้จ้างงาน รัฐบาล และประชาชนทั่วไป (Ajpru, Pasiphol, & Wongwanich, 2011; สำนักงานคณะกรรมการอุดมศึกษา, 2554)

การเรียนรู้โดยใช้ปัญหาเป็นฐาน (Problem Based Learning: PBL) เป็นการเรียนรู้ซึ่งมาจากการขับเคลื่อนของปัญหาที่เป็นบทเรียน โดยผู้เรียนต้องทำการระบุดุลยภาพที่จำเป็นสำหรับการเรียนรู้ จากนั้นจะต้องนำความรู้ที่ได้จากการไปศึกษาค้นคว้าด้วยตนเองมาใช้ในการแก้ไขปัญหา ซึ่งเป้าหมายสำคัญของการเรียนรู้โดยใช้ปัญหาเป็นฐานจะไม่ใช่ประเด็นของการแก้ไขปัญหาเพราะบางปัญหา

อาจจะไม่มีวิธีการแก้ไขโดยตรง แต่จะให้ความสำคัญกับการใช้ปัญหาเป็นจุดเริ่มต้นสำหรับการเรียนรู้ และประโยชน์ที่ได้จากความรู้ใหม่ คือ การเสริมสร้างความเข้าใจและเกิดเป็นความรู้ที่ติดตัวไป (Walsh, 2005) นอกจากนี้การประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานจะใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (tutorial-based assessment) หรือประเมินผลจากการปฏิบัติที่เกิดขึ้นจริงในบริบทของการเรียนรู้ ซึ่งเป็นรูปแบบหนึ่งของการประเมินทางเลือก (alternative assessment) (Tangdhanakanond, Pitiyanuwat, & Archwamety, 2006) ดังนั้นการเรียนรู้โดยใช้ปัญหาเป็นฐานจึงประกอบด้วยวิธีการเรียนรู้และการประเมินผลการเรียนรู้ที่สอดคล้องกับพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 เป็นอย่างดี นอกจากนี้ยังเหมาะสมกับการจัดการเรียนการสอนในศตวรรษที่ 21 ซึ่งต้องการทักษะที่ทำให้มนุษย์ชาติกลายป็นนักสร้างสรรค์และนักอนุรักษ์วัฒนธรรม เป็นผู้สร้างนวัตกรรมทางเทคโนโลยี และผู้ออกแบบวิถีชีวิตและการปกครอง โดยทักษะเหล่านี้ได้แก่ การคิดเชิงวิพากษ์ การแก้ไขปัญหา ความร่วมมือ ความสร้างสรรค์ การชี้นำตนเอง ความเป็นผู้นำ การปรับตัว ความรับผิดชอบ จิตสำนึกต่อโลก และทักษะการสืบค้น (Walser, 2008 cited in Barell, 2010)

การประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อยจากแหล่งผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง (ผู้เรียน) เนื่องจากปฏิสัมพันธ์ระยะยาวระหว่างอาจารย์ เพื่อน และผู้เรียน จะเป็นแหล่งข้อมูลสำคัญที่ช่วยแนะแนวทางการตัดสินใจความสามารถของผู้เรียน แต่จุดอ่อนของวิธีนี้ คือ ปัญหาความแม่นยำในการให้คะแนน (rating accuracy) เมื่อพิจารณาจากความสัมพันธ์ของคะแนนระหว่างการประเมินโดยอาจารย์ เพื่อน และตนเอง ผลการศึกษาเปรียบเทียบความสัมพันธ์ระหว่างการให้คะแนนโดยเพื่อนและโดยอาจารย์ พบว่ามีความสัมพันธ์กันตั้งแต่ระดับต่ำจนถึงระดับปานกลาง นอกจากนี้การให้คะแนนตนเองมีแนวโน้มที่จะมีค่าความสัมพันธ์ระดับต่ำกับการให้คะแนนโดยอาจารย์หรือเพื่อน (Eva, 2001; Papinczak, Young, Groves, & Haynes, 2007) ซึ่งปัญหาเหล่านี้จัดว่ามีความสำคัญมากในทางปฏิบัติ เนื่องจากการวัดและประเมินผลการศึกษาเป็นขั้นตอนหนึ่งที่มีความสำคัญต่อกระบวนการทางการศึกษาดังที่ได้กล่าวมาข้างต้น และข้อมูลที่ได้จากการวัดและประเมินผลทางการศึกษา นอกจากจะนำไปใช้ในการตัดเกรดหรือประเมินพัฒนาการของผู้เรียนแล้ว ยังใช้ในการตรวจสอบว่า รูปแบบหรือวิธีการประเมินผลทางการศึกษานั้นมีความเหมาะสมหรือไม่ และการจัดการเรียนการสอนนั้นมีคุณภาพเพียงไร

การตรวจสอบคุณภาพทางจิตมิติ (psychometric qualities) ของข้อมูลประเภทคะแนนมี 3 รูปแบบ คือ 1. การตรวจสอบคุณภาพทางจิตมิติแบบดั้งเดิม (traditional psychometric) เช่น ความตรง ความเที่ยง ความสอดคล้องระหว่างผู้ประเมิน 2. ความคลาดเคลื่อนของผู้ประเมิน (rater errors) และ

3. ความแม่นยำในการให้คะแนน (rating accuracy) ซึ่ง Murphy & Cleveland (1995) ได้กล่าวสรุปว่า การตรวจสอบคุณภาพทางจิตมิติและความคลาดเคลื่อนของผู้ประเมินเป็นการตรวจสอบทางอ้อมของความแม่นยำในการให้คะแนน (indirect measures of rating accuracy หรือ inaccuracy) เนื่องจากถ้าพบว่าผลการตรวจสอบคุณภาพทางจิตมิติหรือผลการวิเคราะห์ที่ไม่พบความคลาดเคลื่อนของผู้ประเมินอาจกล่าวได้ว่าการให้คะแนนนั้นมีความแม่นยำ (rating accuracy) ส่วนการตรวจสอบความแม่นยำในการให้คะแนนเป็นการตรวจสอบทางตรง (direct measures of rating accuracy) และจากผลการศึกษามากพบว่าการตรวจสอบคุณภาพทางจิตมิติ ซึ่งได้แก่ การตรวจสอบความเที่ยงอาจพบว่ามีค่าที่ต่ำ เนื่องจากการให้คะแนนการปฏิบัติมักมีความหลากหลายของวัตถุประสงค์ ส่วนการตรวจสอบความตรงตามเกณฑ์สัมพันธ์ (criterion-related validity) ทำได้ยากเนื่องจากมีความแตกต่างระหว่างเกณฑ์ที่ใช้มาก และการตรวจสอบความตรงตามทฤษฎี (construct validity) ทำได้ยากเช่นกัน โดยในงานวิจัยส่วนใหญ่ มักศึกษาความตรงแบบลู่เข้า (convergent validity) และความตรงในการจำแนก (discriminant validity) แทน ดังนั้นการตรวจสอบคุณภาพของการให้คะแนนการปฏิบัติจึงควรมุ่งเน้นไปที่การวิเคราะห์ความคลาดเคลื่อนและความแม่นยำในการให้คะแนน ซึ่งการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยจากผู้ประเมินหลายกลุ่ม คือ อาจารย์ เพื่อน และตนเอง ในการเรียนรู้โดยใช้ปัญหาเป็นฐาน ยังไม่เคยมีงานวิจัยที่ศึกษาวิเคราะห์ความคลาดเคลื่อนและความแม่นยำในการให้คะแนนอย่างละเอียด มีแต่เพียงการศึกษาความสัมพันธ์ระหว่างผลการให้คะแนนของกลุ่มผู้ประเมินเท่านั้น

การวิเคราะห์ความคลาดเคลื่อนและความแม่นยำในการให้คะแนน พบว่า มีความแตกต่างกันตามแนวคิดและทฤษฎีที่นำมาใช้ ซึ่งผลจากการศึกษาความสัมพันธ์ระหว่าง ความคลาดเคลื่อนและความแม่นยำในการให้คะแนนที่ใช้วิธีการวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) คือ การวิเคราะห์ความคลาดเคลื่อนจากการกระจายของความคลาดเคลื่อน (distributional errors) และความสัมพันธ์ของความคลาดเคลื่อน (correlational errors) และการวิเคราะห์ความแม่นยำในการให้คะแนนโดยวิธีของ Cronbach ในปี 1955 พบว่า การวิเคราะห์ความคลาดเคลื่อนไม่มีความสัมพันธ์กับการวิเคราะห์ความแม่นยำในการให้คะแนน (Becker & Cardy, 1986; Bernadin & Pence, 1980; Murphy & Balzer, 1989 cited in Murphy & Cleveland, 1995) หรืออาจเรียกว่า halo-accuracy paradox ซึ่ง Murphy & Cleveland (1995) ได้กล่าวไว้ว่า มีนักวิจัยบางส่วนได้ให้เหตุผลว่าเกิดจากการให้นิยามการวัดที่แตกต่างกัน และนักวิจัยอีกส่วนหนึ่งก็ไม่ได้ให้คำอธิบายใดๆ แต่ก็ได้สนับสนุนว่าเกิดจากการให้นิยามการวัดที่แตกต่างกัน และ Murphy & Cleveland ก็ไม่ได้อธิบายถึง

สาเหตุของกรณีนี้ แต่มีการอ้างถึง Engelhard (Engelhard, 1994 cited in Murphy & Cleveland, 1995) ว่าได้ประยุกต์ใช้โมเดลหลายองค์ประกอบของราสช์ (Many-facet Rasch measurement: MFRM) ในการวิเคราะห์ความคลาดเคลื่อนของผู้ประเมินและความแม่นยำในการให้คะแนน ซึ่งเมื่อพิจารณาจากบทความของ Engelhard (Engelhard, 1994) จะพบว่ามีการกล่าวถึงประเด็นที่ใกล้เคียง คือ สาเหตุของการที่ยังไม่มีเกณฑ์สำหรับการตรวจสอบความแม่นยำของการประเมินผลการปฏิบัติในบริบททางการศึกษา คือ การไม่มีนิยามของคะแนนที่แท้จริง (true scores) สำหรับใช้เป็นเกณฑ์มาตรฐานในการเปรียบเทียบ ดังนั้นจากคำอธิบายนี้เมื่อประกอบกับแนวคิดของการทดสอบแนวใหม่ (modern test theories) ในส่วนของของราสช์โมเดล (Rasch model) และทฤษฎีการตอบสนองข้อสอบ (item response theory) มาใช้อธิบายถึงสาเหตุของการเกิด halo-accuracy paradox จะได้ว่า เกิดจากข้อจำกัดของการใช้ทฤษฎีการทดสอบแบบดั้งเดิม คือ ค่าพารามิเตอร์ของข้อสอบเป็นค่าเฉพาะของกลุ่มผู้ถูกประเมินและเป็นอิสระจากความสามารถของผู้ถูกประเมิน ซึ่งในทฤษฎีการตอบสนองข้อสอบได้ปรับแก้ไขจุดอ่อน โดยให้ค่าพารามิเตอร์ของข้อสอบและความสามารถจริงของผู้ถูกประเมินมีความสัมพันธ์กัน ทำให้สามารถวิเคราะห์ความสามารถที่แท้จริงของแต่ละบุคคลได้ (ศิริชัย ศิริชัย กาญจนวาสี, 2550) ดังนั้น MFRM จึงสามารถวิเคราะห์ทั้งความคลาดเคลื่อนและความแม่นยำในการให้คะแนนจาก true scores ของผู้ถูกประเมินได้อย่างเที่ยงตรง และน่าเชื่อถือมากกว่าการวิเคราะห์โดยใช้ทฤษฎีการทดสอบแบบดั้งเดิม

การแก้ไขปัญหาการขาดความแม่นยำในการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยจากผู้ประเมิน 3 กลุ่ม ได้แก่ อาจารย์ เพื่อน และตนเอง พบว่ามีบางงานวิจัยได้แนะนำให้ใช้เทคนิคในการประเมิน เช่น การฝึกอบรมการใช้เกณฑ์การประเมิน และการสร้างแบบประเมินที่ตรงตามคุณลักษณะที่จะใช้ประเมิน แต่ทั้งนี้ก็ยังพบว่าไม่สามารถแก้ไขปัญหานี้ได้ทั้งหมด (Eva, 2001) โดยการแก้ปัญหาที่ดีควรทำอย่างเป็นระบบเพื่อให้ตรงกับสาเหตุของการเกิดปัญหานั้นๆ ซึ่งจากที่ผู้วิจัยได้ทำการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า ความสนใจในการศึกษาปัจจัยเชิงสาเหตุต่างๆ ที่ส่งผลต่อความแม่นยำในการให้คะแนนจะปรากฏอยู่ในศาสตร์ทางจิตวิทยาอุตสาหกรรมและองค์การ โดยอยู่ในส่วนของการประเมินผลการปฏิบัติ (Performance Appraisal; PA) โดยมีแนวคิดว่าการตรวจสอบความแม่นยำในการให้คะแนนจะประเมินได้จากความคลาดเคลื่อนระหว่างคะแนนที่ได้รับ (evaluation score) กับคะแนนจริง (true performance score) ของผู้ที่ถูกประเมิน แต่ในทางปฏิบัติการตรวจสอบคะแนนจริงอาจเป็นเรื่องที่ยากและมีความจำเป็นน้อยสำหรับการปฏิบัติใน

องค์กร ดังนั้นจึงมีการพัฒนาวิธีการวัดทางอ้อม คือ การสร้างตัวบ่งชี้ในการวัดพฤติกรรมการให้คะแนนที่มีความแม่นยำ (Kasten & Weintraub, 1999; Murphy & Cleveland, 1995)

ดังที่ได้กล่าวมานี้ผู้วิจัยจึงสนใจที่จะพัฒนาโมเดลปัจจัยเชิงสาเหตุของคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ซึ่งใช้การวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนโดยใช้โมเดลหลายองค์ประกอบของราสส์ ร่วมกับตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ซึ่งผลที่ได้จากงานวิจัยนี้คือ องค์ความรู้เกี่ยวกับปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนในแต่ละกลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง ตลอดจนข้อมูลเกี่ยวกับความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของผู้ประเมินกลุ่มอาจารย์ เพื่อน และตนเอง ซึ่งผู้วิจัยจะนำองค์ความรู้นี้มาใช้ในการพัฒนาโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และนำไปทดลองใช้โดยเปรียบเทียบคุณภาพการให้คะแนนระหว่างกลุ่มควบคุมกับกลุ่มทดลอง ซึ่งผลจากการทดลองทำให้ได้โมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานที่สามารถนำไปใช้ปฏิบัติได้จริงอย่างมีคุณภาพ

คำถามการวิจัย

1. ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเองในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา มีความแตกต่างกันหรือไม่เมื่อวิเคราะห์โดยโมเดลหลายองค์ประกอบของราสส์
2. โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ควรมีลักษณะอย่างไร และมีความตรงกับข้อมูลเชิงประจักษ์หรือไม่
3. โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ที่พัฒนาขึ้นแปรเปลี่ยนระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเองหรือไม่
4. โมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ควรเป็นอย่างไร และโมเดลที่พัฒนาขึ้นนี้ใช้ได้ดีเพียงใด

วัตถุประสงค์การวิจัย

1. วิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา โดยใช้โมเดลหลายองค์ประกอบของราสส์
2. พัฒนาและตรวจสอบโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา
3. ตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง
4. พัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

ขอบเขตการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยได้ทำการศึกษาโดยใช้ระเบียบวิธีวิจัยและพัฒนา (research and development) ซึ่งกำหนดขอบเขตของการวิจัยไว้ดังนี้

1. การพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ผู้วิจัยได้สังเคราะห์มาจากแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ ซึ่งประกอบด้วย

1.1 โมเดลการให้คะแนนการปฏิบัติ (performance rating mode) ของ Landy & Farr (1980) ถือเป็นแนวคิดแรกๆ ที่อธิบายปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนน ซึ่งในแต่ละองค์ประกอบได้รวบรวมมาจากผลงานวิจัยที่ใช้ทฤษฎีบุคลิกภาพ (personality theory) และจัดองค์ประกอบของโมเดลในลักษณะของกระบวนการที่มีปฏิสัมพันธ์ต่อกัน ภายในโมเดลนอกจากจะสนใจตัวแปรเชิงสาเหตุที่เน้นด้านสถานการณ์ ได้แก่ บทบาท บริบทของการให้คะแนน เครื่องมือกระบวนการให้คะแนน และผลลัพธ์ของการให้คะแนน ที่จะส่งผลต่อการให้คะแนนการปฏิบัติแล้ว ยังมีจุดเน้นที่กระบวนการคิดของผู้ประเมิน และกระบวนการตัดสินใจขององค์กร

1.2 โมเดลสี่องค์ประกอบ (four-Component model) ของ Murphy & Cleveland (1995) ซึ่งพัฒนาขึ้นจากแนวคิดและงานวิจัยต่างๆ ลักษณะของโมเดลจะเน้นความเป็นองค์รวม ประกอบด้วย

องค์ประกอบ 4 องค์ประกอบ ได้แก่ บริบทของการให้คะแนน (rating context) การตัดสินการปฏิบัติ (performance judgment) การให้คะแนนการปฏิบัติ (performance rating) และการประเมินระบบประเมินผล (evaluation of the appraisal system) ทั้งนี้โมเดลได้เน้นถึงความสำคัญขององค์ประกอบด้านบริบทขององค์กรที่จะส่งผลต่อองค์ประกอบอื่นๆ และในองค์ประกอบการประเมินระบบประเมินผลได้มีการอธิบายในส่วนของการประเมินความคลาดเคลื่อนและความแม่นยำในการให้คะแนน

1.3 โมเดลสามขั้นตอน (three-stage model) ของ Heidemeier & Moser (2009) สร้างขึ้นจากการวิเคราะห์ทอิกิมา (meta-analysis) โมเดลการให้คะแนนผลการปฏิบัติงานที่ศึกษาความสอดคล้องระหว่างการให้คะแนนโดยตนเองและผู้บังคับบัญชาจากงานวิจัยจำนวน 102 เรื่อง ที่ตีพิมพ์ระหว่างปี ค.ศ. 1955 – 2007 ซึ่งรายงานเป็นขนาดตัวอย่าง 128 หน่วย โดยโมเดลได้สรุปกระบวนการให้คะแนนออกเป็น 3 ขั้นตอน ได้แก่ การรวบรวมข้อมูล การคัดเลือกและบูรณาการข้อมูล หรืออาจเรียกว่าการพิจารณาตัดสินข้อมูล และการสื่อสารข้อมูล นอกจากนี้ภายในโมเดลยังประกอบด้วยตัวแปรกำกับที่จะส่งผลต่อทั้ง 3 ระยะเวลาของกระบวนการให้คะแนน โดยมีการจำแนกตัวแปรกำกับเป็น 5 กลุ่ม คือ 1) ประเภทของงานและคุณลักษณะของตำแหน่ง 2) คุณลักษณะต่างๆของผู้ประเมินและผู้ถูกประเมิน 3) คุณสมบัติของมาตรวัด ซึ่งรวมถึงรูปแบบและรายละเอียดของมาตรวัด 4) เงื่อนไขของการรายงาน และ 5) ภูมิหลังทางวัฒนธรรม

1.4 งานวิจัยจำนวน 18 เรื่อง ซึ่งมีความเกี่ยวข้องกับการศึกษาปัจจัยเชิงสาเหตุที่ส่งผลต่อความแม่นยำในการให้คะแนน ได้แก่ Bernardin, Tyler, & Villanova (2009), Curtis, Harvey, & Ravden (2005), Gorman & Rentsch (2009), Hedge & Teachout (2000), Heidemeier & Moser (2009), Houry & Analoui (2004), Kim (2011), Murphy, et al. (2004), Ostroff, Atwater, & Feinberg (2004), Park (2006), Payne, et al. (2009), Roch (2007), Salvemini, Reilly, & Smither (1993), Simsek, Pakdil, Dengiz, & Testik (2013), Tziner et al. (2008), Vecchio & Anderson (2009), Wang, Wong & Kwong (2010), และ Wong & Kwong (2007)

จากแนวคิดทฤษฎี และงานวิจัยที่เกี่ยวข้องดังได้กล่าวมาในข้างต้น ผู้วิจัยได้ทำการสังเคราะห์และจัดกลุ่มตัวแปรใหม่เพื่อพัฒนาเป็นโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยแบ่งออกเป็น 2 องค์ประกอบ ได้แก่ องค์ประกอบบริบทของผู้ประเมิน ซึ่งวัดจาก 5 ตัวบ่งชี้ ได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้ ที่จะส่งผลต่อองค์ประกอบคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้

โดยใช้ปัญหาเป็นฐาน ซึ่งวัดจาก 3 ตัวบ่งชี้ ได้แก่ การรับรู้มาตรฐาน กระบวนการเปรียบเทียบ และความคลาดเคลื่อน นอกจากนี้ยังมีตัวแปรกำกับ 1 ตัวแปร คือ กลุ่มผู้ประเมิน ซึ่งได้แก่ อาจารย์ เพื่อน และตนเอง ซึ่งตัวแปรทั้งหมดนี้มีงานวิจัยสนับสนุนจำนวนไม่ต่ำกว่า 3 งานวิจัย ดังนั้นผู้วิจัยจึงได้คัดเลือกตัวแปรเหล่านี้มาใช้ในการศึกษาครั้งนี้

2. ประชากรที่ใช้ในการวิจัยครั้งนี้ คือ อาจารย์และนักศึกษาในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา เนื่องจากการเรียนรู้โดยใช้ปัญหาเป็นฐานส่วนใหญ่จะใช้ใน ระดับอุดมศึกษา และการศึกษาในระดับอื่นๆ ที่ใช้การเรียนรู้โดยใช้ปัญหาเป็นฐานจะมีลักษณะของการประยุกต์ใช้ ซึ่งมีกระบวนการแบบไม่ครบทุกขั้นตอน

คำจำกัดความที่ใช้ในการวิจัย

โมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน หมายถึง รูปแบบการประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานซึ่งมีความแม่นยำและลดความคลาดเคลื่อนของการให้คะแนน ที่พัฒนาขึ้นจากผลของการวิเคราะห์โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และมีความเหมาะสมกับแต่ละกลุ่มของผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง

โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน หมายถึง รูปแบบที่แสดงลักษณะความสัมพันธ์ระหว่างตัวแปรอิสระที่เป็นสาเหตุ คือ องค์ประกอบด้านบริบทของผู้ประเมิน ที่จะส่งผลต่อองค์ประกอบด้านคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน

ความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมิน หมายถึง แบบจำลองความสัมพันธ์เชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน มีลักษณะไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทั้งจำนวนตัวแปรและเส้นทางอิทธิพล ระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง

โมเดลหลายองค์ประกอบของราสส์ หมายถึง รูปแบบสมการเชิงเส้นตรงทางคณิตศาสตร์ที่ใช้ในการแปลงค่าคะแนนที่ได้จากการสังเกตในทุกองค์ประกอบให้มีหน่วยเดียวกัน เพื่อใช้ในการเปรียบเทียบระหว่างองค์ประกอบ ตลอดจนสามารถคำนวณความน่าจะเป็นในการตอบถูกและผิด และวิเคราะห์

ความคลาดเคลื่อนและความแม่นยำในการให้คะแนนของผู้ประเมิน ซึ่งมาจากการประมาณค่าพารามิเตอร์ในโมเดล

การเรียนรู้โดยใช้ปัญหาเป็นฐาน หมายถึง รูปแบบของการเรียนการสอนที่ใช้การขับเคลื่อนของปัญหาเป็นจุดเริ่มต้นสำหรับการเรียนรู้ เพื่อที่จะเสริมสร้างความเข้าใจและพัฒนาเป็นความรู้ที่ติดตัวไป โดยมีกระบวนการเรียนรู้ใน 9 ขั้นตอน ได้แก่ 1) การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ 2) การกำหนดประเด็นปัญหา 3) การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้จากความรู้และประสบการณ์ที่มีอยู่ 4) การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล 5) การระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา 6) การศึกษาค้นคว้าด้วยตนเอง 7) การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้เรียนรู้มาต่อกลุ่ม 8) การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ และ 9) การสะท้อนความรู้ในบทเรียน โดยทำการสรุปทั้งในด้านประเด็นปัญหา ผลจากการไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน ตลอดจนกระบวนการในการแก้ปัญหา ส่วนการวัดและประเมินผลการเรียนรู้ จะใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อยจากกลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง

การประเมินกระบวนการเรียนรู้ในกลุ่มย่อย หมายถึง การพิจารณาตัดสินเกี่ยวกับความสามารถหรือทักษะการปฏิบัติของผู้เรียนในรายวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานตามกระบวนการเรียนรู้ทั้ง 9 ขั้นตอน ร่วมกับการประเมินคุณลักษณะอันพึงประสงค์ที่ได้จากการเรียนรู้โดยใช้ปัญหาเป็นฐาน ได้แก่ ภาวะผู้นำ ความเคารพต่อสมาชิกในกลุ่ม และเจตคติต่อการเรียนรู้โดยใช้ปัญหาเป็นฐาน

กลุ่มผู้ประเมิน หมายถึง บุคคลผู้ทำหน้าที่ในการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ซึ่งได้แก่ อาจารย์ เพื่อนนักศึกษา และตัวนักศึกษาเอง ซึ่งบุคคลทั้งสามฝ่ายนี้จะมีความแตกต่างกันในด้านของระดับความสัมพันธ์และบทบาทหน้าที่ต่อผู้ถูกประเมิน

อาจารย์ หมายถึง บุคลากรวิชาชีพที่ทำหน้าที่หลักทางด้านการเรียนการสอน และการส่งเสริมให้เกิดการเรียนรู้ผ่านวงจรของกระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐาน ในสถาบันระดับอุดมศึกษา

เพื่อน หมายถึง นักศึกษาระดับอุดมศึกษาในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งจะต้องทำหน้าที่ในการประเมินนักศึกษาที่อยู่ภายในกลุ่มเดียวกัน

ตนเอง หมายถึง ผู้เรียนซึ่งเป็นนักศึกษาระดับอุดมศึกษาในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งจะต้องทำหน้าที่ในการประเมินตนเอง

เป้าหมาย หมายถึง ลักษณะความคาดหวังของผู้ประเมินที่มีต่อกระบวนการและผลลัพธ์ของการประเมิน โดยลักษณะของเป้าหมายอาจมีความซับซ้อน หรือเป็นเป้าหมายในเชิงตรงกันข้าม ซึ่งเป้าหมายที่สำคัญของผู้ประเมินภายในบริบทของการศึกษามี 2 ประเภท ได้แก่ 1) เป้าหมายระหว่างบุคคล คือ ความคาดหวังของผู้ประเมินที่เกี่ยวข้องกับการใช้ผลการประเมินเพื่อพัฒนาความสัมพันธ์ระหว่างบุคคล เช่น การให้คะแนนประเมินเท่ากันเพื่อให้เกิดความเสมอภาคในกลุ่ม หรือความพยายามในการรักษาระดับความสัมพันธ์ระหว่างบุคคล และ 2) เป้าหมายภายในบุคคล คือ ความคาดหวังเกี่ยวกับการประเมินอันเป็นผลมาจากความเชื่อและการให้คุณค่าภายในของผู้ประเมิน เช่น ผู้ประเมินที่มีความเชื่อมั่นต่อกระบวนการในการประเมิน หรือผู้ประเมินที่เชื่อว่าผลการประเมินที่แม่นยำจะมีประโยชน์ต่อการนำไปใช้

ความสามารถ หมายถึง ศักยภาพของผู้ประเมินในการทำความเข้าใจคุณลักษณะที่ต้องการประเมิน โดยความแตกต่างของศักยภาพนี้จะขึ้นกับการได้รับความรู้หรือประสบการณ์จากการฝึกอบรมเกี่ยวกับการประเมินผลการปฏิบัติหรือการประเมินผลบุคคล ตลอดจนโอกาสของผู้ประเมินในการสังเกตพฤติกรรมหรือคุณลักษณะที่ต้องการประเมินจากผู้ถูกประเมิน

บุคลิกภาพแบบมีจิตสำนึก หมายถึง ลักษณะของบุคคลที่แสดงถึงการตระหนักในการควบคุมหรือกระตุ้นตนเองเพื่อให้เกิดการตอบสนองอย่างมีประสิทธิภาพ เช่น มีพลังในการทำงาน ต้องการที่จะประสบความสำเร็จ ทำงานเป็นระบบ ยึดมั่นในหน้าที่ความรับผิดชอบ

แรงจูงใจ หมายถึง สิ่งที่มีผลผลักดันให้ผู้ประเมินเกิดความพยายามในการประเมินเพื่อให้บรรลุเป้าหมายแห่งความถูกต้องของผลการประเมิน โดยคนที่มีแรงจูงใจสูงจะมีความพยายามอย่างไม่ลดละ ต่างกับคนที่มีแรงจูงใจต่ำที่อาจล้มเลิกการกระทำก่อนที่จะบรรลุเป้าหมาย

ความรับผิดชอบที่ตรวจสอบได้ หมายถึง ความพร้อมที่จะรายงานหรือรับการตรวจสอบจากผู้ที่เกี่ยวข้องในผลจากการปฏิบัติหน้าที่เป็นผู้ประเมิน

การรับรู้มาตรฐาน หมายถึง การตีความของผู้ประเมินเกี่ยวกับหลักเกณฑ์หรือแนวทางการปฏิบัติที่เหมาะสมสำหรับผู้เรียนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน

กระบวนการเปรียบเทียบ หมายถึง การใช้ความคิดอย่างเป็นลำดับของผู้ประเมินเพื่อให้ได้มาซึ่งสารสนเทศสำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ซึ่งประกอบด้วย 2 ขั้นตอน คือ 1) การรวบรวมข้อมูลเพื่อใช้ในการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย โดยข้อมูลจะได้รับการสังเกตพฤติกรรม หรือผลของการปฏิบัติ และ 2) การคัดเลือกและบูรณาการข้อมูล หรือการพิจารณาตัดสินข้อมูล โดยผู้ประเมินจะพิจารณาคัดเลือก จำแนก และสรุปข้อมูลเฉพาะที่ตรงกับวัตถุประสงค์การประเมิน

ความแม่นยำ หมายถึง ผลการให้คะแนนตามแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยของผู้ประเมินมีความถูกต้องใกล้เคียงตามความสามารถที่แท้จริงของผู้ถูกประเมิน และในการวิเคราะห์ความแม่นยำจะใช้โมเดลหลายองค์ประกอบของราสซ์ โดยพิจารณาจากผลการประมาณค่าของผู้ประเมินจากโปรแกรม Facets ในค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐาน ทั้งนี้ถ้ามีค่าอยู่ระหว่าง 0.6 ถึง 1.4 จะแสดงว่ามีความแม่นยำในการให้คะแนน (Wright, Linacre, Gustafson, & Martin-Löf, 1994)

ความคลาดเคลื่อน หมายถึง ผลการให้คะแนนตามแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยของผู้ประเมินที่แตกต่างจากค่าความสามารถที่แท้จริงหรือค่าที่ยอมรับของผู้ถูกประเมิน ซึ่งมาจากการวิเคราะห์ความคลาดเคลื่อนชนิดกด/ปล่อยคะแนน โดยใช้โมเดลหลายองค์ประกอบของราสซ์จากโปรแกรม Facets ที่พิจารณาจากการประมาณค่าของผู้ประเมิน ในค่า Measure ซึ่งมีหน่วยเป็นโลจิท (logits) ทั้งนี้ถ้ามีค่าเป็นบวกจะแสดงถึงความคลาดเคลื่อนชนิดกดคะแนน ค่าเป็นศูนย์จะแสดงถึงการไม่มี ความคลาดเคลื่อน และค่าเป็นลบจะแสดงถึงความคลาดเคลื่อนชนิดปล่อยคะแนน (Linacre, 2012)

ประโยชน์ที่ได้รับจากการวิจัย

1. ได้ข้อมูลเกี่ยวกับความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมิน อาจารย์ เพื่อน และตนเอง เป็นรายบุคคลเพื่อใช้ในการวางแผนพัฒนาความสามารถ ในการให้คะแนนของผู้ประเมินในรายวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน
2. ได้โมเดลเชิงสาเหตุของคุณภาพการให้คะแนน ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งเป็นโมเดลที่มีความครอบคลุมและทันสมัย โดยสามารถอธิบายความสัมพันธ์ระหว่างตัวแปรต่างๆ

ที่ส่งผลต่อคุณภาพในการให้คะแนนของแต่ละกลุ่มผู้ประเมิน ซึ่งได้แก่ อาจารย์ เพื่อน และตนเอง เพื่อนำไปใช้ปรับปรุงหรือพัฒนาให้เกิดคุณภาพในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน

3. ได้โมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในรูปแบบของคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา สำหรับอาจารย์ และสำหรับนักศึกษา ซึ่งพัฒนามาจากโมเดลเชิงสาเหตุของคุณภาพการให้คะแนน อันจะทำให้เกิดความชัดเจนเป็นรูปธรรม และได้ประโยชน์อย่างสูงสุดจากการใช้งานโมเดล โดยมีเป้าหมายสูงสุด คือ ช่วยในการปรับปรุงหรือพัฒนาให้เกิดคุณภาพในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน

4. ได้เครื่องมือวัดที่มีคุณภาพเกี่ยวกับปัจจัยเชิงสาเหตุที่ส่งผลต่อการให้คะแนน และแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ซึ่งสามารถนำไปใช้ในกลุ่มผู้ประเมินที่ต่างกัน คือ อาจารย์ เพื่อน และตนเองได้



บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เนื้อหาสาระในตอนนี้เป็นการนำเสนอแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง โดยแบ่งกลุ่มเนื้อหาออกเป็น 4 ตอน ได้แก่ ตอนที่ 1 การเรียนรู้โดยใช้ปัญหาเป็นฐานและการประเมินผลการเรียน ตอนที่ 2 องค์ความรู้เกี่ยวกับการให้คะแนนการปฏิบัติ ตอนที่ 3 การวิเคราะห์ความไม่แปรเปลี่ยนของโมเดล และตอนที่ 4 กรอบแนวคิดการวิจัย รายละเอียดของแต่ละตอนมีดังนี้

ตอนที่ 1 การเรียนรู้โดยใช้ปัญหาเป็นฐานและการประเมินผลการเรียน

การเรียนรู้จากปัญหาเป็นแนวทางพื้นฐานอย่างหนึ่งสำหรับการดำรงชีวิตของมนุษย์ ซึ่งความพยายามในการแก้ไขปัญหาเป็นเรื่องที่พบได้ในชีวิตประจำวัน โดยประสบการณ์ที่ได้จากการเรียนรู้ในสถานการณ์ปัญหาจะก่อให้เกิดความรู้และข้อมูลต่างๆ ที่สามารถนำไปประยุกต์ใช้กับปัญหาอื่นๆ ได้ในอนาคต และการได้มีโอกาสใช้ความรู้ต่างๆ นี้อยู่ครั้งยิ่งจะทำให้เกิดเป็นความจำระยะยาว ซึ่งจะต่างจากความรู้ที่ได้จากการอ่านที่มักเป็นความจำระยะสั้น ดังนั้นการเรียนรู้โดยใช้ปัญหาเป็นฐานจึงเป็นการเรียนรู้ที่เกิดจากกระบวนการทำความเข้าใจหรือแก้ไขปัญหาในสถานการณ์ต่างๆ ซึ่งไม่จัดเป็นวิธีการเรียนรู้ที่แปลกใหม่จากเดิม (Barrows & Tamblyn, 1980) ทั้งนี้ความหมายของการเรียนรู้โดยใช้ปัญหาเป็นฐาน (Problem Based Learning: PBL) คือ การเรียนรู้ซึ่งมาจากการขับเคลื่อนของปัญหาที่เป็นบทเรียนโดยผู้เรียนต้องทำการระบุดูวัตถุประสงค์ที่จำเป็นสำหรับการเรียนรู้ จากนั้นจะต้องนำความรู้ที่ได้จากการไปศึกษาค้นคว้าด้วยตนเองมาใช้ในการแก้ไขปัญหา และเป้าหมายที่สำคัญจะไม่ใช้การแก้ไขปัญหาเพราะบางปัญหาอาจจะไม่มีวิธีการแก้ไขโดยตรงแต่จะเป็นเรื่องของการใช้ปัญหาเป็นจุดเริ่มต้นสำหรับการเรียนรู้ และประโยชน์ที่ได้จากความรู้ใหม่ คือ การเสริมสร้างความเข้าใจและเกิดเป็นความรู้ที่ติดตัวไป (Walsh, 2005)

1.1 ความเป็นมาและประโยชน์ของการเรียนรู้โดยใช้ปัญหาเป็นฐาน

การเรียนรู้โดยใช้ปัญหาเป็นฐานได้มีการริเริ่มใช้ในหลักสูตรแพทยศาสตร์ของ McMaster University ประเทศแคนาดา ในปี 1960 ซึ่งในช่วงแรกยังไม่เป็นที่ยอมรับมากนักแต่ภายหลังได้มีผล

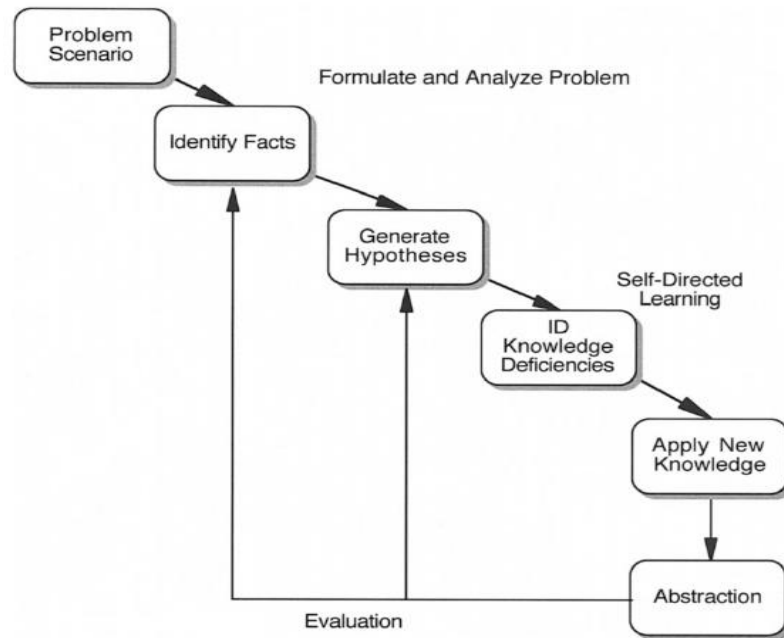
การศึกษาเปรียบเทียบพบว่า ประสิทธิภาพของการเรียนและความพึงพอใจของผู้เรียนและผู้สอนจากการเรียนรู้โดยใช้ปัญหาเป็นฐานมีมากกว่าระบบการเรียนการสอนแบบดั้งเดิม (Kassebaum, 1989) นอกจากนี้ยังมีการนำมาใช้ในการเรียนการสอนของสาขาต่างๆ ตั้งแต่ระดับโรงเรียนจนถึงมหาวิทยาลัย ทั้งนี้การเรียนรู้โดยใช้ปัญหาเป็นฐานเป็นรูปแบบที่สอดคล้องกับแนวคิดการเรียนรู้ที่เน้นผู้เรียนเป็นศูนย์กลาง (student-centered learning) โดยประกอบด้วยการเรียนรู้เป็นรายบุคคล (individualized learning) หรือการเรียนรู้แบบนำตนเอง (self-directed learning) และการเรียนรู้แบบอาจารย์เป็นศูนย์กลาง (teacher-centered learning) ซึ่งในการเรียนรู้แบบอาจารย์เป็นศูนย์กลางจะเกี่ยวข้องกับการสอนแบบบรรยาย (lectures) ดังนั้นในการเรียนรู้โดยใช้ปัญหาเป็นฐานจึงอาจมีการจัดบรรยายเสริมในหัวข้อที่ผู้เรียนต้องการและพิจารณาว่าเป็นเรื่องเฉพาะที่สำคัญ หรือการที่อาจารย์อาจเป็นผู้กำหนดหัวข้อและวัตถุประสงค์ที่สำคัญให้ผู้เรียนในการศึกษาด้วยตนเอง และอาจจัดสอบเมื่อเสร็จสิ้นเพื่อประเมินความรู้ของผู้เรียนว่าตรงตามวัตถุประสงค์หรือไม่ (Barrows & Tamblyn, 1980) เงื่อนไขสำคัญสำหรับแนวคิดการเรียนรู้โดยใช้ปัญหาเป็นฐานมี 4 ข้อ ได้แก่ 1) เน้นองค์ความรู้ที่ดี 2) เป็นการเรียนเชิงรุก (active learning) 3) มีปฏิสัมพันธ์ระหว่างผู้เรียน 4) ลักษณะของบริบทจะเป็นตัวกระตุ้นให้เกิดแรงจูงใจภายในเพื่อไปสู่เป้าหมายของการปฏิบัติ (Margetson 1994; cited in Papinczak et al., 2007; 122) ในส่วนของประโยชน์จากการเรียนรู้โดยใช้ปัญหาเป็นฐานที่ผู้เรียนจะได้รับ คือ ได้เรียนรู้ตัวอย่างจากสถานการณ์ปัญหาซึ่งมีความใกล้เคียงกับสถานการณ์จริง ฝึกการทำงานร่วมกันอย่างมีประสิทธิภาพ พัฒนาทักษะการแก้ปัญหาและเกิดความรู้ที่ได้จากการลงมือกระทำ มีฐานความรู้ที่กว้าง ทราบวิธีการประยุกต์ใช้ความรู้ในสถานการณ์ต่างๆ เกิดความรู้สึกตื่นตัว มีแรงจูงใจ มีพัฒนาการของการกำหนด/ตัดสินใจได้โดยตนเอง (self-determination) มีความมั่นใจในทักษะการเรียนรู้และการให้เหตุผลของตน เกิดความประทับใจต่อองค์ความรู้ และจากการพัฒนาทักษะในการเรียนรู้ต่างๆ นี้ จะช่วยส่งเสริมให้เกิดเป็นการเรียนรู้ตลอดชีวิต (lifelong learning) นอกจากนี้ยังเกิดประโยชน์ในส่วนของอาจารย์ เช่น มีความใกล้ชิดกับผู้เรียน บรรยายภาคเป็นธรรมชาติ คุ่มค่า และลดการใช้เวลาลง ส่วนจุดอ่อนของการเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ การจะประสบผลสำเร็จในการเรียนจะขึ้นอยู่กับวินัยของผู้เรียนที่จะเผชิญกับสิ่งที่ไม่รู้หรือปัญหาที่ยังไม่เข้าใจ ซึ่งเป็นเรื่องที่ทำลายทักษะในการแก้ปัญหาและกระตุ้นให้เกิดการเรียนรู้แบบนำตนเอง และอาจารย์จำเป็นต้องมีทักษะในการแนะนำหรือให้คำแนะนำและออกแบบหรือรวบรวมส่วนประกอบสำหรับการเรียน ส่วนจุดอ่อนอื่นที่อาจเกิดขึ้น เช่น ในระยะแรกอาจรู้สึกเครียดหรือกดดันซึ่งจะสามารถแก้ไขได้โดยการปฏิบัติที่ถูกต้อง (Barrows & Tamblyn, 1980; Hmelo-Silver, 2004; Lee & Kwan, 1997)

1.2 กระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐาน

Bridges (Bridges, 1992) ได้กล่าวถึงสิ่งสำคัญในการเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ 1) จุดเริ่มต้นสำหรับการเรียนรู้ คือ ปัญหา 2) ปัญหาเหล่านี้จะเป็นสิ่งที่ผู้เรียนจะได้นำไปใช้ในอนาคต 3) ผู้เรียนจะได้สัมผัสกับทฤษฎีและงานวิจัยที่เกี่ยวข้องกับปัญหานั้นๆ 4) ผู้เรียนเป็นผู้ที่ตัดสินใจว่าจะใช้ความรู้นี้ในการแก้ไขปัญหาได้อย่างไร และ 5) การเรียนรู้ส่วนใหญ่จะเกิดขึ้นภายในบริบทของกลุ่มย่อยมากกว่าการบรรยาย สอดคล้องกับ Hmelo-Silver (Hmelo-Silver, 2004) ที่กล่าวว่า ลักษณะที่สำคัญของการเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ ผู้เรียนจะทำงานร่วมกันในกลุ่มขนาดเล็กและเรียนรู้ในประเด็นต่างๆ ที่พวกเขาได้กำหนดขึ้นสำหรับการแก้ไขปัญหา และอาจารย์จะเป็นผู้ให้ความช่วยเหลือเพื่อให้ผู้เรียนเกิดการเรียนรู้ผ่านวงจรของกระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐานดังภาพที่ 1

วงจรของกระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐานประกอบด้วยรายละเอียดของกิจกรรม คือ (Hmelo-Silver, 2004) ขั้นตอนแรกผู้เรียนจะได้รับสถานการณ์ปัญหาที่มีความซับซ้อนโดยที่มีข้อมูลให้เพียงเล็กน้อย จากนั้นผู้เรียนจะต้องตั้งคำถามอาจารย์เพื่อที่จะได้รับข้อมูลเพิ่มเติมเกี่ยวกับปัญหานั้น และจะต้องคิดวิเคราะห์ ทบทวนและกำหนดประเด็นปัญหาโดยใช้การจำแนกจากข้อเท็จจริงต่างๆ ที่เกี่ยวข้องในสถานการณ์ ซึ่งขั้นตอนการจำแนกข้อเท็จจริงนี้จะช่วยในการกำหนดปัญหาและทำความเข้าใจปัญหาได้ดีขึ้น จากนั้นผู้เรียนจะต้องสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้ เพื่อที่จะแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูลที่เขาได้รวบรวมมา และส่วนที่สำคัญของวงจรคือ กลุ่มช่วยกันระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหาที่พวกเขายังขาดอยู่ ซึ่งจะต้องไปศึกษาค้นคว้าด้วยตนเองหรือมีการเรียนรู้แบบนำตนเองต่อไป โดยภายหลังจากการศึกษาค้นคว้าด้วยตนเองแล้วผู้เรียนจะกลับมารวมกลุ่มอีกครั้งเพื่อแลกเปลี่ยนสิ่งที่ได้ไปเรียนรู้มาและนำความรู้ใหม่ที่ได้อมาทดสอบสมมติฐานที่กำหนดไว้หรือทำการเปลี่ยนสมมติฐานขึ้นใหม่ตามความรู้ใหม่ที่ได้รับ ซึ่งในขั้นตอนสุดท้าย ผู้เรียนจะต้องมีการสะท้อนถึงบทเรียนที่ได้มาโดยสรุปเกี่ยวกับปัญหาในแต่ละประเด็นและผลจากการไปศึกษาเพิ่มเติมด้วยตนเองรวมถึงกระบวนการแก้ปัญหาของกลุ่ม ทั้งนี้ขณะที่กลุ่มกำลังระดมความคิดจะต้องมีการบันทึกข้อมูลบนกระดานหรือไวท์บอร์ด โดยบันทึกอาจทำเป็นตาราง 4 คอลัมน์ เพื่อสะดวกในการวิเคราะห์แก้ปัญหา ดังภาพที่ 2 เป็นตัวอย่างการบันทึกของนักศึกษาวิศวกรรม ในขณะที่วิเคราะห์สาเหตุที่ทำให้สารเคมีรั่วไหล (Hmelo-Silver et al, 1995 cited in Hmelo-Silver, 2004) ซึ่งในคอลัมน์ที่ 1 จะแสดงข้อมูลที่รวบรวมได้จากสถานการณ์ของปัญหา (facts) คอลัมน์ที่ 2 จะแสดงสมมติฐานที่ได้พัฒนาขึ้น (ideas) คอลัมน์ที่ 3 จะแสดงคำถามที่กลุ่มสร้างขึ้นเพื่อใช้สำหรับ

กำหนดประเด็นสำหรับนำไปศึกษา (learning issues) และคอลัมน์ที่ 4 จะเป็นการวางแผนเกี่ยวกับข้อมูลที่จะนำมาใช้แก้ไขปัญหา (action plan)



ภาพที่ 1 วงจรการเรียนรู้โดยใช้ปัญหาเป็นฐาน (Hmelo-Silver, 2004: 237)

ข้อมูล	สมมติฐาน	ประเด็นปัญหาสำหรับการเรียนรู้	แผนการปฏิบัติ
<ul style="list-style-type: none"> • อันตรายจากสารเคมี • ความใกล้ชิดกับศูนย์ประชากร 	<ul style="list-style-type: none"> • ลดขนาดของสถานที่ในการจัดเก็บ • จัดเตรียมการฝึกอบรมในด้านความปลอดภัย • พัฒนาระบบเตือนภัย 	<ul style="list-style-type: none"> • อะไรคือมาตรฐานความปลอดภัยเกี่ยวกับการจัดเก็บไซยาไนด์ • มีเทคโนโลยีอะไรที่ใช้เกี่ยวกับการเก็บสารเคมีที่อันตราย 	<ul style="list-style-type: none"> • มาตรฐานของ EPA

ภาพที่ 2 ตัวอย่างการบันทึกข้อมูลบนกระดานหรือไวท์บอร์ด (Hmelo-Silver, 2004: 237)

บทบาทของกลุ่มเป็นส่วนหนึ่งที่มีความสำคัญต่อการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งข้อตกลงเบื้องต้นข้อหนึ่งในการเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ โครงสร้างของกลุ่มต้องมีขนาดเล็กเพื่อที่จะช่วยให้เกิดการกระจายของความคิดและช่วยเหลือกันในการแก้ไขปัญหา ภายในกลุ่มจะมีการแบ่งประเด็นสำหรับการไปศึกษาค้นคว้าด้วยตนเองทำให้เกิดเป็นความเชี่ยวชาญในประเด็นนั้นๆ ซึ่งผลจากงานวิจัยพบว่า การอภิปรายโต้ตอบกันในกลุ่มขนาดเล็กจะเพิ่มการแก้ไขปัญหาและเกิดทักษะการคิดขั้นสูง (higher order thinking) และเป็นการส่งเสริมการแลกเปลี่ยนความรู้ (Blumenfeld et al., 1996; Brown, 1995; Vye et al., 1997 cited in Hmelo-Silver, 2004) ส่วนมากในการอภิปรายร่วมกันของสมาชิกจะต้องมีผู้ประสานงานกลุ่ม (facilitator) หรืออาจใช้วิธีการอื่น เช่น จัดกำหนดการ การแบ่งหน้าที่ การร่วมช่วยกันทุกฝ่าย เป็นต้น ซึ่งจะทำให้เกิดการจัดสรรการมีส่วนร่วมของสมาชิกทุกคนและเกิดความครอบคลุมของทุกประเด็นการศึกษา

บทบาทของอาจารย์ในการเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ การช่วยให้ผู้เรียนได้เรียนรู้ทักษะทางการคิดที่จำเป็นสำหรับการแก้ไขปัญหาและการทำงานร่วมกัน (Hmelo-Silver, 2004) นอกจากนี้กรณีที่ผู้เรียนเพิ่งจะเริ่มต้นในการเรียนรู้โดยใช้ปัญหาเป็นฐานอาจารย์ควรจัดช่วงเวลาสำหรับแนะนำเกี่ยวกับหลักสูตร การบริหารจัดการสำหรับกลุ่ม วงจรของกระบวนการเรียนรู้ ทักษะที่ช่วยให้ประสบความสำเร็จในการเรียนรู้โดยใช้ปัญหาเป็นฐาน และควรมีตัวอย่างให้เห็นหรือลองทำ ทั้งนี้อาจแบ่งบทบาทหน้าที่ของอาจารย์เป็น 3 ระยะ ได้แก่ (Bridges, 1992)

1) ก่อนเริ่มสถานการณ์ปัญหา (prior) อาจารย์จะมีกิจกรรมที่สำคัญ คือ การทบทวนกระบวนการสำหรับการเรียนรู้โดยใช้ปัญหาเป็นฐานสำหรับกรณีที่คุณเรียนยังไม่คุ้นเคย การรวบรวมแหล่งข้อมูลตลอดจนเครื่องมือและวัสดุอุปกรณ์ที่จำเป็นต้องใช้สำหรับการเรียนรู้โดยอาจทำเป็นรายการตรวจสอบ (checklist) และการช่วยในการกำหนดบทบาทหน้าที่ของผู้เรียนภายในกลุ่ม เช่น ผู้นำกลุ่ม ผู้ประสานงาน และผู้จัดบันทึก

2) ระหว่างการทำกิจกรรม (during) อาจารย์จะเป็นผู้สังเกตรูปแบบการมีส่วนร่วมของสมาชิกในกลุ่มโดยพิจารณาจาก 3 ประเด็น ได้แก่ ประเด็นแรก คือ เนื้อหาที่พูด ซึ่งจะแสดงให้เห็นว่าผู้เรียนมีความเข้าใจหรือมีความรู้จากการที่ได้ไปศึกษาค้นคว้าด้วยตนเองมา ประเด็นที่สอง คือ กระบวนการ ซึ่งจะพิจารณาว่าใครเป็นคนพูด มีการแสดงความคิดเห็นอย่างไร และมีความเกี่ยวข้องกับเนื้อหาหรือไม่ และประเด็นที่สาม คือ ความถี่ ซึ่งจะพิจารณาว่าการมีส่วนร่วมในกลุ่มของผู้เรียน มีมากน้อยเพียงไร ทั้งนี้อาจารย์อาจต้องมีการแทรกแซงกลุ่มถ้าผู้เรียนนำเสนอข้อมูลที่ไม่ตรงกับเนื้อหา และในกรณีที่ต้องการตั้งคำถามให้เกิดการคิดในประเด็นที่สำคัญหรือเพื่อตรวจสอบการรับรู้ของผู้เรียน ตลอดจนการให้คำแนะนำเพื่อปรับปรุงกระบวนการในกลุ่ม และในระหว่างที่ทำกลุ่มอาจารย์อาจให้ความรู้แก่ผู้เรียน

ในกรณีที่ผู้เรียนต้องการให้อาจารย์ช่วยอธิบายในประเด็นที่ยังไม่เข้าใจ หรืออาจให้ช่วยเป็นแหล่งข้อมูล เหมือนกับการเป็นผู้เชี่ยวชาญ นอกจากนี้อาจารย์ต้องมีการให้กำลังใจผู้เรียนเพื่อเป็นการสนับสนุนเขา หรือในกรณีที่ผู้เรียนบางคนที่ขาดความมั่นใจในตนเองอาจารย์จะต้องไม่กดดันแต่ควรสร้างบรรยากาศ แห่งการช่วยเหลือภายในกลุ่มเพื่อให้ผู้เรียนรู้สึกผ่อนคลายใน การพูดปัญหา ที่เกิดขึ้นของเขาและอาจชี้ ว่าความล้มเหลวที่เกิดจะเป็นโอกาสที่ดีสำหรับการเรียนรู้

3) ช่วงท้าย (following) ในช่วงสิ้นสุดของการทำกิจกรรมอาจารย์ต้องทำหน้าที่สำคัญ ได้แก่ การชักชวนให้ผู้เรียนพูดสะท้อนความคิดเห็นเกี่ยวกับสิ่งที่ตนเองได้ทำและแนวทางในการพัฒนา ตนเองต่อไป การให้ผู้เรียนเขียนแสดงความคิดเห็นเพื่อเป็นข้อมูลสะท้อนกลับเกี่ยวกับการกระทำและ ผลที่เกิดขึ้นทั้งของตนเองและเพื่อนในกลุ่ม และการตรวจสอบข้อมูลสะท้อนกลับของผู้เรียนและใช้ ข้อมูลเหล่านี้เพื่อการแก้ไขและปรับปรุงให้ดีขึ้น

กระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐานสามารถสรุปได้เป็น 9 ขั้นตอน ได้แก่ 1) การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ 2) การกำหนดประเด็นปัญหา 3) การสร้าง สมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้ จากความรู้และประสบการณ์ที่มีอยู่ 4) การสรุป เป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล 5) การระบุประเด็นสำหรับการเรียนรู้เพื่อ ใช้ในการแก้ไขปัญหา 6) การศึกษาค้นคว้าด้วยตนเอง 7) การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้ ไปเรียนรู้มาในกลุ่ม 8) การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตาม ความรู้ใหม่ที่ได้รับ และ 9) การสะท้อนความรู้ในบทเรียน โดยทำการสรุปทั้งในด้านประเด็นปัญหา ผลจากการไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน และกระบวนการในการแก้ปัญหา

1.3 การวัดและประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐาน

การวัดและประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานในทางปฏิบัติสามารถประเมินได้หลาย รูปแบบ โดยรวมอยู่ในการประเมินเพื่อปรับปรุง (formative evaluation) และการประเมินเพื่อ ตัดสิน (summative evaluation) และสถานะของผู้ประเมินอาจเป็นการประเมินตนเองของผู้เรียน (self-assessment) การประเมินโดยเพื่อน (peer assessment) และการประเมินโดยอาจารย์ (instructor assessment)

1.3.1 แนวคิดทางการวัดและประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐาน

รูปแบบการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานที่ผ่านมามักจะเน้นประเมินในด้านสภาพแวดล้อมของการทำงานร่วมกัน (cooperative environment) สภาพแวดล้อมที่เอื้อให้เกิดการเรียนรู้ (supportive environment) กระบวนการเรียนรู้ (learning process) ลักษณะบุคลิกภาพ (personal characteristics) และกลไกการทำกลุ่ม (group dynamics) เพื่อหลีกเลี่ยงปัญหาการแข่งขันชิงเกรด ซึ่ง Norman ในปีค.ศ. 1991 (Norman, 1991 cited in Eva, 2001) กล่าวว่า การประเมินโดยใช้วัตถุประสงค์เชิงพฤติกรรมเป็นหลักถือเป็นแนวคิดที่ขัดแย้งกับการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการเรียนรู้แบบนำตนเองเนื่องจากการประเมินรูปแบบนี้จะสนับสนุนให้เกิดการจัดสอบซึ่งไม่ได้เป็นการสนองตอบต่อวัตถุประสงค์ในการเรียนรู้โดยใช้ปัญหาเป็นฐาน แต่รูปแบบการประเมินที่น่าจะเหมาะสมกับการเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ การประเมินภายในบริบทของการปฏิบัติ (Norman, 1988 cited in Eva, 2001) เนื่องจากการเรียนรู้โดยใช้ปัญหาเป็นฐานได้ถูกออกแบบมาเพื่อเน้นในเรื่องของการวิเคราะห์และแก้ไขปัญหา (Barrows and Tamblyn, 1976 cited in Eva, 2001) ดังนั้นเครื่องมือสำหรับใช้ในการประเมินควรจะมุ่งเน้นในด้านทักษะ กระบวนการ และทัศนคติของผู้เรียนแต่ละคนภายในบริบทของบทเรียนหรือสถานการณ์นั้นๆ ซึ่งรูปแบบการประเมินที่กล่าวมาจึงตรงกับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (tutorial-based assessment) หรือประเมินจากการปฏิบัติที่เกิดขึ้นจริงในบริบทของการเรียนรู้ (Hay, 1995 cited in Eva, 2001) แต่จากผลการศึกษาที่ไม่ได้ตีพิมพ์เผยแพร่ของ Black, Norman, & Mueller ใน McMaster University ปีค.ศ. 1994 พบว่า การประเมินการเรียนรู้ดังกล่าวไม่มีความไวในการวัดการความรู้ของผู้เรียน (acquisition of knowledge) (Lee & Kwan, 1997)

การประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานในปัจจุบันนิยมแบ่งเป็น 2 ประเภท ได้แก่ 1) การประเมินเพื่อปรับปรุง (formative evaluation) คือ การประเมินในด้านทักษะการปฏิบัติระหว่างการทำกลุ่ม ผลลัพธ์ที่เกิดขึ้นกับผู้เรียนหรือกลุ่ม และการเรียนรู้ที่เกิดขึ้นตามวัตถุประสงค์ โดยเป็นการประเมินแบบใช้เวลาสั้นๆ เพื่อให้ข้อมูลสะท้อนกลับซึ่งจะเป็นประโยชน์สำหรับการปรับปรุงแก้ไขในครั้งต่อไปและควรทำการประเมินทุกครั้งหลังเสร็จสิ้นในแต่ละบทเรียนหรือสถานการณ์ และ 2) การประเมินเพื่อตัดสิน (summative evaluation) คือ การประเมินเมื่อสิ้นสุดรายวิชาโดยผลการประเมินอาจเป็นแบบผ่านหรือตก ทั้งนี้สถานะของผู้ประเมินอาจเป็นการประเมินตนเองของผู้เรียน (self-assessment) การประเมินโดยเพื่อน (peer assessment) และการประเมินโดยอาจารย์ (instructor assessment) (Bridges, 1992; Walsh, 2005) รูปแบบของการประเมินกระบวนการเรียนรู้ในกลุ่ม

ย่อยจัดว่ามีบทบาทสำคัญและอาจเป็นรูปแบบที่ดีที่สุดสำหรับใช้ประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐาน เนื่องจากช่วยให้ข้อมูลที่สามารถนำมาใช้ในการประเมินเพื่อปรับปรุง (formative evaluation) ได้ดี แต่ทั้งนี้อาจพิสูจน์ได้ยากกว่าเป็นการบรรลุตามเป้าหมายของการเรียนหรือไม่ ซึ่งต่างจากรูปแบบของการจัดสอบที่แสดงให้เห็นถึงการวัดในเรื่องความรู้ที่ชัดเจน นอกจากนี้การประเมินกระบวนการเรียนรู้อาจใช้ในการประเมินความสามารถด้านอื่นๆ เช่น การให้เหตุผล การเรียนรู้แบบนำตนเอง การสื่อสาร แรงจูงใจภายใน การคิดอย่างมีวิจารณญาณ และความสามารถทำงานร่วมกัน เป็นทีม แต่จุดอ่อนของการประเมินตนเองและการประเมินโดยเพื่อนก็ยังเป็นเรื่องที่ได้แย้งกันอยู่ เนื่องจากการประเมินตนเองของผู้เรียนอาจไม่ได้บ่งชี้ถึงความสามารถที่แท้จริง และเป็นเรื่องปกติที่ความสัมพันธ์ของบุคคลจะทำให้เกิดมุมมองด้านบวกซึ่งจะทำให้เกิดการล่งเลในการให้คะแนนตามวัตถุประสงค์การประเมิน แต่ในทางตรงข้ามข้อมูลเหล่านี้จะมีประโยชน์ในด้านการใช้เป็นข้อมูลสะท้อนกลับเพื่อให้เกิดการพัฒนาขึ้น (Eva, 2001; Walsh, 2005)

1.3.2 หลักการปฏิบัติในการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐาน

Moore & Poikela (Moore & Poikela, 2011) ได้กล่าวถึงหลักหรือองค์ประกอบที่สำคัญในการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐาน 9 ข้อ ได้แก่ 1) การประเมินควรสอบถามข้อมูลจากอาจารย์ นักศึกษา และผู้ที่เกี่ยวข้องเพื่อให้ได้ข้อมูลเกี่ยวกับประสิทธิภาพของการเรียน โดยเฉพาะนักศึกษาควรจะมีบทบาทเชิงรุกในการประเมินเกี่ยวกับพฤติกรรม ทักษะ ตลอดจนการเรียนรู้ที่เกิดขึ้นกับตนเอง 2) ทักษะสำคัญที่ทำให้เกิดการเรียนรู้ที่มีประสิทธิภาพ คือ การประเมินตนเอง (self-evaluation) และการสะท้อนความคิด (critical reflection) ซึ่งทั้งสองทักษะนี้เกิดขึ้นในการเรียนรู้โดยใช้ปัญหาเป็นฐาน และในการประเมินด้านกระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐานไม่ควรจะมุ่งเน้นที่การได้ทำหรือไม่ได้ทำเนื่องจากไม่ช่วยให้เกิดการเสริมพลังแก่ผู้เรียน แต่ควรให้ความสำคัญในการประเมินเกี่ยวกับกับกลยุทธ์และผลลัพธ์ที่เกิดขึ้น 3) การประเมินควรจะให้ผู้ที่มีส่วนได้ส่วนเสียได้มีส่วนร่วมในการประเมินอย่างครอบคลุมและควรมีการแลกเปลี่ยนความรู้ที่เกิดขึ้นจากการประเมิน 4) การประเมินควรแยกระหว่างการประเมินเพื่อปรับปรุง (formative evaluation) และการประเมินเพื่อตัดสิน (summative evaluation) โดยการประเมินเพื่อปรับปรุงควรให้ความสำคัญกับกระบวนการเรียน และการประเมินเพื่อตัดสินควรให้ความสำคัญกับผลกระทบและประสิทธิภาพที่เกิดขึ้นจากการเรียน 5) การประเมินโดยมุ่งเน้นผลลัพธ์ควรแยกจากการติดตามพฤติกรรมในการเรียน ความก้าวหน้าของกระบวนการ และบทบาทหน้าที่ในการเรียน ซึ่งการประเมินโดยมุ่งเน้น

ผลลัพธ์จะช่วยในการจำแนกลักษณะของการเปลี่ยนแปลงที่เกิดขึ้นจริงกับสิ่งที่เป็นมุมมองของผู้เรียนได้ 6) การประเมินควรเป็นไปอย่างเปิดเผยและซื่อสัตย์ซึ่งทำได้โดยทุกคนในกลุ่มมีการปฏิบัติตามกฎหรือมีความคุ้นเคยในการปฏิบัติตามระเบียบข้อตกลงร่วมกัน อันจะทำให้การประเมินได้ข้อมูลจริงที่เป็นประโยชน์ต่อการปรับปรุง 7) การประเมินที่มีประสิทธิภาพจะให้ข้อมูลป้อนกลับที่ต่อเนื่องเป็นประโยชน์ต่อการพัฒนา ปรับปรุง และช่วยในการตัดสินใจ นอกจากนี้การประเมินควรจะแทรกสอดและรวมอยู่ในทุกๆ กระบวนการของการเรียนรู้โดยใช้ปัญหาเป็นฐาน 8) สัดส่วนของประเด็นที่ใช้ในการประเมินควรสอดคล้องกับกิจกรรมที่เกิดขึ้นจริง และ 9) การประเมินควรเป็นไปอย่างมีจรรยาบรรณและสุภาพ

1.3.3 การประเมินกระบวนการเรียนรู้ในกลุ่มย่อย

ในงานวิจัยเกี่ยวกับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยผู้ที่ทำหน้าที่ประเมินส่วนมากจะเป็นผู้เรียนที่ทำการประเมินตนเองเนื่องจากผู้เรียนมีบทบาทเชิงรุก (active role) ในกระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐาน รวมถึงอาจารย์และเพื่อนควรมีบทบาทในการร่วมประเมินด้วยเนื่องจากเป็นผู้ที่ทราบดีถึงจุดแข็งและจุดอ่อนของสมาชิกในกลุ่มจึงสามารถที่จะสะท้อนข้อมูลได้อย่างถูกต้องและสร้างสรรค์ (Van Rosendaal and Jennett, 1992 cited in Eva, 2001) ซึ่งรายละเอียดของการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยโดยผู้ประเมินทั้ง 3 กลุ่มมีดังนี้

1) การประเมินโดยอาจารย์ มีงานวิจัยจำนวนมากเกี่ยวกับการประเมินตนเองและการประเมินโดยเพื่อนที่ใช้รูปแบบของการให้คะแนนโดยอาจารย์เป็นมาตรฐานต้นแบบ (gold standard) ซึ่งจุดแข็งของการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย คือ การประเมินที่อยู่บนฐานของการปฏิบัติที่เกิดขึ้นจริงในบริบทของการเรียนรู้ ดังนั้นจึงมองว่าการประเมินรูปแบบนี้จะประเมินความสามารถของผู้เรียนได้ถูกต้องกว่าการประเมินโดยการจัดสอบ แต่ทั้งนี้ก็ยังขาดหลักฐานที่จะพิสูจน์ว่าผู้เรียนมีการเรียนรู้ตรงตามวัตถุประสงค์หรือไม่ ซึ่ง Graesser & Person (Graesser & Person, 1994 cited in Eva, 2001) ได้อธิบายว่า เนื่องจากคำถามที่ใช้ประเมินกระบวนการเรียนรู้มักจะเป็นแบบปลายเปิด จึงทำให้อาจารย์ไม่สามารถประเมินหรือเปรียบเทียบระดับความเข้าใจของผู้เรียนได้เป็นรายคน (individual assessment) ดังนั้นการแก้ปัญหาจึงควรอยู่ที่เครื่องมือที่จะใช้ในการประเมิน ทั้งนี้ Hebert & Bravo (Hebert & Bravo, 1996 cited in Eva, 2001) ได้เป็นผู้เริ่มออกแบบเครื่องมือมาตรฐานที่ใช้ประเมินทัศนคติและทักษะที่เป็นผลจากการเรียนรู้โดยใช้ปัญหาเป็นฐานของนักศึกษาแพทย์ จำนวน 273 คน

ประกอบด้วยข้อคำถามจำนวน 44 ข้อ ค่าความเที่ยงของเครื่องมือเท่ากับ 0.86 ค่าสหสัมพันธ์ของคะแนนจากแบบวัดนี้กับ tutor's global evaluation มีค่า $r = 0.64$ และค่าสหสัมพันธ์ของคะแนนจากแบบวัดนี้กับผลการสอบข้อเขียนของผู้เรียน มีค่า $r = 0.39$ ผลการวิเคราะห์องค์ประกอบ (factor analysis) สามารถจำแนกความแตกต่างของผู้เรียนได้ 4 องค์ประกอบ คือ ประสิทธิภาพ (effectiveness) การสื่อสารและภาวะผู้นำ (communication and leadership) ความสนใจทางวิทยาศาสตร์ (scientific curiosity) และความเคารพต่อสมาชิกในกลุ่ม (respect for peers) เช่นเดียวกับ Valle et al. และ Des Marchais & Vu (Valle et al., 1999; Des Marchais & Vu, 1996 cited in Eva, 2001) ที่พัฒนาเครื่องมือสำหรับประเมินการปฏิบัติของผู้เรียนระหว่างกระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐาน ทั้งนี้ Des Marchais & Vu ได้กล่าวว่าทั้งผู้เรียนและอาจารย์ต่างมีความลังเลที่จะใช้รูปแบบการให้คะแนนโดยอาจารย์ (tutor rating form) แต่ในขณะนั้นก็ยังไม่มียุทธศาสตร์ที่มีความตรงสำหรับการใช้ในการประเมินความสามารถของผู้เรียน ดังนั้นจึงมีการนำรูปแบบการให้คะแนนโดยอาจารย์มาใช้แทน ซึ่ง Eva (Eva, 2001) ได้กล่าวสรุปว่า ถึงแม้การให้คะแนนในกระบวนการเรียนรู้โดยอาจารย์จะมีประโยชน์ในด้านการประเมินทักษะของผู้เรียน แต่ก็เป็นเรื่องยากที่จะอ้างอิงถึงความน่าเชื่อถือของการประเมิน เช่น การเข้าถึงแหล่งข้อมูลที่ใช้ประเมิน จึงจำเป็นที่จะต้องมียุทธศาสตร์มาตรฐานสำหรับใช้ประเมิน และข้อดีอีกด้านของการประเมินโดยใช้เครื่องมือมาตรฐาน คือ มีค่าใช้จ่ายน้อยกว่าการประเมินโดยการจัดสอบ แต่ทั้งนี้ในการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความตรงของเครื่องมืออาจเป็นเพียงวิธีการหนึ่งซึ่งไม่ได้ครอบคลุมปัจจัยอื่นๆ ที่อาจส่งผลต่อความน่าเชื่อถือของการประเมินในรูปแบบนี้ ดังนั้นการประเมินโดยเพื่อน (peer assessment) ซึ่งมีการให้ข้อมูลสะท้อนกลับทุกๆ ไป อาจมีประโยชน์สำหรับใช้เปรียบเทียบความแม่นยำและความครอบคลุมของการประเมินโดยอาจารย์ได้

2) การประเมินโดยเพื่อน ในขณะที่รูปแบบของการให้คะแนนโดยอาจารย์ได้รับการคาดหวังว่ามีความตรงเพราะเป็นผลมาจากปฏิสัมพันธ์ระหว่างผู้เรียนและอาจารย์ Van Rosendaal & Jennett (Van Rosendaal & Jennett, 1994 cited in Eva, 2001) ได้โต้แย้งว่า การประเมินโดยเพื่อนจะมีความครอบคลุมมากกว่าทั้งในด้าน ระยะเวลา ความใกล้ชิด ความต่อเนื่อง และเป็นข้อมูลที่มีการจัดฉากของผู้ถูกประเมินน้อยกว่า นอกจากนี้ยังมีข้อดีของการเป็นคะแนนเฉลี่ยที่มาจากความคิดเห็นของผู้ประเมินจำนวนหลายคนทำให้คะแนนมีลักษณะของการประมาณค่าที่มีความคงที่มากกว่า (stable estimate) และจากการที่มีลักษณะเป็นคะแนนรวม (global ratings) จึงทำให้การประเมินความสามารถมีความน่าเชื่อถือเมื่อมีจำนวนผู้ประเมินที่มากพอ (Carline et al., 1992 cited in

Eva, 2001) ซึ่งการเพิ่มจำนวนอาจารย์อาจเป็นเรื่องที่ยากและมีค่าใช้จ่ายมากดังนั้นการให้คะแนนโดยเพื่อนอาจเป็นทางเลือกที่ดีในการเพิ่มจำนวนผู้ประเมิน แต่ทั้งนี้ก็ต้องคำนึงถึงผลในทางลบของการประเมินโดยเพื่อน เช่น ผลการศึกษาเปรียบเทียบความสัมพันธ์ระหว่างการให้คะแนนโดยเพื่อนและโดยอาจารย์ พบว่ามีความสัมพันธ์กันตั้งแต่ระดับต่ำจนถึงระดับปานกลาง (Burnett & Cavaye 1980; Van Rosendaal & Jennett, 1992; Pond & ul-Haq, 1997; Sullivan, 1999; Rudy et al. 2001; Minion et al., 2002; Reiter et al., 2002 cited in Papinczak et al., 2007) ตลอดจนมีการศึกษาพบว่า มีหลายปัจจัยที่ส่งผลเชิงลบต่อความแม่นยำในการประเมินโดยเพื่อน ซึ่งได้แก่สัมพันธภาพของบุคคล ความโดดเด่นในการช่วยเหลือกลุ่ม เป็นต้น รวมถึงการศึกษาของ Van Rosendaal & Jennett (Van Rosendaal & Jennett, 1992 cited in Eva, 2001) ได้รายงานว่าการแพทย์ประจำบ้านสาขาอายุรกรรมมีความเห็นด้วยเป็นอย่างยิ่งว่าการประเมินโดยเพื่อนอาจเป็นสาเหตุหลักของการแทรกแซงความสัมพันธ์ระหว่างเพื่อน แต่รายงานนี้ก็ยังมีจุดอ่อนในแง่ที่ไม่ได้คำนึงถึงประเด็นที่มีความอ่อนไหว เช่น การประเมินการขาดประสิทธิภาพของผู้ร่วมงานย่อมจะส่งผลต่อความสัมพันธ์ระหว่างเพื่อนได้ และสอดคล้องกับข้อมูลที่แพทย์ประจำบ้านสาขาอายุรกรรมจำนวน 89% มองว่าการประเมินเป็นเรื่องของการใช้อำนาจมากกว่าจะมองว่าเป็นการประเมิน แต่ทั้งนี้งานวิจัยส่วนมากจะจำกัดอยู่ที่บริบทของการเรียนรู้โดยใช้ปัญหาเป็นฐานของนักศึกษาแพทย์ซึ่งในบริบทของการเรียนการสอนแพทย์จะมีลักษณะของความสูงสำหรับการมีส่วนร่วมได้ส่วนเสียที่สูง (high stake setting) ดังนั้นจึงอาจพบได้ว่าเป็นเรื่องปกติที่การให้คะแนนโดยเพื่อนจะมีลักษณะของคะแนนที่สูง (Papinczak et al., 2007)

ถึงแม้ว่าการประเมินโดยเพื่อนจะยังคงมีจุดอ่อนอยู่แต่ก็มีหลักฐานเชิงประจักษ์ที่พบว่าการให้คะแนนโดยเพื่อนมีความเที่ยงและความตรงค่อนข้างสูง ดังตัวอย่างงานวิจัยของ Korman & Stubblefield (Korman & Stubblefield, 1971 cited in Eva, 2001) ศึกษาตัวแปรที่ใช้ในการทำนายความสำเร็จของการฝึกปฏิบัติของนักศึกษาแพทย์ปีสุดท้าย (internship success) ซึ่งได้แก่ เกรดเฉลี่ยความสามารถในช่วงเรียน และการให้คะแนนจากเพื่อน พบว่า ตัวแปรทั้งสามมีความสัมพันธ์กันในระดับต่ำถึงปานกลาง คือ 0.3 – 0.4 ซึ่งก็ไม่สามารถอธิบายความแปรปรวนส่วนที่เหลืออยู่ได้ แต่ก็พบว่าการให้คะแนนโดยเพื่อนจัดเป็นตัวทำนายที่ดีที่สุด Sullivan et al. (Sullivan et al., 1999 cited in Eva, 2001) ได้ทำการสังเคราะห์ข้อมูลการให้คะแนนโดยเพื่อน พบว่า 1) เป็นตัวทำนายความสามารถในอนาคตได้ดี 2) มีความเที่ยงภายในสูง (internal consistency) และ 3) ให้ข้อมูลสารสนเทศที่ต่างจากวิธีการประเมินแบบเดิม นอกจากนี้ยังพบว่าการให้คะแนน โดยอาจารย์และเพื่อนมีความสัมพันธ์

กันสูงในบริบทของการสอบเพื่อรับใบอนุญาตเฉพาะทางของแพทย์ (National Board of Medical Examiners: NBME) (Eva, 2001) อย่างไรก็ตามจากงานวิจัยที่กล่าวมาจะพบว่า การประเมินโดยเพื่อนจะไม่ถูกใช้สำหรับวัตถุประสงค์ในการประเมินเพื่อตัดสินผู้เรียนเป็นรายบุคคล (summative evaluation) ซึ่งในเรื่องนี้ Hay (Hay, 1995 cited in Eva, 2001) ได้กล่าวไว้ว่า การประเมินโดยเพื่อนมีแนวโน้มว่าจะมีการให้คะแนนสูงกว่าค่าเฉลี่ยเนื่องจากจะส่งผลโดยตรงต่อเกรด แต่เมื่อลองพิจารณาในทางกลับกันว่า ถ้าการให้คะแนนโดยเพื่อนมีความสัมพันธ์กับความสามารถในอนาคตของผู้ถูกประเมิน ดังนั้นการให้คะแนนที่สูงกว่าค่าเฉลี่ยก็ควรมีความสัมพันธ์ในระดับที่ต่ำจนถึงสูงกับความสามารถในอนาคตของผู้ถูกประเมินด้วยเช่นกัน แต่การประเมินผู้เรียนโดยมีการจำแนกความสามารถระดับต่ำยังคงมีประโยชน์ในระบบการประเมินแบบอิงกลุ่มเนื่องจากจะช่วยให้การพิจารณาหาทางแก้ไขปรับปรุงผู้เรียนให้ดีขึ้น นอกจากนี้การประเมินโดยเพื่อนยังมีประเด็นที่สำคัญคือ ความสามารถในการประเมินเพื่อนได้อย่างมีความเที่ยงและความตรงของผู้เรียนจะส่งเสริมให้ผู้เรียนนั้นมีทักษะและความสามารถที่จะประสบความสำเร็จต่ออาชีพในอนาคต และรวมถึงการมีความสามารถด้านการประเมิน แต่มีประเด็นคำถามอยู่ว่าความสามารถในการประเมินผู้อื่นจะสามารถถ่ายโอนไปยังการประเมินตนเองได้หรือไม่

3) การประเมินโดยตนเอง การประเมินตนเองจัดเป็นเรื่องที่สำคัญมากที่สุดในบริบทของการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยเพราะไม่ได้เป็นเพียงแค่รูปแบบหนึ่งของการประเมินเท่านั้น แต่ยังเป็นทักษะและความสำคัญอย่างหนึ่งของบุคคลซึ่งทักษะนี้จะพัฒนามาจากการฝึกฝนอบรม นอกจากนี้การที่ผู้เรียนได้มีการพัฒนาความสามารถในการสะท้อนข้อมูลเกี่ยวกับจุดแข็งและจุดอ่อนของตนก็เป็นส่วนสำคัญของการเรียนรู้แบบนำตนเอง (Eva et al. 2004 cited in Papinczak et al., 2007; 122) มีผลการศึกษาเกี่ยวกับการประเมินตนเองจำนวนมากที่พบความแตกต่างกัน เช่น ผลการศึกษาวิเคราะห์อิทธิพลของการประเมินตนเองในระดับอุดมศึกษาซึ่งพบว่าการประเมินตนเองมีความแม่นยำในระดับดี (Sluijmans et al. 1999 cited in Papinczak et al., 2007; 123) การให้คะแนนตนเองมีแนวโน้มที่จะมีค่าความสัมพันธ์ระดับต่ำกับการให้คะแนนโดยอาจารย์หรือเพื่อน (Papinczak et al., 2007; 128; Sclabassi & Woelfel, 1984; Morton and MacBeth, 1977 cited in Eva, 2001) การให้คะแนนตนเองมีความน่าเชื่อถือน้อยกว่าผู้บังคับบัญชา (Gordon, 1991 cited in Eva, 2001) การศึกษาความสัมพันธ์ระหว่างคะแนนจากการประเมินตนเองกับผลการสอบพบว่ามีนัยสำคัญทางสถิติ (Tousignant & Des Marchais 2002; Eva et al. 2004 cited in Papinczak et al., 2007) และรวมถึงหลักสูตรการศึกษาของแพทย์ที่ไม่มีการเรียนรู้โดยใช้ปัญหา

เป็นฐานก็พบว่า การประเมินตนเองและการประเมินโดยอาจารย์หรือผลการสอบพบว่ามีความสัมพันธ์ในระดับต่ำ (Arnold et al. 1985; Woolliscroft et al. 1993; Rudy et al. 2001; Fitzgerald et al. 2003 cited in Papinczak et al., 2007) ยกเว้นผลการศึกษาของ Burnett and Cavaye (Burnett and Cavaye, 1980 cited in Papinczak et al., 2007) ที่ศึกษาจากนักศึกษาแพทย์ปี 5 ซึ่งไม่ได้เรียนหลักสูตรการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยพบว่ามีความสัมพันธ์ที่สูงระหว่างคะแนนจากการประเมินตนเองกับคะแนนที่ประเมินโดยเพื่อน ส่วนการศึกษาเกี่ยวกับตัวแปรความแตกต่างระหว่างเพศก็พบว่า มีผลเพียงเล็กน้อยต่อความแม่นยำทั้งของการประเมินตนเองและการประเมินโดยเพื่อน (Falchikov & Magin 1997; Mattheos et al. 2004 cited in Papinczak et al., 2007)

ในด้านการวิจัยที่นำอุปกรณ์เทคโนโลยีมาช่วยในการประเมิน เช่น การประเมินความสามารถ ภายหลังการปฏิบัติโดยใช้การบันทึกวิดีโอ ก็พบว่า ผู้เรียนมีความสามารถในการประเมินตนเองเพิ่มขึ้นเล็กน้อยเมื่อเปรียบเทียบกับ การประเมินโดยเพื่อน (Palmer et al., 1985 cited in Eva, 2001) หรือในกรณีที่ผู้เรียนดูเหมือนจะมีความสามารถในการประเมินความสามารถที่สอดคล้องกับความคาดหวังของอาจารย์ แต่เมื่อปฏิบัติกลับพบว่าผู้เรียนมีความสามารถในการประเมินตนเองได้น้อยลงเมื่อเปรียบเทียบกับ การประเมินเพื่อน ทำให้เกิดความสงสัยว่าผู้เรียนอาจมีความเข้มงวดหรือปล่อยคะแนนตนเองมากไปเมื่อเปรียบเทียบกับอาจารย์ (Calhoun et al., 1988 cited in Eva, 2001) และจากการศึกษาอื่นๆ พบว่าผู้เรียนมีความลังเลในการให้คะแนนตนเองสูงทั้งที่สมควรจะได้รับ (Stuart et al., 1980 cited in Eva, 2001) ซึ่งผลการศึกษาเหล่านี้ได้สร้างทั้งความแปลกใจและความยุ่งยากในการตั้งสมมติฐานเบื้องต้นที่ว่า ผู้เรียนแต่ละคนไม่มีข้อบกพร่องตามความเข้าใจของตนเอง ทั้งที่ความเข้าใจนี้วัดจากการบันทึกรวบรวมของสิ่งที่ผู้เรียนมีความบกพร่อง ทั้งนี้การอธิบายที่มีความเป็นไปได้อย่างหนึ่ง คือ ผู้เรียนต้องการเวลาสำหรับการพัฒนาทักษะในการประเมินตนเอง และการพัฒนานี้จะเกิดขึ้นจากการได้รับประสบการณ์เท่านั้น ดังนั้นความสัมพันธ์ระหว่างการให้คะแนนตนเองและการให้คะแนนโดยอาจารย์อาจถูกคาดหวังว่าจะมีมากขึ้น เมื่อผู้เรียนมีประสบการณ์ซึ่งได้รับการให้ข้อมูลสะท้อนกลับของอาจารย์ Das et al.'s (Das et al.'s, 1998 cited in Eva, 2001) ได้สนับสนุนในสมมติฐานนี้โดยกล่าวว่า 80% ของผู้เรียนและ 70% ของอาจารย์มีความเชื่อว่าการเปรียบเทียบผลการประเมินโดยอาจารย์กับผู้เรียนถือสิ่งที่ควรพิจารณาซึ่งจะนำไปสู่การพัฒนาทักษะการประเมิน นอกจากนี้ยังมีผลการศึกษาเชิงประจักษ์ส่วนหนึ่งที่พบว่า การประเมินตนเองมีแนวโน้มที่จะมีความสัมพันธ์น้อยกับการประเมินแบบอื่นๆ ภายในบริบทของการประเมินที่มุ่งเน้นกระบวนการเรียนรู้ ยกเว้นมีการจัดฝึกอบรมที่ชัดเจนเกี่ยวกับเกณฑ์การประเมิน รวมถึงการพัฒนา

การรับรู้ความสามารถของตนเอง (self-efficacy) (Eva, 2001; Papinczak et al., 2007) และมีการศึกษาหลายเรื่องที่เกี่ยวข้องกับวิธีแก้ปัญหาความแม่นยำของการประเมินโดยใช้การประเมินร่วมกัน (co-assessment) ระหว่างอาจารย์ เพื่อน และผู้เรียน (Sluijmans et al. 1999 cited in Papinczak et al., 2007)

การประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (tutorial-based assessment) เป็นแนวคิดหลักที่ใช้ในการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐาน เนื่องจากปฏิสัมพันธ์ระยะยาวระหว่างอาจารย์ เพื่อน และผู้เรียน เป็นแหล่งข้อมูลที่สำคัญที่ช่วยแนะแนวการตัดสินใจในเรื่องความสามารถของผู้เรียน และการประเมินที่มุ่งเน้นกระบวนการเรียนรู้ยังช่วยทำให้เกิดการสะท้อนกลับของข้อมูลเกี่ยวกับหลักสูตรเพื่อให้การปรับปรุงพัฒนา อีกทั้งอาจารย์ยังสามารถนำข้อมูลที่ได้จากการประเมินมาใช้พัฒนาการสอนหรือความเหมาะสมของบทเรียน ซึ่งถือเป็นการประเมินเพื่อปรับปรุง (formative evaluation) ส่วนปัญหาด้านคุณสมบัติทางจิตมิติของการประเมินที่มุ่งเน้นกระบวนการเรียนรู้ที่พบว่า ขาดความเที่ยงและความตรงในส่วนของประเมินโดยเพื่อนและตนเอง ได้มีงานวิจัยบางส่วนที่แนะนำให้แก้ไขโดยใช้เทคนิคของการประเมิน ได้แก่ การฝึกอบรมการใช้เกณฑ์การประเมิน และการสร้างเครื่องมือวัดที่ตรงตามคุณลักษณะที่จะใช้ประเมิน (Eva, 2001) แต่ทั้งนี้ก็ยังไม่มีงานวิจัยใดที่มีการใช้หรือการเสนอแนะให้วิเคราะห์ความแม่นยำจากผู้ประเมิน ตลอดจนยังไม่มีงานวิจัยใดที่ทำการศึกษารวสอบยืนยันโมเดลปัจจัยเชิงสาเหตุของความแม่นยำในการให้คะแนนโดยวิเคราะห์ปัจจัยต่างๆ พร้อมกัน

ตอนที่ 2 องค์ความรู้เกี่ยวกับคุณภาพการให้คะแนนการปฏิบัติ

องค์ความรู้เกี่ยวกับคุณภาพการให้คะแนนการปฏิบัติ (performance rating) ในที่นี้จะกล่าวถึง 2 ส่วน คือ ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนการปฏิบัติ และปัจจัยเชิงสาเหตุต่างๆ ที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ

2.1 ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนการปฏิบัติ

การศึกษาในเรื่องความแม่นยำและความคลาดเคลื่อนในการให้คะแนนการปฏิบัติได้มีจุดเริ่มต้นมาจากศาสตร์ทางจิตวิทยาอุตสาหกรรมและองค์การ โดยอยู่ในส่วนของการประเมินผลการทำงาน (Performance Appraisal; PA) เนื่องจากความสำคัญในด้านของการนำผลคะแนนไปใช้ต่อ เช่น การเลื่อนขั้น หรือการจัดโปรแกรมสำหรับพัฒนาบุคลากร รวมถึงการใช้เพื่อตรวจสอบว่ารูปแบบหรือวิธีการประเมินผลการปฏิบัติงานนั้นดีเพียงพอหรือไม่ ทั้งนี้ Murphy & Cleveland

(1995) ได้กล่าวว่า การตรวจสอบคุณภาพทางจิตมิติ (psychometric qualities) ของข้อมูลประเภทคะแนนแบ่งเป็น 3 ประเภท คือ 1. การตรวจสอบคุณภาพทางจิตมิติแบบดั้งเดิม (traditional psychometric) เช่น ความตรง (validity) ความเที่ยง (reliability) ความสอดคล้องระหว่างผู้ประเมิน (interrater agreement) 2. ความคลาดเคลื่อนของผู้ประเมิน (rater errors) และ 3. ความแม่นยำในการให้คะแนน (rating accuracy) ทั้งนี้การศึกษาในส่วนของความคลาดเคลื่อนของผู้ประเมิน (rater errors) และความแม่นยำในการให้คะแนน (rating accuracy) พบว่า ยังคงมีความแตกต่างกันตามแนวคิดและทฤษฎีที่นำมาใช้ (ดังจะกล่าวถึงโดยละเอียดต่อไป) Balzer & Sulsky (Balzer & Sulsky, 1990 cited in Murphy & Cleveland, 1995) ได้สำรวจเกณฑ์ที่ใช้สำหรับประเมินประสิทธิภาพในการให้คะแนนจากงานวิจัยที่เกี่ยวกับการประเมินผลการปฏิบัติงานในช่วงปี ค.ศ. 1976 – 1986 พบว่า ในช่วงปี ค.ศ. 1976 มีการตรวจสอบความคลาดเคลื่อนของผู้ประเมิน (54%) การตรวจสอบความเที่ยงและความตรง (31%) การตรวจสอบความแม่นยำในการให้คะแนน (0%) และการตรวจสอบอื่นๆ (16%) ส่วนในปี ค.ศ. 1986 มีการตรวจสอบความคลาดเคลื่อนของผู้ประเมิน (30%) การตรวจสอบความเที่ยงและความตรง (10%) การตรวจสอบความแม่นยำในการให้คะแนน (25%) และการตรวจสอบอื่นๆ (70%) ดังนั้น Murphy & Cleveland (1995) จึงได้กล่าวสรุปว่า การตรวจสอบคุณภาพทางจิตมิติและความคลาดเคลื่อนของผู้ประเมินเป็นการตรวจสอบทางอ้อมของความแม่นยำในการให้คะแนน (indirect measures of rating accuracy หรือ inaccuracy) เนื่องจากถ้าพบว่า ผลการตรวจสอบคุณภาพทางจิตมิติดีหรือผลการวิเคราะห์ไม่พบความคลาดเคลื่อนของผู้ประเมินอาจกล่าวได้ว่าการให้คะแนนนั้นมีความแม่นยำ (rating accuracy) ส่วนการตรวจสอบความแม่นยำในการให้คะแนนเป็นการตรวจสอบทางตรง (direct measures of rating accuracy) และจากผลการศึกษามากพบว่า ในการตรวจสอบคุณภาพทางจิตมิติ ได้แก่ การตรวจสอบความเที่ยงอาจพบว่ามีค่าต่ำ เนื่องจากการประเมินผลการปฏิบัติจะมีองค์ประกอบหลายมิติ ส่วนการตรวจสอบความตรงตามเกณฑ์สัมพันธ์ (criterion-related validity) จะเป็นเรื่องที่ยากเพราะมีความแตกต่างของเกณฑ์ที่นำมาใช้มาก และการตรวจสอบความตรงตามทฤษฎี (construct validity) พบว่าเป็นเรื่องที่ยากเช่นเดียวกัน โดยงานวิจัยที่พบมักจะเป็นศึกษาความตรงแบบลู่อเข้า (convergent validity) และความตรงในการจำแนก (discriminant validity) ซึ่งไม่จัดเป็นความตรงตามทฤษฎี ดังนั้นการตัดสินความถูกต้องของการให้คะแนนจึงมุ่งเน้นไปที่การวิเคราะห์ความคลาดเคลื่อนของผู้ประเมินและความแม่นยำในการให้คะแนน

2.1.1 ความคลาดเคลื่อนของผู้ประเมิน (rater errors)

ความคลาดเคลื่อนของผู้ประเมิน (rater errors) หมายถึง ความคลาดเคลื่อนที่เป็นคุณลักษณะของผู้ประเมินโดยไม่เกี่ยวข้องกับความสามารถของผู้ถูกประเมิน ซึ่งผู้ประเมินมีแนวโน้มที่จะให้คะแนนสูงหรือต่ำกว่าความสามารถจริงของผู้ถูกประเมิน ทำให้เกิดการบิดเบือนในการวัดผลและส่งผลต่อความตรงหรือความยุติธรรมในการประเมิน (Eckes, 2009; Engelhard, 1994 cited in Kassim, 2011) ส่วนคำอื่นๆ ที่นิยมใช้และมีความหมายเช่นเดียวกัน คือ rater variability, rater effects, rater bias, และ rater severity (Eckes, 2009) ปัญหาความคลาดเคลื่อนของผู้ประเมินอาจเกิดจากการไม่เข้าใจการใช้มาตรประเมิน หรือการมีมาตรฐานหรือความคาดหวังที่ต่างกันของผู้ประเมิน โดยในช่วงปี ค.ศ. 1980 – 1990 ผลการศึกษาส่วนใหญ่พบว่า ไม่มีการลดลงของความคลาดเคลื่อนของผู้ประเมิน ((Becker, Hess, & Gibney, 1993; Blok, 1985; Braun, 1988; Engelhard Jr., 1992; Lane & Sabers, 1989; Lunz & Stahl, 1990; Lunz, Wright, & Linacre, 1990; Michael, Cooper, Shaffer, & Wallis, 1980 cited in MacMillan, 2000) ถึงแม้ว่าจะมีการฝึกฝนอบรมสำหรับการให้คะแนนที่มีเป้าหมายเพื่อการพัฒนาาระบบผู้ประเมินที่มีความสามารถเทียบเคียงกันได้ (Welch & Miller, 1995 cited in MacMillan, 2000)

1. ประเภทความคลาดเคลื่อนของผู้ประเมิน ความคลาดเคลื่อนของผู้ประเมินโดยทั่วไปนิยมแบ่งเป็น 5 ประเภท ได้แก่ ผลจากฮาโล (halo effect) การกดหรือปล่อยคะแนน (severity-lenency) การให้คะแนนที่มีแนวโน้มเข้าสู่ส่วนกลาง (central tendency effect) การจำกัดช่วงการให้คะแนน (restriction of range) และความสอดคล้องระหว่างผู้ประเมิน (interrater agreement) (Eckes, 2009) ซึ่งมีรายละเอียดดังนี้

ก) ผลจากฮาโล (Halo effect) หมายถึง ความไม่แน่นอนของผู้ประเมิน โดยจะมีลักษณะของการขาดความรอบคอบในการแยกแยะระดับความสามารถของผู้ถูกประเมินที่แตกต่างกันอันเนื่องมาจากผลของความประทับใจโดยรวมต่อผู้ถูกประเมิน หรือเป็นการให้คะแนนจากมุมมองส่วนตัว ซึ่งไม่ตรงกับความสามารถจริงของผู้ถูกประเมิน หรือแนวโน้มที่จะให้คะแนนผู้ถูกประเมินในระดับเดียวกันทั้งที่ระดับความสามารถต่างกัน ซึ่งเกิดจากอิทธิพลของความประทับใจในภาพรวมหรือจากการเห็นว่าคุณลักษณะนั้นมีความสำคัญอย่างมาก (Eckes, 2009; Saal, Downey, & Lahey, 1980) ข) การกดหรือปล่อยคะแนน (severity-lenency) เป็นลักษณะความคลาดเคลื่อนของผู้ประเมินที่พบได้บ่อยที่สุด โดยสามารถอธิบายใน 3 ลักษณะ คือ 1) แนวโน้มที่ผู้ประเมินจะให้คะแนนสูงหรือต่ำกว่าความสามารถจริงของผู้ถูกประเมิน 2) ลักษณะการให้คะแนนของผู้ประเมินที่ให้คะแนนง่ายหรือยาก โดยมีการให้คะแนนที่สม่ำเสมอแบบสูงหรือต่ำกว่าเกณฑ์ทั่วไปของระดับความสามารถที่แท้จริง และ 3) การให้

คะแนนที่ห่างจากจุดกึ่งกลางไปในทางที่สูงหรือต่ำ ซึ่งเกิดขึ้นเมื่อเปรียบเทียบกับผู้ประเมินคนอื่นหรือเมื่อเทียบกับเกณฑ์มาตรฐาน (Saal, Downey, & Lahey, 1980; Eckes, 2009)

ค) การให้คะแนนที่มีแนวโน้มเข้าสู่ส่วนกลาง (central tendency) หมายถึง ลักษณะความคลาดเคลื่อนของผู้ประเมินที่หลีกเลี่ยงการให้คะแนนในส่วนปลายของมาตรวัดทั้งที่แบบสูงหรือต่ำ แต่จะให้คะแนนกระจุกอยู่ในช่วงกลางแทน ซึ่งสถานการณ์นี้อาจเกิดจากการที่ผู้ประเมินกังวลว่าตนจะให้คะแนนต่างจากผู้ประเมินคนอื่น ๆ ในกลุ่ม จึงต้องทำเพื่อความปลอดภัยไว้ก่อน (Eckes, 2009; Saal et al., 1980; Wolfe, Chiu, & M., 2000)

ง) การจำกัดช่วงการให้คะแนน (restriction of range) หมายถึง ลักษณะความคลาดเคลื่อนของผู้ประเมินที่มีการจำกัดขอบเขตการให้คะแนนผู้ถูกประเมินให้อยู่เฉพาะส่วนซึ่งอาจเป็นส่วนปลายมาตรวัดที่สูงหรือต่ำ ซึ่งต่างจากการให้คะแนนที่มีแนวโน้มเข้าสู่ส่วนกลางที่เป็นการให้คะแนนเฉพาะส่วนกลางของมาตรวัดเท่านั้น (Iramaneerat & Yudkowsky, 2007; Saal et al., 1980)

จ) ความสอดคล้องระหว่างผู้ประเมิน (interrater agreement) หรือความเที่ยงระหว่างผู้ประเมิน (interrater reliability) หรือในบางแห่งอาจใช้คำที่มีความหมายตรงกันข้าม เช่น random effect หรือ rater inconsistency ทั้งนี้ความสอดคล้องระหว่างผู้ประเมิน หมายถึง ผู้ประเมินจำนวน 2 คน หรือมากกว่าที่เป็นอิสระกัน มีผลการให้คะแนนที่เหมือนกันในข้อสอบข้อเดียวกัน ซึ่งจัดเป็นความตรงเชิงฉันทามติ (consensual validity) หรือความตรงแบบลู่เข้า (convergent validity) (Saal, Downey, & Lahey, 1980; Iramaneerat, & Yudkowsky, 2007)

2. แนวทางป้องกันปัญหาความคลาดเคลื่อนของผู้ประเมิน มาตรฐานทั่วไปที่ใช้เพื่อป้องกันปัญหาความคลาดเคลื่อนของผู้ประเมิน และพบว่ามีใช้ในการทดสอบที่สำคัญ ซึ่งประกอบด้วย 3 องค์ประกอบ คือ การฝึกอบรมการตรวจให้คะแนน การตรวจให้คะแนนโดยผู้ประเมินอย่างน้อย 2 คน ในข้อสอบข้อเดียวกัน และการรายงานค่าความเที่ยงระหว่างผู้ประเมิน (Eckes, 2009)

ก) การฝึกอบรมการตรวจให้คะแนน (rater training) มีเป้าหมายเพื่อให้เกิดความคุ้นเคยกับลักษณะของแบบสอบและเกณฑ์การให้คะแนน โดยเฉพาะผู้ประเมินจะถูกฝึกให้มีความเข้าใจตรงกันในเรื่องของ 1) คุณลักษณะที่ต้องการวัด (construct) 2) ระดับของความสามารถที่เป็นเป้าหมาย 3) เกณฑ์และคำอธิบายที่แสดงถึงคุณลักษณะของความสามารถในแต่ละระดับ 4) ชนิดของการให้คะแนนหรือมาตรวัด และ 5) ระดับความยากโดยรวมของแบบสอบ

ข) การตรวจให้คะแนนโดยผู้ประเมินอย่างน้อย 2 คน ในข้อสอบข้อเดียวกัน (independent ratings of the same performance by two or more raters) จัดเป็นวิธีการสำคัญที่ใช้ป้องกันการเกิดความคลาดเคลื่อนของผู้ประเมินซึ่งในกรณีที่มีความเห็นไม่ตรงกันผู้ที่มีหน้าที่ควบคุมดูแลการตรวจจะเป็นผู้ที่ตัดสินคะแนนในท้ายสุด โดยอาจใช้วิธีการหาค่าเฉลี่ยของคะแนนซึ่งเป็นอิสระจากกัน การลงความเห็นร่วมกัน การขอความช่วยเหลือจากผู้ประเมินที่มีประสบการณ์ การตัดสินโดยผู้ประเมินคนที่ 3 เป็นต้น

ค) การรายงานค่าความเที่ยงระหว่างผู้ประเมิน (interrater reliability) ถ้ามีค่าสูงจะอธิบายได้ว่า กลุ่มผู้ประเมินมีมุมมองที่คล้ายกันเกี่ยวกับคุณลักษณะ (construct) ในข้อคำถามนั้น ซึ่งผลที่ตามมาคือ คะแนนที่ได้จากการตรวจนี้จะมีความใกล้เคียงกับคะแนนความสามารถที่จริงของผู้ถูกประเมิน (true score) แต่ทั้งนี้การรายงานค่าความเที่ยงระหว่างผู้ประเมินก็ยังคงขาดนิยามที่มีความชัดเจน ทำให้เกิดความพยายามที่จะสร้างแนวคิดของความเที่ยงระหว่างผู้ประเมินซึ่งมีจุดเน้นต่างกัน คือ วิธีการ บุคคล และวัตถุประสงค์ ซึ่งเป็นผลให้เกิดความสับสนเกี่ยวกับตัวบ่งชี้ที่จะใช้ (Bramley, 2007; Hayes & Krippendorff, 2007; LeBreton & Senter, 2008; Shoukri, 2004; von Eye & Mun, 2005; Zegers, 1991 cited in Eckes, 2009) และเกิดความซับซ้อนในความหมาย ดังนั้นจึงมีการจัดตัวบ่งชี้ของความเที่ยงระหว่างผู้ประเมินออกเป็น 2 ประเภท คือ 1) ตัวบ่งชี้ฉันทามติ (consensus indices) หรืออาจเรียกว่า ความสอดคล้องระหว่างผู้ประเมิน (interrater agreement) ซึ่งหมายถึงผู้ประเมินแต่ละคนมีการให้คะแนน (rating) ที่เท่ากัน (absolute correspondence of ratings) และ 2) ตัวบ่งชี้ความเข้ากัน (consistency indices) มีความหมายที่ต่างจากตัวบ่งชี้ฉันทามติ คือ ผู้ประเมินแต่ละคนมีการจัดระดับ (ordering or ranking) ได้ตรงกัน (relative correspondence of ratings) แต่ทั้งนี้ก็ยังคงมีการเรียกชื่อที่สลับกันระหว่างตัวบ่งชี้ทั้งสองชนิดนี้ ทำให้มีความหมายที่คลาดเคลื่อนไปหรือบางครั้งอาจเกิดความขัดแย้งในการรายงานผล ตัวอย่างเช่น ผู้ประเมินอาจให้คะแนนพิเศษแก่ผู้เข้าสอบคนหนึ่งเพียงหนึ่งหรือสองคะแนนซึ่งต่ำกว่าคะแนนพิเศษที่ให้กับผู้เข้าสอบอีกคนที่อยู่ในกลุ่มเดียวกัน (rating) แต่ทั้งนี้เมื่อมีการจัดระดับผู้เข้าสอบ คะแนนของผู้เข้าสอบทั้งสองคนนี้ก็จะต้องอยู่ในระดับที่เท่ากัน (ordering or ranking) ซึ่งจากตัวอย่างนี้ถือว่าเป็นเพียงลักษณะของความเข้ากันของผู้ประเมิน (high consistency estimates) แต่ไม่ใช่ความสอดคล้องกันอย่างแท้จริง (exact agreement) (Stemler & Tsai, 2008; Tinsley & Weiss, 1975, 2000 cited in Eckes, 2009)

3. การวิเคราะห์ความคลาดเคลื่อนของผู้ประเมิน โมเดลการวิเคราะห์ความคลาดเคลื่อนของผู้ประเมินมี 3 โมเดลหลัก คือ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical test theory: CTT) ทฤษฎี

การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability theory: GT) และโมเดลหลายองค์ประกอบของราสช์ (Many-facet Rasch measurement: MFRM) (MacMillan, 2000) นอกจากนี้ยังมีการพัฒนาโมเดลอื่นๆ เช่น Hierarchical rater model (HRM), Hierarchical rater model with latent class signal detection theory (HRM-SDT) เป็นต้น ซึ่งใน CTT จะใช้วิธีการคำนวณค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (standard deviation) หรือในกรณีหาค่าความเที่ยงของผู้ประเมิน (rater reliability) จะใช้สูตรของ Spearman-Brown (MacMillan, 2000) ส่วน GT จะเป็นการประเมินความคลาดเคลื่อนในระดับกลุ่มของผู้ประเมิน (group-level rater effects) โดยเป้าหมายของ GT คือ การประมาณค่า error variance ที่สัมพันธ์กับค่าคะแนนดิบ (raw score) ของผู้ถูกประเมิน แต่ไม่ได้สนใจในเรื่องความยากของข้อสอบหรือความคลาดเคลื่อนของผู้ประเมิน และผลการวิเคราะห์ขั้นสุดท้ายของ GT คือ ค่าสัมประสิทธิ์ความเที่ยง (reliability coefficient) หรือค่าสัมประสิทธิ์การสรุปอ้างอิง (generalizability coefficient) ซึ่งต่างจาก MFRM ที่วิเคราะห์ผู้ถูกประเมินและผู้ประเมินเป็นรายคน หรือวิเคราะห์ข้อสอบเป็นรายข้อ และผลการวิเคราะห์ขั้นสุดท้ายของ MFRM เช่น ระดับความสามารถของผู้ถูกประเมิน และ fit statistics ต่างๆ ที่สามารถอธิบายถึงความคลาดเคลื่อนชนิดต่างๆ ของผู้ประเมิน หรือความสามารถของผู้ถูกประเมิน หรือความยากง่ายของข้อสอบแต่ละข้อ (J. M. Linacre, 1996) นอกจากนี้ยังมีโปรแกรมคอมพิวเตอร์สำหรับใช้ในการวิเคราะห์ข้อมูล เช่น Facets, ConQuest เป็นต้น ส่วน Hierarchical Rater Model (HRM) ที่เริ่มพัฒนาโดย Patz ในปี ค.ศ. 1996 และ 2002 เป็นโมเดลการวิเคราะห์ 2 ระดับ คือ rater's score และ latent categories ซึ่งพัฒนาจากแนวคิด คือ คะแนนที่ได้จากการตรวจจะเป็นตัวบ่งชี้ของการจำแนกระดับความสามารถซึ่งเป็นลักษณะคะแนนแบบรูบริกส์ (scoring rubric) และการจำแนกระดับความสามารถนี้จะเป็นตัวบ่งชี้ของความสามารถของผู้ถูกประเมิน (θ) แต่ทั้งนี้ HRM ก็ยังคงมีข้อบกพร่อง คือ สามารถวิเคราะห์ความคลาดเคลื่อนของผู้ประเมินได้เพียงชนิดการกด-ปล่อยคะแนน (severe-lenient) ดังนั้นจึงมีการพัฒนา Hierarchical rater model with latent class signal detection theory (HRM-SDT) โดย DeCarlo, Kim, & Johnson ในปี ค.ศ. 2011 เพื่อให้สามารถวิเคราะห์ความคลาดเคลื่อนของผู้ประเมินชนิดอื่นๆ ได้ ซึ่งเป็นการปรับเปลี่ยนจาก HRM โดยการเปลี่ยน first level model จาก signal detection model เป็น latent class extension of signal detection theory ส่วน second level model ยังคงเป็น IRT model ดังนั้นโมเดลพื้นฐานของ HRM-SDT จึงประกอบด้วย HRM และ latent class SDT (DeCarlo, Kim, & Johnson, 2011) แต่จากงานวิจัยส่วนใหญ่ในปัจจุบันพบว่า MFRM เป็นโมเดลที่ได้รับความนิยมในการใช้งานมาก

นอกจากนี้ MFRM ยังสามารถใช้วิเคราะห์ความแม่นยำในการให้คะแนนอีกด้วย ส่วน HRM-SDT เป็นโมเดลที่มีการพัฒนาขึ้นมาใหม่ในปี 2011 แต่ยังไม่ค่อยมีการใช้งานหรือมีงานวิจัยที่กล่าวถึงค่อนข้างน้อย (รายละเอียดจะกล่าวถึงต่อไป)

2.1.2 ความแม่นยำในการให้คะแนน (rating accuracy)

การวิเคราะห์ความแม่นยำในการให้คะแนนโดยตรง (direct measures of rating accuracy) จำเป็นจะต้องใช้เกณฑ์หรือมาตรฐานสำหรับการเปรียบเทียบ ซึ่งมาตรฐานในที่นี้ก็คือ คะแนนจริง (true score) ภายใต้งैอนไขที่เหมาะสมและปราศจากอคติต่างๆ (biase) ดังนั้นการวัดความแม่นยำในการให้คะแนนจึงอาจมีความแตกต่างกันได้ขึ้นกับนิยามหรือแนวคิดทฤษฎีที่ใช้ ซึ่งรูปแบบที่นิยมใช้ในการวัดความแม่นยำในการให้คะแนนมี 2 แนวคิด คือ

Cronbach (Cronbach, 1955) ได้นำเสนอ 4 องค์ประกอบของการวัดความแม่นยำในการให้คะแนน ได้แก่ องค์ประกอบความถูกต้องของคะแนนเฉลี่ยรวมของผู้ถูกประเมินทุกคนและทุกมิติ (Elevation) องค์ประกอบความถูกต้องของคะแนนเฉลี่ยในทุกมิติระหว่างผู้ถูกประเมินแต่ละคน (Differential elevation) องค์ประกอบความถูกต้องของคะแนนเฉลี่ยระหว่างผู้ถูกประเมินทุกคนในแต่ละมิติ (Stereotype accuracy) และองค์ประกอบความถูกต้องของคะแนนเฉลี่ยของผู้ถูกประเมินแต่ละคนในแต่ละมิติ (Differential accuracy) ซึ่ง Borman (Borman, 1977 cited in Murphy & Cleveland, 1995) ได้แนะนำให้ใช้ค่าสหสัมพันธ์ระหว่างคะแนนจริง (true scores) กับองค์ประกอบความถูกต้องของคะแนนเฉลี่ยของผู้ถูกประเมินแต่ละคน ในแต่ละมิติ (Differential accuracy ที่พัฒนาโดย Cronbach) ส่วน Murphy, Garcia and colleagues (Murphy, Garcia & colleagues, 1982 cited in Murphy & Cleveland, 1995) ได้โต้แย้งว่า องค์ประกอบความถูกต้องของคะแนนเฉลี่ยในทุกมิติระหว่างผู้ถูกประเมินแต่ละคน (Differential elevation) มีความสำคัญมากกว่าองค์ประกอบความถูกต้องของคะแนนเฉลี่ยระหว่างผู้ถูกประเมินทุกคนในแต่ละมิติ (Stereotype accuracy) ซึ่ง Murphy & Cleveland (1995) ได้กล่าวสรุปว่า ยังไม่มีฉันทามติของนักวิจัยเกี่ยวกับสิ่งที่ดีที่สุดในการวัดความแม่นยำในการให้คะแนน ซึ่งเขาได้แนะนำว่า ถ้าเป็นไปได้ควรใช้การวัดหลายๆ องค์ประกอบ (multiple measures)

Engelhard (Engelhard, 1994 cited in Murphy & Cleveland, 1995) ได้ประยุกต์ใช้โมเดลหลายองค์ประกอบของราสซ์ และโปรแกรม Facets ในการตรวจสอบความคลาดเคลื่อนของ

ผู้ประเมินและความแม่นยำในการให้คะแนน โดยวิธีนี้จำเป็นต้องมีการออกแบบโดยมีผู้ประเมินหลายคนซึ่งอาจไม่เหมาะสมกับบริบทของการประเมินผลการปฏิบัติงาน (performance appraisal) แต่ถ้าได้นำมาใช้ก็อาจมีประโยชน์เนื่องจากมีการประมาณค่าโดยตรงของ true scores และมีการปรับเทียบมาตรวัด (รายละเอียดของโมเดลหลายองค์ประกอบของราสส์จะกล่าวถึงต่อไป)

2.1.3 ความสัมพันธ์ระหว่างความคลาดเคลื่อนของผู้ประเมินและความแม่นยำในการให้คะแนน

เนื่องจากความคลาดเคลื่อนของผู้ประเมิน (rater errors) เป็นการตรวจสอบทางอ้อมของความแม่นยำในการให้คะแนน (indirect measures of rating accuracy หรือ inaccuracy) โดยถ้าผลการวิเคราะห์ไม่พบความคลาดเคลื่อนของผู้ประเมินอาจกล่าวได้ว่าการให้คะแนนนั้นมีความแม่นยำ (rating accuracy) ส่วนการตรวจสอบความแม่นยำในการให้คะแนนเป็นการตรวจสอบทางตรง (direct measures of rating accuracy) ซึ่งจากแนวคิดนี้ได้มีหลายงานวิจัยในช่วงปี 1977 – 1989 ที่ตรวจสอบความสัมพันธ์ระหว่างความคลาดเคลื่อนของผู้ประเมินและความแม่นยำในการให้คะแนนซึ่งการวัดความคลาดเคลื่อนของผู้ประเมินและความแม่นยำในการให้คะแนนจะใช้หลักของการคำนวณค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (standard deviation) โดยการวัดความคลาดเคลื่อนของผู้ประเมินจะมาจากการวัดกระจายของความคลาดเคลื่อน (distributional errors) และค่าสหสัมพันธ์ของความคลาดเคลื่อน (correlational errors) ส่วนการวัดความแม่นยำในการให้คะแนนจะใช้แนวคิดของ Cronbach ปี 1955 ซึ่งมี 4 องค์ประกอบ คือ ได้แก่ องค์ประกอบความถูกต้องของคะแนนเฉลี่ยรวมของผู้ถูกประเมินทุกคนและทุกมิติ (Elevation) องค์ประกอบความถูกต้องของคะแนนเฉลี่ยในทุกมิติระหว่างผู้ถูกประเมินแต่ละคน (Differential elevation) องค์ประกอบความถูกต้องของคะแนนเฉลี่ยระหว่างผู้ถูกประเมินทุกคนในแต่ละมิติ (Stereotype accuracy) และองค์ประกอบความถูกต้องของคะแนนเฉลี่ยของผู้ถูกประเมินแต่ละคนในแต่ละมิติ (Differential accuracy) ซึ่งผลจากการตรวจสอบความสัมพันธ์ระหว่างความคลาดเคลื่อนของผู้ประเมินและความแม่นยำในการให้คะแนนพบว่า การวัดความคลาดเคลื่อนของผู้ประเมินไม่มีความสัมพันธ์กับการวัดความแม่นยำในการให้คะแนน (Becker & Cardy, 1986; Bernadin & Pence, 1980; Murphy & Balzer, 1989 cited in Murphy & Cleveland, 1995) หรืออาจเรียกว่าเป็น halo-accuracy paradox ซึ่ง Murphy & Cleveland (1995) ได้กล่าวว่า มีนักวิจัยบางส่วนได้ให้เหตุผลว่าเกิดจากการให้นิยามการวัดที่แตกต่างกัน และนักวิจัยอีกส่วนหนึ่งก็ไม่ได้ให้คำอธิบายใดๆ แต่ก็ไม่ได้สนับสนุนว่าเกิด

จากการให้นิยามการวัดที่แตกต่างกัน และ Murphy & Cleveland ก็ไม่ได้อธิบายถึงสาเหตุของกรณีนี้ แต่มีการอ้างถึง Engelhard (Engelhard, 1994 cited in Murphy & Cleveland, 1995) ว่าได้ประยุกต์ใช้โมเดลหลายองค์ประกอบของราสช์ (Many-facet Rasch measurement: MFRM) ซึ่งเมื่อพิจารณาจากบทความของ Engelhard (Engelhard, 1994) จะพบว่ามีกรกล่าวถึงประเด็นที่ใกล้เคียง คือ สาเหตุของการที่ยังไม่มีเกณฑ์สำหรับการตรวจสอบความแม่นยำของการประเมินผลการปฏิบัติในบริบททางการศึกษา คือ การไม่มีนิยามของคะแนนที่แท้จริง (true scores) ในการใช้เป็นเกณฑ์มาตรฐานสำหรับการเปรียบเทียบ ดังนั้นจากคำอธิบายนี้เมื่อประกอบกับแนวคิดของการทดสอบแนวใหม่ (modern test theories) ในส่วนของของราสช์โมเดล (Rasch model) และทฤษฎีการตอบสนองข้อสอบ (item response theory) มาใช้อธิบายถึงสาเหตุของการเกิด halo-accuracy paradox จะได้ว่าเกิดจากข้อจำกัดของการใช้ทฤษฎีการทดสอบแบบดั้งเดิม คือ ค่าพารามิเตอร์ของข้อสอบเป็นค่าเฉพาะของกลุ่มผู้ถูกประเมินและเป็นอิสระจากความสามารถของผู้ถูกประเมิน ซึ่งในทฤษฎีการตอบสนองข้อสอบได้ปรับแก้ไขจุดอ่อน โดยให้ค่าพารามิเตอร์ของข้อสอบและความสามารถจริงของผู้ถูกประเมินมีความสัมพันธ์กัน ทำให้สามารถวิเคราะห์ความสามารถที่แท้จริงของแต่ละบุคคลได้ (ศิริชัย กาญจนวาสี, 2550) ซึ่งทำให้สามารถวิเคราะห์ ความคลาดเคลื่อนของผู้ประเมิน (rater errors) และความแม่นยำในการให้คะแนน (accuracy) จาก true scores ของผู้ถูกประเมินได้ โดยแนวคิดนี้สอดคล้องกับคำอธิบายของ Engelhard (Engelhard, 1996) ที่ได้กล่าวถึงความแตกต่างของแนวคิดการใช้ MFRM ในการศึกษาความแม่นยำในการให้คะแนน (rating accuracy) กับแนวคิดอื่นๆ ที่มีอยู่ในช่วงแรก ดังนี้ 1) MFRM จะถือว่าความแม่นยำเป็นตัวแปรแฝงระหว่างผู้ประเมินระดับบุคคล และเกณฑ์การให้คะแนนก็เป็นระดับหนึ่งในการวิเคราะห์ ซึ่งในแนวคิดอื่น ดัชนีหรือตัวบ่งชี้ในเชิงปริมาณที่วัดความแม่นยำในการให้คะแนนจะเป็นข้อมูลระดับกลุ่ม 2) MFRM เป็นกรอบแนวคิดที่มีระบบสำหรับการตรวจสอบและวิเคราะห์ทางสถิติระหว่างความแตกต่างของความแม่นยำในการให้คะแนนของผู้ประเมิน โดยผู้ประเมินจะถือเป็นองค์ประกอบ (facet) หนึ่งภายในโมเดล ต่างจากแนวคิดอื่นที่ตัวบ่งชี้ในการวัดความแม่นยำในการให้คะแนนจะใช้การวิเคราะห์ 2 ระดับ คือ คำนวณความแม่นยำในการให้คะแนนสำหรับผู้ประเมินเป็นรายคน จากนั้นจะวิเคราะห์แยกความแตกต่างระหว่างผู้ประเมิน 3) แนวคิดจากโมเดลอื่นๆ จะไม่วิเคราะห์ความแตกต่างของผู้ประเมินที่เป็นผู้เชี่ยวชาญ โดยจะถือว่าผู้ประเมินกลุ่มนี้มีความแม่นยำมากกว่าผู้ประเมินอื่นๆ 4) MFRM เป็นแนวคิดเช่นเดียวกับ IRT ที่มีความก้าวหน้าทางสถิติและจิตมิติ (psychometrics) และ 5) MFRM เป็นแนวคิดที่พัฒนาต่อจากแนวคิดของ Cronbach ที่กล่าวว่า

ธรรมชาติของความแม่นยำในการให้คะแนนเป็นแบบหลายองค์ประกอบ แต่ส่วนที่สำคัญและแตกต่างจาก Cronbach คือ การวิเคราะห์สามารถจะขยายหรือเพิ่มองค์ประกอบอื่นๆ (Facets) ได้ตามต้องการ

2.1.4 โมเดลหลายองค์ประกอบของราสช์ (Many-facet Rasch measurement: MFRM)

จากที่ได้กล่าวในข้างต้นว่า งานวิจัยส่วนใหญ่ในปัจจุบันนิยมใช้โมเดลหลายองค์ประกอบของราสช์สำหรับการวิเคราะห์ความคลาดเคลื่อนของผู้ประเมิน (rater errors) อยู่เป็นจำนวนมาก นอกจากนี้โมเดลหลายองค์ประกอบของราสช์ยังสามารถใช้วิเคราะห์ความแม่นยำในการให้คะแนน (accuracy) ได้ด้วย ทั้งนี้ John M. Linacre ได้พัฒนาโมเดลหลายองค์ประกอบของราสช์ขึ้นในปี ค.ศ. 1989 และยังมีชื่อเรียกอื่นๆ เช่น multi-faceted หรือ many-faceted Rasch measurement, many-faceted conjoint measurement, multifacet Rasch modeling, Facets model เป็นต้น โดยจัดเป็นโมเดลการวัดรูปแบบหนึ่งที่มีเป้าหมายเพื่อการวิเคราะห์ตัวแปรหลายระดับในการทดสอบหรือการประเมินผลลัพธ์ โดยโมเดลนี้พัฒนามาจาก basic Rasch ให้สอดคล้องกับตัวแปรหรือฟาเซท (facet) ที่มีมากกว่าสอง (ผู้ถูกประเมินและข้อสอบ) ขึ้นไป รวมถึงในสถานการณ์ของการสอบประเภทเขียนตอบ ดังนั้นจึงจัดอยู่ในกลุ่มตระกูลของราสช์โมเดล (family of Rasch models) ซึ่งได้แก่ rating scale model (RSM), partial credit model (PCM), linear logistic test model (LLTM), mixed Rasch model เป็นต้น ทั้งนี้โมเดลหลายองค์ประกอบของราสช์เริ่มเป็นที่นิยมใช้มากขึ้นอย่างต่อเนื่องในด้านการทดสอบภาษา การศึกษา การวัดทางจิตวิทยา วิทยาศาสตร์สุขภาพ และด้านอื่นๆ (Eckes, 2009; J. M. (2012). Linacre, 2012)

1. สมการโมเดลหลายองค์ประกอบของราสช์ :

$$\ln \left(\frac{P_{nij k}}{P_{nij (k-1)}} \right) = B_n - D_{gi} - C_j - F_{gk}$$

\ln = natural logarithm

$P_{nij k}$ = ความน่าจะเป็นของผู้ถูกประเมินคนที่ n ซึ่งมีความสามารถ B_n จะได้รับ

คะแนนจากผู้ให้คะแนนคนที่ j ในระดับ k ตามเกณฑ์ในข้อที่ i ที่มี

ความยาก D_{gi} ซึ่งจะตรงข้ามกับ $P_{nij(k-1)}$

$P_{nij(k-1)}$ = ความน่าจะเป็นของผู้ถูกประเมินคนที่ n จะได้รับคะแนนจากผู้ประเมินคนที่ j ในระดับ $k-1$ ตามเกณฑ์ในข้อที่ i จากผู้ประเมินคนที่ j ที่มีความยาก D_{gi}

B_n = ความสามารถของผู้ถูกประเมินคนที่ n

D_{gi} = ค่าความยากข้อที่ i

C_j = ความคลาดเคลื่อนของผู้ประเมินคนที่ j

F_{gk} = ตัวแปรแฝงที่เป็น Rasch-Andrich threshold หรือเรียกว่า step calibration

หรือ step difficulty ซึ่งมีความสัมพันธ์กับ D_{gi} โดยเป็นจุดที่ความน่าจะเป็นของการตอบระดับ k เท่ากับความน่าจะเป็นของการตอบระดับ $k-1$ ซึ่งจะมีค่าเท่ากันทุกข้อสำหรับชุดข้อสอบนั้น หรืออาจอธิบายว่าเป็น ความยากที่จะจำแนกระดับที่ k หรือระดับ $k-1$ สำหรับข้อสอบสอบในชุด g

ค่าพารามิเตอร์ที่จะใช้ในการวิเคราะห์ความมาจากมาตราประเมินค่า (rating scale) จากสมการตัวอย่างจะพบว่าเป็นลักษณะของโมเดลที่มี 3 ฟาเซต (three-facet rating scale model) ซึ่งสมการพื้นฐานของโมเดลหลายองค์ประกอบของราสซ์จะมีลักษณะของโมเดลเชิงเส้นตรง (additive linear model) โดยมาจากการแปลงจากคะแนนที่ได้จากการสังเกตไปเป็นมาตรวัตโลจิท (logits) หรือ log-odds unit จึงทำให้ตัวแปรตามมีลักษณะหลายฟาเซต เช่น ผู้ถูกประเมิน ผู้ประเมิน และเกณฑ์ ในการเริ่มต้นของมาตรวัตโลจิทและการระบุโมเดลจะมีการตั้งศูนย์ (centered) ของฟาเซตให้มีค่าเฉลี่ยเป็น 0 และผลรวมของค่าสัมประสิทธิ์ (sum of the category coefficients) มีค่าเท่ากับ 0 (ยกเว้นฟาเซตที่ 1 โดยทั่วไปคือ person ability ซึ่งจะเป็น noncenter) นอกจากนี้โมเดลหลายองค์ประกอบของราสซ์ยังสามารถใช้ได้กับรูปแบบการให้คะแนนที่ไม่สมบูรณ์ (incomplete rating designs) เช่น มีข้อมูลไม่ครบรวมถึงข้อมูลจากการจำลองหรือไม่ได้มาจากการสังเกต (Eckes, 2009) ในการวิเคราะห์จะเริ่มจากการแปลงคะแนนที่ได้จากการสังเกต (observed scores) ในทุกฟาเซตให้เป็นหน่วยเดียวกันคือ โลจิท (logits) เพื่อใช้ในการเปรียบเทียบและคำนวณความน่าจะเป็นในการตอบถูก

(probability of success) และความน่าจะเป็นในการตอบผิด (probability of failure) ตลอดจนการประมาณค่าพารามิเตอร์ในโมเดล จากนั้นจะมีการคำนวณค่าส่วนเหลือ (residual; R_{ni}) ซึ่งเป็นค่าที่ไม่ต้องการให้เกิดขึ้น (unexpected values) จากคะแนนที่สังเกตได้ (observation; X_{ni}) ลบด้วยคะแนนที่คาดหวัง (expectation; E_{ni}) นอกจากนี้ยังมีการคำนวณค่าสถิติต่างๆ เช่น sum-of-squared-residuals, model residual variance, standardized residual, และ Chi-square fit statistics เพื่อใช้ในการตรวจสอบคุณภาพโมเดล หรือองค์ประกอบต่างๆ ภายในโมเดล

2. สถิติที่ใช้ตรวจสอบโมเดล ค่าสถิติที่ใช้ตรวจสอบความสอดคล้องระหว่างข้อมูลกับการประมาณค่าจากโมเดลได้แก่

1) Standardized residual (Z_{ni}) คือ ค่าส่วนเหลือมาตรฐาน ซึ่งมีลักษณะของการแจกแจงแบบปกติ โดยมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 โดยสามารถคำนวณได้จากสมการที่ (1)

$$Z_{ni} = R_{ni} / \sqrt{V_{ni}} \quad ; \quad (1)$$

V_{ni} = model residual variance มีค่าเท่ากับ $\sum(R_{ni}^2)$

R_{ni} = residual ซึ่งเป็นค่าที่ไม่ต้องการให้เกิดขึ้น (unexpected values) มีค่าเท่ากับคะแนนที่สังเกตได้ (observation; X_{ni}) ลบด้วยคะแนนที่คาดหวัง (expectation; E_{ni})

Z_{ni} สามารถทดสอบความมีนัยสำคัญทางสถิติ (z-test) โดย $Z_{ni} > |\pm 1.96|$ สำหรับ $p < .05$ และ $Z_{ni} > |\pm 2.58|$ สำหรับ $p < .01$ ซึ่งการมีนัยสำคัญทางสถิติจะหมายถึง ข้อมูลมีความสอดคล้องกับโมเดล (fitting the Rasch model) หรือโมเดลมีความถูกต้องในการประมาณค่าพารามิเตอร์ ส่วนค่าติดลบจะหมายความว่า ทำได้ต่ำกว่าที่คาดหวัง ซึ่งก็ควรมีการพิจารณาหาสาเหตุที่ผิดปกติ เช่น เครื่องมือวัด กลุ่มตัวอย่าง ผู้ประเมิน เป็นต้น

2) Chi-square fit statistics ใช้ประโยชน์สำหรับตรวจสอบความมีนัยสำคัญทางสถิติของความแตกต่างระหว่างโดเมนภายในฟาเซต ซึ่งการนำผลบวกรวมของ standardized residual ยกกำลังสอง ($\sum[Z_{ni}^2]$) จะได้การประมาณค่าใกล้เคียงกับ Chi-square distribution และมีค่าเฉลี่ย

เท่ากับ Z_{ni} นอกจากนี้ Mean square (MnSq) ซึ่งแสดงปริมาณความคลาดเคลื่อนของการวัด จะมีค่าเท่ากับ Chi-square หารด้วย degrees of freedom และควรมีค่าที่ใกล้เคียงกับ 1 ดังสมการที่ (2)

$$\text{MnSq} = \frac{\chi^2}{\text{d.f.}} \quad ; \quad (2)$$

ถ้า χ^2 มีค่าสูง ดังนั้น Z_{ni} จะมีค่าห่างจาก 0 มาก หรือ observation มีค่าห่างมากจาก prediction ที่ได้จาก Rasch model แสดงว่า ข้อมูลมีการประมาณค่าได้ต่ำ (unpredictable หรือ underfit) ใน Rasch model ส่วน χ^2 ที่มีค่าต่ำ ดังนั้น Z_{ni} จะมีค่าใกล้เคียงกับ 0 หรือ observation มีค่าใกล้กับ prediction ที่ได้จาก Rasch model แสดงว่า ข้อมูลชุดนั้นมีการประมาณค่าสูงเกินไป (too predictable หรือ overfit)

3) Outfit Mean - Square และ Infit Mean - Square เป็น fit statistics สำหรับการตรวจสอบข้อมูลว่ามีค่าที่ไม่ต้องการ (unexpected values) หรือไม่ โดยจะมีค่าอยู่ระหว่าง 0 ถึง อนันต์ (infinity) ซึ่งค่าที่คาดหวังจะเท่ากับ 1 และสามารถคำนวณได้จากสมการที่ (3), (4)

$$\text{Outfit Mean - Square} = \frac{\sum(R_{ni}^2 / \sum V_{ni})}{N} \quad ; \quad (3)$$

$$\text{Infit Mean - Square} = \frac{\sum(R_{ni}^2)}{\sum V_{ni}} \quad ; \quad (4)$$

ถ้า MnSq มีค่าต่ำกว่า 1 มากๆ แสดงว่าข้อมูล overfit และ MnSq มีค่ามากกว่า 1 มากๆ แสดงว่าข้อมูล underfit หรือ $\text{MnSq} > 2$ = การวัดมีการบิดเบือนมาก, MnSq มีค่าระหว่าง 1.5 ถึง 2 = ลักษณะของการวัดยังไม่ดีนัก, MnSq มีค่าระหว่าง 0.6 ถึง 1.4 = มีการวัดที่ดี, และ $\text{MnSq} < 0.5$ = ลักษณะของการวัดยังไม่ดีนัก อาจเกิดจากข้อมูลมีการกระจายมากไปทำให้มี reliabilities ต่ำ ทั้งนี้ค่า MnSq ที่สูงจะมีความบิดเบือนมากกว่าค่าที่ต่ำ

ความแตกต่างระหว่าง Outfit Mean - Square และ Infit Mean - Square คือ Infit จะมีลักษณะของการตรวจสอบความสอดคล้องระหว่างข้อมูลกับโมเดลที่ไวต่อคุณลักษณะภายใน (inlier-sensitive fit) โดย overfit มักพบในกรณีของข้อสอบที่เน้นวัดบุคคล และ Guttman patterns เป็น

ต้น และ underfit มักพบกรณีที่มีหลายทางเลือกหรือกลุ่มที่แปลก ส่วน Outfit จะมีลักษณะของการตรวจสอบความสอดคล้องระหว่างข้อมูลกับโมเดลที่ไวต่อคุณลักษณะภายนอก (outlier-sensitive fit) โดย overfit มักพบในกรณีการวัดแบบผสมผสาน (imputed response) และ underfit มักพบในกรณีที่ผู้ถูกประเมินเดาถูกหรือผู้ที่มีความสามารถแต่ตอบผิดเพราะไม่ระวัง

4) Mean – Squares standardized (Zstd) สำหรับการตรวจสอบสมมติฐานว่าข้อมูลมีความสอดคล้องเหมาะสม (fit) กับราสซ์โมเดลหรือไม่ ซึ่งเป็นการตรวจสอบความแม่นยำของการประมาณค่า (accuracy) ภายในโมเดล โดยมีลักษณะคล้ายกับ Z-statistics แยกการคำนวณเป็น Infit Zstd และ Outfit Zstd มีการแปลความหมาย คือ $|Z_{std}| \geq 2.0 = \text{statistical significance } (\alpha = .05)$ และ $|Z_{std}| \geq 2.6 = \text{highly significance } (\alpha = .01)$

5) Model Standard Error (Model S.E.) เป็นความคลาดเคลื่อนมาตรฐานที่เกิดจากการวัดทั้งที่ตรวจสอบพบว่าข้อมูลมีความสอดคล้องพอดีกับโมเดล (fit) ซึ่งใช้ในการตรวจสอบความถูกต้องในการตัดสิน (precision) หรือความเที่ยงในการวัดซ้ำ (repeatability) ของตัวแปรแฝงในกรณีที่ข้อมูลมีความคล้ายคลึงกัน ซึ่งสถานการณ์ที่มักพบว่ามีค่าถูกต้องในการตัดสินสูง ได้แก่ 1. มีการสังเกตหรือวัดหลายครั้ง เช่น จำนวนข้อมาก ผู้ตรวจหลายคน 2. มีเครื่องมือที่ดี เช่น แบบสอบที่เหมาะสมไม่ง่ายหรือไม่ยากเกินไป และ 3. มีระดับการจำแนกหลายระดับ เช่น ใช้มาตรฐานประมาณค่า 5 ช่วงแทน 3 ช่วง เป็นต้น

3. การวิเคราะห์ความคลาดเคลื่อนของผู้ประเมิน และความแม่นยำในการให้คะแนน การใช้โมเดลหลายองค์ประกอบของราสซ์ (MFRM) วิเคราะห์ความคลาดเคลื่อนของผู้ประเมินและความแม่นยำในการให้คะแนน ซึ่งมีรายละเอียดในการวิเคราะห์และแปลผลทางสถิติ ดังนี้ (Barrett, 2005; J. M. (2012). Linacre, 2012; Wolfe et al., 2000)

1) การกดหรือปล่อยคะแนน (severity-lenency) แสดงถึงแนวโน้มของผู้ประเมินที่จะให้คะแนนผู้ถูกประเมินสูงหรือต่ำกว่าความเป็นจริง ซึ่งตัวบ่งชี้ในการวัดความคลาดเคลื่อนของผู้ประเมิน (rater error) จะพิจารณาจากคะแนนดิบ (raw scores) และคะแนนในหน่วยโลจิทของฟาเซตผู้ประเมิน (rater logit) ซึ่งจะพบคะแนนต่ำ ถ้าผู้ประเมินมีลักษณะของการปล่อยคะแนน (leniency) และจะพบคะแนนสูงถ้าผู้ประเมินมีลักษณะของการกดคะแนน (severity)

2) อิทธิพลฮาโล (halo effect) จะเกิดขึ้นเมื่อผู้ประเมินไม่สามารถแยกแยะความแตกต่างระหว่างความคิดและคำตอบจากผู้ถูกประเมิน (Thorndike, 1920 cited in Barrett, 2005) เช่น

ผู้ประเมินอาจให้คะแนนบนพื้นฐานของการประทับใจในภาพรวมของแต่ละคำตอบ ซึ่งแสดงว่าผู้ประเมินไม่สามารถแยกความแตกต่างระหว่างแนวคิดที่จำเป็นหรือไม่จำเป็น หรือผู้ประเมินไม่สามารถประเมินในโดเมนหรือเกณฑ์การวัดตามคุณลักษณะที่เฉพาะ (Engelhard, 1994 cited in Barrett, 2005) นอกจากนี้แนวคิดของการให้คะแนนแบบภาพรวมอาจเกิดจากการสร้างลักษณะเทียมที่สัมพันธ์กันระหว่างข้อ ดังนั้นจึงมีหลักของการให้คะแนนระหว่างข้อที่ควรเป็นอิสระจากกัน และการให้คะแนนที่เป็นอิสระจากกันระหว่างข้อจะสามารถตรวจสอบได้จากราสช์โมเดล (Rasch model) โดยการประมาณค่าผู้ประเมิน (rater estimates) เช่น mean square error statistics หรือจาก fit MnSq ถ้าค่าสถิติเหล่านี้มีค่าน้อยกว่า 0.6 หรือการให้คะแนนเท่าๆ กัน ในหลายๆ ข้อ แสดงว่าผู้ประเมินอาจจะมีการให้คะแนนในแต่ละข้อไม่เป็นอิสระจากกัน

3) แนวโน้มการให้คะแนนค่ากลาง (central tendency effect) เป็นสถานการณ์ของการให้คะแนนที่กระจุกอยู่รอบๆ คะแนนกึ่งกลาง และสะท้อนถึงการที่ผู้ประเมินไม่ต้องการให้คะแนนส่วนปลายของมาตรวัด ซึ่งเป็นปัญหาที่มักพบในกรณีที่ใช้การให้คะแนนแบบหลายค่า นอกจากนี้ยังพบว่าการให้คะแนนค่ากลางมักมีความสัมพันธ์กับผู้ประเมินที่ไม่มีประสบการณ์หรือมีคุณภาพน้อย ซึ่งความคลาดเคลื่อนชนิดนี้พบได้โดยการตรวจสอบเบื้องต้น เช่น ค่าเฉลี่ย ค่ามัธยฐาน พิสัย และส่วนเบี่ยงเบนมาตรฐาน (raw SD จะมีค่าต่ำ) จากคะแนนของผู้ประเมินแต่ละคน แต่การพิจารณาเบื้องต้นนี้อาจเกิดความผิดพลาดในการแปลผลได้ จึงมีการใช้ราสช์โมเดลในการตรวจสอบทั้งจากการประมาณค่าข้อสอบ (item estimates) ได้แก่ mean square error statistics หรือ unweighted fit MnSq และ unweighted fit t ซึ่งถ้าค่าสถิติเหล่านี้มีค่าสูง (เช่น weighted fit MnSq มีค่ามากกว่า 1.5 และ unweighted fit t มีค่ามากกว่า 1) แสดงว่ามีแนวโน้มการให้คะแนนค่ากลาง และการประมาณค่าของผู้ประเมิน (rater estimates) ที่พบว่ามีความ rater fit ต่ำ ยกเว้น ในบางกรณีจะพบว่ามีความ rater fit สูง เช่น ช่วงมาตรประมาณค่ากว้าง หรือมีความแตกต่างระหว่างผู้ถูกประเมินมาก

4) การจำกัดช่วงการให้คะแนน (restriction of range) จะมีความสัมพันธ์กับแนวโน้มการให้คะแนนค่ากลาง แต่มีลักษณะเฉพาะกว่าในขอบเขตของการจำแนกคะแนนที่สูงหรือต่ำในผู้ถูกประเมินที่มีระดับความสามารถต่างกัน (Engelhard 1994; Engelhard & Stone, 1998 cited in Barrett, 2005) ซึ่งการวิเคราะห์ในราสช์โมเดลจะพิจารณาจาก mean square error statistics หรือ unweighted fit MnSq ถ้าพบว่า weighted fit MnSq มีค่ามากกว่า 1.30 หรือน้อยกว่า 0.77 (หรือมีค่ามากกว่า 1.2 หรือน้อยกว่า 0.4; Wright & Linacre, 1994 cited in Iramaneerat, et al., 2008) จะแสดงว่ามีความคลาดเคลื่อนของผู้ประเมินชนิดนี้ หรืออาจแสดงว่า ข้อคำถามนั้นที่ไม่สามารถ

จำแนกระหว่างผู้ถูกประเมินที่มีความสามารถต่างกันได้นอกจากนี้ความคลาดเคลื่อนชนิดนี้ยังสัมพันธ์กับผู้ประเมินที่ไม่มีประสบการณ์หรือมีคุณภาพน้อย

5) ความเที่ยงหรือความสอดคล้องระหว่างผู้ประเมิน (inter-rater reliability or agreement) มาจากแนวคิดว่าการให้คะแนนจะมีคุณภาพสูงถ้ามีผู้ประเมินซึ่งเป็นอิสระจำนวนสองคนหรือมากกว่านั้นได้มีการให้คะแนนที่ตรงกัน และความคลาดเคลื่อนชนิดนี้จะสะท้อนเกี่ยวกับฉันทามติหรือความตรงแบบลู่อู่เข้า (convergent validity) ซึ่งสถิติในราสชโมเดลที่จะบ่งบอกถึงความคลาดเคลื่อนชนิดนี้คือ model fit ได้แก่ mean square errors ของทั้งผู้ประเมินและข้อสอบ โดยค่าที่ต่ำจะแสดงว่ามีความสอดคล้องกันหรือมีความเห็นตรงกันมากหรือเป็นการตรวจที่มีคุณภาพสูง แต่ถ้ามีค่าสูงจะแสดงว่ามีความสอดคล้องกันหรือมีความเห็นตรงกันน้อยหรือเป็นการตรวจที่มีคุณภาพต่ำ ซึ่งค่าที่เป็นอุดมคติของ weighted fit MnSq คือ 1.00 และ weighted fit t คือ 0.00 ถ้าค่าของ weighted fit MnSq มากกว่า 1.5 แสดงว่าผู้ประเมินมีการให้คะแนนไม่สอดคล้องกัน ซึ่งค่าของ MnSq คือ ความชัน ณ ตำแหน่งจุดเปลี่ยนของ Person Characteristics Curves ซึ่งในทางอุดมคติความชันนี้ควรจะติดลบ 1.00 และการเพิ่มความชันที่เบี่ยงเบนไปจากนี้แสดงถึงการมีความสอดคล้องน้อยและเป็นการให้คะแนนที่มีความเที่ยงน้อย (Engelhard & Stone, 1998 cited in Barrett, 2005)

6) ความแม่นยำในการให้คะแนน (accuracy) การวัดการให้คะแนนที่มีความแม่นยำจะประเมินจากความคลาดเคลื่อนระหว่างคะแนนที่ได้รับ (evaluation score) กับคะแนนจริง (true score) ของผู้ที่ถูกประเมิน ซึ่งมาจาก Rasch model โดยถ้าข้อมูลมีความสอดคล้อง (fit) กับโมเดล ดังนั้นพารามิเตอร์ที่ได้จากการประมาณค่าจะมีความแม่นยำ (accuracy) หรือเป็นค่าที่แท้จริง (true value) สำหรับการวัด การวิเคราะห์ความแม่นยำในการให้คะแนนโดยใช้โมเดลหลายองค์ประกอบของราสช สามารถพิจารณาจากคะแนนในหน่วยโลจิตของฟาเซตผู้ประเมิน (rater logit) และค่า Outfit Mean – Square, Infit Mean – Square โดย MnSq ควรมีค่าระหว่าง 0.5 ถึง 1.5 นอกจากนี้ อาจพิจารณาจากค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนดิบ (raw SD ที่มีค่าต่ำ)

ตารางที่ 1 ลักษณะการวิเคราะห์ความคลาดเคลื่อนของผู้ประเมินด้วยวิธีโมเดล (Barrett, 2005: 164-5)

ความคลาดเคลื่อนของผู้ประเมิน	ลักษณะที่พบจากค่าสถิติ
Severity - leniency	คะแนนดิบ (raw scores) และคะแนนในหน่วยโลจิทของ ฟาเซตผู้ประเมิน (rater logit) จะมีค่าสูงถ้าเป็นลักษณะของ การกตคะแนน (severity) และจะพบค่าต่ำ ถ้าเป็นลักษณะ ของการปล่อยคะแนน (leniency) และค่าความคลาด เคลื่อนที่ต่ำจะแสดงถึงความคงที่ในการให้คะแนน
Halo effect	Rater estimates : Weighted fit MnSq < 1
Central tendency	Item estimates : Unweighted fit MnSq >> 1 Unweighted fit t >> 0
Restriction of range	Item estimates : Weighted fit $0.77 < \text{MnSq} < 1.3$
Reliability	Rater estimates : Weighted fit MnSq >> 1 Weighted fit t >> 0

หมายเหตุ: unweighted fit หมายถึง OUTFIT และ weighted fit หมายถึง INFIT

>> หมายถึง มากกว่ามาก (much greater than)

2.1.5 ตัวอย่างงานวิจัยที่เกี่ยวข้องกับคุณภาพการให้คะแนนการปฏิบัติโดยใช้โมเดลหลายองค์ประกอบของราสซ์

การศึกษาคูณภาพการให้คะแนนการปฏิบัติในด้านของความแม่นยำและความคลาดเคลื่อนในการให้คะแนนโดยใช้โมเดลหลายองค์ประกอบของราสซ์ (MFRM) มีตัวอย่างงานวิจัยดังนี้

Congdon & McQueen (Congdon & McQueen, 2000) ศึกษาความคงที่ (stability) ของการเกิดความคลาดเคลื่อนของผู้ประเมินโดยใช้ MFRM ในการประเมินขนาดใหญ่ (large-scale assessment) ซึ่งเป็นการทดสอบระดับรัฐเกี่ยวกับความสามารถในการอ่าน เขียน และคิดเลขในปี ค.ศ. 1996 โดยข้อมูลที่น่ามาวิเคราะห์จะใช้เฉพาะส่วนของการประเมินทักษะการเขียนของนักศึกษาจำนวน 8,285 คน

ลักษณะของข้อสอบจะให้เลือกเขียนข่าวที่น่าสนใจในปัจจุบัน 1 เรื่อง (จำนวน 2 หน้า) มีการให้คะแนนซึ่งเป็นลักษณะของมาตรประเมินค่า 6 ระดับ ซึ่งคำตอบที่ได้จะมีการสุ่มจัดเป็นชุดโดยมี 15 คำตอบ/ชุด ผู้ประเมินมีจำนวน 16 คน (ซึ่งได้รับการฝึกอบรมการให้คะแนนเป็นเวลาครึ่งวันมาก่อน) แบ่งเป็น 2 กลุ่มทำงานแยกกัน กลุ่มละ 8 คน ซึ่งในแต่ละกลุ่มจะมีหัวหน้าเพิ่มต่างหากอีก 1 คน ซึ่งไม่ได้เป็นผู้ที่ให้คะแนนโดยตรง และในการตรวจให้คะแนนจะใช้เวลาทั้งหมด 7 วัน (จันทร์-ศุกร์, จันทร์-อังคาร) ในแต่ละวันจะแบ่งเวลาในการตรวจเป็น 2 ช่วง (ภายในกลุ่มเดียวกันจะต้องตรวจช่วงเช้าวันที่ 1, 3, 5 และช่วงบ่ายวันที่ 2, 4, 8) โดยวันสุดท้ายจะเป็นการให้คะแนนซ้ำในงานของวันแรก คิดเป็นค่าเฉลี่ยในการตรวจจำนวน 173 ข้อ/ผู้ตรวจ 1 คน/วัน ซึ่งการออกแบบการให้คะแนนไม่ได้เป็นแบบไขว้ระหว่างกลุ่ม (no cross-over between groups) แต่จะเป็นแบบเชื่อมโยงภายในกลุ่ม (directly linked) โดยให้ผู้ตรวจ 2 คนภายในกลุ่มเดียวกันให้คะแนน/คำตอบ 1 ข้อ ส่วนในแต่ละกลุ่มจะเชื่อมโยงกันโดยช่วงเวลา ผลการวิเคราะห์ข้อมูลโดยใช้โปรแกรม ConQuest พบว่า สัดส่วนความแปรปรวนของคะแนนไม่ได้มาจากความคลาดเคลื่อนในการวัด (measurement error) การประมาณค่าความคลาดเคลื่อนของผู้ประเมินในแต่ละวันมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p < .0001$) และพบว่าวันที่ 5 มีความคลาดเคลื่อนของผู้ประเมินน้อยที่สุด ผู้ประเมินคนที่ 2 มีความคลาดเคลื่อนของผู้ประเมินชนิดกตคะแนนมากที่สุด และผู้ประเมินคนที่ 5 มีความคลาดเคลื่อนของผู้ประเมินชนิดกตคะแนนน้อยที่สุด ความคลาดเคลื่อนของผู้ประเมินจะมีมากใน 3 วันแรกและลดลงใน 3 วันต่อมา และเพิ่มขึ้นในวันสุดท้าย จากวันที่ 1-8 มีผู้ประเมินที่มีความคลาดเคลื่อนในการให้คะแนนชนิด severe ตามลำดับจากมากไปหาน้อย คือ คนที่ 8, 11, 7, 2, และ 6 นอกจากนี้พบว่าแนวโน้มการให้คะแนนจะมีความคงที่ (stability) เมื่อผ่านช่วงวันหยุด (เสาร์, อาทิตย์) โดยจะกลับมาเป็นเหมือนกับวันแรก และการที่วันสุดท้ายเป็นการให้คะแนนซ้ำในงานของวันแรก พบว่า การประมาณค่าความคลาดเคลื่อนของผู้ประเมินชนิด severe มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p < .0001$) และความคลาดเคลื่อนของผู้ประเมินภายในกลุ่มจะมีการเปลี่ยนแปลงในแต่ละวัน และจากวันที่ 1-9 พบว่าผู้ประเมินจำนวน 9 คน มีลักษณะของคลาดเคลื่อนของผู้ประเมินชนิดกตคะแนน และ 1 คน มีลักษณะคลาดเคลื่อนของผู้ประเมินชนิดปล่อยคะแนน

McManus, Thompson, & Mollon (McManus, Thompson, & Mollon, 2006) ศึกษาความคลาดเคลื่อนของผู้ประเมินชนิดกตคะแนนและปล่อยคะแนนโดยใช้ MFRM ในการประเมินทักษะทางคลินิกของราชวิทยาลัยแพทย์ ประเทศอังกฤษ (Membership of the Royal Colleges of Physicians of the UK; MRCP (UK)) การประเมินทักษะทางคลินิกประกอบด้วย 5 สถานี ซึ่งในสถานีที่ 2 (ซักประวัติ)

และ 4 (ทักษะการสื่อสารอย่างมีจรรยาบรรณ) จะประเมินด้านการติดต่อสื่อสารและทั้ง 2 สถานีจะใช้เวลา 20 นาที/สถานี ส่วนสถานีที่ 1 (ระบบหายใจและระบบในช่องท้อง) และสถานีที่ 3 (ระบบไหลเวียนเลือดและระบบประสาท) ในแต่ละสถานีจะแบ่งย่อยเป็น 2 ส่วน แต่ละส่วนใช้เวลา 10 นาที และสถานีที่ 5 (ระบบผิวหนัง, ระบบการเคลื่อนไหว ตา และต่อมไร้ท่อ) ซึ่งแบ่งย่อยเป็น 4 ส่วน แต่ละส่วนใช้เวลา 5 นาที ซึ่งคะแนนรวมทั้งหมดมี 14 ส่วน (เฉพาะในสถานีที่ 5 จะคิดเป็น 8 ส่วน) ในแต่ละสถานีจะมีผู้ประเมินที่เป็นอิสระจากกันจำนวน 2 คน และจะไม่มี การพูดคุยหรืออภิปรายกัน ภายหลังที่ผู้สอบออกจากห้อง การตัดสินจะเป็นแบบอิงเกณฑ์ โดยในแต่ละส่วนจากทั้งหมด 14 ส่วน จะมีการประเมินผลใน 4 ระดับ คือ 4 = ผ่านชัดเจน 3 = ผ่าน 2 = ตก และ 1 = ตกชัดเจน ซึ่งจะมีคะแนนรวมระหว่าง 14 – 56 คะแนน ซึ่งเกณฑ์ที่จะสอบผ่าน คือ ตั้งแต่ 41 คะแนน ผลการวิเคราะห์ ข้อมูลในภาพรวมพบว่า ในช่วงระหว่างเดือนมิถุนายน 2001-มีนาคม 2004 มีการจัดสอบ 9 ครั้ง จำนวนผู้สอบรวม 10,145 คน (เฉลี่ย 1,127 คน/ครั้ง) ซึ่งในจำนวนนี้มีคนที่สอบมากกว่าหนึ่งครั้ง รวมอยู่ด้วย ผลการสอบผ่านในภาพรวมอยู่ที่ 46.6% และโดยเฉลี่ยผู้ประเมิน 1 คนจะทำการ ประเมินผู้สอบ 113 คน ส่วนการวิเคราะห์ด้วยโปรแกรม Facets พบว่า มาตรฐานวัด 4 ระดับมีช่วงที่ เท่ากัน สถานการณ์สอบเรียงตามลำดับความยากจากมากไปน้อย คือ ระบบไหลเวียนเลือด ระบบประสาท มีความยากเท่ากันกับระบบในช่องท้อง การซักประวัติ ส่วนสถานการณ์ในสถานีที่ 5 พบว่า มีความ ยากเท่ากันกับสถานีที่ 4 ส่วนสถานการณ์ที่ 1 มีความง่ายมากที่สุด ซึ่งค่าความเที่ยงของการประเมิน ความยากของสถานีเหล่านี้มีสูงถึง 0.99 และพบว่ามี ความแตกต่างระหว่างผู้สอบ 87% ความแตกต่าง ระหว่างสถานี 1% และความแตกต่างระหว่างผู้ประเมิน 12% ซึ่งแสดงถึงความคลาดเคลื่อน ในการให้คะแนนของผู้ประเมินชนิดกต/ปล่อยคะแนน และเมื่อวิเคราะห์ multiple regression พบว่า ผู้ประเมินที่มีความคลาดเคลื่อนในการให้คะแนนชนิดกตคะแนน จะมีความสัมพันธ์กับการมี ประสบการณ์ในการประเมินมาก ส่วนความแตกต่างระหว่างเพศของผู้ประเมินไม่ส่งผลต่อ ความคลาดเคลื่อนของการให้คะแนนชนิดปล่อยคะแนน

Iramaneerat & Yodkosky (Iramaneerat & Yudkowsky, 2007) ศึกษาความคลาดเคลื่อน ของผู้ประเมินในการประเมินทักษะทางคลินิกของนักศึกษาแพทย์ปี 4 จาก Midwestern medical school ในประเทศสหรัฐอเมริกา กลุ่มตัวอย่างประกอบด้วย นักศึกษาแพทย์จำนวน 173 คน และผู้ป่วย สมมติจำนวน 17 คน ซึ่งได้รับการฝึกฝนในด้านการแสดงบทบาทผู้ป่วยและการให้คะแนนแก่นักศึกษา แพทย์เป็นเวลา 4 ชั่วโมง การสอบภาคปฏิบัติ (Objective Structured Clinical Examination; OSCE) ประกอบด้วย 6 สถานี (stations) ซึ่งเป็นสถานการณ์ต่างๆ ของผู้ป่วย ได้แก่ การปฏิเสธการรับเลือด

ถ่ายเหลว ปวดท้อง น้ำหนักลด เจ็บหน้าอก และการปรึกษาทางโทรศัพท์จากผู้ปกครองของเด็กป่วย ซึ่งในสถานการณ์ที่ 5 จะแสดงโดยผู้ป่วยสมมติ 2 คน ส่วนที่เหลือ 5 สถานการณ์ จะแสดงโดยผู้ป่วยสมมติ 3 คน/สถานการณ์ ผู้ป่วยสมมติจะเป็นผู้ทำการประเมินให้คะแนนนักศึกษาแพทย์เมื่อเสร็จสิ้นสถานการณ์โดยใช้แบบประเมินที่เฉพาะในแต่ละสถานการณ์แต่มีครอบคลุมในเรื่องของทักษะ ดังนี้ 1) ชักประวัติ 2) ตรวจร่างกาย 3) ทักษะด้านมนุษยสัมพันธ์ 4) ทักษะการสื่อสาร 5) ทักษะการให้คำปรึกษา และ 6) ความสุภาพในการตรวจร่างกาย ผลการวิเคราะห์โดยใช้โปรแกรม Facets พบว่า ในภาพรวมส่วนมากนักศึกษาแพทย์จะมีความสามารถแตกต่างกัน ($p < .05$) สถานการณ์มีความยากแตกต่างกัน ($p < .05$) โดยสถานการณ์จากง่ายที่สุดไปถึงยากที่สุดเรียงตามลำดับ คือ สถานการณ์ที่ 6, 3, 1, 2, 4, และ 5 ทักษะที่ใช้ประเมินมีความยากแตกต่างกัน ($p < .05$) โดยทักษะจากง่ายที่สุดไปถึงยากที่สุดเรียงตามลำดับ คือ ทักษะด้านมนุษยสัมพันธ์ ทักษะด้านการให้คำปรึกษา ความสุภาพในการตรวจร่างกาย ชักประวัติ ตรวจร่างกาย และทักษะการสื่อสาร และผู้ป่วยสมมติมีการให้คะแนนที่แตกต่างกัน ($p < .05$) โดยส่วนมากเป็นแบบปล่อยคะแนน ส่วนผลการวิเคราะห์ MNSQ พบว่า ผู้ป่วยสมมติจำนวน 4 คน (คนที่ 3, 4, 7, และ 8) คิดเป็น 24% ไม่มีความคงที่ในการให้คะแนน (inconsistency) เนื่องจากมีค่า infit MNSQ มากกว่า 1.2 และทั้งหมด 17 คน มีค่า infit MNSQ ไม่ต่ำกว่า 0.4 ซึ่งแสดงว่าไม่มีลักษณะของการให้คะแนนแบบ restriction of range และ halo effect แต่เมื่อพิจารณาค่า infit MNSQ ที่มากกว่า 1.2 อาจแสดงถึงลักษณะของการให้คะแนนแบบ restriction of range และ halo effect จึงทำการตรวจสอบลักษณะการกระจายความถี่ (frequency distribution) ของผู้ป่วยสมมติทั้ง 4 คน (คนที่ 3, 4, 7, และ 8) และพบว่า ทั้ง 4 คน มีแนวโน้มของการให้คะแนนแบบ restriction of range ส่วนผู้ป่วยสมมติที่มีลักษณะของการให้คะแนนแบบ halo effect มีเพียงผู้ป่วยคนที่ 3

Basturk (Basturk, 2008) ประยุกต์ใช้ MFRM ในการประเมินภาคปฏิบัติจากการนำเสนอด้วยโปรแกรมคอมพิวเตอร์ (PowerPoint presentation) กลุ่มตัวอย่างเป็นนักศึกษาปี 1 คณะศึกษาศาสตร์ เอกวิทยาศาสตร์ มหาวิทยาลัยในประเทศตุรกี โดยสุ่มเลือกมา 1 ห้องจากทั้งหมด 5 ห้อง แต่ละห้องมีจำนวนนักศึกษา 30 คน ซึ่งทุกห้องจะมีผู้สอนคนเดียวกัน ในการนำเสนองานห้องหนึ่ง จะแบ่งเป็นกลุ่มละ 5 คน (6 กลุ่ม) ให้เวลา 20 นาที และสมาชิกในกลุ่มต้องมีการพูดอย่างน้อยคนละ 1 ครั้ง ผู้ประเมินจะประกอบด้วย ผู้สอน 1 คน และนักศึกษา 6 คน (มาจากการสุ่มตัวอย่างกลุ่มละ 1 คน) โดยใช้แบบประเมินที่มีจำนวน 9 ข้อ (มาตรฐานประเมินค่า 0-6 คะแนน) และมีเกณฑ์ ดังนี้ ดีมาก = 50 – 54 คะแนน พอใช้ = 45 – 49 คะแนน และต้องปรับปรุง < 45 คะแนน ผลการวิเคราะห์โดยใช้โปรแกรม

Facets พบว่า ผู้ประเมินคนที่ 5 มีความคลาดเคลื่อนในการให้คะแนนชนิดปล่อยคะแนนมากที่สุด รองลงมาคือ คนที่ 1 ส่วนผู้ประเมินคนที่ 3 มีความคลาดเคลื่อนในการให้คะแนนชนิด severe มากที่สุด และในภาพรวม พบว่าค่า fit statistics ทั้งหมดอยู่ในเกณฑ์ที่ยอมรับ (0.6 -1.4) แสดงว่าการให้คะแนนทั้งหมดนี้สามารถยอมรับได้ และค่า separation index = 2.66 (ค่าปกติ > 2) แสดงว่าผู้ประเมิน แต่ละคนสามารถให้คะแนนได้เป็นอย่างดีและมีความคงที่ในการให้คะแนน (consistency) แบบประเมิน 3 ใน 9 ข้อมีความยากมาก และ 2 ใน 9 ข้อ มีความง่าย และในภาพรวมพบว่า แบบประเมินสามารถจำแนกความแตกต่างในความสามารถของแต่ละกลุ่มได้ ส่วนการพิจารณาในแต่ละกลุ่ม พบว่า กลุ่มนำเสนอก่อนกลุ่มที่ 3 มีความสามารถสูงสุด รองลงมาคือ กลุ่มที่ 5 และกลุ่มที่มีความสามารถน้อยสุดคือ กลุ่มที่ 4

Iramaneerat, et al. (Iramaneerat, Yudkowsky, Myford, & Downing, 2008) ศึกษาการควบคุมคุณภาพของการสอบภาคปฏิบัติ (Objective Structured Clinical Examination; OSCE) โดยใช้การวิเคราะห์ GT และ MFRM กลุ่มตัวอย่างประกอบด้วยแพทย์ฝึกหัดเฉพาะทาง (residents) จำนวน 79 คนจาก Midwestern medical school ในประเทศสหรัฐอเมริกา โดย 68 คนมาจากแผนกอายุรกรรม (ปี 2 จำนวน 35 คน และปี 3 จำนวน 33 คน) และ 11 คน มาจากแผนกเวชกรรมครอบครัว (ปี 2 จำนวน 6 คน และปี 3 จำนวน 5 คน) ประเมินผลจากทักษะการสื่อสารโดยใช้ 6 สถานี (stations) แต่ละสถานีประกอบด้วยสถานการณ์ทางคลินิกที่จะต้องแสดงทักษะการสื่อสารที่เฉพาะ ดังนี้ 1) ให้ความรู้แก่ผู้ป่วย 2) การยินยอมรับการรักษา 3) การแก้ไขปัญหาเมื่อผู้ป่วยปฏิเสธการรักษา 4) การให้คำปรึกษาในผู้ป่วยสูงวัยที่ถูกทำร้าย 5) การแจ้งข่าวร้ายแก่ผู้ป่วย และ 6) การอธิบายการตรวจร่างกาย โดยผู้ป่วยสมมติ 1 คน/สถานการณ์ และจะเป็นผู้ทำการประเมินเมื่อเสร็จสิ้นสถานการณ์ โดยใช้แบบวัดความพึงพอใจของ American Board of Internal Medicine ซึ่งประกอบด้วยข้อคำถามเชิงบวก 18 ข้อ เป็นมาตรฐานประเมิน 1-5 คะแนน (1 = ไม่เห็นด้วยอย่างยิ่ง ... 5 = เห็นด้วยอย่างยิ่ง) ผลการศึกษาจาก GT พบว่า มีแหล่งความแปรปรวนมากในพาเซทของผู้ป่วยสมมติ ดังนั้นจึงควรลดความแปรปรวนโดย 1) การเพิ่มจำนวนผู้ป่วยสมมติ และ 2) การฝึกอบรมผู้ป่วยสมมติด้านการแสดงบทบาทและ/หรือด้านการให้คะแนนอย่างคงที่ ซึ่ง GT ไม่ช่วยในด้านการตัดสินใจเลือกระหว่าง 2 วิธีนี้ และผลการศึกษาจาก MFRM โดยใช้โปรแกรม Facets พบว่า ผู้ป่วยสมมติมีการให้คะแนนแบบปล่อยคะแนน การสอบมีความง่ายมากสำหรับแพทย์ฝึกหัด และแพทย์ฝึกหัดมีความสามารถมาก นอกจากนี้พบว่า MFRM สามารถระบุถึงรายละเอียดในระดับบุคคล คือ ผู้ป่วยสมมติในสถานการณ์ที่ 3 และ 5

มีการให้คะแนนที่ไม่คงที่ ทำให้สามารถเลือกวิธีการแก้ปัญหาที่ตรงประเด็น คือ ส่งผู้ประเมินทั้งสองคนที่มีปัญหาไปฝึกอบรมการใช้เกณฑ์การให้คะแนน

Jie & Lianzhen (Jie & Lianzhen, 2008) ศึกษาแหล่งความแปรปรวนของคะแนนการสอบภาคปฏิบัติ speaking test of Public English Testing System (PETS) Band 3 ในประเทศจีน โดยใช้การวิเคราะห์ MFRM ผู้สอบจะถูกแบ่งกลุ่มขนาด 2 หรือ 3 คน การทดสอบจะประกอบด้วย 3 ส่วน คือ ส่วนที่ 1 ให้กล่าวแนะนำตัวอย่างสั้นๆ และตอบคำถาม ส่วนที่ 2 ในกลุ่มแต่ละคนจะได้รับแผ่นภาพและมีเวลาให้ 3 นาที สำหรับอภิปรายในกลุ่ม ส่วนที่ 3 ผู้ถามจะให้รูปภาพที่ต่างกันแก่ผู้สอบและพูดถามให้ตอบเกี่ยวกับสิ่งที่เห็น ใช้เวลา 1.5 นาที หลังจากนั้นจะให้ผู้สอบพูดวิจารณ์เกี่ยวกับสมาชิกในกลุ่มเป็นเวลา 0.5 นาที ผู้เข้าสอบมีจำนวน 322 คน (เดือนมีนาคม 2007) และจำนวนผู้ประเมิน 34 คน โดยผู้สอบจะได้รับการสุ่มเข้าแต่ละห้องสอบที่มีผู้ประเมินจะมีจำนวน 2 คน/ห้อง จากทั้งหมด 17 ห้อง ในผู้ประเมินหนึ่งคนจะประเมินเกี่ยวกับทักษะการพูดและการมีปฏิสัมพันธ์ภายในกลุ่ม ใช้การให้คะแนนโดยภาพรวม (holistic scoring) เป็นมาตรฐานประเมินค่า (rating scale) ส่วนอีกหนึ่งคนจะประเมินโดยแยกวิเคราะห์ (analytic scoring) ใน 4 โดเมน คือ ไวยากรณ์และคำศัพท์ การบริหาร การสนทนา การออกเสียงพูด และการมีปฏิสัมพันธ์ โดยใช้การให้คะแนนเป็นมาตรฐานประเมินค่า ผลการวิเคราะห์ข้อมูลโดยใช้โปรแกรม Facets พบว่า กลุ่มที่ให้คะแนนโดยภาพรวม มีความคลาดเคลื่อนของผู้ประเมินจากห้องที่ 9 เป็นแบบกดคะแนนอย่างมาก (extremely severe) และห้องที่ 11 เป็นแบบปล่อยคะแนนอย่างมาก (extremely lenient) ส่วนห้องที่ 6 เป็นแบบให้คะแนนที่มีแนวโน้มเข้าสู่ส่วนกลาง (central tendency) หรือการจำกัดช่วงการให้คะแนน (restriction of range) ส่วนกลุ่มที่ให้คะแนนโดยแยกวิเคราะห์ มีผลการวิเคราะห์ ที่คล้ายกับกลุ่มที่ให้คะแนนโดยภาพรวม ในด้านการกด-ปล่อยคะแนน (severe-lenient) นอกจากนี้ผลการวิเคราะห์พบว่า มีผู้ให้คะแนนจำนวน 5 คน (infit > 1.4) ที่มีการให้คะแนนไม่คงที่ (inadequate consistency) และเมื่อเปรียบเทียบระหว่างผลการให้คะแนนโดยภาพรวมกับการแยกวิเคราะห์ พบว่า การให้คะแนนโดยแยกวิเคราะห์มีการจำแนกผู้สอบได้ดีกว่า การให้คะแนนโดยภาพรวม ซึ่งพิจารณาจากค่า separation ratio และ reliability

Haiyang (Haiyang, 2010) ศึกษาความเที่ยงของแบบสอบภาษาอังกฤษและความเที่ยงระหว่างผู้ประเมินโดยใช้ CTT และ MFRM ในการวิเคราะห์ กลุ่มตัวอย่างประกอบด้วย นักศึกษาชั้นปี 1 ที่กำลังจะขึ้นปี 2 จำนวน 56 คน เครื่องมือที่ใช้เป็นแบบสอบชนิดเลือกตอบและเขียนตอบ ซึ่งในส่วนของ การวิเคราะห์ใช้เฉพาะข้อสอบเขียนตอบที่ประกอบด้วย 3 ส่วน คือ การแปลจากภาษาจีนไปเป็นภาษาอังกฤษ การแปลจากภาษาอังกฤษไปเป็นภาษาจีน และการเขียนภาษาอังกฤษ การตรวจจะ

ใช้ผู้ประเมินจำนวน 2 คน ที่เป็นอิสระจากกันและมีการให้คะแนนโดยภาพรวม (holistic scoring) ผลการวิเคราะห์ในส่วนของ CTT พบว่าด้านความเที่ยงระหว่างผู้ประเมิน (interrater reliability) ได้ค่า Spearman's rho (ρ) จากข้อสอบทั้ง 3 ส่วน มีค่าเท่ากับ 0.853, 0.774, และ 0.678 ตามลำดับ ($p < .000$) ซึ่งแสดงว่ามีความเที่ยงระหว่างผู้ประเมินทั้งสองคนมาก ส่วนการวิเคราะห์ด้านความเที่ยงแบบความสอดคล้องภายในของแบบสอบ (internal consistency) โดยแบ่งเป็นการวิเคราะห์ 3 ครั้ง คือ รวมทั้งสามส่วน เฉพาะการแปลทั้ง 2 ส่วน และเฉพาะการเขียนภาษาอังกฤษ ได้ค่า Cronbach's alpha (α) เท่ากับ 0.366, 0.712, และ 0.201 ตามลำดับ ซึ่งแสดงว่ามีความเที่ยงแบบความสอดคล้องภายในต่ำในส่วนของการวิเคราะห์ แบบรวมทั้งสามส่วนกับเฉพาะการเขียนภาษาอังกฤษ ส่วนการวิเคราะห์เฉพาะการแปลทั้ง 2 ส่วนพบว่ามีค่าความเที่ยงแบบสอดคล้องภายในสูง ผลการวิเคราะห์จาก MFRM โดยใช้โปรแกรม Facets พบว่า มีความคลาดเคลื่อนของการให้คะแนนชนิด severity เกิดขึ้นในผู้ประเมินคนที่ 2 มากกว่าคนที่ 1 ซึ่งผลการวิเคราะห์นี้ขัดแย้งกับ CTT ส่วนในด้านความยากของข้อสอบพบว่า ส่วนที่ 1 มีความง่าย ส่วนที่ 2 ค่อนข้างยาก และส่วนที่ 3 มีความยากมาก และในด้านของผู้สอบพบว่า มีความสามารถที่แตกต่างกัน และเมื่อเปรียบเทียบระหว่าง 2 โมเดลพบว่า ใน CTT ไม่สามารถวิเคราะห์ความเที่ยงภายในของผู้ประเมิน (intra-rater reliability) โดยปราศจากความคลาดเคลื่อน เนื่องจากเป็นการวิเคราะห์โดยการให้คะแนนซ้ำในข้อสอบเดิม แต่ใน MFRM สามารถทำได้โดยประเมินจาก infit หรือ outfit mean squares ของผู้ประเมินแต่ละคน นอกจากนี้ CTT ยังเป็นการวิเคราะห์แบบให้ภาพแบบกว้างๆ เกี่ยวกับความเที่ยงของแบบสอบและผู้ประเมิน แต่ MFRM สามารถวิเคราะห์ถึงสาเหตุของความไม่เที่ยงได้ และยังให้ผลการวิเคราะห์ที่มีความถูกต้องมากกว่า CTT

Farrokhi & Esfandiari (Farrokhi & Esfandiari, 2011) ศึกษาความคลาดเคลื่อนในการให้คะแนนแบบฮาโล (halo effect) ในผู้ประเมิน 3 ประเภท คือ ตนเอง เพื่อน และอาจารย์ กลุ่มตัวอย่างจำนวน 194 คน ซึ่งเป็นอาจารย์และนักศึกษาในวิชา Advanced Writing จาก 2 มหาวิทยาลัย โดยประกอบด้วยนักศึกษาชาวอิหร่านวิชาเอกภาษาอังกฤษ จำนวน 188 คน และอาจารย์ชาวอิหร่านจำนวน 6 คน ใช้การให้คะแนนโดยแยกวิเคราะห์ (analytic scoring) เป็นมาตรฐานประเมิน 6 ช่วง ข้อสอบจะให้นักศึกษาเขียนบรรยายเป็นภาษาอังกฤษจำนวน 5 ย่อหน้า ความยาว 500 – 700 คำ ในหัวข้อเดียวกันซึ่งนำมาจากข้อสอบ TOEFL ผลการวิเคราะห์โดยใช้โปรแกรม Facets พบว่า ไม่มีความคลาดเคลื่อนในการให้คะแนนแบบฮาโลในกลุ่มของผู้ประเมินทั้ง 3 กลุ่ม (group level) (fixed chi-square $p < .05$, trait separation ratio = 12.33, trait separation index = 16.77, reliability of the trait separation index = .99) แต่เมื่อวิเคราะห์ในรายบุคคล (individual level) ของผู้ประเมินทั้ง

3 กลุ่ม พบว่า ผู้ประเมินตนเองมีความคลาดเคลื่อนในการให้คะแนนแบบฮาโลมากกว่าผู้ประเมินที่เป็นเพื่อนและอาจารย์ เนื่องจากผล positive & negative t.score ของผู้ประเมินตนเองมี 9 ราย ส่วนผู้ประเมินที่เป็นเพื่อนและอาจารย์มีจำนวนเท่ากันคือ 5 ราย ซึ่งข้อที่มีความคลาดเคลื่อนในการให้คะแนนแบบฮาโล คือ ข้อที่ 2, 3, 7, และ 10

Farrokhi, Esfandiari, & Dalili (Farrokhi, Esfandiari, & Dalili, 2011) ศึกษาความคลาดเคลื่อนในการให้คะแนนแบบแนวโน้มการให้คะแนนค่ากลาง (central tendency effect หรือ Centrality) ในผู้ประเมิน 3 ประเภท คือ ตนเอง เพื่อน และอาจารย์ ซึ่งใช้ข้อมูลเดียวกับ Farrokhi, & Esfandiari (2011) ผลการวิเคราะห์โดยใช้โปรแกรม Facets พบว่า ไม่มีความคลาดเคลื่อนในการให้คะแนนแบบแนวโน้มการให้คะแนนค่ากลางในกลุ่มของผู้ประเมินทั้ง 3 กลุ่ม (group level) (fixed chi-square $p < .05$, reliability of separation index = .92, Infit MnSq $> .5$ ทั้ง 15 ข้อ) และเมื่อวิเคราะห์ในรายบุคคล (individual level) ของผู้ประเมินทั้ง 3 กลุ่ม พบว่า ไม่มีความคลาดเคลื่อนในการให้คะแนนแบบแนวโน้มการให้คะแนนค่ากลางของผู้ประเมินทั้ง 3 กลุ่ม เนื่องจากค่า infit & outfit mean squares ทั้งหมดมีค่าใกล้ 1 (self-assessor = .94 & .95, peer-assessor = .91 & .91, teacher-assessor = 1.02 & 1.02)

Kassin (Kassim, 2011) ศึกษาพฤติกรรมการให้คะแนนและความคลาดเคลื่อนของการให้คะแนน โดยใช้การวิเคราะห์ MFRM ศึกษาในกลุ่มนักศึกษาที่เข้าใหม่ของสถาบัน และผู้ประเมินมีจำนวน 34 คน ซึ่งเป็นอาจารย์สอนภาษาอังกฤษในสถาบันการศึกษาระดับอุดมศึกษาประเทศมาเลเซีย ซึ่งมีประสบการณ์การสอนไม่ต่ำกว่า 1 ปี ข้อสอบจะให้นักศึกษาเขียนบรรยายเป็นภาษาอังกฤษจำนวน 12 ย่อหน้า (ความยาวย่อหน้าละ 100 – 120 คำ) ในหัวข้อ “My favorite game” และมีการสุ่มตรวจโดยผู้ประเมินที่ต่างคนกัน มีการให้คะแนนโดยภาพรวม (holistic scoring) ซึ่งเป็นมาตรฐานค่า 1-10 คะแนน ผลการวิเคราะห์โดยใช้โปรแกรม Facets พบว่า นักศึกษามีความสามารถในระดับที่แตกต่างกัน (พิสัย = 6 logits) ผู้ประเมินคนที่ 6, 10, และ 11 มีความคลาดเคลื่อนในการให้คะแนนแบบ severe ส่วนผู้ประเมินคนที่ 25 มีความคลาดเคลื่อนในการให้คะแนนแบบปล่อยคะแนน และผู้ประเมิน 3 คน คือ คนที่ 3, 4, 13 มีความคลาดเคลื่อนในการให้คะแนนแบบจำกัดช่วงการให้คะแนน และผู้ประเมินคนที่ 18, 21, 11, 23, และ 7 มีการให้คะแนนที่ไม่คงที่ (internal consistency) ส่วนผู้ประเมินคนที่ 33 มีความคงที่ในการให้คะแนน

Classen, et al. (Classen et al., 2012) ศึกษาความเที่ยงของผู้ประเมิน (จาก 3 กลุ่ม คือ ผู้สูงวัย สมาชิกในครอบครัวหรือผู้ดูแล และผู้ประเมินการขับรถ) และความคลาดเคลื่อนในการให้คะแนน (จาก 2 กลุ่ม คือ สมาชิกในครอบครัวหรือผู้ดูแล และผู้ประเมินการขับรถ) สำหรับการวัดพฤติกรรมการขับรถอย่างปลอดภัย โดยใช้แบบวัดที่มีจำนวน 68 ข้อ แต่ละข้อเป็นมาตราประเมิน 1 – 5 คะแนน (1 = ทำไม่ได้ ... 5 = ทำได้ง่าย) กลุ่มตัวอย่างมี 3 กลุ่ม คือ กลุ่มผู้สูงวัย (อายุ 65 – 85 ปี) จำนวน 80 คน ซึ่งมาจากการสุ่มตามความสะดวก (convenience sampling) ที่มีคุณสมบัติตามเกณฑ์ คือ มีใบอนุญาตขับขี่ ปัจจุบันสามารถขับรถได้ สามารถตอบแบบประเมินได้ (การสัมภาษณ์ทางโทรศัพท์) และสามารถสอบปฏิบัติโดยการขับรถ กลุ่มสมาชิกในครอบครัวหรือผู้ดูแลมีจำนวน 80 คน ต้องมีอายุระหว่าง 18 – 85 ปี สามารถที่จะเก็บข้อมูลโดยการสังเกตและรายงานผลเกี่ยวกับพฤติกรรม การขับรถของผู้สูงวัย และกลุ่มผู้ประเมิน ซึ่งประกอบด้วย ผู้เชี่ยวชาญในด้านการฟื้นฟูการขับรถที่มีประสบการณ์ 7 ปี 1 คน และผู้ประเมินการขับรถที่เป็นอาจารย์สอนขับรถประจำจังหวัดซึ่งมีประสบการณ์มากกว่า 10 ปี 1 คน ผลจากการวิเคราะห์ค่าความสัมพันธ์ระหว่างผู้ประเมิน (interclass correlation; ICC) ในแต่ละกลุ่ม พบว่า มีความสอดคล้องระหว่างผู้ประเมิน (rater agreement) ทั้ง 3 กลุ่ม โดยมีค่า ICC = .256, $p < .001$ และผลการวิเคราะห์จาก MFRM โดยใช้โปรแกรม Facets พบว่า ผู้ประเมินทั้งสองกลุ่ม (สมาชิกในครอบครัวหรือผู้ดูแล และผู้ประเมินการขับรถ) ไม่มีความคลาดเคลื่อนในการให้คะแนน (fixed chi-square $p < .001$) แต่กลุ่มผู้ประเมินมีแนวโน้มที่จะเกิดความคลาดเคลื่อนชนิดกตคะแนนมากกว่ากลุ่มสมาชิกในครอบครัวหรือผู้ดูแล

Farrokhi, Esfandiari, & Schaefer (Farrokhi, Esfandiari, & Schaefer, 2012) ศึกษาความคลาดเคลื่อนในการให้คะแนนแบบกตหรือปล่อยคะแนน (severity/leniency) ในผู้ประเมิน 3 ประเภท คือ อาจารย์ เพื่อน และตนเอง ซึ่งใช้ข้อมูลเดียวกับ Farrokhi, & Esfandiari (2011) ผลการวิเคราะห์โดยใช้โปรแกรม Facets พบว่า ผู้เรียนซึ่งให้คะแนนตนเองจะมีลักษณะของการปล่อยคะแนนในข้อที่ 2, 3, และ 4 ซึ่งเป็นลักษณะที่ตรงข้ามกับการให้คะแนนโดยอาจารย์ที่เป็นของการกตคะแนน ส่วนในข้อที่ 7 และ 10 ผู้เรียนซึ่งให้คะแนนตนเองจะมีลักษณะของการกตคะแนน แต่อาจารย์จะมีลักษณะของการปล่อยคะแนน และพบว่ามีสัดส่วนของ severity/leniency ในการให้คะแนนโดยตนเอง เพื่อน และอาจารย์ เท่ากับ 5/4 3/2 และ 3/2 ตามลำดับ

น้ำผึ้ง อินทเนตร (น้ำผึ้ง น้ำผึ้ง อินทเนตร, 2554) ศึกษาคุณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์เมื่อจำนวนผู้ตรวจต่างกัน 3 ลักษณะ คือ 2 คน 3 คน และ 4 คน และรูปแบบการตรวจให้คะแนนต่างกัน 3 ลักษณะ คือ รูปแบบที่ 1 ผู้ตรวจตรวจข้อสอบทุกข้อของ

ผู้สอบทุกคน รูปแบบที่ 2 ผู้ตรวจตรวจข้อสอบทุกข้อของผู้สอบบางคน และรูปแบบที่ 3 ผู้ตรวจตรวจข้อสอบบางข้อของผู้สอบทุกคน โดยใช้โมเดลการสุรุปอ้างอิง (GT) และโมเดลหลายองค์ประกอบของรารส์ช (MFRM) เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ ระดับมัธยมศึกษาตอนต้น จำนวน 12 ข้อ มีการให้คะแนนโดยภาพรวม (holistic scoring) ซึ่งชุดการให้คะแนนเป็นแบบรูบิกส์ ข้อละ 0-4 คะแนน กลุ่มตัวอย่าง คือ นักศึกษาระดับชั้นมัธยมศึกษาปีที่ 4 จากโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาจังหวัดน่าน ปีการศึกษา 2552 ที่ได้จากการสุ่มแบบสองขั้นตอน (two-stage Random Sampling) จำนวน 180 คน และผู้ประเมินเป็นอาจารย์ผู้สอนที่มีวุฒิการศึกษาระดับปริญญาตรีขึ้นไปทางด้านคณิตศาสตร์ และมีประสบการณ์ในการสอนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 ไม่น้อยกว่า 5 ปี จำนวน 4 คน โดยเป็นอาจารย์ที่เต็มใจให้ความร่วมมือผลการวิจัย เมื่อวิเคราะห์ด้วยโมเดลการสุรุปอ้างอิง พบว่า เมื่อใช้รูปแบบการตรวจให้คะแนนเดียวกันในทุกเงื่อนไขจำนวนผู้ตรวจ ความแปรปรวนขององค์ประกอบเดียวกันมีค่าใกล้เคียงกัน และค่าสัมประสิทธิ์การสุรุปอ้างอิงในรูปแบบการตรวจที่ 2 มีค่าสูงสุด รองลงมาคือรูปแบบการตรวจที่ 1 และรูปแบบการตรวจที่ 3 มีค่าต่ำสุด ค่าสัมประสิทธิ์การสุรุปอ้างอิงในรูปแบบการตรวจที่ 1 มีค่าสูงขึ้นเมื่อจำนวนผู้ตรวจเพิ่มขึ้น คะแนนในทุกเงื่อนไขที่ต่างกันมีความเที่ยงตรงตามสภาพสูงและแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนการวิเคราะห์ด้วยโมเดลหลายองค์ประกอบของรารส์ชโดยใช้โปรแกรม Facets พบว่า ในทุกเงื่อนไขจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนน ความแปรปรวนของผู้สอบมีค่าสูงที่สุดรองลงมาคือ ความแปรปรวนของข้อสอบ และความแปรปรวนของผู้ตรวจมีค่าต่ำที่สุด แสดงว่าความสามารถของผู้สอบมีความแตกต่างกันมากกว่าความแตกต่างของค่าความยากของข้อสอบและค่าความคลาดเคลื่อนของการให้คะแนน และค่าความเชื่อมั่นแยกส่วนของผู้สอบของรูปแบบการตรวจที่ 1 มีค่าสูงสุดในทุกเงื่อนไขจำนวนผู้ตรวจ ค่าความเชื่อมั่นแยกส่วนของผู้สอบของรูปแบบการตรวจที่ 1 มีค่าสูงขึ้นเมื่อจำนวนผู้ตรวจเพิ่มขึ้น คะแนนความสามารถในทุกเงื่อนไขที่ต่างกันมีความเที่ยงตรงตามสภาพสูงและแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

จากตัวอย่างงานวิจัยดังที่ได้กล่าวในรายละเอียดข้างต้น ซึ่งเกี่ยวข้องกับการศึกษาความคลาดเคลื่อนของผู้ประเมินซึ่งเป็นการตรวจสอบทางอ้อมของความแม่นยำในการให้คะแนนโดยใช้ MFRM พบว่า ส่วนใหญ่เป็นการประเมินทางการศึกษาซึ่งได้แก่ ภาษา แพทย์ และคณิตศาสตร์ ซึ่งมีทั้งการทดสอบขนาดใหญ่และขนาดเล็ก นอกจากนี้ยังมีการนำมาใช้ในบริบท ของการประเมินการขับรถ เป็นต้น และโปรแกรมที่ใช้ในการวิเคราะห์ที่เกือบทั้งหมดจะใช้ Facets ซึ่งผู้วิจัยได้สรุปเป็นตารางที่ 2 สันเคราะห์งานวิจัย ดังนี้

ตารางที่ 2 การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับวิเคราะห์ความคลาดเคลื่อนของผู้ประเมิน

งานวิจัย	สิ่งที่ประเมิน	ประเภทความคลาด เคลื่อนของผู้ประเมิน	โมเดล	โปรแกรม วิเคราะห์
Congdon & McQueen (2000)	Writing assessment	severe-lenient	MFRM	ConQuest
McManus, Thompson, & Mollon (2006)	OSCE (performance tasks)	severe-lenient	MFRM	Facets
Iramaneerat & Yodkosky (2007)	OSCE (performance tasks)	halo effect, lenient, restriction of range, inconsistency	MFRM	Facets
Basturk (2008)	PowerPoint presentation	severe-lenient	MFRM	Facets
Iramaneerat, et al. (2008)	OSCE	severe-lenient, interrater reliability	GT, MFRM	Facets
Jie & Lianzhen (2008)	Speaking test	severe-lenient, inadequate consistency	MFRM	Facets
Haiyang (2010)	Language	severity, interrater reliability	CTT, MFRM	Facets
Farrokhi & Esfandiari (2011)	language rating of self, peer, teacher	halo effect	MFRM	Facets
Farrokhi, Esfandiari, & Dalili (2011)	language rating of self, peer, teacher	central tendency	MFRM	Facets

งานวิจัย	สิ่งที่ประเมิน	ประเภทความคลาดเคลื่อนของผู้ประเมิน	โมเดล	โปรแกรมวิเคราะห์
Kassim (2011)	Language	severity, restriction of range, internal consistency	MFRM	Facets
Classen, et al. (2012)	safe driving behavior : rating of self, caregiver, rater	severe-lenient, internal consistency, interrater reliability	MFRM	Facets
Farrokhi, Esfandiari, & Schaefer (2012)	language rating of self, peer, teacher	severity-lenienicy	MFRM	Facets
น้ำผึ้ง อินทเนตร (2554)	open-ended mathematics test	raters variance	GT, MFRM	Facets

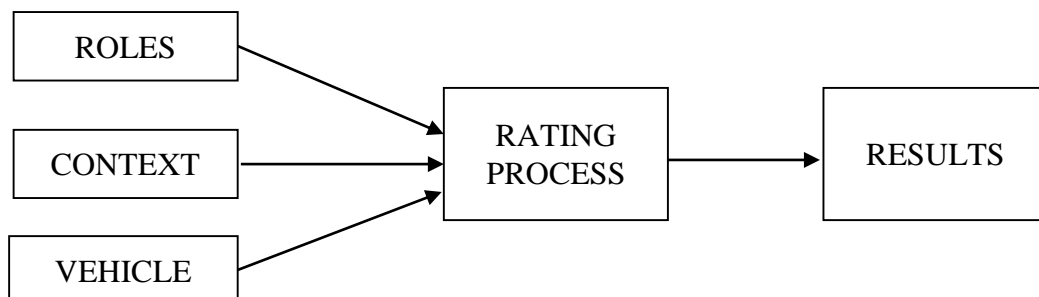
2.2 ปัจจัยเชิงสาเหตุที่ส่งผลกระทบต่อคุณภาพการให้คะแนนการปฏิบัติ

ปัจจัยเชิงสาเหตุที่ส่งผลกระทบต่อคุณภาพการให้คะแนนการปฏิบัติในด้านของความแม่นยำและความคลาดเคลื่อนในการให้คะแนนส่วนมากจะปรากฏอยู่ในศาสตร์ทางจิตวิทยาอุตสาหกรรมและองค์การ โดยอยู่ในส่วนของการประเมินผลการปฏิบัติ (Performance Appraisal; PA) และได้มีหลายงานวิจัยที่ได้นำตัวแปรเชิงสาเหตุเหล่านี้มาประยุกต์ใช้ในบริบททางการศึกษา ซึ่งในการนำมาใช้ในบริบทที่ต่างกันนี้ Murphy, Cleveland, Skattebo, & Kinney (Murphy, Cleveland, Skattebo, & Kinney, 2004) ได้แนะนำให้พิจารณาถึงความแตกต่างที่อาจเกิดขึ้น เช่น บริบททางการศึกษาจะแตกต่างจากบริบทการประเมินผลการปฏิบัติงานทั่วไป คือ ผลลัพธ์จากการประเมินไม่ได้มีผลต่อผู้ประเมินโดยตรง หรือผู้ประเมินไม่ค่อยมีส่วนได้ส่วนเสียกับคะแนนที่ประเมิน แต่ทั้งนี้ประเด็นสำคัญจะอยู่ที่เมื่อทราบถึงปัจจัยที่มีผลต่อความแตกต่างแล้ว ควรมีการออกแบบการวิจัยเพื่อควบคุมปัญหาเหล่านั้น โดยรายละเอียดของการศึกษาเกี่ยวกับปัจจัยเชิงสาเหตุที่ส่งผลกระทบต่อคุณภาพการให้คะแนนการปฏิบัติมีดังนี้

2.2.1 แนวคิดทฤษฎีที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ

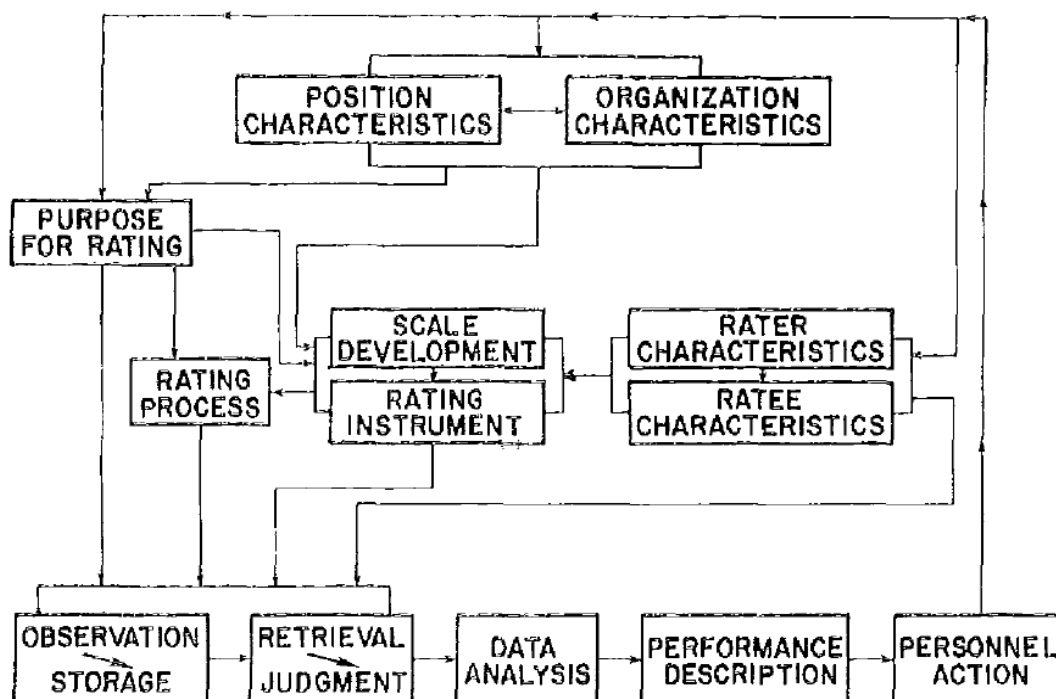
จากการศึกษาแนวคิดทฤษฎีที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุที่ส่งผลต่อการให้คะแนนการปฏิบัติ ผู้วิจัยได้ทำการศึกษาแนวคิดทฤษฎีซึ่งได้แก่ โมเดลการให้คะแนนการปฏิบัติ (performance rating mode) ของ Landy & Farr โมเดลสี่องค์ประกอบ (Four-component Model) ของ Murphy & Cleveland และโมเดลสามขั้นตอน (Three-stage model) ของ Heidemeier & Moser โดยมีรายละเอียดดังนี้

1) โมเดลการให้คะแนนการปฏิบัติ (performance rating model) พัฒนาโดย Landy & Farr (Landy & Farr, 1980) แบ่งเป็น 2 โมเดล คือ โมเดลองค์ประกอบในการให้คะแนนการปฏิบัติ และโมเดลกระบวนการให้คะแนนการปฏิบัติ โดยมีวัตถุประสงค์เพื่ออธิบายโครงสร้างหรือองค์ประกอบที่มีผลต่อการตัดสินให้คะแนนการปฏิบัติ ซึ่งในแต่ละองค์ประกอบได้รวบรวมมาจากผลงานวิจัยที่ใช้ทฤษฎีบุคลิกภาพ (personality theory) และจัดองค์ประกอบของโมเดลในลักษณะของกระบวนการที่มีปฏิสัมพันธ์ต่อกันภายในบริบทของการตัดสินผลการปฏิบัติ โดยโมเดลองค์ประกอบในการให้คะแนนการปฏิบัติ ดังภาพที่ 3 มี 5 องค์ประกอบได้แก่ 1) บทบาท (role) ประกอบด้วย ตัวแปรคุณลักษณะทางประชากร และบุคลิกภาพของผู้ประเมินและผู้ถูกประเมิน โดยตัวแปรของผู้ประเมินที่ส่งผลต่อการให้คะแนนการปฏิบัติ ได้แก่ เชื้อชาติ อายุ ระดับการศึกษา ประสบการณ์ และกลุ่มของผู้ประเมิน ส่วนตัวแปรเพศ และความรู้ยังไม่แสดงผลสนับสนุนที่ชัดเจน และตัวแปรของผู้ถูกประเมินส่งผลต่อการให้คะแนนการปฏิบัติ ได้แก่ เพศ อาชีพ 2) บริบทของการให้คะแนน (rating context) ประกอบด้วยประเภทขององค์กร และวัตถุประสงค์ของการประเมิน ซึ่งผลการวิจัยส่วนใหญ่พบว่า การประเมินที่มีวัตถุประสงค์เพื่อการบริหารงานจะมีความคลาดเคลื่อนชนิดปล่อยคะแนนมากที่สุด 3) เครื่องมือ (vehicle) ประกอบด้วยแบบประเมินรูปแบบต่างๆ เช่น กราฟิก ตัวเลือกคู่ (forced-choice items) มาตรฐานค่ากลุ่มพฤติกรรม (Behaviorally Anchored Rating Scale; BARS) โดยผลการวิจัยส่วนใหญ่พบว่า ผู้ประเมินชอบแบบประเมินที่มีความหลากหลาย ส่วนชนิดของแบบประเมินส่งผลเพียงเล็กน้อยต่อการให้คะแนน 4) กระบวนการให้คะแนน (rating process) ประกอบด้วย ข้อจำกัดต่างๆ กลยุทธ์ในการให้คะแนน การฝึกอบรมการให้คะแนน เป็นต้น ซึ่งผลจากงานวิจัยส่วนใหญ่พบว่า การฝึกอบรมการให้คะแนนจะช่วยลดความคลาดเคลื่อนในการให้คะแนน และ 5) ผลลัพธ์ของการให้คะแนน (results of the rating) ประกอบด้วยการนำข้อมูลผลการประเมินไปใช้ต่อ ซึ่งผลงานวิจัยโดยรวมพบว่า ผลการให้คะแนนมีประโยชน์ในด้านการนำไปใช้ถึงแม้ว่าผลคะแนนนั้นจะมีความคลาดเคลื่อน



ภาพที่ 3 โมเดลองค์ประกอบในการให้คะแนนการปฏิบัติ (ที่มา: Landy & Farr, 1980: 73)

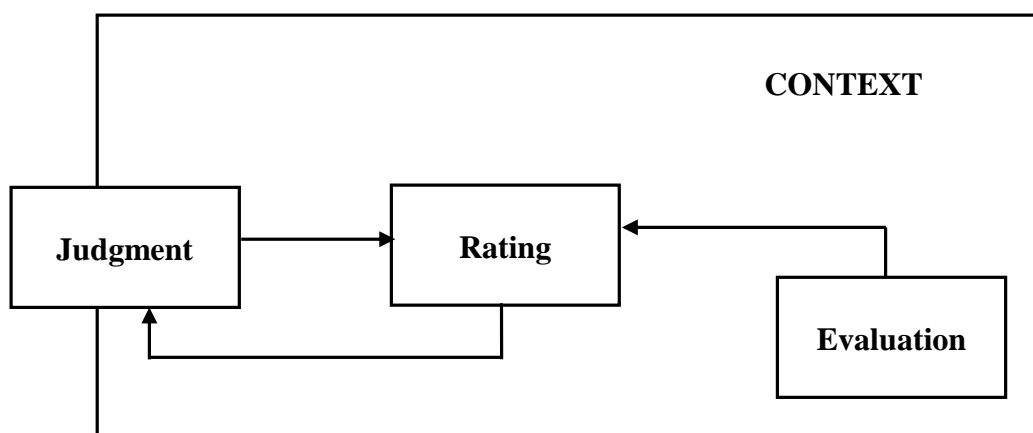
ส่วนโมเดลกระบวนการให้คะแนนการปฏิบัติ ดังภาพที่ 4 แสดงกระบวนการการตัดสินใจที่เกี่ยวข้องกับโมเดลองค์ประกอบในการให้คะแนนการปฏิบัติใน 2 องค์ประกอบ ได้แก่ องค์ประกอบบริบทของการให้คะแนน ซึ่งประกอบด้วย 3 ส่วน คือ ลักษณะของตำแหน่งงาน (position characteristics) ลักษณะขององค์กร (organization characteristics) และวัตถุประสงค์ในการให้คะแนน (purpose for rating) และองค์ประกอบกระบวนการให้คะแนน ประกอบด้วย 2 ส่วน คือ กระบวนการคิดของผู้ประเมิน (rater's cognitive process) ซึ่งได้แก่ การสังเกต (observation) การบันทึก (storage) การทบทวน (retrieval) และการตัดสินใจ (judgment) และกระบวนการตัดสินใจขององค์กร ซึ่งได้แก่ การวิเคราะห์ข้อมูล (data analysis) การอภิปรายผล (performance description) และการนำข้อมูลผลการประเมินไปใช้ต่อ (personnel action) ส่วนองค์ประกอบบทบาท ซึ่งได้แก่ คุณลักษณะต่างๆ ของผู้ประเมินและผู้ถูกประเมิน และองค์ประกอบเครื่องมือจะอยู่ในส่วนของสารสนเทศของการปฏิบัติงานที่จะส่งผลต่อกระบวนการให้คะแนน



ภาพที่ 4 โมเดลกระบวนการให้คะแนนการปฏิบัติ (ที่มา: Landy & Farr, 1980: 94)

โมเดลการให้คะแนนการปฏิบัติของ Landy & Farr นอกจากจะสนใจตัวแปรเชิงสาเหตุที่เน้นด้านสถานการณ์ ได้แก่ บทบาท บริบทของการให้คะแนน เครื่องมือ กระบวนการให้คะแนน และผลลัพธ์ของการให้คะแนนที่จะส่งผลต่อการให้คะแนนการปฏิบัติแล้ว ยังมีจุดเน้นที่กระบวนการคิดของผู้ประเมิน และกระบวนการตัดสินใจขององค์กร

2) โมเดลสี่องค์ประกอบ (Four-component Model) พัฒนาโดย Murphy & Cleveland ในปี 1995 ซึ่งพัฒนาขึ้นจากแนวคิดและงานวิจัยต่างๆ ลักษณะของโมเดลจะเน้นความเป็นองค์รวม ประกอบด้วยองค์ประกอบ 4 องค์ประกอบ ได้แก่ บริบทของการให้คะแนน (rating context) การตัดสินใจการปฏิบัติ (performance judgment) การให้คะแนนการปฏิบัติ (performance rating) และการประเมินระบบประเมินผล (evaluation of the appraisal system) ดังภาพที่ 5 โดยมีรายละเอียดของแต่ละองค์ประกอบดังนี้



ภาพที่ 5 โมเดลสื่อองค์ประกอบ (ที่มา: Murphy & Cleveland, 1995: 19)

1.บริบทของการให้คะแนน ได้แก่ บริบทขององค์กร (organizational context) โดยจะส่งผลต่อการให้คะแนนทั้งในส่วนของการตัดสินใจ (judgment process) กระบวนการให้คะแนน (rating process) กระบวนการประเมินผล (evaluation process) และการใช้ข้อมูลคะแนน (uses of rating data) ซึ่งผลงานวิจัยที่เกี่ยวข้องกับบริบทของการให้คะแนนมี 2 ประเด็นหลัก คือ 1) ปัจจัยภายในองค์กร เช่น การให้คุณค่าขององค์กร บรรยากาศหรือวัฒนธรรมองค์กร การแข่งขันภายในองค์กร ความแตกต่างของสถานะการงาน ผลลัพธ์ของการประเมินผลการทำงาน 2) ประเด็นเกี่ยวกับสิ่งแวดล้อมขององค์กร เช่น การปฏิบัติงานที่มีลักษณะเฉพาะขององค์กร การแข่งขันระหว่างองค์กร เศรษฐกิจทั่วไปหรือนโยบายทางการเมืองที่เกี่ยวข้องกับองค์กร

2.การตัดสินใจการปฏิบัติ (performance judgment) การตัดสินใจเป็นส่วนหนึ่งของกระบวนการให้คะแนน แต่มีความแตกต่างกันโดยการตัดสินใจจะเกี่ยวข้องกับทักษะส่วนบุคคล ส่วนการให้คะแนนจะมีลักษณะของความชัดเจน และพบว่าการตัดสินใจมักจะไม่มีการเปรียบเทียบ แต่บริบทอาจใช้เป็นตัวกำหนดสำหรับการตัดสินใจที่เฉพาะเจาะจง โดยการตัดสินใจที่ปราศจากความเกี่ยวข้องของบริบท โดยเฉพาะบริบทจำเพาะของงาน (job-specific context) จะขึ้นอยู่กับรูปแบบการคิด (cognitive schema) และเกณฑ์การประเมิน (evaluation criteria) ส่วนบุคคล งานวิจัยที่เกี่ยวข้องกับการประเมินผลการปฏิบัติ ส่วนมากจะศึกษากระบวนการตัดสินใจ (judgment process) ใน 2 ประเด็น คือ

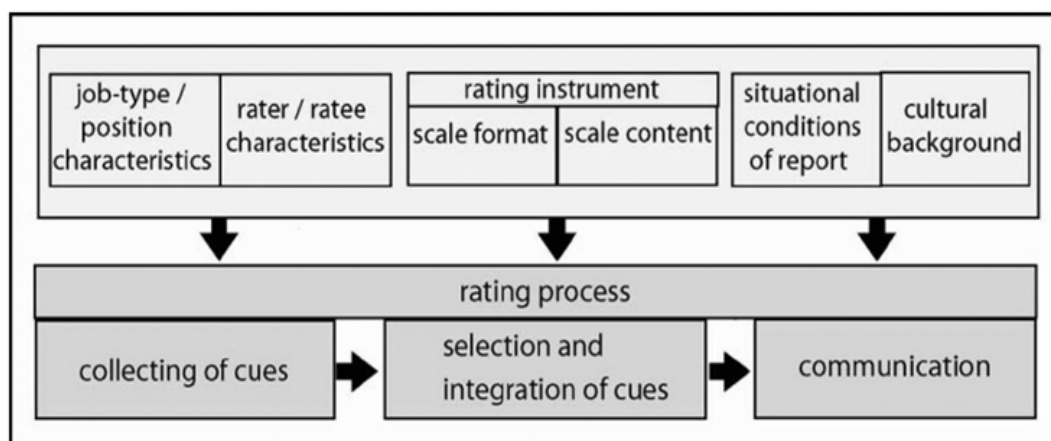
ข้อมูลเกี่ยวกับการทำงานที่จำเป็นสำหรับใช้ในการประเมินผล และวิธีการใช้ข้อมูลเหล่านี้สำหรับการตัดสินใจ

3.การให้คะแนนการปฏิบัติ (performance rating) เนื่องจากโมเดลนี้กำหนดให้มีความแตกต่างระหว่างการตัดสิน (judgments) และการให้คะแนน (ratings) ดังนั้นการบันทึกข้อมูลในแบบบันทึกการให้คะแนนอาจไม่ได้สะท้อนถึงการตัดสินของผู้ประเมิน ตัวอย่างเช่น ผู้ที่ได้คะแนนสูงไม่จำเป็นต้องเป็นผู้ที่ปฏิบัติได้ดี การให้คะแนนจะเป็นเพียงการบ่งบอกว่า ผู้ประเมินต้องการจะสื่อสารอะไร ซึ่งการสื่อสารที่ถูกต้องการคือการสื่อสารที่ตรงกับเป้าหมายและเกี่ยวข้องกับบริบท นอกจากนี้การให้คะแนนจะมีความสัมพันธ์สูงกับการตัดสินเมื่อ 1) องค์กรมีการระบุลักษณะของงานที่ชัดเจน 2) ผู้ประเมินรับรู้ถึงความเชื่อมโยงระหว่างการให้คะแนนกับผลลัพธ์ที่ตามมา 3) ผู้ประเมินเชื่อว่าผลลัพธ์ควรขึ้นอยู่กับ การปฏิบัติในปัจจุบัน และ 4) ปริมาณผลลัพธ์ทั้งหมดมีมากกว่าปริมาณผลลัพธ์ในเชิงลบ และถ้าการให้คะแนนมีความสัมพันธ์ใกล้เคียงกับการตัดสิน จะทำให้ไม่เกิดผลลัพธ์ที่ไม่พึงประสงค์สำหรับผู้ประเมินหรือผู้ถูกประเมิน

4.การประเมินระบบประเมินผล (evaluation of the appraisal system) การให้คะแนนการปฏิบัติจะถูกประเมินในด้านความลำเอียงทางจิตมิติ (psychometric biases) หรือความคลาดเคลื่อนของผู้ประเมิน (rater errors) แต่สาเหตุส่วนหนึ่งของความคลาดเคลื่อนของผู้ประเมิน เช่น การให้คะแนนสูงหรือต่ำเกินจริง อาจมาจากการจัดกระทำขึ้นเพื่อที่จะกระตุ้นให้เกิดการปรับเปลี่ยนพฤติกรรมไปในทางที่ต้องการ ซึ่งในกรณีนี้อาจไม่จำเป็นต้องวิเคราะห์ ความคลาดเคลื่อนของผู้ประเมิน นอกจากนี้ยังมีการประเมินในเรื่องความแม่นยำของการให้คะแนนการปฏิบัติ (accuracy in performance rating) ซึ่งความแตกต่างระหว่างคะแนนการปฏิบัติที่เกิดขึ้นจริงกับคะแนนที่ได้จากการประเมินอาจไม่ได้มาจากความคลาดเคลื่อนของการวัด (measurement error) เพียงอย่างเดียว แต่อาจเกิดจากปัจจัยต่างๆ เช่น แรงขับจากองค์กร หรือการสนับสนุนจากองค์กร เป็นต้น

โมเดลสี่องค์ประกอบของ Murphy & Cleveland ประกอบด้วย บริบทของการให้คะแนน การตัดสินการปฏิบัติ การให้คะแนนการปฏิบัติ และการประเมินระบบประเมินผล ซึ่งโมเดลได้เน้นถึงความสำคัญขององค์ประกอบด้านบริบทขององค์กรที่จะส่งผลกระทบต่อองค์ประกอบอื่นๆ และในองค์ประกอบ การประเมินระบบประเมินผลได้มีการอธิบายในส่วนของการประเมินความคลาดเคลื่อนและความแม่นยำในการให้คะแนน ซึ่งการนำไปปฏิบัติจะขึ้นกับการพิจารณาด้านการใช้ประโยชน์จากข้อมูลคะแนนประเมินผลการปฏิบัติ

3) โมเดลสามขั้นตอน (Three-stage model) พัฒนาโดย Heidemeier & Moser ในปี ค.ศ. 2009 (Heidemeier & Moser, 2009) จากการศึกษาวิเคราะห์อภิมาน (meta-analysis) โมเดลการให้คะแนนผลการปฏิบัติงาน โดยศึกษาความสอดคล้องระหว่างการให้คะแนนโดยตนเองและผู้บังคับบัญชาจากงานวิจัยจำนวน 102 เรื่อง ที่ตีพิมพ์ระหว่างปี ค.ศ. 1955 – 2007 ซึ่งรายงานเป็นขนาดตัวอย่าง 128 หน่วย ทำการเปรียบเทียบโดยใช้ค่า correlational agreement, mean-level agreement และ effect size indexes ผลการศึกษาพบว่า จากงานวิจัยส่วนใหญ่สามารถสรุปกระบวนการในการให้คะแนนเป็นรูปแบบของโมเดล โดยมีชื่อว่าโมเดลสามขั้นตอน (three-stage model) และมีหลักฐานยืนยันว่าทั้งสามขั้นตอนภายในโมเดล มีความสัมพันธ์กัน ตลอดจนช่วยในการทำความเข้าใจถึงความสอดคล้องในการให้คะแนน นอกจากนี้โมเดลยังช่วยให้แนวทางสำหรับการศึกษาชุดของตัวแปรที่ส่งผลต่อความสอดคล้องของผู้ประเมิน และผลการวิเคราะห์ความสัมพันธ์การให้คะแนนโดยตนเองและผู้บังคับบัญชามีค่า $r = .22$ ($p = .34$) และพบว่ามี การปล่อยคะแนนในกลุ่มที่มีการให้คะแนนโดยตนเองมากกว่ากลุ่มที่มีการให้คะแนนโดยผู้บังคับบัญชา



ภาพที่ 6 โมเดลสามขั้นตอน (ที่มา: Heidemeier & Moser, 2009: 355)

กระบวนการคิดทั้ง 3 ระยะในกระบวนการให้คะแนน (three cognitive stage of rating process) ได้แก่ การรวบรวมข้อมูล (collecting of cues) การคัดเลือกและบูรณาการข้อมูล (selection and integration of cues) หรืออาจเรียกว่าการพิจารณาตัดสินข้อมูล (judgment) และการสื่อสารข้อมูล (communication) นอกจากนี้ภายในโมเดลยังประกอบด้วยตัวแปรกำกับ (moderator) ที่จะ

ส่งผลต่อทั้ง 3 ระยะของกระบวนการให้คะแนน โดยตัวแปรกำกับแบ่งออกเป็น 5 กลุ่ม คือ 1) ประเภทของงานและคุณลักษณะของตำแหน่ง (job type and position characteristics) 2) คุณลักษณะต่างๆ ของผู้ประเมินและผู้ถูกประเมิน (rater and ratee characteristics) 3) คุณสมบัติของมาตรวัด (properties of the scales) ซึ่งรวมถึงรูปแบบและรายละเอียด (format and content) ของมาตรวัด 4) เงื่อนไขของการรายงาน (conditions of report) และ 5) ภูมิหลังทางวัฒนธรรม (cultural background) โดยอิทธิพลจากบริบทซึ่งเป็นตัวแปรกำกับเหล่านี้จะมีผลต่อความสอดคล้องของผู้ประเมิน (rater agreement) รายละเอียดของโมเดลมีดังนี้

1.การรวบรวมข้อมูล (collecting of cues) การรวบรวมข้อมูลจัดเป็นระยะเริ่มต้นในกระบวนการให้คะแนน (rating process) โดยข้อมูลอาจได้มาจากการสังเกตพฤติกรรม ผลการปฏิบัติ ข้อมูลการจัดประเภท (เช่น ตำแหน่งงานของผู้ถูกประเมิน) หรือจากแหล่งอื่น เช่น ผลคะแนนย้อนหลัง ข้อมูลที่ได้จากผู้ประเมินคนอื่นๆ โดยกลุ่มตัวแปรกำกับที่จะมีผลต่อระยะนี้ได้แก่ ประเภทของงานและคุณลักษณะของตำแหน่ง (job type and position characteristics) ซึ่งวัดได้จากตัวแปร เช่น ความซับซ้อนของงาน (job complexity) โอกาสในการสังเกตของผู้ประเมิน (rater's opportunity to observe) แรงจูงใจในการสังเกตของผู้ประเมิน (rater's motivation to observe) และการปฏิบัติงาน (job performance) ซึ่งตัวแปรเหล่านี้ควรจะมีผลต่อความไม่สอดคล้องระหว่างผู้ประเมินและความยากในการให้คะแนน และผลการวิเคราะห์ correlational agreement และ mean-level agreement พบว่า ตัวแปรกำกับที่มีนัยสำคัญทางสถิติ ($p < .01$) คือ ประเภทของงาน (job type) ระดับการศึกษา (level of education) ความซับซ้อนของงาน (job complexity) และตำแหน่งบริหาร (managerial position)

2.การคัดเลือกและบูรณาการข้อมูล (selection and integration of cues) เป็นการพิจารณาตัดสินข้อมูล (judgment) ซึ่งในการประเมินผลการปฏิบัติงานจะพิจารณาจากพฤติกรรมที่เกิดขึ้นตามเป้าหมาย ซึ่งความสอดคล้องกันระหว่างผู้ประเมินจะขึ้นอยู่กับ 1) ผู้ประเมินมีความสามารถในการจำแนกข้อมูลได้ตรงกัน และ 2) ผู้ประเมินมีการใช้ข้อมูลตรงกัน โดยกลุ่มตัวแปรกำกับสำคัญที่จะมีผลต่อระยะนี้ คือ เครื่องมือสำหรับการให้คะแนน (rating instrument) เกี่ยวข้องกับรูปแบบ (format) และรายละเอียด (content) ของมาตรวัด ได้แก่ ตัวบ่งชี้การปฏิบัติงาน (performance indicators) การวัดโดยใช้ข้อคำถามข้อเดียว (single-item measures) การวัดแบบรวมคุณลักษณะ (aggregation) ความยาวของมาตรวัด (scale length) การใช้มาตรวัดการเปรียบเทียบระหว่างตนเองกับผู้อื่น (social comparison scales) มาตรวัดที่ระบุพฤติกรรมชัดเจน (behavioral item labels)

และมาตรวัดแยกมิติการทำงานและบริบททางสังคม (scale content) โดยผลการวิเคราะห์ correlational agreement พบว่า มีเพียง 3 ตัวแปรกำกับ คือ ตัวบ่งชี้การปฏิบัติงาน การวัดแบบรวมคุณลักษณะ และมาตรวัดแยกมิติการทำงานและบริบททางสังคมที่มีนัยสำคัญทางสถิติ ($p < .01$) ส่วนผลการวิเคราะห์ mean-level agreement พบว่า มีเพียง 2 ตัวแปรกำกับ คือ การใช้มาตรวัดการเปรียบเทียบระหว่างตนเองกับผู้อื่น และการวัดโดยใช้ข้อความข้อเดียวที่มีนัยสำคัญทางสถิติ ($p < .01$) ส่วนคุณลักษณะต่างๆ ของผู้ประเมินและผู้ถูกประเมิน (rater and ratee characteristics) ซึ่งอาจเป็นตัวแปรกำกับที่ส่งผลต่อระยะนี้ ได้แก่ เพศ อายุ ระดับสติปัญญา หรือการได้รับการฝึกอบรม การให้คะแนน จะไม่อยู่ในการวิเคราะห์เนื่องจากไม่ได้รับอนุญาตให้ใช้ข้อมูล (Heidemeier & Moser, 2009: 356)

3. การสื่อสารข้อมูล (communication) ระยะที่สามของกระบวนการให้คะแนนมาจากแนวคิดที่ว่า การประเมินผลการปฏิบัติอาจเป็นพฤติกรรมรวมที่เกิดขึ้นอย่างมีเป้าหมาย เช่น เพื่อความถูกต้อง ป้องกันปัญหา หรือให้รางวัล ซึ่งผู้ประเมินจะต้องมีการสื่อสารข้อมูลภายหลังการให้คะแนน และพบว่าผู้ที่ให้คะแนนตนเองจะมีการสื่อสารข้อมูลเพื่อหลีกเลี่ยงผลลัพธ์ที่ไม่พึงประสงค์ โดยกลุ่มตัวแปรกำกับสำคัญที่จะมีผลต่อระยะนี้ คือ เงื่อนไขของการรายงาน (conditions of report) และภูมิหลังทางวัฒนธรรม (cultural background) โดยผลการวิเคราะห์ correlational agreement และ mean-level agreement พบว่า มีตัวแปรกำกับที่มีนัยสำคัญทางสถิติ ($p < .05$, $p < .01$) คือ วัตถุประสงค์ของการให้คะแนน (rating purpose) และวัฒนธรรมของแต่ละประเทศ (culture) ส่วนตัวแปรการคาดหวังความถูกต้องในการให้คะแนน (expectation that ratings may be validated) มีนัยสำคัญทางสถิติ ($p < .05$) เฉพาะการวิเคราะห์ correlational agreement

โมเดลสามขั้นตอนของ Heidemeier & Moser สร้างขึ้นจากการวิเคราะห์อภิมาน (meta-analysis) โมเดลการให้คะแนนผลการปฏิบัติงานที่ศึกษาความสัมพันธ์ระหว่างการให้คะแนนโดยตนเองและผู้บังคับบัญชาจากงานวิจัยจำนวน 102 เรื่อง ที่ตีพิมพ์ระหว่างปี ค.ศ. 1955 – 2007 ซึ่งรายงานเป็นขนาดตัวอย่าง 128 หน่วย โดยโมเดลจะสรุปเป็นกระบวนการคิดทั้ง 3 ระยะ ในกระบวนการให้คะแนน ได้แก่ การรวบรวมข้อมูล การคัดเลือกและบูรณาการข้อมูล หรืออาจเรียกว่าการพิจารณาตัดสินข้อมูล และการสื่อสารข้อมูล นอกจากนี้ภายในโมเดลยังประกอบด้วยตัวแปรกำกับที่จะส่งผลต่อทั้ง 3 ระยะของกระบวนการให้คะแนน โดยมีการจำแนกตัวแปรกำกับเป็น 5 กลุ่ม คือ 1) ประเภทของงานและคุณลักษณะของตำแหน่ง 2) คุณลักษณะต่างๆ ของผู้ประเมินและผู้ถูกประเมิน 3) คุณสมบัติของ

มาตรวัด ซึ่งรวมถึงรูปแบบและรายละเอียดของมาตรวัด 4) เงื่อนไขของการรายงาน และ 5) ภูมิหลังทางวัฒนธรรม

2.2.2 งานวิจัยที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ

ตัวอย่างของงานวิจัยที่ศึกษาปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติในด้านของความแม่นยำและความคลาดเคลื่อนในการให้คะแนนมีดังนี้

Bernardin, Tyler, & Villanova (Bernardin, Tyler, & Villanova, 2009) ศึกษาตัวแปรบุคลิกภาพของผู้ประเมินแบบประเมิน (agreeableness) และบุคลิกภาพแบบมีจิตสำนึก (conscientiousness) ที่ส่งผลต่อระดับการให้คะแนน (rating level) และความแม่นยำในการให้คะแนน (accuracy) กลุ่มตัวอย่างเป็นนักศึกษาระดับปริญญาตรีจำนวน 126 คน ในวิชาการบริหารทรัพยากรมนุษย์ที่มี 5 ตอนเรียน (sections) ตอนเรียนละ 18 – 32 คน สอนโดยอาจารย์คนเดียวกันซึ่งเป็นผู้ที่มีประสบการณ์สอนในวิชานี้กว่า 2 ปี ในการทดสอบความแตกต่างของกลุ่มตัวอย่างทั้ง 5 ตอนเรียน พบว่า ไม่มีความแตกต่างกันในคุณลักษณะส่วนบุคคล บุคลิกภาพแบบห้าองค์ประกอบ เกรดเฉลี่ย และค่าเฉลี่ยของคะแนนประเมินจากเพื่อน ในขั้นตอนของการวิจัยนักศึกษาแต่ละคนจะต้องทำงานกลุ่มโดยมีการสุมเข้ากลุ่มๆ ละ 4 – 6 คน รวม 4 – 6 กลุ่ม ซึ่งลักษณะของงานจะประกอบด้วยงานเดี่ยว คือ ให้เขียนตอบเกี่ยวกับหน้าที่ของตนและการให้นิยามเกี่ยวกับตำแหน่งงานที่เกี่ยวข้องกับการบริหารทรัพยากรมนุษย์ก่อนเริ่มคอร์สเรียน และงานที่ทำร่วมกับกลุ่ม คือ ทุกคนจะต้องเขียนถึงสมาชิกคนอื่นๆ ในกลุ่มและประเด็นที่เป็นฉันทามติร่วมกัน เครื่องมือที่ใช้ประกอบด้วย 1) แบบวัดบุคลิกภาพ NEO-FFI จำนวน 60 ข้อ โดยวัดภายใน 3 สัปดาห์แรกของการเรียน 2) แบบประเมินการปฏิบัติใน 5 มิติ ได้แก่ การคิดวิเคราะห์ภาวะผู้นำ การวางแผนและจัดลำดับความสำคัญ การพูดสื่อสาร และการเขียนสื่อสาร ซึ่งเป็นมาตรฐานค่า 7 ช่วง โดยคะแนนเฉลี่ยจากแบบวัดนี้จะเป็นค่าของตัวแปรระดับการให้คะแนนของนักศึกษาแต่ละคน และจะนำไปใช้คิดค่าของตัวแปรความแม่นยำในการให้คะแนนตามองค์ประกอบของ Borman (1977) และคนอื่นๆ คือ ความถูกต้องของคะแนนเฉลี่ยของผู้ถูกประเมินแต่ละคนในแต่ละมิติ (Differential accuracy) โดยนักศึกษาแต่ละคนจะต้องทำการประเมินเพื่อนในกลุ่ม (ประเมินเพื่อนเฉลี่ย 5.3 คน/ผู้ประเมิน) ซึ่งคะแนนจากการประเมินนี้จะไม่ได้นำมาใช้ในการตัดเกรด แต่จะใช้รายงานผลแบบภาพรวมแก่นักศึกษาผู้ถูกประเมินเพื่อให้ข้อมูลป้อนกลับ (feedback) ผลการศึกษาเมื่อออกแบบการวิจัยโดยกำหนดให้การให้คะแนนมีความรับผิดชอบที่ตรวจสอบได้แบบ

ต่ำ (low accountability) พบว่า คะแนนการประเมินมีความสัมพันธ์กับค่าเฉลี่ยของระดับการให้คะแนน (average rating level) โดยมีค่า $r = 0.18$ ($p < .05$) และทั้งความแม่นยำในการให้คะแนน (accuracy) กับคะแนนบุคลิกภาพแบบมีจิตสำนึก (conscientiousness) มีความสัมพันธ์กับระดับการให้คะแนน ($r = 0.27$ และ -0.20 ตามลำดับ $p < .05$)

Curtis, Harvey, & Ravden (Curtis, Harvey, & Ravden, 2005) ศึกษาวิจัยถึงทดลองโดยมีการสร้างตัวแทรกแซง (intervention) ที่เป็นตัวแปรบริบทของการประเมินผลการปฏิบัติงาน 2 ตัวแปร คือ ตัวแปรวัตถุประสงค์การประเมิน (appraisal purpose) ประกอบด้วย 2 กลุ่ม คือ การประเมินเพื่อพัฒนา และการประเมินเพื่อบริหารงาน และตัวแปรความรับผิดชอบที่ตรวจสอบได้ (accountability) มี 4 กลุ่ม คือ ความรับผิดชอบที่ตรวจสอบได้ต่อผู้ที่มีตำแหน่งน้อยกว่าตน (downwardly accountability) ความรับผิดชอบที่ตรวจสอบได้ต่อผู้ที่มีตำแหน่งสูงกว่าตน (upward accountability) ความรับผิดชอบทั้งสองแบบ (downwardly & upward accountability) และกลุ่มปกติ (none) ร่วมกับตัวแปรเพศ ที่จะส่งผลต่อความแม่นยำของผลการประเมิน กลุ่มตัวอย่างเป็นนักศึกษาสาขาจิตวิทยาใน Midwestern University จำนวน 123 คน โดยกำหนดให้นักศึกษาต้องทำโครงการซึ่งเป็นข้อกำหนดในรายวิชา คือ การช่วยทำหน้าที่ประเมินผลการปฏิบัติงานของพนักงานขายในบริษัทบัตรเครดิตแห่งหนึ่ง ซึ่งมีนโยบายเพิ่มกลุ่มลูกค้าที่เป็นนักศึกษา และในการวัดตัวแปรตาม คือ ความแม่นยำของผลการประเมิน จะพิจารณาจากความคลาดเคลื่อนชนิดปล่อยคะแนน (leniency) คือ คะแนนประเมิน (actual scores) มีค่าน้อยกว่าคะแนนอิงกลุ่ม (norm scores) อย่างมีนัยสำคัญทางสถิติ ($p < .001$) เมื่อทดสอบด้วยสถิติ t-test และผลการวิจัยพบว่า เมื่อวิเคราะห์ด้วยสถิติ three-way ANOVA ไม่พบอิทธิพลของปฏิสัมพันธ์ของทั้งสามตัวแปร ($F = 1.078$, $p = .361$) ส่วนอิทธิพลหลัก (main effect) พบว่ามีนัยสำคัญทางสถิติที่ตัวแปรความรับผิดชอบที่ตรวจสอบได้ ($F = 3.077$, $p < .05$) ซึ่งการเปรียบเทียบรายคู่ด้วย Tukey post hoc analysis พบว่า คู่ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติมี 1 คู่ คือ กลุ่มที่มีความรับผิดชอบที่ตรวจสอบได้ต่อผู้ที่มีตำแหน่งน้อยกว่าตนและกลุ่มที่มีความรับผิดชอบที่ตรวจสอบได้ต่อผู้ที่มีตำแหน่งสูงกว่าตน (ค่าเฉลี่ย = 1.87, 1.14 ตามลำดับ) ส่วนกลุ่มที่มีความรับผิดชอบทั้งสองแบบและกลุ่มปกติ ซึ่งมีค่าเฉลี่ยเท่ากับ 1.62, 1.40 ตามลำดับ ส่วนตัวแปรวัตถุประสงค์การประเมินเมื่อพิจารณาค่าเฉลี่ยจะพบว่าวัตถุประสงค์การประเมินเพื่อบริหารงานจะมีความคลาดเคลื่อนชนิดปล่อยคะแนนมากกว่าการประเมินเพื่อพัฒนาในกลุ่มที่มีความรับผิดชอบที่ตรวจสอบได้ต่อผู้ที่มีตำแหน่งน้อยกว่าตน และกลุ่มที่มีความรับผิดชอบที่ตรวจสอบได้ต่อผู้ที่มีตำแหน่งสูงกว่าตน แต่ผลนี้จะไม่พบในกลุ่มที่มีความรับผิดชอบทั้งสองแบบ และกลุ่มปกติ นอกจากนี้ในการทดสอบ 2 Purpose (Developmental

& Administrative) * 2 Accountability (Downward & Upward) ผลพบว่า อิทธิพลหลักของวัตถุประสงค์ การประเมินมีนัยสำคัญทางสถิติ ($F=5.371, p=.05$) โดยวัตถุประสงค์การประเมินเพื่อพัฒนาจะมีความ คลาดเคลื่อนชนิดปล่อยคะแนนน้อยกว่าการประเมินเพื่อบริหารงาน (ค่าเฉลี่ย = 1.24, 1.83 ตามลำดับ)

Gorman & Rentsch (Gorman & Rentsch, 2009) ศึกษาเกี่ยวกับผลของการฝึกอบรม ผู้ประเมินในการใช้กรอบการอ้างอิง (frame-of-reference training) กลุ่มตัวอย่างเป็นนักศึกษาจำนวน 144 คน แบ่งเป็นเพศชาย 56% และเพศหญิง 44% มีประสบการณ์ในการทำงานพิเศษ 60% และไม่เคยมีประสบการณ์ในการประเมินผลการปฏิบัติงานให้แก่ผู้อื่น 77% ทั้งนี้กลุ่มตัวอย่างจะได้รับการสุ่ม เข้ากลุ่ม คือ กลุ่มที่ได้รับการฝึกอบรมผู้ประเมินในการใช้กรอบการอ้างอิง ($n = 73$) และกลุ่มควบคุม ($n = 71$) โดยจะได้รับการแจ้งข้อมูลเกี่ยวกับวัตถุประสงค์ในการศึกษา ในกลุ่มที่ได้รับการฝึกอบรม จะมีการวัดผลการเรียนรู้ที่ได้รับ (schema accuracy) ซึ่งเป็นลักษณะที่สัมพันธ์กับความแม่นยำ ในการให้คะแนน (rating accuracy) ลักษณะของการฝึกอบรมจะใช้เวลาประมาณ 45 นาที ประกอบด้วยการให้อ่านออกเสียงในนิยามองค์ประกอบ มาตรฐานวัดคะแนน และตัวอย่างของพฤติกรรม จากนั้นจะเป็นการสนทนาโต้ตอบข้อมูลกันโดยเริ่มต้นที่ผู้ให้การอบรมจะนำเสนอตัวอย่างการให้ คะแนนของในแต่ละองค์ประกอบ ในการเก็บข้อมูลกลุ่มตัวอย่างจะได้รับการดูวิดีโอ 4 ชุด ซึ่งเป็น วิดีโอที่ใช้ฝึกในศูนย์พัฒนาการประเมินผลการปฏิบัติงานของผู้บริหาร และวิดีโอจะมีการควบคุมตัวแปร เพศที่อาจส่งผลกระทบต่อผลการประเมินผลการปฏิบัติงาน โดยภายในวิดีโอจะประกอบด้วยสถานการณ์ ที่ผู้ได้บังคับบัญชาที่มีความสามารถในการปฏิบัติงานสูง 2 ชุด คือ ผู้แสดงที่เป็นเพศชายและเพศหญิง และสถานการณ์ที่ผู้ได้บังคับบัญชาที่มีความสามารถในการปฏิบัติงานต่ำ 2 ชุด คือ ผู้แสดงที่เป็นเพศชาย และเพศหญิง วิดีโอแต่ละชุดมีความยาว 15 นาที โดยลำดับของวิดีโอจะเป็นแบบสุ่ม และในขณะที่ดู วิดีโอกลุ่มตัวอย่างจะต้องสวมบทบาทของผู้บังคับบัญชาเพื่อทำการประเมินผลการปฏิบัติงานของ ผู้ได้บังคับบัญชาโดยจะต้องมีการสังเกตพฤติกรรมที่เกิดขึ้นตามแบบประเมินและทำการบันทึกแบบ ประเมินเมื่อสิ้นสุดการดูวิดีโอแต่ละชุด ซึ่งแบบประเมินจะมี 2 ส่วน ส่วนแรกในแต่ละข้อของการ ประเมินพฤติกรรมจะให้ใส่ผลการประเมินจาก 3 แบบ คือ เป็นบวก (positive) ติดลบ (negative) และเป็นศูนย์ (neutral) และส่วนที่สองจะเป็นการประเมินโดยใช้มาตราประมาณค่า 11 ช่วง (Likert-type scale) ตั้งแต่ 1 (extremely weak), 1.7 (very weak), 2 (weak), 2.5 (moderately weak),..., จนถึง 5 (exceptional) จำนวน 24 ข้อ (จากวิดีโอทั้ง 4 ชุด) ผลการวิจัยพบว่า ผลการเรียนรู้ที่ได้รับ (schema accuracy) จากการฝึกอบรมมีความสัมพันธ์ทางบวกกับความแม่นยำในการให้คะแนน

Hedge & Teachout (Hedge & Teachout, 2000) ศึกษาการยอมรับในผลการประเมินในด้าน 1)การพัฒนาเกณฑ์ที่ใช้ 2)ความสัมพันธ์ระหว่างการยอมรับในผลการประเมินกับความไว้วางใจและแรงจูงใจของผู้ประเมิน 3)ความสัมพันธ์ระหว่างแหล่งที่มาของคะแนนและการยอมรับผลการประเมิน และ 4)ความแตกต่างของการยอมรับในผลการประเมินในผู้ประเมินที่ใช้รูปแบบการให้คะแนนต่างกัน กลุ่มตัวอย่างเป็นเจ้าหน้าที่ในกองทัพอากาศ 1,608 คน ผู้บังคับบัญชา 534 คน และเครื่องมือทั้งหมดเป็นมาตรประมาณค่าซึ่งประกอบด้วย แบบวัดชุดตัวแปรรวมจำนวน 2 ชุดสำหรับเจ้าหน้าที่และผู้บังคับบัญชา แต่ละชุดประกอบด้วย 30 ข้อ (ชุดของเจ้าหน้าที่: แรงจูงใจในการให้คะแนนอย่างเที่ยงตรง 5 ข้อ ความพึงพอใจต่องาน 5 ข้อ การยอมรับในผลการประเมิน 6 ข้อ ข้อจำกัดในการปฏิบัติงาน 3 ข้อ ความไว้วางใจต่อผู้ประเมินคนอื่น 3 ข้อ การสนับสนุนของผู้บังคับบัญชา 2 ข้อ ความไว้วางใจในกระบวนการประเมินผลการปฏิบัติงาน 2 ข้อ แรงจูงใจอื่นที่มีผลต่อการให้คะแนน 2 ข้อ และความไว้วางใจต่อการเปิดเผยข้อมูล 2 ข้อ และชุดของผู้บังคับบัญชา: แรงจูงใจในการให้คะแนนอย่างเที่ยงตรง 5 ข้อ ความพึงพอใจต่องาน 5 ข้อ การยอมรับในผลการประเมิน 6 ข้อ การสนับสนุนของผู้บังคับบัญชา 2 ข้อ ความไว้วางใจต่อผู้ประเมินคนอื่น 3 ข้อ ความไว้วางใจต่อการเปิดเผยข้อมูล 2 ข้อ ข้อจำกัดในการปฏิบัติงาน 3 ข้อ ความไว้วางใจในกระบวนการประเมินผลการปฏิบัติงาน 2 ข้อ และความรักหมั่นคณะ 2 ข้อ) และแบบประเมินผลการปฏิบัติงานที่ต่างกัน 4 รูปแบบ ซึ่งประเมินเกี่ยวกับเทคนิคการทำงานและปฏิสัมพันธ์กับผู้อื่น ผลการวิเคราะห์การถดถอยแบบพหุพบว่า ในระดับของเจ้าหน้าที่พบว่าใน 6 ตัวแปร ได้แก่ แรงจูงใจในการให้คะแนนอย่างเที่ยงตรง ความไว้วางใจต่อการเปิดเผยข้อมูล แรงจูงใจอื่นที่มีผลต่อการให้คะแนน ความไว้วางใจต่อผู้ประเมินคนอื่น ความไว้วางใจในกระบวนการประเมินผลการปฏิบัติงาน และข้อจำกัดในการปฏิบัติงาน ร่วมกันทำนายความแปรปรวนของการยอมรับในผลการประเมินได้ 31% ส่วนในระดับของผู้บังคับบัญชา พบว่าทั้ง 6 ตัวแปร ได้แก่ แรงจูงใจในการให้คะแนนอย่างเที่ยงตรง ความไว้วางใจต่อการเปิดเผยข้อมูล ความไว้วางใจต่อผู้ประเมินคนอื่น ความไว้วางใจในกระบวนการประเมินผลการปฏิบัติงาน ความรักหมั่นคณะ และข้อจำกัดในการปฏิบัติงานเป็นตัวทำนายความแปรปรวนของการยอมรับในผลการประเมินได้ 25%

Heidemeier & Moser (Heidemeier & Moser, 2009) ได้ทำการศึกษาวิเคราะห์อภิมาน (meta-analysis) โมเดลการให้คะแนนผลการปฏิบัติงาน โดยการศึกษาความสัมพันธ์ระหว่างการให้คะแนนโดยตนเองและผู้บังคับบัญชาจากงานวิจัยจำนวน 102 เรื่อง ที่ตีพิมพ์ระหว่างปี ค.ศ. 1955 – 2007 ซึ่งรายงานเป็นขนาดตัวอย่าง 128 หน่วย ทำการเปรียบเทียบโดยใช้ค่า correlational agreement, mean-level agreement และ effect size indexes ผลการศึกษาพบว่า จากงานวิจัย

ส่วนใหญ่สามารถสรุปกระบวนการในการให้คะแนนเป็นรูปแบบของโมเดล โดยมีชื่อว่าโมเดลสามขั้นตอน (3-stage model) ประกอบด้วยกระบวนการคิดทั้ง 3 ระยะ (three cognitive stage of rating process) ได้แก่ การรวบรวมข้อมูล (collecting of cues) การคัดเลือกและบูรณาการข้อมูล (selection and integration of cues) และการสื่อสารข้อมูล (communication) นอกจากนี้ภายในโมเดลยังประกอบด้วยตัวแปรกำกับ (moderator) 5 กลุ่ม คือ 1) ประเภทของงานและคุณลักษณะของตำแหน่ง (job type and position characteristics) 2) คุณลักษณะต่างๆ ของผู้ประเมินและผู้ถูกประเมิน (rater and ratee characteristics) 3) คุณสมบัติของมาตรวัด (properties of the scales) ซึ่งรวมถึงรูปแบบและรายละเอียด (format and content) ของมาตรวัด 4) เงื่อนไขของการรายงาน (conditions of report) และ 5) ภูมิหลังทางวัฒนธรรม (cultural background) ผลการวิเคราะห์ความสัมพันธ์การให้คะแนนโดยตนเองและผู้บังคับบัญชา มีค่า $r = .22$ ($p = .34$) และพบว่ามีการปล่อยคะแนนในกลุ่มที่มีการให้คะแนนโดยตนเองมากกว่ากลุ่มที่มีการให้คะแนนโดยผู้บังคับบัญชา และตัวแปรกำกับทั้ง 5 กลุ่ม ส่งผลต่อทั้ง 3 ระยะของกระบวนการให้คะแนน

Khoury & Analoui (Khoury & Analoui, 2004) ได้พัฒนาโมเดลการประเมินผลการปฏิบัติงานกรณีศึกษาที่มหาวิทยาลัยของรัฐในประเทศปาเลสไตน์ ในงานวิจัยนี้ได้แบ่งเป็น 2 ระยะ คือ ระยะแรกเป็นการวิจัยเชิงปริมาณและการวิจัยเชิงคุณภาพเพื่อสำรวจการรับรู้เกี่ยวกับกระบวนการประเมินผลการปฏิบัติงานทั้งจุดเด่นและจุดด้อย และระยะที่สองเป็นการสร้างโมเดลการประเมินผลการปฏิบัติงาน กลุ่มตัวอย่าง คือ ผู้ปฏิบัติงานเฉพาะด้านการสอนและการวิจัยในมหาวิทยาลัยของรัฐทั้ง 5 แห่ง แบ่งเป็นกลุ่มที่ตอบแบบสอบถามในงานวิจัยเชิงปริมาณ จำนวน 451 คน โดยใช้วิธีการสุ่ม ซึ่งผลการส่งคืนของแบบสอบถามมีจำนวน 265 คน ซึ่งคิดเป็น 59% ของกลุ่มตัวอย่าง และกลุ่มที่ใช้สัมภาษณ์ในงานวิจัยเชิงคุณภาพมีจำนวน 25 คน ซึ่งผลการวิจัยพบว่า กลุ่มตัวอย่างยังไม่พึงพอใจเกี่ยวกับประสิทธิผลของกระบวนการประเมินผลการปฏิบัติงาน โดยมีสาเหตุ ได้แก่ กระบวนการประเมินผลใช้วิธีการแบบเก่า ให้ความสำคัญกับผลการประเมินของนักศึกษาเกินไป ผู้ประเมินขาดการรับรู้เกี่ยวกับมาตรฐานการปฏิบัติงาน ผู้ประเมินไม่มีความสามารถผู้ประเมินไม่มีการให้ข้อมูลสะท้อนกลับในผลการประเมิน ไม่มีขั้นตอนการสัมภาษณ์ และเกิดความล้มเหลวในการนำผลการประเมินไปใช้ในการบริหารงาน ส่วนกระบวนการประเมินผลการปฏิบัติงานที่พึงประสงค์ ได้แก่ การมีมาตรฐานการปฏิบัติงานที่ชัดเจน มีการให้ข้อมูลสะท้อนกลับทั้งแบบเป็นทางการและไม่เป็นทางการ เพื่อพัฒนาผลการปฏิบัติงาน และผู้ประเมินควรได้รับการฝึกอบรมวิธีการประเมินที่หลากหลาย เพื่อนำมาใช้ในการประเมินผลการปฏิบัติงาน ส่วนงานวิจัยในระยะที่สองเป็นการสร้างโมเดลการประเมิน

ผลการปฏิบัติงาน โดยมีการใช้ข้อมูลจากงานวิจัยในระยะแรก รวมถึงแนวคิด ทฤษฎี และงานวิจัยต่างๆ ที่เกี่ยวข้อง ซึ่งโมเดลที่พัฒนาขึ้นมีชื่อว่า SOFIA ประกอบด้วย 5 ขั้นตอน คือ 1. การกำหนดมาตรฐาน การปฏิบัติงานและเผยแพร่ให้ผู้ปฏิบัติงานได้ทราบและเข้าใจตรงกัน (Setting the ground rules-standards) 2. การประเมินผลการปฏิบัติงานแบบไม่เป็นทางการ (On-going informal appraisal) ซึ่งในขั้นตอนนี้จะไม่มีจุดเริ่มต้นหรือสิ้นสุดที่ชัดเจนเพราะจะขึ้นอยู่กับผู้ประเมินและผู้ถูกประเมินในการตั้งเป้าหมายและพัฒนาการปฏิบัติงานตามที่กำหนดไว้ร่วมกัน และผู้ประเมินควรมีการสะท้อนข้อมูลผลการปฏิบัติงานแก่ผู้ถูกประเมินอย่างต่อเนื่อง 3. การประเมินผลการปฏิบัติงานแบบเป็นทางการ (Formal appraisal) ในขั้นตอนนี้อาจทำก่อนสิ้นสุดปลายปีการศึกษา โดยให้ผู้ถูกประเมินได้ทำการประเมินตนเอง ก่อนเกี่ยวกับภาระงานหรือกิจกรรมที่ได้ทำ ความสำเร็จหรือผลที่เกิดขึ้น รวมถึงจุดแข็ง จุดอ่อน และปัญหาต่างๆ ที่ส่งผลต่อการปฏิบัติงานในระหว่างปีการศึกษานั้น นอกจากนี้ในขั้นตอนนี้ควรมีการประเมินทั้งปัจจัยนำเข้า (input) และผลลัพธ์ (output) และควรมีการรวบรวมข้อมูลจากหลายๆ แหล่ง เช่น ข้อมูลแบบ 360 องศา จากหัวหน้า เพื่อน และลูกน้องหรือนักศึกษา เป็นต้น 4. การสัมภาษณ์ ข้อมูล (Interview) เป็นการให้ข้อมูลผลการประเมินผลการปฏิบัติงานแบบเป็นทางการและสอบถามความคิดเห็นของผู้ถูกประเมิน ซึ่งในขั้นตอนนี้ผู้ประเมินควรได้รับการฝึกอบรมมาก่อน และขั้นตอนนี้ถือว่าเป็นความลับระหว่างผู้ประเมินและผู้ถูกประเมิน และ 5. การปฏิบัติ (Action) เป็นขั้นตอนนี้สุดท้ายซึ่งผู้บริหารได้นำข้อมูลการประเมินผลการปฏิบัติงานแบบเป็นทางการไปใช้ประโยชน์ นอกจากนี้ควรมีการพิจารณาถึงปัจจัยภายนอกและภายในที่อาจส่งผลกระทบต่อการทำงานของผู้ถูกประเมินแต่ละคน เช่น วัฒนธรรมองค์กร สถานะขององค์กร สถานการณ์ทางการเมืองและเศรษฐกิจภายในประเทศ เป็นต้น และผู้วิจัยได้กล่าวถึงข้อจำกัด 3 ข้อ ในการนำโมเดล SOFIA ไปใช้ คือ ลักษณะทางวัฒนธรรม สถานการณ์ที่แวดล้อม และปัจจัยทางการเงิน

Kim (Kim, 2011) ศึกษาความแตกต่างระหว่างภูมิภาคหลังที่ส่งผลต่อการพัฒนาความสามารถในขณะการฝึกอบรมการให้คะแนน กลุ่มตัวอย่างแบ่งเป็น กลุ่มผู้ประเมินและกลุ่มผู้สอบ ซึ่งกลุ่มตัวอย่างที่เป็นผู้ประเมินจะเป็นกลุ่มตัวอย่างหลักของงานวิจัย ส่วนข้อมูลจากกลุ่มผู้สอบในที่นี้จะเป็นการนำข้อมูลที่บันทึกไว้จากการจัดสอบครั้งก่อนมาใช้ คือ การจัดสอบในช่วงฤดูร้อน ปี ค.ศ. 2008 ที่มีผู้เข้าสอบประมาณ 700 คน และได้สุ่มตัวอย่างแบบแบ่งชั้น (stratified random sampling) โดยใช้คะแนนการสอบพูด Community English Program (CEP) ซึ่งเป็นส่วนหนึ่งของ Teaching English to Speakers of Other Languages (TESOL) มาใช้ในการแบ่งระดับความสามารถคือ ระดับต้น กลาง และสูง ได้ข้อมูลของผู้เข้าสอบกลุ่มนี้จำนวน 18 คน มาใช้ในงานวิจัย คือ ระดับต้นจำนวน 6 คน ระดับกลางจำนวน 6 คน

และระดับสูงจำนวน 6 คน ซึ่งผู้เรียนในแต่ละระดับๆ ละ 2 คน จะถูกสุ่มอีกครั้งเพื่อเข้ากลุ่มจำนวน 3 กลุ่ม เพื่อใช้เป็นการทดสอบการให้คะแนนในแต่ละช่วงเวลาของการอบรมผู้ประเมิน (มี 3 ช่วงตามระยะของการฝึกอบรม ซึ่งใช้เวลาอบรมช่วงละ 1 เดือน) จากนั้นได้ทดสอบความแตกต่างของค่าเฉลี่ยความสามารถในการพูดของ แต่ละกลุ่มด้วย ANOVA พบว่าค่าเฉลี่ยในแต่ละกลุ่มมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ จากนั้นได้ทดสอบ post hoc test ซึ่งผลได้ยืนยันความแตกต่างภายในแต่ละกลุ่ม ส่วนกลุ่มตัวอย่างของผู้ประเมินมีจำนวน 9 คน ได้จากการคัดเลือกแบบเจาะจงตามเป้าหมายที่กำหนด (purposeful sampling) คือ มีภูมิหลังซึ่งวัดจากประสบการณ์การให้คะแนนใน L2 speaking assessments ประสบการณ์การสอน TESOL ประสบการณ์จากการอบรมการให้คะแนน และจำนวนวิชาและภาระหน้าที่ที่สอน ซึ่งจะแบ่งระดับผู้ประเมินเป็น 3 กลุ่ม คือ เป็นผู้เริ่มต้น (novice) ผู้กำลังพัฒนา (developing) และผู้เชี่ยวชาญ (expert) โดยใช้การทดสอบและพิจารณาจากข้อมูลต่างๆ เครื่องมือที่ใช้ในงานวิจัยนี้ประกอบด้วย 1) CEP speaking placement test 2) analytic scoring rubric ใช้เป็นเกณฑ์สำหรับการให้คะแนนผู้สอบ และ 3) แบบสอบถามภูมิหลังเพื่อใช้ในการจำแนกกลุ่มของผู้ประเมิน ทั้งนี้ผู้ประเมินทั้ง 9 คนจะต้องให้คะแนนผู้สอบจำนวน 6 คนตามกลุ่มที่จัดไว้ในแต่ละช่วงเวลา และการวัดพฤติกรรมการให้คะแนนจะใช้เกณฑ์การให้คะแนนโดยการแยกวิเคราะห์ (analytic scoring criteria) ในชุดทักษะการพูดที่แตกต่างกัน และข้อมูลนี้จะนำมาใช้ในการแก้ไขพฤติกรรมกรให้คะแนนที่ไม่พึงประสงค์และพัฒนาความสามารถในการให้คะแนน ซึ่งการวิเคราะห์พฤติกรรมกรให้คะแนนจะมีทั้งแบบเชิงปริมาณและเชิงคุณภาพ ซึ่งการวิเคราะห์เชิงปริมาณ ได้แก่ ความคลาดเคลื่อนของผู้ประเมิน (ใช้ MFRM ในการวิเคราะห์) ความเที่ยงภายใน (internal consistency) และปฏิสัมพันธ์ระหว่างความสามารถของผู้สอบ แบบสอบ และเกณฑ์การให้คะแนน ขณะที่การวิเคราะห์เชิงคุณภาพจะเป็นการตรวจสอบแนวโน้มของผู้ประเมินในด้านความเข้าใจเกี่ยวกับคำอธิบายในเกณฑ์การให้คะแนนโดยการแยกวิเคราะห์ (analytic scoring criteria) ซึ่งพฤติกรรมกรให้คะแนนนี้จะใช้เป็นข้อมูลสะท้อนกลับ (feedback) ที่เป็นส่วนหนึ่งของการฝึกอบรมการให้คะแนน ผลการวิจัยพบว่า ผู้ประเมินกลุ่มผู้เริ่มต้นและผู้กำลังพัฒนาไม่มีความแตกต่างของรูปแบบการให้คะแนน ขณะที่กลุ่มผู้เชี่ยวชาญจะมีความสามารถในการให้คะแนนระดับสูงทั้ง 3 ช่วงเวลาของการฝึกอบรม และในช่วงที่ 2 ของการฝึกอบรมจากการวิเคราะห์ข้อมูลเชิงคุณภาพที่มุ่งวิเคราะห์การใช้เกณฑ์การให้คะแนน พบว่าผู้ประเมินทั้ง 3 กลุ่มมีรูปแบบการให้คะแนนที่แตกต่างกันและมีความแตกต่างทางการพัฒนาความสามารถทั้งใน 3 ช่วงเวลาของการฝึกอบรม ดังนั้นจึงสรุปได้ว่า การฝึกอบรมทั้งรายบุคคลหรือรายกลุ่มจะช่วยพัฒนาความสามารถและลดความคลาดเคลื่อนของผู้ประเมินได้

Murphy, Cleveland, Skattebo, & Kinney (Murphy et al., 2004) ศึกษาเป้าหมายของผู้ประเมิน (rater goals) ที่ส่งผลต่อการให้คะแนนที่แตกต่างกันในบริบทของการศึกษา วิธีการศึกษาจะให้ให้นักศึกษาทั้งหมดทำการประเมินการสอนของอาจารย์ในหัวข้อที่เหมือนกัน และอาจารย์จะทราบถึงหัวข้อที่ถูกประเมิน เพื่อที่จะสามารถอธิบายผลของสมมติฐานเกี่ยวกับ เป้าหมายของผู้ประเมินที่ต่างกันจะมีผลต่อการให้คะแนนต่างกัน เนื่องจากถ้าผลการวิจัยพบว่านักศึกษามีการให้คะแนนแตกต่างกันทั้งที่มีการใช้ตัวบ่งชี้ในการวัดแบบเดียวกันก็จะแสดงว่านักศึกษามีเป้าหมายที่ต่างกัน เครื่องมือที่ใช้มี 2 ชุด ประกอบด้วย แบบประเมินการสอนของอาจารย์ ซึ่งประกอบด้วย 4 ข้อคำถาม แต่ละข้อเป็นมาตราประมาณค่า 5 ช่วง (Likert-type scale) ตั้งแต่ 1 (ไม่เห็นด้วยอย่างยิ่ง) ถึง 5 (เห็นด้วยอย่างยิ่ง) ค่า Cronbach's alpha = .94 และแบบวัดเป้าหมายของผู้ประเมินที่มีข้อคำถามจำนวน 19 ข้อ ซึ่งประกอบด้วยเป้าหมาย 19 อย่างที่จะส่งผลต่อการประเมินการสอน ในแต่ละข้อเป็นมาตราประมาณค่า 5 ช่วง (Likert-type scale) ตั้งแต่ 1 (ไม่เห็นด้วยอย่างยิ่ง) ถึง 5 (เห็นด้วยอย่างยิ่ง) โดยแบบวัดนี้ได้ปรับปรุงมาจากงานวิจัยของ Kerst (Kerst, 1993 cited in Murphy, Cleveland, Skattebo, & Kinney, 2004: 159) และได้มีการนำแบบวัดเป้าหมายของผู้ประเมินไปทดลองใช้กับนักศึกษาปีที่ 1-3 วิเคราะห์ข้อมูลโดยใช้การวิเคราะห์องค์ประกอบแบบยืนยัน (confirmatory factor analysis) พบว่าโมเดลการวัดมีความสอดคล้องกับข้อมูลเชิงประจักษ์ (NFI = .97, CFI = .98, RMSEA = .09) ซึ่งประกอบด้วย 4 องค์ประกอบ คือ 1) จุดอ่อนของอาจารย์ผู้สอน (instructor's weaknesses) ประกอบด้วย 8 ข้อคำถาม (Cronbach's alpha = .91) 2) จุดแข็งของอาจารย์ผู้สอน (instructor's strengths) ประกอบด้วย 6 ข้อคำถาม (Cronbach's alpha = .74) 3) มาตรฐานการทำงานของอาจารย์ (performance standards) ประกอบด้วย 3 ข้อคำถาม (Cronbach's alpha = .60) และ 4) แรงจูงใจของผู้ประเมิน (motivating rates) ประกอบด้วย 2 ข้อคำถาม (Cronbach's alpha = .67) ค่าสหสัมพันธ์ระหว่าง 4 องค์ประกอบมีค่าอยู่ระหว่าง .26 ถึง .53 ส่วนผลการวิเคราะห์หลักได้เก็บข้อมูลเป้าหมายของผู้ประเมินในช่วงสัปดาห์แรกของภาคเรียน (N = 303) และแบบประเมินการสอนของอาจารย์ได้เก็บข้อมูลในช่วงก่อนปิดภาคเรียนซึ่งห่างจากเวลาที่เก็บข้อมูลเป้าหมายของผู้ประเมินนาน 14 สัปดาห์ (N = 232) ซึ่งมีจำนวน 186 คนที่สามารถเก็บข้อมูลได้ทั้ง 2 ชุด และมีการเก็บข้อมูลเป้าหมายของผู้ประเมิน เพื่อวิเคราะห์การเปลี่ยนแปลงของเป้าหมายของผู้ประเมินจากกลุ่มตัวอย่างซ้ำโดยเก็บเพิ่มในช่วงก่อนปิดภาคเรียน (N = 154 ซึ่งจับคู่กับข้อมูลชุดแรกได้จำนวน 125 คน) ผลการวิเคราะห์ข้อมูลพบว่าอาจารย์จำนวน 5 คน ที่ได้คะแนนสูงสุด โดยมีค่าเฉลี่ย 4.15 ส่วนคะแนนเฉลี่ยของอาจารย์โดยรวมมีค่าอยู่ระหว่าง 4.03 ถึง 4.22 และค่าความสัมพันธ์ระหว่างเป้าหมายของผู้ประเมินกับการให้คะแนน

การปฏิบัติงานจากกลุ่มตัวอย่างมีขนาดปานกลาง คือ incremental R-square = .08 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ผลการวิเคราะห์การเปลี่ยนแปลงของเป้าหมายของผู้ประเมินที่เก็บข้อมูลในช่วงสัปดาห์แรกและสัปดาห์สุดท้ายของภาคเรียนจำนวน 125 คน ยืนยันสมมติฐานที่ว่า เป้าหมายของผู้ประเมินที่เก็บจากสัปดาห์สุดท้ายจะเป็นตัวทำนายการให้คะแนนการปฏิบัติงานได้ดีกว่าชุดที่เก็บจากสัปดาห์แรกของภาคเรียน โดยมี incremental $R^2 = .09$ และ $.01$ ตามลำดับ และค่าสหสัมพันธ์ระหว่างเป้าหมายของผู้ประเมินในช่วงสัปดาห์แรกกับสัปดาห์สุดท้ายมีความสอดคล้องกันปานกลาง แต่พบว่าผู้ประเมินจำนวน 18 จาก 19 คน ที่ให้คะแนนในช่วงสัปดาห์แรกและสัปดาห์สุดท้ายมีความสัมพันธ์เป็นบวก (r อยู่ระหว่าง $.22$ ถึง $.29$) ค่าสหสัมพันธ์ระหว่างองค์ประกอบจุดแข็ง จุดอ่อน มาตรฐานการทำงาน และแรงจูงใจระหว่างสัปดาห์แรกกับสัปดาห์สุดท้ายมีค่า $.30$, $.60$, $.27$, และ $.06$ ตามลำดับ และพบว่าแรงจูงใจในช่วงสัปดาห์แรกลดลงจากสัปดาห์สุดท้าย ($d = -.09$) ส่วนองค์ประกอบจุดแข็ง จุดอ่อน และมาตรฐานในช่วงสัปดาห์สุดท้ายจะมากกว่าสัปดาห์แรก ($d = .09$, $.08$, และ $.37$ ตามลำดับ) และจากการวิเคราะห์ถดถอย (hierarchical regression) พบว่า เป้าหมายของผู้ประเมินทั้งในช่วงสัปดาห์แรกและสัปดาห์สุดท้าย และปัจจัยเกี่ยวกับอาจารย์ (instructor effects) สามารถทำนายคะแนนการปฏิบัติงานของอาจารย์ได้ $R^2 = .07$ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อแยกวิเคราะห์ในแต่ละเป้าหมายทั้ง 19 ข้อ ในช่วงสัปดาห์แรกและการเปลี่ยนแปลงของ R^2 เมื่อเพิ่มเป้าหมายของผู้ประเมินทั้งในช่วงสัปดาห์สุดท้าย พบว่า การเปลี่ยนแปลงของเป้าหมายของผู้ประเมินเป็นตัวทำนายที่ดีสำหรับการปฏิบัติงานของอาจารย์ และเป้าหมาย 10 ข้อจาก 19 ข้อ พบว่า ไม่มีการเปลี่ยนแปลงของ R^2 และพบว่าไม่มีเป้าหมาย 5 ข้อ ซึ่งประกอบด้วย 3 จุดอ่อนและ 2 จุดแข็งที่มีการเปลี่ยนแปลงของ R^2 มากกว่าหรือเท่ากับ $.01$

Ostroff, Atwater, & Feinberg (2004) ได้ศึกษาความสัมพันธ์ระหว่างคุณลักษณะของผู้ประเมินและผู้ถูกประเมิน (เพศ อายุ เชื้อชาติ ระดับการศึกษา และประสบการณ์) ปัจจัยทางบริบท (จำนวนผู้ใต้บังคับบัญชา และขอบเขตภาระงาน) และความสอดคล้องระหว่างผู้ประเมิน กลุ่มตัวอย่างเป็นผู้บริหารจำนวน 3,217 คน จาก 527 องค์กร ซึ่งเข้าร่วมโปรแกรมการพัฒนาผู้นำระหว่างปี ค.ศ. 1991- 2000 ใช้แบบวัดพฤติกรรมผู้นำ โดยผู้ประเมินจะประกอบด้วยผู้บังคับบัญชา เพื่อน และผู้ใต้บังคับบัญชา วิเคราะห์ข้อมูลโดยใช้สถิติ multivariate regression ผลการวิจัยพบว่า เพศของผู้ประเมินส่งผลต่อการให้คะแนนโดยตนเองและผู้อื่น (ผู้บังคับบัญชา เพื่อนและผู้ใต้บังคับบัญชา) อย่างมีนัยสำคัญทางสถิติ เช่น ผู้ใต้บังคับบัญชาจะมีการให้คะแนนผู้บังคับบัญชาที่เป็นเพศหญิงสูงกว่าเพศชาย เชื้อชาติไม่มีผลต่อการให้คะแนนโดยผู้ใต้บังคับบัญชาหรือผู้บังคับบัญชา แต่จะมีผลต่อการให้

คะแนนตนเองโดยกลุ่มที่มีผิวสีจะให้คะแนนตนเองสูงกว่ากลุ่มที่มีผิวขาวอย่างมีนัยสำคัญทางสถิติ อายุส่งผลต่อการให้คะแนนโดยตนเองและผู้อื่น (ผู้บังคับบัญชา และผู้ใต้บังคับบัญชา) อย่างมีนัยสำคัญทางสถิติ เช่น ผู้ที่อายุมากมีแนวโน้มที่จะให้คะแนนตนเองสูงกว่าผู้ที่อายุน้อยเมื่อเทียบกับการให้คะแนนโดยเพื่อนและผู้ใต้บังคับบัญชา ประสบการณ์การบริหารมีผลต่อการให้คะแนนโดยตนเองและผู้อื่น (ผู้บังคับบัญชา เพื่อนและผู้ใต้บังคับบัญชา) อย่างมีนัยสำคัญทางสถิติ เช่น ผู้บังคับบัญชาที่มีประสบการณ์การบริหารมากจะมีการให้คะแนนตนเองสูงกว่าการให้คะแนนโดยเพื่อน ระดับการศึกษา มีผลต่อการให้คะแนนโดยตนเองและผู้อื่น (ผู้บังคับบัญชา เพื่อนและผู้ใต้บังคับบัญชา) อย่างมีนัยสำคัญทางสถิติ เช่น ผู้ที่มีการศึกษาสูงจะมีการให้คะแนนตนเองและผู้อื่นสูง และผู้ที่มีการศึกษาสูงจะได้รับคะแนนประเมินที่สูงจากเพื่อนและผู้บังคับบัญชา ปัจจัยทางบริบท คือ ผู้ประเมินที่มีลักษณะงานแบบบริการจะมีการให้คะแนนตนเองสูงกว่าผู้ประเมินที่มีลักษณะงานแบบเทคนิค และผู้ที่มีจำนวนผู้ใต้บังคับบัญชามากจะได้รับคะแนนสูงจากการประเมินโดยตนเองและผู้อื่น (ผู้บังคับบัญชา เพื่อนและผู้ใต้บังคับบัญชา)

Park (Park, 2006) ศึกษาผลของความสัมพันธ์ระหว่างบุคคลต่อการให้คะแนนผลการปฏิบัติงาน กลุ่มตัวอย่างเป็นพนักงานทุกระดับในบริษัทประกันภัยจำนวน 390 คน ซึ่งทุกตำแหน่งงานจะมีรายละเอียดของการทำงานที่มีความเกี่ยวข้องกัน โดยกลุ่มตัวอย่างจะได้รับการฝึกอบรมการประเมินผลการปฏิบัติงาน 360 องศา ก่อนที่จะทำการประเมินผลการปฏิบัติงานของเพื่อนร่วมงาน จำนวนผลการประเมินมี 1638 รายงาน ซึ่งใน 605 คู่จะถูกคัดออกเนื่องจากข้อมูลไม่ครบถ้วน ดังนั้นจึงเหลือ 1033 รายงาน ซึ่งประกอบด้วยผู้ถูกประเมิน 278 คน และผู้ประเมิน 210 คน ซึ่งเฉลี่ยผู้ถูกประเมิน 1 คนต่อผู้ประเมิน 3.7 คน และผู้ประเมิน 1 คนต่อผู้ถูกประเมิน 4.9 คน เครื่องมือที่ใช้วัดประกอบด้วย แบบวัดความสัมพันธ์ระหว่างบุคคลซึ่งพัฒนาโดย Tsui & Barry ปี 1986 ซึ่งเป็นมาตรประมาณค่า 7 ช่วง (Cronbach's alpha = .90) และแบบ ประเมินพฤติกรรมกรรมการทำงานทั่วไป (contextual work behaviors) ซึ่งเกี่ยวข้องกับวิสัยทัศน์และการให้คุณค่าขององค์กร จำนวน 20 ข้อ เป็นมาตรประมาณค่า 5 ช่วง วิเคราะห์องค์ประกอบแบบ principal component และ varimax rotation ได้ 3 องค์ประกอบ คือ การคำนึงถึงผู้ร่วมงานจำนวน 9 ข้อ (Cronbach's alpha = .88) การสื่อสารกับผู้ร่วมงาน จำนวน 4 ข้อ (Cronbach's alpha = .80) และความรับผิดชอบในงานของตน จำนวน 4 ข้อ (Cronbach's alpha = .76) ที่เหลือจำนวน 3 ข้อได้ตัดออกเนื่องจากไม่จัดเข้ากลุ่มองค์ประกอบใด ผลการวิเคราะห์ข้อมูลพบว่า ความสัมพันธ์ระหว่างบุคคลมีความสัมพันธ์กับการให้คะแนนโดยเพื่อน และความรับผิดชอบในงานของตนอย่างมีนัยสำคัญทางสถิติ ($r = .37, .37$ ตามลำดับ $p < .001$) ส่วนการวิเคราะห์ regression

(ordinary least squares) โดยเพิ่มตัวแปรสถานภาพ (ผู้ประเมินและผู้ถูกประเมิน) พบว่าความสัมพันธ์ระหว่างบุคคลส่งผลทางบวกต่อการให้คะแนน โดยมีค่า coefficients = .235 ($p < .001$) และเมื่อพิจารณาเป็นรายองค์ประกอบ ได้แก่ การคำนึงถึงผู้ร่วมงานจำนวน การสื่อสารกับผู้ร่วมงาน และความรับผิดชอบในงานของตน มีค่า coefficients = .235, .228, และ .204 ตามลำดับ ($p < .001$)

Payne, et al. (Payne, Horner, Boswell, Schroeder, & Stine-Cheyne, 2009) ศึกษาเปรียบเทียบผลลัพธ์ของการประเมินผลการปฏิบัติงานระหว่างรูปแบบออนไลน์กับรูปแบบดั้งเดิม (paper-and-pencil approach) โดยใช้การวิจัยกึ่งทดลอง กลุ่มตัวอย่างเป็นพนักงานเจ้าหน้าที่ในมหาวิทยาลัยซึ่งจะต้องทำการประเมินผลการปฏิบัติงานประจำปีโดยแบ่งเป็น 2 กลุ่ม คือ กลุ่มที่ใช้การประเมินแบบออนไลน์จำนวน 83 คน และกลุ่มที่ใช้การประเมินแบบดั้งเดิมจำนวน 152 คน ซึ่งทั้งสองกลุ่มพบว่าการตอบแบบประเมินไม่ครบถ้วน โดยมีจำนวนผู้ที่ตอบครบถ้วนในกลุ่มประเมินแบบออนไลน์จำนวน 102 คน และกลุ่มประเมินแบบดั้งเดิมจำนวน 62 คน ด้านเครื่องมือการประเมินผลการปฏิบัติงานทั้งสองรูปแบบจะมีรายละเอียดเหมือนกัน คือ มาตรฐานวัดประมาณค่า 4 ระดับ (4=มีประสิทธิภาพมาก...1=ต่ำกว่าที่คาดไว้) ส่วนแบบวัดผลลัพธ์ของการประเมินซึ่งประกอบด้วย 6 ตัวแปร คือ ความรับผิดชอบที่ตรวจสอบได้ต่อผู้ที่มีตำแหน่งสูงกว่าตน (supervisor accountability) การรับรู้ในเรื่องความปลอดภัยของข้อมูล คุณภาพของการประเมิน ความพึงพอใจต่อการประเมิน การใช้ประโยชน์ และการมีส่วนร่วมในการประเมิน ได้มีการวิเคราะห์องค์ประกอบเชิงยืนยัน วิเคราะห์ข้อมูลโดยใช้สถิติ MANOVA ผลการวิจัยพบว่า พนักงานที่ใช้การประเมินแบบออนไลน์มีระดับความรับผิดชอบที่ตรวจสอบได้ต่อผู้ที่มีหน้าที่สูงกว่าตน และการมีส่วนร่วมในการประเมินมากกว่าพนักงานที่ใช้การประเมินแบบดั้งเดิมอย่างมีนัยสำคัญทางสถิติ (ความรับผิดชอบที่ตรวจสอบได้ มีค่าเฉลี่ย = 3.55, 3.27 และการมีส่วนร่วมมีค่าเฉลี่ย = 0.17, -0.08 ที่ $p < .05$) และในด้านคุณภาพของการประเมิน พบว่าการประเมินแบบออนไลน์มีระดับต่ำกว่าการประเมินแบบดั้งเดิมอย่างมีนัยสำคัญทางสถิติ (ค่าเฉลี่ย = 3.12, 3.51 ที่ $p < .05$) ส่วนการรับรู้ในเรื่องความปลอดภัยของข้อมูล การใช้ประโยชน์ และความพึงพอใจต่อการประเมิน พบว่า ไม่มีความแตกต่างกันระหว่างการประเมินทั้งสองรูปแบบ

Roch (Roch, 2007) ศึกษาความสำคัญของการประชุมทีม กฎการตัดสินใจร่วมกัน และแรงจูงใจของผู้ประเมิน ที่ส่งผลต่อความแม่นยำในการให้คะแนนของทีมผู้ประเมิน กลุ่มตัวอย่างเป็นผู้ประเมินจำนวน 384 คน ซึ่งในจำนวนนี้เป็นเพศชาย 40% โดยแบ่งผู้ประเมินเป็น 111 ทีม แต่ละทีมมีสมาชิกประมาณ 3-4 คน ซึ่งจะมีการสุ่มเข้ากลุ่มใน 6 สถานการณ์ (3 กลุ่มทดลอง ได้แก่ กลุ่มควบคุม กลุ่มอภิปราย ซึ่งจะมีเฉพาะการอภิปรายแลกเปลี่ยนความคิดเห็นร่วมกันในการให้คะแนน และกลุ่มฉันทมติที่มี

การแลกเปลี่ยนความคิดเห็นแล้วหาข้อสรุปที่ใช้ร่วมกัน และ 3 กลุ่มที่แบ่งตามช่วงเวลาการประเมิน ได้แก่ การจำแนกพฤติกรรม การให้คะแนน และการทำแบบประเมินแรงจูงใจ) เครื่องมือที่ใช้ประกอบด้วย 1)คลิปวิดีโอบรรยายความผิดปกติในการนอน 2 คลิป โดยผู้บรรยายแต่ละคลิปเป็นนักศึกษาระดับบัณฑิตศึกษาจำนวน 1 คน/คลิป 2)แบบบันทึกพฤติกรรม 6 ด้าน โดยบันทึกว่า มี/ไม่มีพฤติกรรมนั้นๆ และใช้เกณฑ์ คือ hit rate และ false alarm rate 3)แบบประเมินการปฏิบัติ 3 ประเด็น คือ ความลึกซึ้งของเนื้อหา การจัดลำดับ และความสามารถในการพูด ซึ่งเป็นมาตรฐานค่า 5 ช่วง 3)แบบวัดแรงจูงใจของผู้ประเมิน 5 ข้อ ซึ่งปรับมาจาก Hedge & Teachout (2000) และ 4)แบบวัดคุณลักษณะทั่วไป เช่น เพศ อายุ ส่วนการวิเคราะห์ความแม่นยำในการให้คะแนนจะคำนวณจากองค์ประกอบความถูกต้องของคะแนนเฉลี่ยของผู้ถูกประเมินแต่ละคนในแต่ละมิติ (Differential accuracy) และองค์ประกอบความถูกต้องของคะแนนเฉลี่ยในทุกมิติระหว่างผู้ถูกประเมินแต่ละคน (Differential elevation) ตาม Cronbach ปี 1955 และใช้ true scores ตามวิธีคำนวณของ Sulsky & Balzer (Sulsky & Balzer, 1988) ผลการวิเคราะห์ข้อมูลด้วย ANOVA พบว่า 1) มีปฏิสัมพันธ์ระหว่างกลุ่มทดลองและช่วงเวลาการประเมินที่ส่งผลต่อความถูกต้องของการบันทึกพฤติกรรม (ประเมินจากความสามารถในการจำของผู้ประเมินเป็นรายคน) โดยมีค่า $F=2.98, p=.05$ และเมื่อวิเคราะห์ simple main effects พบว่า ในกลุ่มอภิปรายจะมีความถูกต้องของการบันทึกพฤติกรรมมากกว่ากลุ่มควบคุมและกลุ่มฉันทามติ 2) ไม่พบปฏิสัมพันธ์ร่วมและอิทธิพลหลักระหว่างกลุ่มทดลองและช่วงเวลาการประเมินที่ส่งผลต่อแรงจูงใจของผู้ประเมิน แสดงว่า การอภิปรายแลกเปลี่ยนความคิดเห็นระหว่างกันของผู้ประเมินไม่มีอิทธิพลต่อแรงจูงใจของผู้ประเมิน (ส่วนการวิเคราะห์ความสัมพันธ์ระหว่างแรงจูงใจของผู้ประเมินกับความถูกต้องของการบันทึกพฤติกรรม พบว่า กลุ่มอภิปรายมีค่า $r=.23, p<.05$ ส่วนกลุ่มควบคุมและกลุ่มฉันทามติไม่พบว่ามีนัยสำคัญทางสถิติ $r=-.04, -.01$ ตามลำดับ)

Salvemini, Reilly, & Smither (Salvemini, Reilly, & Smither, 1993) ศึกษาผลของแรงจูงใจในผู้ประเมินที่มีต่อความแม่นยำในการประเมินผลการปฏิบัติงาน กลุ่มตัวอย่างเป็นนักศึกษาอาสาสมัครจำนวน 108 คน และมีการให้ค่าตอบแทนคนละ 6\$ โดยกลุ่มตัวอย่างจะได้รับการสุ่มเข้ากลุ่ม 9 กลุ่ม (การสร้างแรงจูงใจแก่ผู้ประเมินโดยบอกว่า ถ้าสามารถประเมินผลการปฏิบัติงานของผู้ถูกประเมินได้อย่างแม่นยำจะได้รับเงินค่าตอบแทน ซึ่งลักษณะของการสร้างแรงจูงใจจะมี 3 รูปแบบ คือ บอกก่อนการดูวิดีโอ บอกหลังดูวิดีโอ และไม่มีการบอก * การให้ข้อมูลการทำงานของผู้ถูกประเมินมี 3 ระดับ คือ เชิงบวก เชิงลบ และไม่ให้ข้อมูล) ในการวิจัยจะให้กลุ่มตัวอย่างดูวิดีโอ 2 เรื่อง คือ เรื่องแรกจะให้ข้อมูลเกี่ยวกับลักษณะการทำงานใน 3 ด้าน คือ การขาย การติดตั้ง และการแก้ไขปัญหา และมี

ตัวอย่างของพนักงานชายที่มีการปฏิบัติงานทั้ง 3 ด้าน อยู่ในระดับปานกลาง ส่วนเรื่องที่สองจะมีความยาว 8-13 นาที ประกอบด้วยผู้แสดง 5 คนในสถานการณ์ของการปฏิบัติงานชาย โดยผู้ประเมินต้องทำการประเมินลักษณะการให้บริการทั้ง 3 ด้าน ส่วนการวิเคราะห์ความแม่นยำในการให้คะแนนใช้แนวคิดของ Cronbach ปี 1955 โดยเลือกเฉพาะ 3 องค์ประกอบ คือ องค์ประกอบความถูกต้องของคะแนนเฉลี่ยรวมของผู้ถูกประเมินทุกคนและทุกมิติ (Elevation) องค์ประกอบความถูกต้องของคะแนนเฉลี่ยในทุกมิติระหว่างผู้ถูกประเมินแต่ละคน (Differential elevation) และองค์ประกอบความถูกต้องของคะแนนเฉลี่ยของผู้ถูกประเมิน แต่ละคนในแต่ละมิติ (Differential accuracy) และมีการเพิ่มองค์ประกอบความถูกต้องของคะแนนเฉลี่ยของผู้ประเมินแต่ละคนในทุกผู้ถูกประเมินและทุกมิติ (overall average rating) ซึ่งในองค์ประกอบนี้จะใช้เป็นตัวแปรตามในงานวิจัย ผลการวิจัยเมื่อวิเคราะห์ด้วย ANOVA พบปฏิสัมพันธ์ระหว่างการสร้างแรงจูงใจแก่ผู้ประเมิน และการให้ข้อมูลการทำงานของผู้ถูกประเมิน ($F=7.80, p<.01$) และเมื่อทำ simple effects tests ในกรณีที่ไม่มีการสร้างแรงจูงใจ พบว่าการให้ข้อมูลการทำงานของผู้ถูกประเมินจะส่งผลต่อความแม่นยำในการให้คะแนน ($F=21.80, p<.01$) โดยกลุ่มที่ได้รับข้อมูลเชิงบวกจะให้คะแนนที่แม่นยำมากกว่ากลุ่มที่ไม่ให้ข้อมูลอย่างมีนัยสำคัญทางสถิติ (ค่าเฉลี่ย=120.92 และ 95.37 ตามลำดับ, $F=17.26, p<.01$) ส่วนกลุ่มที่ได้รับข้อมูลเชิงลบจะมีการให้คะแนนที่แม่นยำน้อยกว่ากลุ่มที่ไม่ให้ข้อมูลอย่างมีนัยสำคัญทางสถิติ (ค่าเฉลี่ย=81.13 และ 95.37 ตามลำดับ, $F=5.40, p<.05$) แสดงว่าลักษณะของการให้ข้อมูลมีผลต่อความแม่นยำในการให้คะแนน และพบว่า การให้ข้อมูลไม่ส่งผลต่อความแม่นยำในการให้คะแนนเมื่อมีการสร้างแรงจูงใจก่อนดูวิดีโอ ($F=.27, p>.05$) และหลังดูวิดีโอ ($F=.73, p>.05$)

Simsek, Pakdil, Dengiz, & Testik (Simsek, Pakdil, Dengiz, & Testik, 2013) ได้ศึกษาการประเมินผลการปฏิบัติงานของพนักงานขับรถ กลุ่มตัวอย่าง คือ พนักงานขับรถจำนวน 8 คน ซึ่งประจำรถบรรทุกสินค้าจำนวน 8 คัน โดยในงานวิจัยนี้ได้พัฒนารอบการประเมินผลการปฏิบัติงาน 7 ขั้นตอน ได้แก่ 1) การกำหนดวัตถุประสงค์เฉพาะสำหรับการประเมินผลการปฏิบัติงาน ซึ่งในงานวิจัยนี้จะกำหนดในบริบทของบริษัทขนส่งสินค้า ได้แก่ การประเมินพฤติกรรมการขับขี่ตามข้อกำหนด สาเหตุของการละเมิด ข้อกำหนดและแนวทางการแก้ไข การประเมินความจำเป็นสำหรับการฝึกอบรม การติดตามพัฒนาการและประสิทธิผลในการทำงาน การให้ข้อมูลสะท้อนกลับเกี่ยวกับผลการปฏิบัติงาน และการให้ข้อมูลในการตัดสินใจ 2) การกำหนดเกณฑ์การประเมินผลการปฏิบัติงาน ในงานวิจัยนี้จะเน้นกำหนดเกณฑ์ที่มีความยุติธรรมและเป็นปรนัย ซึ่งจะลดอคติและทำให้มีความน่าเชื่อถือ โดยวัดจากเกณฑ์ 3 ข้อ คือ ความเร็วในการขับรถตามที่กำหนด อัตราการหยุดรถกรณีที่ติด

เครื่องยนต์แต่ไม่ได้เคลื่อนที่ซึ่งจะมีผลต่อต้นทุนการใช้น้ำมันเชื้อเพลิงและการบำรุงรักษาเครื่องยนต์ และค่าใช้จ่ายของน้ำมันเชื้อเพลิง 3) การกำหนดวิธีการประเมินผลการปฏิบัติงาน โดยทั่วไปในการให้คะแนนการปฏิบัติงานจะเป็นลักษณะข้อมูลเชิงคุณภาพและเป็นอัตนัย ซึ่งทำให้พนักงานคนเดียวกันอาจได้รับคะแนนแตกต่างกันในผู้ประเมินที่ต่างกัน ดังนั้นในงานวิจัยนี้จึงได้มีการนำ statistical process control (SPC) ซึ่งเป็นเครื่องมือในการรวบรวมข้อมูลของผู้ขับขี่แต่ละคนโดยแสดงผลเป็นตัวเลขและกราฟ ทำให้กระบวนการให้คะแนนมีความคงที่และลดความคลาดเคลื่อน 4) การรวบรวมข้อมูล (Collecting data) โดยวัดจากเครื่อง GPS (Global Positioning System) ที่ติดไว้กับรถยนต์ ซึ่งจะส่งสัญญาณผ่านดาวเทียมไปยังเครื่องคอมพิวเตอร์ ที่จะมีการบันทึกแบบออนไลน์ทั้งระยะเวลาจอดและความเร็วในการขับรถของพนักงานแต่ละคน และ 5) การวิเคราะห์ข้อมูลและประเมินผลการปฏิบัติงาน จากการใช้ SPC ในการเก็บรวบรวมและวิเคราะห์ข้อมูลของผู้ขับขี่แต่ละคน จะแสดงผลข้อมูลใน 5 รูปแบบ ได้แก่ กราฟแท่งแสดงอัตราความเร็วของรถยนต์ กราฟเส้นแสดงสัดส่วนของการขับรถเร็วเกินอัตราที่กำหนดต่อเดือน กราฟแท่งแสดงจำนวนครั้งที่ขับเร็วเกินอัตราที่กำหนดต่อวัน box-plots แสดงปริมาณการขับรถเร็วเกินอัตราที่กำหนดเปรียบเทียบในผู้ขับขี่แต่ละคน และ ข้อมูลการวิเคราะห์ค่าใช้จ่ายของน้ำมันเชื้อเพลิงกรณีติดเครื่องยนต์แต่ไม่ได้เคลื่อนที่ของผู้ขับขี่แต่ละคน 6) การนำผลการประเมินไปใช้ ซึ่งในงานวิจัยนี้มี 2 ประเด็น คือ ใช้สำหรับกระบวนการคัดเลือกบรรจุพนักงานขับรถในช่วงทดลองงาน และใช้ในการอบรมและพัฒนาตามความจำเป็นของพนักงานแต่ละคน และ 7) การสัมภาษณ์ความคิดเห็น ภายหลังจากเสร็จสิ้นการประเมินผู้ขับรถทุกคนจะมีการพูดคุยสัมภาษณ์ความคิดเห็นระหว่างผู้ประเมินกับผู้ขับขี่ที่ละคน โดยจะเป็นการให้ข้อมูลการประเมินผล และพูดคุยเกี่ยวกับเหตุผลในการขับรถเร็วเกินอัตราที่กำหนดและการจอดขณะที่ติดเครื่องยนต์ ซึ่งจะ เป็นประโยชน์สำหรับพัฒนาการปฏิบัติงานต่อไป ซึ่งผลสรุปในงานวิจัยนี้พบว่า การใช้ข้อมูลที่วัดได้จากเครื่อง GPS มีประโยชน์มากสำหรับการประเมินผลการปฏิบัติงานของผู้ขับขี่รถยนต์

Tziner et al. (Tziner, Murphy, Cleveland, Yavo, & Hayoon, 2008) ศึกษาปัจจัยที่เกี่ยวข้องกับพฤติกรรมการให้คะแนน ได้แก่ ความเชื่อเกี่ยวกับระบบการประเมินผลการปฏิบัติงาน และบุคลิกภาพ กลุ่มตัวอย่างเป็นพนักงานเขียนโปรแกรมคอมพิวเตอร์จำนวน 100 คน โดยแต่ละคนจะต้องทำแบบสอบถามซึ่งประกอบด้วยข้อมูลทั่วไปและตัวแปรที่เกี่ยวข้อง ได้แก่ ตัวแปรความเชื่อเกี่ยวกับระบบการประเมินผลการปฏิบัติงาน ประกอบด้วย 1) การรับรู้ความสามารถของตน (self-efficacy) ในการประเมินให้คะแนนการปฏิบัติงาน ซึ่งวัดจากข้อคำถามจำนวน 9 ข้อที่ปรับปรุงมาจาก Napier & Latham ปี1986 (Cronbach's alpha = .87) 2) ความเชื่อมั่นในระบบการประเมินผล

การทำงาน (confidence in the appraisal system) ซึ่งวัดจากข้อคำถามจำนวน 17 ข้อที่นำมาจาก Tziner, Prince, and Murphy's ปี 1997 (Cronbach's alpha = .85) และ 3) ความคิดเห็นเกี่ยวกับความสะดวกในการประเมินผลการทำงาน ซึ่งวัดจากข้อคำถามจำนวน 9 ข้อที่นำมาจาก Villanova et al. ปี 1993 (Cronbach's alpha = .83) และตัวแปรบุคลิกภาพซึ่งวัดจากบุคลิกภาพแบบกำกับตนเอง (self-monitoring) จำนวน 6 ข้อที่นำมาจาก Gagenstad & Snyder's ปี 1985 (Cronbach's alpha = .69) และกลุ่มตัวอย่างจะต้องทำการประเมินให้คะแนนการปฏิบัติงานของเพื่อนจำนวน 3 คน ซึ่งเป็นมาตรฐานค่า 6 ช่วง (Likert-type scale) สำหรับความถี่ของการเกิดพฤติกรรม จำนวน 13 พฤติกรรม โดยมีค่าตั้งแต่ 1 = ไม่เคย หรือ 0% จนถึง 1 = ทุกครั้ง หรือ 100% (Cronbach's alpha = .91) วิเคราะห์พฤติกรรมการให้คะแนนเป็นรายคนในด้าน ratings level, discrimination among ratees, และ discrimination among dimensions โดยหาค่า M , SD , และ Z-scores ผลการศึกษาพบว่า ความเชื่อเกี่ยวกับระบบการประเมินผลการปฏิบัติงานและบุคลิกภาพแบบ กำกับตนเองมีความสัมพันธ์ในระดับปานกลางกับพฤติกรรมการให้คะแนน โดยผู้ประเมินซึ่งเป็นเพื่อนกับผู้ถูกประเมินและมีความเชื่อว่ระบบการประเมินผลการปฏิบัติงานมีความสำคัญสำหรับการจำแนกผู้ถูกประเมินจะมีแนวโน้มที่จะให้คะแนนประเมินแบบสูง นอกจากนี้เมื่อวิเคราะห์การเป็นตัวแปรกำกับของตัวแปรบุคลิกภาพแบบการกำกับตนเองที่มีผลต่อความสัมพันธ์ระหว่างความเชื่อเกี่ยวกับระบบการประเมินผลการปฏิบัติงานและพฤติกรรมการให้คะแนนโดยเพื่อนซึ่งใช้การวิเคราะห์ถดถอย (moderated regression) พบว่า R^2 -values มีค่าเพิ่มขึ้นอยู่ระหว่าง .12 ถึง .20 ซึ่งแสดงว่าตัวแปรบุคลิกภาพแบบการกำกับตนเองเป็นตัวแปรกำกับความสัมพันธ์ระหว่างความเชื่อเกี่ยวกับระบบการประเมินผลการปฏิบัติงานและพฤติกรรมการให้คะแนน

Vecchio & Anderson (Vecchio & Anderson, 2009) ศึกษาอิทธิพลของความแตกต่างระหว่างบุคคลในด้านคุณลักษณะส่วนบุคคล (demographics) ซึ่งได้แก่ เพศ อายุ และเชื้อชาติ และบุคลิกภาพส่วนบุคคล (personality attributes) ซึ่งได้แก่ บุคลิกภาพแบบใส่ใจสังคม (social sensitivity) และบุคลิกภาพแบบมุ่งอำนาจในสังคม (social dominance) ที่มีผลต่อความสอดคล้องในการให้คะแนนประสิทธิผลของความเป็นผู้นำจากการประเมินโดยตนเองและผู้อื่น (หัวหน้า เพื่อน และผู้ใต้บังคับบัญชา) กลุ่มตัวอย่างมีตำแหน่งงานระดับผู้จัดการจำนวน 1221 คน ซึ่งมาจากหลากหลายสาขาอาชีพ เช่น การศึกษา สุขภาพ การปกครอง ประกันภัย เป็นต้น โดยได้เข้าร่วมในโครงการให้ข้อมูลย้อนกลับแบบ 360 องศา โดยผู้ประเมินซึ่งเป็นเพื่อนและผู้ใต้บังคับบัญชาจะได้รับการสุ่มอย่างละ 1 คน เครื่องมือที่ใช้วัดบุคลิกภาพแบบใส่ใจสังคมมีจำนวน 3 ข้อ (Cronbach's

alpha = .80) บุคลิกภาพแบบมุ่งอำนาจในสังคมมีจำนวน 5 ข้อ (Cronbach's alpha = .82) และประสิทธิผลของความเป็นผู้นำจำนวน 5 ข้อ (Cronbach's alpha = .88) จะมาจากแบบสอบถามย่อย (subscales) ของเครื่องมือที่สร้างโดย Anderson ปี 2006 (Leadership Circle Profile) ซึ่งมีจำนวนรวม 144 ข้อ เป็นมาตราประมาณค่า (Likert-type scale) วิเคราะห์ข้อมูลโดยใช้ polynomial regression analyses ผลการศึกษาพบว่า เพศและอายุมีความสัมพันธ์กับการให้คะแนนทั้งโดยตนเองและผู้อื่น โดยผู้ประเมินอื่นๆ (หัวหน้า เพื่อน และผู้ใต้บังคับบัญชา) ที่เป็นเพศชายและอายุมากมักจะมีการให้คะแนนประเมินแบบสูง ส่วนผู้ถูกประเมินที่มีบุคลิกภาพแบบใส่ใจสังคมจะได้รับคะแนนประเมินแบบสูงจากเพื่อนและผู้ใต้บังคับบัญชา นอกจากนี้ความสอดคล้องในการให้คะแนนระหว่างผู้ประเมินยังมีความสัมพันธ์กับเพศ บุคลิกภาพแบบใส่ใจสังคม และบุคลิกภาพแบบมุ่งอำนาจในสังคมของผู้ประเมิน

Wang, Wong & Kwong (Wang, Wong, & Kwong, 2010) ศึกษาเป้าหมายของผู้ประเมินทั้ง 4 แบบ ได้แก่ เป้าหมายเพื่อการจำแนก (identification goal) ซึ่งเป็นเงื่อนไขควบคุม (control condition) เป้าหมายเพื่อความสามัคคี (harmony goal condition) เป้าหมายเพื่อความยุติธรรม (fairness goal condition) และเป้าหมายเพื่อการจูงใจ (motivating goal condition) ที่ส่งผลต่อการให้คะแนนผู้ถูกประเมินที่มีระดับความสามารถแตกต่างกัน โดยแบ่งการศึกษาออกเป็น 2 เรื่อง เรื่องที่ 1 ใช้ข้อมูลเดียวกับการศึกษาในปี 2007 (103 คน) โดยเป็นการให้คะแนนเพื่อนในกลุ่มและเป้าหมายของผู้ประเมินจะจัดเป็นตัวแปรภายในส่วนบุคคล (within-subject factor) ส่วนเรื่องที่ 2 กลุ่มตัวอย่างเป็นนักศึกษาจำนวน 120 คน ที่ลงทะเบียนเรียนวิชาพฤติกรรมองค์กร แบ่งเป็นเพศหญิงจำนวน 60 คน เพศชายจำนวน 52 คน และ 8 คนไม่ได้ระบุเพศตนเอง มีอายุเฉลี่ย 18.8 ปี ผู้ประเมินจะต้องดูวิดีโอจากซีรีส์ทางโทรทัศน์เรื่อง The Apprentice (Season 2, Episode 11) ที่มีความยาว 15 นาที และได้มีการศึกษานำร่อง (pilot study) เพื่อทดลองสื่อวิดีโอนี้ โดยใช้นักศึกษาจำนวน 13 คน พบว่า ทีม 1 มีคะแนนเฉลี่ย 5.29 ซึ่งสูงกว่าทีม 2 ซึ่งได้คะแนนเฉลี่ย 3.88 จึงแสดงว่าสื่อวิดีโอนี้สามารถจำแนกระดับความสามารถในการปฏิบัติงานของทั้ง 2 กลุ่มได้แตกต่างกัน ทั้งนี้กลุ่มตัวอย่าง 120 คนจะต้องทำการประเมินความสามารถของสมาชิกในกลุ่มทั้ง 6 คนจากในวิดีโอ ซึ่งแบ่งเป็น 2 กลุ่มให้ทำการกิจออกแบบขวดเป๊ปซี่ตั้งแต่การออกแบบและการนำเสนอผลการออกแบบต่อผู้บริหารของบริษัทเป๊ปซี่ แบบประเมินจะเป็นมาตราประมาณค่า 5 ช่วง (Likert-type scale) ตั้งแต่ 1 (ระดับต่ำ) ถึง 5 (ระดับยอดเยี่ยม) จำนวน 10 ข้อ ซึ่งจะถามเกี่ยวกับนวัตกรรมในการออกแบบ การมีส่วนร่วมของสมาชิก ทักษะการนำเสนอ จิตวิญญาณของทีม (team work spirit) และอื่นๆ โดยมีค่าความตรงเท่ากับ .90 (คิดรวมทั้ง 4 เป้าหมาย) โดยผู้ประเมินคะแนนแต่ละคนจะถูกกำหนดเป้าหมายให้โดยทำการสุ่มมา

จากทั้ง 4 เป้าหมาย ซึ่งเป้าหมายของผู้ประเมินจะจัดเป็นตัวแปรระหว่างบุคคล (between-subjects factor) หรือเป็นปัจจัยภายนอก ผลการศึกษาสรุปได้ว่าในการศึกษาเรื่องที่ 1 ผู้ประเมินมีการให้คะแนนแตกต่างกันตามเป้าหมายของตน และพบว่าใน 3 เป้าหมาย คือ เป้าหมายเพื่อความสามัคคี เป้าหมายเพื่อความยุติธรรม และเป้าหมายเพื่อการจูงใจ จะมีการประเมินเพื่อนแบบปล่อยคะแนน และพบว่าจะมีการปล่อยคะแนนมากให้เพื่อนที่มีความสามารถระดับต่ำมากกว่าเพื่อนที่มีความสามารถระดับกลาง ส่วนผลการศึกษาเรื่องที่ 2 ซึ่งเป็นบริบทผู้ประเมินไม่มีความเกี่ยวข้องกับผู้ถูกประเมิน พบว่า ในกลุ่มผู้ประเมินตามเป้าหมายเพื่อความยุติธรรมจะมีการกดคะแนนสำหรับผู้ถูกประเมินที่มีความสามารถระดับสูง และในกลุ่มผู้ประเมินตามเป้าหมายเพื่อการจูงใจจะมีการปล่อยคะแนนสำหรับผู้ถูกประเมินที่มีความสามารถระดับต่ำ

Wong & Kwong (Wong & Kwong, 2007) ศึกษาวิจัยกึ่งทดลองเกี่ยวกับผลของเป้าหมายของผู้ประเมิน (rater goals) ที่ส่งผลต่อรูปแบบของการให้คะแนน โดยกลุ่มตัวอย่างเป็นนักศึกษา มหาวิทยาลัยจำนวน 103 คน ในวิชาการบริหารทรัพยากรมนุษย์ แบ่งเป็นกลุ่มย่อย 14 กลุ่มๆ ละ 7-8 คน เพื่อทำโครงการประเมินทรัพยากรมนุษย์ในบริษัท ในแต่ละกลุ่มจะต้องมีการส่งเค้าโครงของรายงานและนำเสนอโครงการในชั้นเรียนเพื่อเป็นส่วนหนึ่งของการประเมินผลรายวิชาซึ่งคิดเป็น 25% ของการตัดเกรด ทั้งนี้นักศึกษาแต่ละคนจะต้องประเมินให้คะแนนเพื่อน 2 ครั้งโดยเก็บข้อมูล 2 ช่วงเวลาคือ หลังส่งรายงานเค้าโครงเบื้องต้นประมาณกลางเทอม และรายงานฉบับสมบูรณ์ประมาณปลายเทอม โดยคะแนนเฉลี่ยจากทั้ง 2 ช่วงเวลาจะถูกนำไปคิดเป็นคะแนน 20% ของงานกลุ่ม ซึ่งผู้เรียนจะได้รับการอธิบายวิธีการประเมินผลการเรียนและจะได้รับแจ้งผลคะแนนเฉลี่ย จากการประเมินโดยเพื่อนภายหลังการประเมิน 1 สัปดาห์ เพื่อที่จะเป็นการสร้างหรือจัดกระทำกับเป้าหมายของผู้ประเมิน (rater goals) ซึ่งต่างจากปกติที่นักศึกษาจะไม่ทราบผลคะแนนประเมิน ในแบบประเมินที่นักศึกษาใช้จะมีส่วนหนึ่งที่ถามเกี่ยวกับเป้าหมายของผู้ประเมิน ซึ่งลำดับของเป้าหมายจะเป็นแบบสุ่มจากทั้ง 4 แบบ คือ เป้าหมายเพื่อการจำแนก (identification goal) ซึ่งเป็นเงื่อนไขควบคุม (control condition) จะเป็นการถามเกี่ยวกับเป้าหมายของการประเมินเพื่อนเพื่อที่จะจำแนกจุดแข็งและจุดอ่อน เป้าหมายเพื่อความสามัคคี (harmony goal condition) ซึ่งจะเป็นการถามเกี่ยวกับเป้าหมายของการประเมินเพื่อนเพื่อที่จะรักษาสัมพันธ์ภาพและความสามัคคีของกลุ่ม เป้าหมายเพื่อความยุติธรรม (fairness goal condition) ซึ่งจะเป็นการถามเกี่ยวกับเป้าหมายของการประเมินเพื่อนเพื่อที่จะสะท้อนถึงความถูกต้องและยุติธรรม และเป้าหมายเพื่อการจูงใจ (motivating goal condition) ซึ่งจะเป็นการถามเกี่ยวกับเป้าหมายของการประเมินเพื่อนเพื่อที่จะเกิดการจูงใจให้

สมาชิกในกลุ่มมีความตั้งใจ และภายในชุดของแบบประเมินจะมีการระบุไว้ในแบบประเมินประกอบด้วย 4 ชุด ในแต่ละชุดจะต้องเติมข้อมูลเกี่ยวกับเป้าหมายของตนซึ่งจะมีความแตกต่างกันให้สมบูรณ์ และในแบบประเมินจะมีการระบุชื่อเพื่อนที่จะต้องประเมินให้โดยให้นักศึกษาทำการให้คะแนนตั้งแต่ 1 (ไม่มีผลงาน) ถึง 7 (มีผลงานชัดเจน) ทั้งนี้ข้อมูลในส่วนของการประเมินตนเองจะไม่นำมาใช้วิเคราะห์ตัวแปรตามที่ใช้วิเคราะห์ ได้แก่ คะแนนเฉลี่ยที่คำนวณจากสมาชิกกลุ่มทุกคน และค่าเบี่ยงเบนมาตรฐานที่คำนวณจากสมาชิกกลุ่มทุกคนซึ่งจะแยกเป็นกลางเทอมและปลายเทอม ซึ่งทั้งสองค่าพบว่ามีผลสอดคล้องภายใน (internal consistency) ระหว่างเป้าหมายทั้ง 4 แบบ คือ Cronbach's alpha ของคะแนนเฉลี่ยและค่าเบี่ยงเบนมาตรฐานในช่วงกลางภาค มีค่า .86 และ .85 ตามลำดับ ส่วน Cronbach's alpha ของคะแนนเฉลี่ยและค่าเบี่ยงเบนมาตรฐานในช่วงปลายภาค มีค่า .93 และ .94 ตามลำดับ ผลการวิเคราะห์ repeated measures ANOVA พบว่า อิทธิพลหลัก (main effect) ของเป้าหมายของผู้ประเมินส่งผลต่อค่าเบี่ยงเบนมาตรฐานของคะแนนประเมินที่ต่างกัน $F(3, 306) = 32.99, p < .001$ ซึ่งสนับสนุนว่า เป้าหมายของผู้ประเมินส่งผลต่อความแตกต่างในการให้คะแนน

จากตัวอย่างงานวิจัยทั้ง 18 เรื่องดังที่ได้กล่าวมา ซึ่งมีความเกี่ยวข้องกับการศึกษาปัจจัยเชิงสาเหตุที่ส่งผลต่อความแม่นยำในการให้คะแนน สามารถสรุปเป็นตารางสังเคราะห์งานวิจัย (ตารางที่ 3) ดังนี้ (รายชื่อผู้วิจัย 1 = Bernardin, Tyler, & Villanova (2009), 2 = Curtis, Harvey, & Ravden (2005), 3 = Gorman & Rentsch (2009), 4 = Hedge & Teachout (2000), 5 = Heidemeier & Moser (2009), 6 = Khoury & Analoui (2004), 7 = Kim (2011), 8 = Murphy, et al. (2004), 9 = Ostroff, Atwater, & Feinberg (2004), 10 = Park (2006), 11 = Payne, et al. (2009), 12 = Roch (2007), 13 = Salvemini, Reilly, & Smither (1993), 14 = Simsek, Pakdil, Dengiz, & Testik (2013), 15 = Tziner et al. (2008), 16 = Vecchio & Anderson (2009), 17 = Wang, Wong & Kwong (2010), 18 = Wong & Kwong (2007))

ตารางที่ 3 การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุที่ส่งผลต่อความแม่นยำในการให้คะแนน

ตัวแปรที่ ศึกษา	1	2	3	4	5	6	7	8	9	1	1	1	1	1	1	1	1	1	รวม	
										0	1	2	3	4	5	6	7	8		
ผู้ประเมิน (Rater)																				
เป้าหมายของ ผู้ประเมิน (rater goals)		✓		✓	✓			✓		✓					✓		✓	✓		8
คุณลักษณะ ทางประชากร (demographics)		✓			✓				✓			✓				✓				5
ความสามารถใน การให้คะแนน (ability for rating)			✓		✓	✓	✓		✓											5
บุคลิกภาพ แบบมี จิตสำนึก (conscientio usness)	✓				✓										✓	✓				4
แรงจูงใจ (motivation)				✓								✓	✓							3

ตารางที่ 3 (ต่อ)

ตัวแปรที่ ศึกษา	1	2	3	4	5	6	7	8	9	1 0	1 1	1 2	1 3	1 4	1 5	1 6	1 7	1 8	รวม
ความ รับผิดชอบ ที่ตรวจสอบได้ (accountability)	✓	✓		✓							✓								4
การรับรู้ มาตรฐาน (standards)					✓	✓		✓						✓					4
กระบวนการ เปรียบเทียบ (process)					✓	✓								✓					3
กลุ่มผู้ ประเมิน (groups of raters)					✓				✓	✓							✓		4
<u>ผู้ถูกประเมิน</u> (Ratee) คุณลักษณะ ทางประชากร (demographics)					✓				✓										2

จากตารางที่ 3 พบว่า สาเหตุที่ส่งผลต่อความแม่นยำในการให้คะแนนแบ่งเป็นตัวแปรที่วัดจากผู้ประเมินจำนวน 9 ตัวแปร ได้แก่ เป้าหมายของผู้ประเมิน คุณลักษณะทางประชากร

ความสามารถในการให้คะแนน บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ ความรับผิดชอบที่ตรวจสอบได้ การรับรู้มาตรฐาน กระบวนการเปรียบเทียบ และกลุ่มผู้ประเมิน และตัวแปรที่วัดจากผู้ถูกประเมินมีจำนวน 1 ตัวแปร คือ คุณลักษณะทางประชากร ซึ่งพบว่าตัวแปรเกือบทั้งหมดมีงานวิจัยสนับสนุนไม่ต่ำกว่า 3 งานวิจัย ยกเว้นตัวแปรคุณลักษณะทางประชากรของผู้ถูกประเมิน ดังนั้นผู้วิจัยจึงได้คัดเลือกตัวแปรเหล่านี้มาใช้ในการศึกษารวมทั้งสิ้นจำนวน 8 ตัวแปร ส่วนตัวแปรคุณลักษณะทางประชากรของผู้ประเมิน (เช่น เพศ อายุ เชื้อชาติ) ถึงแม้จะมีงานวิจัยสนับสนุนจำนวน 5 งานวิจัย แต่เมื่อพิจารณาประโยชน์ของการนำไปใช้ต่อในด้านการพัฒนาหรือปรับปรุงนั้นจะไม่สามารถกระทำได้ ดังนั้นผู้วิจัยจึงไม่นำมาใช้ในการศึกษาครั้งนี้

ตอนที่ 3 การวิเคราะห์ความไม่แปรเปลี่ยนของโมเดล

กลุ่มตัวอย่างที่มีลักษณะหรือมีตัวแปรบางตัวที่แตกต่างกัน เช่น เพศ ระดับการศึกษา ช่วงอายุ และสถานภาพ เป็นต้น อาจมีผลกระทบต่อผลการสรุปผลการวิจัยหรืออาจทำให้การสรุปผลการวิจัยแตกต่างไปจากการวิเคราะห์ในภาพรวม เนื่องจากการมีพารามิเตอร์หรือโมเดลที่แตกต่างกัน ดังนั้นหากนักวิจัยทำการวิเคราะห์ข้อมูลโดยแยกกลุ่มย่อยตามลักษณะของตัวแปรที่แตกต่างกันนี้ หรือเรียกว่า การวิเคราะห์กลุ่มพหุ (multiple groups) จะทำให้สามารถสรุปผลการวิจัยได้ละเอียดลึกซึ้งตามระดับค่าของตัวแปรที่ใช้ในการแบ่งกลุ่ม

3.1 หลักการวิเคราะห์โมเดลกลุ่มพหุ

การทดสอบความไม่แปรเปลี่ยนของโมเดล (measurement invariance) เป็นการศึกษาความไม่แปรเปลี่ยนของโมเดลตามทฤษฎีระหว่างกลุ่มประชากรที่แตกต่างกันตั้งแต่ 2 กลุ่มขึ้นไปที่มาจากการเก็บข้อมูลในครั้งเดียวกัน (Schmitt & G., 2008) โดยสถิติที่ใช้วิเคราะห์ความแตกต่างระหว่างกลุ่มแบ่งได้เป็น 2 ประเภทใหญ่ คือ สถิติพารามेटริก และสถิตินั้นพารามेटริก นอกจากนี้ยังสถิติได้พัฒนาวิธีการวิเคราะห์ข้อมูลสำหรับตรวจสอบความแตกต่างระหว่างสัมประสิทธิ์สหสัมพันธ์ เกิดเป็นการวิเคราะห์ข้อมูลทางสถิติแบบใหม่ ได้แก่ 1) สถิติวิเคราะห์ข้อมูลหลายระดับ (multi-level analysis) หรือสถิติวิเคราะห์โมเดลเชิงเส้นระดับลดหลั่น (hierarchical linear model) 2) สถิติวิเคราะห์สำหรับการสังเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ในการวิเคราะห์อภิมาน (meta-analysis) และ 3) สถิติการวิเคราะห์กลุ่มพหุ (multi sample analysis) หรือกลยุทธ์กลุ่มพหุ (multiple group strategy) ในการวิเคราะห์ด้วยโปรแกรมลิซเรล ซึ่งสถิติวิเคราะห์ทั้งสามประการนี้มีจุดมุ่งหมายในการพัฒนาคัลล์กันในส่วน และต่างกันในส่วน โดยหลักการในการวิเคราะห์จะคล้ายคลึงกันมาก

โดยเกี่ยวข้องกับเรื่องปฏิสัมพันธ์ (interaction) (นงลักษณ์ นงลักษณ์ วิรัชชัย, 2542) และพบว่าในตัวแปรต่อเนื่อง (continuous variables) มักจะใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยันแบบกลุ่มพหุ (multiple group confirmatory factor analysis) ในการทดสอบความไม่แปรเปลี่ยนของโมเดล (Chen, Sousa, & West, 2005)

การวิเคราะห์กลุ่มพหุ เป็นการศึกษาความสัมพันธ์ของตัวแปรเพื่อมุ่งสรุปผลการวิจัยให้ละเอียดลึกซึ้งตามระดับค่าของตัวแปรที่ใช้ในการแบ่งกลุ่ม หรือเป็นการวิเคราะห์ข้อมูลที่ได้มาจากกลุ่มตัวอย่างหลายๆ กลุ่ม โดยการแยกโมเดลการวิจัยออกเป็นโมเดลย่อยๆ ตามค่าของตัวแปรที่นำมาศึกษา และเป็นรูปแบบหนึ่งของการวิเคราะห์โมเดลสมการโครงสร้าง โดยการวิเคราะห์กลุ่มพหุมีประโยชน์ต่อการสรุปผลการวิจัย 2 ประการ คือ ประการแรก ช่วยตรวจสอบความสัมพันธ์ของตัวแปรในโมเดลการวิจัยว่ามีความเหมือนหรือแตกต่างกันระหว่างกลุ่มหรือไม่อย่างไร และประการที่สอง เพื่อเปรียบเทียบความเหมือนและความต่างของค่าพารามิเตอร์บางค่าในโมเดลการวิจัยตามกลุ่มย่อย ดังนั้นจึงทำให้สามารถสรุปผลการวิจัยได้ทั้งในภาพรวมและภาพย่อยของแต่ละกลุ่มที่จำแนกตามตัวแปรที่นักวิจัยต้องการศึกษา ทั้งนี้ตัวแปรที่ใช้ในการแยกโมเดลการวิจัยออกเป็นกลุ่มย่อยๆ เรียกว่า ตัวแปรกำกับ หรือ ตัวแปรปรับ (moderator variable) ซึ่งมีใช้ตัวแปรต่อเนื่อง หรือถ้าเป็นตัวแปรต่อเนื่องจะต้องนำมาจัดหรือเปลี่ยนรูปให้เป็นตัวแปรจำแนกประเภทก่อน โดยในปี 1989 Joreskog และ Sorbom ได้พัฒนาโปรแกรมลิสเรลให้สามารถวิเคราะห์ข้อมูลที่ประกอบด้วยกลุ่มตัวอย่างหลายกลุ่มพร้อมกันได้ โดยกลุ่มตัวอย่างหรือ กลุ่มประชากรนั้นอาจเป็นกลุ่มที่เกิดจากการจัดแบ่งกลุ่มตามตัวแปรจัดประเภท โดยมีเงื่อนไขในการแบ่งกลุ่มว่าหน่วยตัวอย่างทุกหน่วยต้องเป็นสมาชิกของกลุ่มใดกลุ่มหนึ่งเพียงกลุ่มเดียว โดยไม่เป็นสมาชิกร่วมกันในสองกลุ่ม (mutually exclusive) (นงลักษณ์ นงลักษณ์ วิรัชชัย, 2542; สุขมาส อังศุโชติ, สมถวิล วิจิตรวรรณ, & รัชนิกุล ภิญญานุวัฒน์, 2554)

หัวใจสำคัญของการวิเคราะห์กลุ่มพหุ คือ การวิเคราะห์ข้อมูลที่รวบรวมมาจากกลุ่มตัวอย่างทุกกลุ่ม โดยมีการกำหนดเงื่อนไขบังคับ (constraints) ให้โมเดลสมการโครงสร้างที่เป็นกรอบความคิดในการวิจัยนั้น มีลักษณะเป็นแบบเดียวกันสำหรับการตรวจสอบความสอดคล้องกลมกลืนระหว่างโมเดลและข้อมูลเชิงประจักษ์ หรือเป็นการทดสอบสมมติฐานทางสถิติว่า โมเดลสมการโครงสร้างมีเมทริกซ์พารามิเตอร์ไม่แปรเปลี่ยนระหว่างกลุ่มประชากร โดยพิจารณาจากค่าสถิติทดสอบความกลมกลืน คือ ค่าไค-สแควร์รวม (overall chi-square) ซึ่งเป็นค่าสถิติทดสอบในภาพรวมทุกกลุ่ม ถ้าผลการวิเคราะห์มีค่าต่ำกว่าค่าวิกฤตอย่างไม่มีนัยสำคัญทางสถิติ จะสรุปได้ว่า โมเดลสมการโครงสร้างที่สร้างขึ้นตามทฤษฎีสอดคล้องกับข้อมูลเชิงประจักษ์ทุกกลุ่ม และโมเดลมีลักษณะเป็นแบบเดียวกัน เรียกตามศัพท์

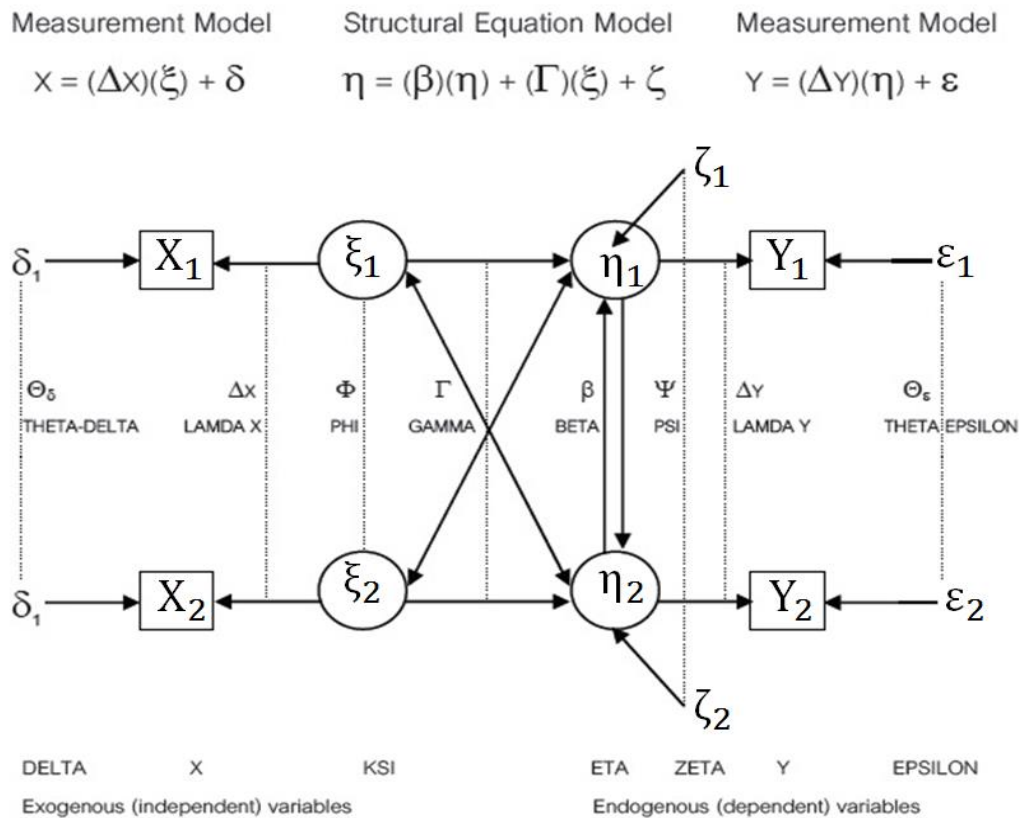
สถิติวิเคราะห์ลิสเรลว่า โมเดลสมการโครงสร้างไม่แปรเปลี่ยน หรือมีความยืดหยุ่นระหว่างกลุ่ม (invariance across groups) ในทางตรงกันข้ามถ้าค่าไค-สแควร์รวมมีค่าสูงกว่าค่าวิกฤตอย่างมีนัยสำคัญทางสถิติ แสดงว่าโมเดลสมการโครงสร้างในภาพรวมของกลุ่มประชากรทุกกลุ่ม หรือกลุ่มพหุไม่สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ทั้งนี้นอกจากจะพิจารณาตีความหมายจากค่าไค-สแควร์รวมแล้ว นักวิจัยยังต้องพิจารณาตีความจากค่าสถิติที่เป็นดัชนีวัดความกลมกลืนของโมเดลด้วย เช่น GFI, RMSE, RMR, และ critical N สำหรับสมมติฐานในการทดสอบทดสอบความไม่แปรเปลี่ยนของโมเดลสมการโครงสร้างระหว่างกลุ่มจะสามารถตั้งได้หลายสมมติฐานตามจำนวนเมทริกซ์พารามิเตอร์ ซึ่งโดยทั่วไปนิยมตั้งเป็นชุดของสมมติฐาน (family of hypotheses) เรียงกันตั้งแต่สมมติฐานที่มีเงื่อนไขบังคับ (constraints) น้อยที่สุดไปจนถึงสมมติฐานที่มีเงื่อนไขบังคับมากที่สุด และการทดสอบสมมติฐานก็ทำการทดสอบเรียงไปตามลำดับ เรียกว่าเป็นการทดสอบเป็นลำดับลดหลั่น (hierarchical testing) ซึ่งช่วยให้นักวิจัยสามารถทำการทดสอบได้ทั้งในภาพรวมและทดสอบค่าพารามิเตอร์แต่ละค่าในเมทริกซ์แต่ละเมทริกซ์เป็นรายตัวหรือเป็นกลุ่มค่าพารามิเตอร์ตามความสนใจของนักวิจัยได้ ส่วนการแปลความหมายผลการทดสอบนอกจากจะใช้ค่าไค-สแควร์รวมสำหรับสมมติฐานแต่ละข้อแล้ว ยังต้องตีความหมายผลต่างระหว่างค่าไค-สแควร์รวมระหว่างสมมติฐานด้วย ซึ่งผลต่างของค่าไค-สแควร์รวมจะมีการแจกแจงแบบไค-สแควร์และตีความหมายเช่นเดียวกับค่าไค-สแควร์โดยทั่วไป คือ ถ้าผลต่างของค่าไค-สแควร์รวม ระหว่างสมมติฐานข้อ 1 และ 2 ไม่มีนัยสำคัญทางสถิติ โดยที่ค่าไค-สแควร์รวมของสมมติฐานข้อ 1 และข้อ 2 ไม่มีนัยสำคัญทางสถิติด้วย แสดงว่าการทดสอบความไม่แปรเปลี่ยนของโมเดลตามสมมติฐาน ข้อ 1 และข้อ 2 ให้ผลเป็นแบบเดียวกัน คือ โมเดลไม่แปรเปลี่ยนระหว่างกลุ่ม (นงลักษณ์ วิรัชชัย, 2542)

3.2 ขั้นตอนการวิเคราะห์โมเดลกลุ่มพหุ

การวิเคราะห์กลุ่มพหุจะสามารถทดสอบความแตกต่างหรือความเหมือนของพารามิเตอร์ในโมเดลสมการโครงสร้าง ได้แก่ สัมประสิทธิ์การถดถอย (regression coefficient) น้ำหนักองค์ประกอบ (factor loading) ความคลาดเคลื่อน (error) และความสัมพันธ์ขององค์ประกอบ (factorial relationship) โดยมีรายละเอียดดังนี้ (นงลักษณ์ วิรัชชัย, 2542)

3.2.1 การทดสอบความไม่แปรเปลี่ยนของรูปแบบโมเดล หมายถึง การทดสอบว่าโมเดลสมการโครงสร้างตามทฤษฎี (ภาพที่ 7) ที่สอดคล้องกับข้อมูลเชิงประจักษ์ของกลุ่มประชากรแต่ละ

กลุ่ม ประกอบด้วยจำนวนตัวแปรในโมเดลและรูปแบบลักษณะโครงสร้างความสัมพันธ์ระหว่างตัวแปร เป็นแบบเดียวกันหมดทุกกลุ่ม เมทริกซ์พารามิเตอร์ทั้งหมดในการวิเคราะห์โมเดลสมการโครงสร้าง มีขนาดเท่ากัน มีรูปแบบเมทริกซ์ (matrix form) และสถานะเมทริกซ์ (matrix mode) เป็นแบบเดียวกันทุกกลุ่ม ซึ่งผลการวิเคราะห์ด้วยโปรแกรมลิสเรลจะรายงานค่าดัชนีวัดระดับความกลมกลืน (goodness of fit indices) ของการทดสอบทุกกลุ่มประชากรเป็นภาพรวม ดัชนีวัดระดับความกลมกลืนรวม (overall chi-square) ซึ่งได้มาจากดัชนีวัดระดับความกลมกลืนจากกลุ่มประชากรแต่ละกลุ่มรวมกัน ถ้าผลการวิเคราะห์ข้อมูลในส่วนนี้ได้ค่าไค-สแควร์รวมไม่มีนัยสำคัญทางสถิติแสดงว่าโมเดลแต่ละกลุ่มประชากรสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ทุกกลุ่ม แต่ถ้าได้ค่าไค-สแควร์รวมมีนัยสำคัญทางสถิติ แสดงว่าโมเดลของประชากรอย่างน้อยหนึ่งกลุ่มไม่สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ซึ่งในกรณีนี้ต้องทำการปรับโมเดลแล้ววิเคราะห์ใหม่ เพื่อให้ได้โมเดลปรับแก้ที่มีลักษณะโครงสร้างความสัมพันธ์ระหว่างตัวแปรสอดคล้องกับข้อมูลเชิงประจักษ์ โดยการปรับแก้โมเดลนี้ทำได้โดยใช้ข้อเสนอแนะที่โปรแกรมลิสเรลรายงาน ในส่วนของดัชนีดัดแปลง (modification indices) หรือปรับแก้ตามข้อสังเกตของนักวิจัยบนพื้นฐานทางทฤษฎี เช่นเดียวกับการปรับแก้โมเดลในการวิเคราะห์ลิสเรลทั่วไป เมื่อโมเดลสอดคล้องกับข้อมูลเชิงประจักษ์แล้ว จึงจะดำเนินการวิเคราะห์ขั้นตอนต่อไป



ภาพที่ 7 โมเดลในโปรแกรม LISREL (นงลักษณ์ วิรัชชัย, 2542: 26)

ในที่นี้

X = เวกเตอร์ตัวแปรภายนอกสังเกตได้ X ขนาด $(N \times 1)$

Y = เวกเตอร์ตัวแปรภายในสังเกตได้ Y ขนาด $(N \times 1)$

ξ = เวกเตอร์ตัวแปรแฝงภายนอก K ขนาด $(N \times 1)$

η = เวกเตอร์ตัวแปรแฝงภายใน E ขนาด $(N \times 1)$

δ = เวกเตอร์ความคลาดเคลื่อน d ในการวัดตัวแปร X ขนาด $(N \times 1)$

ε = เวกเตอร์ความคลาดเคลื่อน e ในการวัดตัวแปร Y ขนาด $(N \times 1)$

ζ = เวกเตอร์ความคลาดเคลื่อน z ในการวัดตัวแปร E ขนาด $(N \times 1)$

ΔX = เมทริกซ์สัมประสิทธิ์การถดถอยของ X บน K ขนาด (NX x NK)

ΔY = เมทริกซ์สัมประสิทธิ์การถดถอยของ Y บน E ขนาด (NYx NE)

Γ = เมทริกซ์อิทธิพลเชิงสาเหตุจาก K ไป E ขนาด (NE x NK)

β = เมทริกซ์อิทธิพลเชิงสาเหตุระหว่าง E ขนาด (NE x NE)

ϕ = เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมระหว่างตัวแปรแฝงภายนอก K ขนาด (NK x NK)

Ψ = เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมระหว่างความคลาดเคลื่อน Z ขนาด (NE x NE)

Θ_{δ} = เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมระหว่างความคลาดเคลื่อน d ขนาด (NX x NX)

Θ_{ϵ} = เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมระหว่างความคลาดเคลื่อน e ขนาด (NY x NY)

3.2.2 การทดสอบความไม่แปรเปลี่ยนของพารามิเตอร์ในโมเดล หมายถึง การทดสอบต่อจากการทดสอบความไม่แปรเปลี่ยนของรูปแบบโมเดล คือ จากที่ทราบว่าเมทริกซ์พารามิเตอร์ทั้งหมดมีรูปแบบเมทริกซ์และสถานะเมทริกซ์ เป็นแบบเดียวกันสำหรับทุกกลุ่มประชากรแล้ว ขั้นตอนนี้จะเป็นการทดสอบว่า ค่าพารามิเตอร์ในแต่ละเมทริกซ์มีค่าเท่ากันทุกกลุ่มประชากร โดยหลักการทดสอบจะทดสอบความไม่แปรเปลี่ยนของเมทริกซ์พารามิเตอร์ที่มีความเข้มงวดน้อยที่สุด (least restriction) ไปจนถึงการทดสอบความไม่แปรเปลี่ยนของเมทริกซ์พารามิเตอร์ที่มีความเข้มงวดมากที่สุด (most restriction) จากการกำหนดเงื่อนไขบังคับ (constrain) พารามิเตอร์ให้เท่ากันทุกกลุ่ม ซึ่งเป็นการตรวจสอบชุดของสมมติฐานที่ซ้อนกันเป็นระดับลดหลั่น (hierarchical nested hypotheses) โดยในการทดสอบนอกจากจะทดสอบนัยสำคัญของสมมติฐานแต่ละข้อแล้ว ยังต้องทดสอบความแตกต่างระหว่างสมมติฐานแต่ละข้อด้วย เช่น เมื่อได้ค่าสถิติไค-สแควร์ในการทดสอบสมมติฐานข้อ 2 และข้อ 3 (ตารางที่ 4) นำค่าไค-สแควร์มาลบกัน ผลต่างของไค-สแควร์ ซึ่งมีการแจกแจงแบบไค-สแควร์ด้วย จะเป็นค่าสถิติสำหรับการทดสอบนัยสำคัญของความไม่แปรเปลี่ยนของเมทริกซ์ BE และ GA ที่เพิ่มเข้ามา

ในการทดสอบสมมติฐาน ทั้งยังเป็นการทดสอบว่า ความไม่แปรเปลี่ยนของโมเดลตามสมมติฐานที่ 2 และ 3 แตกต่างกันหรือไม่อย่างไร ซึ่งถ้าค่าไค-สแควร์ของโมเดลที่ยังไม่บังคับพารามิเตอร์ให้เท่ากันไม่แตกต่างจากค่าไค-สแควร์ของโมเดลที่บังคับพารามิเตอร์ให้เท่ากัน จะสรุปได้ว่าโมเดลไม่แปรเปลี่ยนระหว่างกลุ่ม (invariance across groups) และการวัดความไม่แปรเปลี่ยนมีระดับความเข้มของการตรวจสอบหลายระดับขึ้นอยู่กับพารามิเตอร์ที่บังคับให้เท่ากัน โดยลักษณะของชุดสมมติฐานมีดังนี้

ตารางที่ 4 การทดสอบชุดของสมมติฐานที่ซ้อนกันเป็นระดับลดหลั่นในการทดสอบกลุ่มพหุ

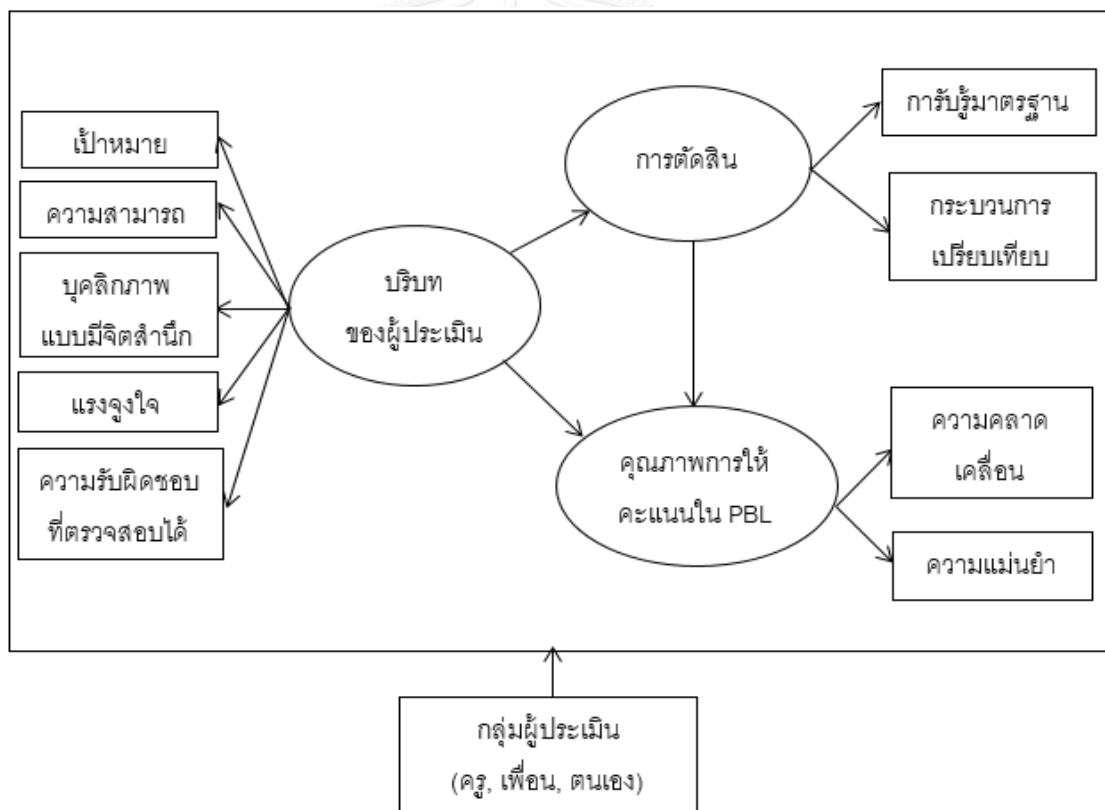
สมมติฐาน (Ho)	สัญลักษณ์
1) ความไม่แปรเปลี่ยนของรูปแบบขององค์ประกอบ (factor form invariance)	Ho: Form 1 = Form 2 = Form 3
2) ความไม่แปรเปลี่ยนของน้ำหนักองค์ประกอบ และสมมติฐานข้อ 1	Ho: $\Lambda_X (1) = \Lambda_X (2) = \Lambda_X (3)$ Ho: $\Lambda_Y (1) = \Lambda_Y (2) = \Lambda_Y (3)$
3) ความไม่แปรเปลี่ยนของเส้นทางอิทธิพล และสมมติฐานข้อ 2	Ho: $\Lambda_X (1) = \Lambda_X (2) = \Lambda_X (3)$ Ho: $\Lambda_Y (1) = \Lambda_Y (2) = \Lambda_Y (3)$ Ho: $\beta (1) = \beta (2) = \beta (3)$ Ho: $\Gamma (1) = \Gamma (2) = \Gamma (3)$
4) ความไม่แปรเปลี่ยนของพารามิเตอร์ในเมทริกซ์ PH, PS และสมมติฐานข้อ 3	Ho: $\Lambda_X (1) = \Lambda_X (2) = \Lambda_X (3)$ Ho: $\Lambda_Y (1) = \Lambda_Y (2) = \Lambda_Y (3)$ Ho: $\beta (1) = \beta (2) = \beta (3)$ Ho: $\Gamma (1) = \Gamma (2) = \Gamma (3)$ Ho: $\phi (1) = \phi (2) = \phi (3)$ Ho: $\Psi (1) = \Psi (2) = \Psi (3)$

สมมติฐาน (Ho)	สัญลักษณ์
5) ความไม่แปรเปลี่ยนของเทอมความคลาดเคลื่อน และสมมติฐานข้อ 4	$H_0: \Lambda_X (1) = \Lambda_X (2) = \Lambda_X (3)$ $H_0: \Lambda_Y (1) = \Lambda_Y (2) = \Lambda_Y (3)$ $H_0: \beta (1) = \beta (2) = \beta (3)$ $H_0: \Gamma (1) = \Gamma (2) = \Gamma (3)$ $H_0: \phi (1) = \phi (2) = \phi (3)$ $H_0: \Psi (1) = \Psi (2) = \Psi (3)$ $H_0: \theta_X (1) = \theta_X (2) = \theta_X (3)$ $H_0: \theta_Y (1) = \theta_Y (2) = \theta_Y (3)$

3.2.3 การวิเคราะห์สรุป เป็นการวิเคราะห์คำนวณหาผลต่างของดัชนีวัดระดับความกลมกลืนที่ได้จากการทดสอบสมมติฐานในขั้นที่สอง ระหว่างคู่ที่มีเงื่อนไขบังคับน้อยกับมีเงื่อนไขบังคับมาก จากผลต่างของดัชนีวัดระดับความกลมกลืนนี้ จะนำมาแปลความหมายสรุปผลการวิเคราะห์เกี่ยวกับโมเดลกลุ่มพหุทั้งหมด โดยถ้าจุดมุ่งหมายของการวิจัยมุ่งตอบปัญหาว่ามีความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มประชากรหรือไม่ อย่างไร การตีความหมายจะเน้นที่ลักษณะผลการทดสอบสมมติฐานว่าโมเดลที่ไม่แปรเปลี่ยนมีลักษณะอย่างไร พารามิเตอร์ใดที่แปรเปลี่ยนและพารามิเตอร์ใดไม่แปรเปลี่ยนระหว่างกลุ่มประชากร แต่ถ้าจุดมุ่งหมายของการวิจัยมุ่งตอบปัญหาเกี่ยวกับปฏิสัมพันธ์หรือสนใจตอบปัญหาเกี่ยวกับอิทธิพลของตัวแปรปรับจะต้องตีความหมายเพิ่มจากการวิเคราะห์ความไม่แปรเปลี่ยนให้สามารถตอบคำถามวิจัยได้ด้วย

ตอนที่ 4 กรอบแนวคิดการวิจัย

กรอบแนวคิดเบื้องต้นในการวิจัย เรื่อง การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการวิเคราะห์ด้วยโมเดลหลายองค์ประกอบของราสซ์ ผู้วิจัยได้ศึกษาจากแนวคิดทฤษฎี และงานวิจัยที่เกี่ยวข้อง โดยได้คัดเลือกตัวแปรที่เหมาะสมแบ่งเป็น 2 องค์ประกอบ ได้แก่ องค์ประกอบบริบทของผู้ประเมิน ซึ่งวัดจาก 5 ตัวแปร คือ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้ และ องค์ประกอบการตัดสินใจ ซึ่งวัดจาก 2 ตัวแปร คือ การรับรู้มาตรฐาน และกระบวนการเปรียบเทียบ และตัวแปรกำกับ 1 ตัวแปร คือ กลุ่มผู้ประเมิน ซึ่งได้แก่ อาจารย์ เพื่อน และตนเอง ที่จะส่งผลต่อ องค์ประกอบคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งวัดจาก 2 ตัวแปร คือ ความคลาดเคลื่อน และความแม่นยำ ดังภาพที่ 8



ภาพที่ 8 กรอบแนวคิดเบื้องต้นในการวิจัย

บทที่ 3 วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยและพัฒนา (research and development) มีวัตถุประสงค์เพื่อ

- 1) วิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานโดยใช้โมเดลหลายองค์ประกอบของราสส์
- 2) พัฒนาและตรวจสอบโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน
- 3) ตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง
- 4) พัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยแบ่งขั้นตอนการวิจัยออกเป็น 2 ระยะ ได้แก่ ระยะที่ 1 การพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง และระยะที่ 2 การพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยมีรายละเอียดดังนี้

ตอนที่ 1 การพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง

ในการพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง มีรายละเอียดของการดำเนินงานดังนี้

1.1 ประชากรและตัวอย่าง

ประชากร คือ อาจารย์และนักศึกษาของสถาบันอุดมศึกษาต่างๆ ในวิชาต่างๆ ที่จัดการสอนในรูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐาน

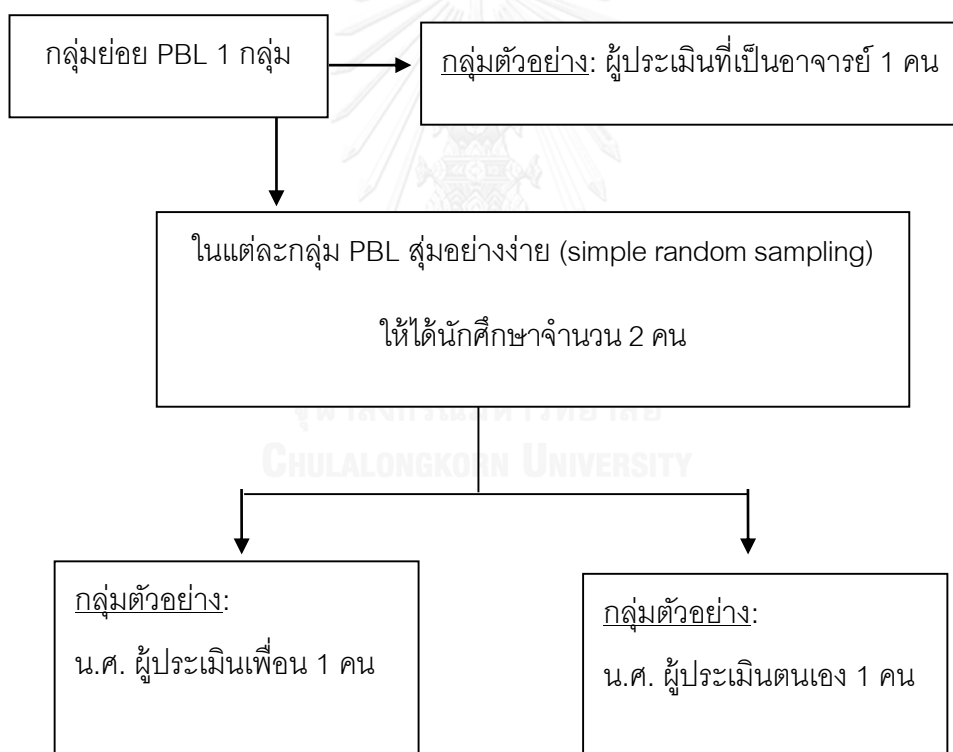
กลุ่มตัวอย่าง คือ อาจารย์และนักศึกษาของสถาบันอุดมศึกษาต่างๆ ในวิชาต่างๆ ที่จัดการสอนในรูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยมีเกณฑ์ในการคัดเลือก คือ 1) เป็นสถาบันอุดมศึกษา

ของรัฐหรือเอกชนที่มีรายวิชาใดก็ได้ที่ใช้รูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐานครบทั้ง 9 ขั้นตอน ได้แก่ ก)การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ ข)การกำหนดประเด็นปัญหา ค)การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้ จากความรู้และประสบการณ์ที่มีอยู่ ง)การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล จ)การระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา ฉ)การศึกษาค้นคว้าด้วยตนเอง ช)การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้ไปเรียนรู้มาในกลุ่ม ซ)การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ และ ฉ)การสะท้อนความรู้ในบทเรียน โดยทำการสรุปทั้งในด้านประเด็นปัญหา ผลจากการไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน และกระบวนการในการแก้ปัญหา 2)ผู้บริหารสถาบันหรือผู้ที่มีอำนาจให้การอนุมัติในการศึกษาวิจัยได้ และ 3) อาจารย์และนักศึกษาสมัครใจเข้าร่วมการวิจัยโดยการตอบแบบสอบถาม ส่วนเกณฑ์ในการคัดกลุ่ม PBL ออก คือ มีอาจารย์ประจำกลุ่มไม่ประสงค์เข้าร่วมการวิจัย และหรือมีนักศึกษาในกลุ่มแสดงความประสงค์เข้าร่วมการวิจัยน้อยกว่า 2 คน ทั้งนี้ได้กำหนดขนาดของตัวอย่างสำหรับการวิเคราะห์โมเดลสมการโครงสร้างโดยมีหลัก 2 ประการ คือ 1) ขนาดของตัวอย่างควรมากกว่า 100 คน และ 2) อัตราส่วนระหว่างหน่วยตัวอย่างและจำนวนพารามิเตอร์ควรจะเป็น 20:1 หรือสัดส่วนต่ำสุดที่พอใช้ได้คือ 5:1 และในกรณีที่กลุ่มตัวอย่างมีลักษณะแตกต่างกันมากต้องเพิ่มขนาดของสัดส่วนสูงขึ้น นอกจากนี้ขนาดตัวอย่างที่มากเกินไปทำให้ปฏิเสธโมเดล และทำให้มีค่าใช้จ่ายสูงจากการเก็บและวิเคราะห์ข้อมูล (นงลักษณ์ วิรัชชัย, 2542) ซึ่งการวิจัยในครั้งนี้ประกอบด้วยตัวแปรสังเกตได้จำนวน 9 ตัวแปร และมีจำนวนพารามิเตอร์ที่จะต้องประมาณค่า 19 ค่า ดังนั้นกลุ่มตัวอย่าง (ผู้ประเมินกลุ่มอาจารย์เพื่อน และตนเอง) ควรมีขนาด กลุ่มละ 100-380 คน เนื่องจากกลุ่มตัวอย่างที่ศึกษามีความใกล้เคียงกับงานของนุชนารถ วงศ์จำปา (นุชนารถ นุชนารถ วงศ์จำปา, 2553) ซึ่งศึกษาเกี่ยวกับการพัฒนาตัวบ่งชี้การจัดการเรียนรู้แบบใช้ปัญหาเป็นฐานในระดับอุดมศึกษา โดยกลุ่มตัวอย่างเป็นผู้สอนที่มีการจัดการเรียนรู้แบบใช้ปัญหาเป็นฐานจาก 6 มหาวิทยาลัยของรัฐ ซึ่งพบว่ามีอัตราการตอบกลับร้อยละ 55.87 และจากขั้นตอนในการเก็บข้อมูลของงานวิจัยนี้ซึ่งผ่านการรับรองโครงการวิจัยจากคณะกรรมการพิจารณาจริยธรรมการวิจัยในคน กลุ่มสหสถาบัน ชุดที่ 1 จุฬาลงกรณ์มหาวิทยาลัย (รหัสชุดโครงการ COA No. 070/2557) ที่มีการขอความยินยอมในการเข้าร่วมการวิจัยและกลุ่มตัวอย่างสามารถยกเลิกการเข้าร่วมวิจัยได้ตลอดเวลา ดังนั้นเพื่อให้การวิจัยมีข้อมูลเพียงพอที่จะนำมาวิเคราะห์ได้ ผู้วิจัยจึงดำเนินการเก็บข้อมูลจากทุกหน่วยในกลุ่มตัวอย่างที่ตรงตามเกณฑ์ในการคัดเลือก

การคัดเลือกกลุ่มตัวอย่าง

ขั้นที่ 1 ผู้วิจัยจำแนกอาจารย์และนักศึกษาที่จัดการสอนในรูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐานครบทั้ง 9 ชั้นตอน โดยแบ่งตามกลุ่มย่อย PBL (โดยทั่วไปประกอบด้วยอาจารย์ 1 คน และนักศึกษาจำนวน 8-10 คน)

ขั้นที่ 2 ในแต่ละกลุ่มย่อย PBL ประกอบด้วยผู้ประเมิน 3 กลุ่ม ได้แก่ อาจารย์ เพื่อน และตนเอง อย่างละ 1 คน และเพื่อให้สัดส่วนของผู้ประเมินแต่ละกลุ่มเท่าๆ กัน จึงทำการสุ่มอย่างง่ายให้ได้นักศึกษาในแต่ละกลุ่มย่อยจำนวน 2 คน สำหรับเป็นผู้ประเมินเพื่อน และผู้ประเมินตนเอง รายละเอียดดังภาพที่ 9



ภาพที่ 9 ประชากรและกลุ่มตัวอย่างที่ใช้ในงานวิจัย

ในการเก็บรวบรวมข้อมูลครั้งนี้ผู้วิจัยได้รับอนุญาตให้ดำเนินการวิจัยจากกลุ่มตัวอย่างที่ตรงตามเกณฑ์ในการคัดเลือก ซึ่งประกอบด้วยคณะต่างๆ ภายในมหาวิทยาลัยของรัฐทั้งสิ้นจำนวน 5 แห่ง รายละเอียดดังตารางที่ 5

ตารางที่ 5 จำนวนตัวอย่าง อัตราการตอบกลับและอัตราความครบถ้วนของแบบสอบถามที่ได้รับจากกลุ่มตัวอย่าง

สถาบันอุดมศึกษา	คณะ	จำนวน (กลุ่ม PBL)	อัตราการตอบกลับ		อัตราความครบถ้วน	
			จำนวน	ร้อยละ	จำนวน	ร้อยละ
จุฬาลงกรณ์	แพทยศาสตร์	42	40	95.24	40	100.00
	ทันตแพทยศาสตร์	20	20	100.00	19	95.00
ธรรมศาสตร์	ทันตแพทยศาสตร์	14	7	50.00	2	28.57
	สหเวชศาสตร์	17	17	100.00	17	100.00
	พยาบาลศาสตร์	22	20	90.91	17	85.00
สงขลานครินทร์	แพทยศาสตร์	17	17	100.00	12	70.59
วลัยลักษณ์	สหเวชศาสตร์	10	9	90.00	9	100.00
วพ. สุพรรณบุรี	พยาบาลศาสตร์	4	4	100.00	4	100.00
	รวม	146	134	91.78	120	89.55

จากตารางที่ 5 ในการเก็บรวบรวมข้อมูลผู้วิจัยหรือผู้ช่วยนักวิจัยได้ดำเนินการเก็บรวบรวมข้อมูลโดยแจกแบบสอบถามให้กับกลุ่มตัวอย่างที่ตรงตามเกณฑ์ในการคัดเลือกจำนวน 146 กลุ่ม PBL และได้รับข้อมูลกลับคืนมาทั้งสิ้น 134 กลุ่ม PBL คิดเป็นร้อยละ 91.78 ถือได้ว่าเป็นอัตราการตอบกลับในระดับสูง เนื่องจากทั้งหมดเป็นกลุ่มตัวอย่างที่ให้ความยินยอมในการเข้าร่วมการวิจัย ส่วนกลุ่มตัวอย่างที่ไม่อนุญาตให้ทำวิจัยจะไม่ได้ระบุไว้ และแบบสอบถามที่ได้รับกลับคืนมาจากกลุ่มตัวอย่างพบว่ามีครบถ้วนของข้อมูลจำนวน 120 กลุ่ม PBL คิดเป็นร้อยละ 89.55

1.2 ตัวแปรที่ใช้ในการวิจัย

ตัวแปรที่ใช้ในการวิจัยในครั้งนี้ ประกอบด้วย

1. ตัวแปรอิสระ ประกอบด้วยตัวแปรภายนอกแฝง 1 ตัว คือ
 - บริบทของผู้ประเมิน วัดได้จากตัวแปรสังเกตได้ 5 ตัวแปร ได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้
2. ตัวแปรตาม ประกอบด้วยตัวแปรภายในแฝง 2 ตัว คือ
 - การตัดสินใจ วัดได้จากตัวแปรสังเกตได้ 2 ตัวแปร ได้แก่ การรับรู้มาตรฐาน และกระบวนการเปรียบเทียบ
 - คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน วัดได้จากตัวแปรสังเกตได้ 2 ตัวแปร ได้แก่ ความแม่นยำ และความคลาดเคลื่อน
3. ตัวแปรกำกับ จำนวน 1 ตัว คือ กลุ่มผู้ประเมิน ซึ่งแบ่งเป็น อาจารย์ เพื่อน และตนเอง

1.3 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในขั้นตอนนี้มาจากผู้วิจัยพัฒนาขึ้นมาใหม่ และบางส่วนได้ปรับปรุงมาจากเครื่องมือที่มีอยู่เดิม โดยประกอบด้วย แบบสัมภาษณ์ผู้ทรงคุณวุฒิ จำนวน 1 ฉบับ แบบสอบถามคุณลักษณะของผู้ประเมินจำนวน 1 ฉบับ และแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย จำนวน 1 ฉบับ รายละเอียดของเครื่องมือแต่ละฉบับมีดังนี้

1. แบบสำรวจและคัดเลือกองค์ประกอบและตัวบ่งชี้คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ใช้ในการสัมภาษณ์หรือสอบถามจากผู้ทรงคุณวุฒิ ซึ่งจะต้องเป็นผู้ที่มีความรู้และความเชี่ยวชาญในด้านการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยมีลักษณะเป็นคำถามปลายเปิด และแบบตรวจสอบรายการ (check list) แบ่งเป็น 2 ตอน คือ ตอนที่ 1 ข้อมูลทั่วไปของผู้ทรงคุณวุฒิ จำนวน 15 ข้อ เช่น ชื่อ อายุ ประสบการณ์การใช้รูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐาน เป็นต้น และตอนที่ 2 ความคิดเห็นเกี่ยวกับองค์ประกอบและตัวบ่งชี้คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน จำนวน 15 ข้อ ซึ่งเป็นการถามความคิดเห็นเกี่ยวกับองค์ประกอบและตัวบ่งชี้คุณภาพการให้คะแนนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน ตลอดจนความคิดเห็นและข้อเสนอแนะ

เพิ่มเติมเพื่อเป็นแนวทางในการปรับกรอบแนวคิดในการวิจัย โดยมีขั้นตอนในการสร้างและตรวจสอบคุณภาพของเครื่องมือ ดังนี้

ขั้นตอนที่ 1 ผู้วิจัยศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเกี่ยวกับการพัฒนาตัวบ่งชี้ และองค์ประกอบของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน แล้วนำมาสร้างเป็นกรอบแนวคิดเบื้องต้นในการวิจัย

ขั้นตอนที่ 2 ผู้วิจัยสร้างแบบสำรวจฯ จากนั้นนำไปให้อาจารย์ที่ปรึกษาตรวจสอบความถูกต้องและความเหมาะสมของข้อคำถาม พร้อมทั้งขอคำแนะนำและข้อเสนอแนะเพื่อปรับปรุงแก้ไข

ขั้นตอนที่ 3 ผู้วิจัยและอาจารย์ที่ปรึกษาร่วมกันพิจารณาคัดเลือกผู้ทรงคุณวุฒิจำนวน 5 คน ตามเกณฑ์ดังนี้ 1) เป็นอาจารย์ผู้สอนในระดับอุดมศึกษาที่มีตำแหน่งทางวิชาการระดับผู้ช่วยศาสตราจารย์ขึ้นไป และ 2) มีประสบการณ์เป็นวิทยากรบรรยาย หรือมีผลงานทางวิชาการทางด้านการเรียนรู้โดยใช้ปัญหาเป็นฐาน เช่น ตำรา งานวิจัย ไม่ต่ำกว่า 3 ชิ้นงาน

ขั้นตอนที่ 4 ผู้วิจัยติดต่อประสานงานผู้ทรงคุณวุฒิเพื่อขอความอนุเคราะห์ในการให้สัมภาษณ์/ตอบแบบสอบถามเพื่อสำรวจและคัดเลือกองค์ประกอบและตัวบ่งชี้คุณภาพการให้คะแนนในวิชาการเรียนรู้โดยใช้ปัญหาเป็นฐาน

ขั้นตอนที่ 5 ผู้วิจัยทำหนังสือเชิญผู้ทรงคุณวุฒิจำนวน 5 คน จากคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และส่งไปยังหน่วยงานที่เป็นต้นสังกัดของผู้ทรงคุณวุฒิ

ขั้นตอนที่ 6 ผู้วิจัยนำแบบสำรวจฯ ที่ปรับปรุงแก้ไขแล้วไปสัมภาษณ์ผู้ทรงคุณวุฒิ หรือในกรณีที่ผู้ทรงคุณวุฒิไม่สะดวกให้สัมภาษณ์แต่ต้องการทำแบบสอบถามแทน ผู้วิจัยจะส่งแบบสอบถามไปทางไปรษณีย์

ขั้นตอนที่ 7 ผู้วิจัยรวบรวมข้อมูลที่ได้จากการสัมภาษณ์หรือแบบสอบถามที่ได้รับกลับคืนจากผู้ทรงคุณวุฒิ และทำการตรวจสอบความครบถ้วนของข้อมูล ซึ่งในกรณีที่ข้อมูลไม่ชัดเจนผู้วิจัยจะสอบถามจากผู้ทรงคุณวุฒิอีกครั้ง จากนั้นผู้วิจัยทำการวิเคราะห์เนื้อหา (content analysis) และคำนวณค่าดัชนีความสอดคล้อง (index of congruency; IOC) เพื่อสรุปผลการคัดเลือกตัวองค์ประกอบและตัวบ่งชี้คุณภาพการให้คะแนนในวิชาการเรียนรู้โดยใช้ปัญหาเป็นฐาน

ผลการวิเคราะห์เนื้อหาดังแสดงในตารางที่ 6 โดยในคอลัมน์ ก. เป็นความคิดเห็นหรือประสบการณ์ของผู้ทรงคุณวุฒิเกี่ยวกับสาเหตุของความแตกต่าง/ความไม่สอดคล้องกันระหว่าง

คะแนนจากกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ในรายวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ส่วนในคอลัมน์ ข. เป็นการวิเคราะห์ที่ผู้วิจัยได้เชื่อมโยงความสอดคล้องระหว่างความคิดเห็น/ประสบการณ์ของผู้ทรงคุณวุฒิในคอลัมน์ ก. กับนิยามของตัวบ่งชี้หรือตัวแปรที่ศึกษาในงานวิจัยนี้

ตารางที่ 6 สาเหตุของความไม่สอดคล้องกันระหว่างคะแนนจากกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง

คนที่	ก.ความคิดเห็น/ประสบการณ์ของผู้ทรงคุณวุฒิ	ข.ความสอดคล้องกับตัวแปรเชิงสาเหตุ
1	● การไม่เข้าใจแบบประเมิน ใช้แบบประเมินไม่เป็น	ความสามารถ
	● ความลำเอียง	แรงจูงใจ
2	● ขาด Standardization ในการประเมินกระบวนการกลุ่ม	การรับรู้มาตรฐาน กระบวนการเปรียบเทียบ
3	● เพื่อนมักจะประเมินเพื่อนดีกว่าปกติ เพราะกลัวจะทำให้เพื่อนสอบตกหากประเมินตามความจริง ส่วนผู้เรียนจะประเมินตนเองต่ำกว่าปกติ แต่ถ้า descriptions ดีจะทำให้ทุกฝ่ายประเมินได้ใกล้เคียงกันมาก นอกจากนั้นควรมีการฝึกประเมิน	ความรับผิดชอบที่ตรวจสอบได้ การรับรู้มาตรฐาน กระบวนการเปรียบเทียบ
4	● ผู้ประเมินไม่เข้าใจบทบาทของตนเอง	บุคลิกภาพแบบมีจิตสำนึก
	● ผู้ประเมินขาดทักษะในการประเมิน	ความสามารถ
	● ผู้ประเมินมีความเกรงใจผู้ที่รับการประเมิน	เป้าหมาย
5	● ความคาดหวังในผลลัพธ์ต่างกัน	เป้าหมาย
	● ความรู้ ประสบการณ์การประเมิน	ความสามารถ
	● ความคิดเห็นที่แตกต่างกัน	การรับรู้มาตรฐาน

ผลการคำนวณ ค่าดัชนีความสอดคล้อง (index of congruency; IOC) พบว่า องค์กรประกอบและตัวบ่งชี้ทั้งหมดมีค่าอยู่ระหว่าง 0.8 – 1 (ดังตารางที่ 7) ซึ่งมีความมากกว่า 0.8 ขึ้นไป แสดงว่าองค์กรประกอบและตัวบ่งชี้ทั้งหมดมีความเป็นตัวแทนของสิ่งที่ต้องการวัด (ศิริชัย กาญจนวาสี, 2552)

ตารางที่ 7 ผลการคำนวณค่าดัชนีความสอดคล้อง (IOC) และข้อเสนอแนะจากผู้ทรงคุณวุฒิ

รายการ	ความคิดเห็นของผู้ทรงคุณวุฒิ						ข้อเสนอแนะ
	1	2	3	4	5	IOC	
องค์ประกอบที่ 1							
บริบทของผู้ประเมิน (context) หมายถึง สถานการณ์หรือเงื่อนไขที่เป็นตัวกำหนดให้เกิดความแตกต่างของพฤติกรรมทำให้คะแนนระหว่างผู้ประเมิน ซึ่งวัดได้จากตัวแปรสังเกตได้ 5 ตัวแปร ประกอบด้วย เป้าหมายความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้	1	1	1	1	1	1	
ตัวบ่งชี้ที่ 1 ในองค์ประกอบที่ 1							
เป้าหมาย (goals) หมายถึง ลักษณะความคาดหวังของผู้ประเมินที่มีต่อกระบวนการและผลลัพธ์ของการประเมิน โดยลักษณะของเป้าหมายอาจมีความซับซ้อน หรือเป็นเป้าหมายในเชิงตรงกันข้าม (conflicting goals) เป้าหมายที่สำคัญของผู้ประเมินภายในบริบทของการศึกษามี 2 ประเภท คือ 1) เป้าหมายระหว่างบุคคล (interpersonal goals) หมายถึง ลักษณะความคาดหวังของผู้ประเมินที่เกี่ยวข้องกับการใช้ผลการประเมินเพื่อพัฒนาความสัมพันธ์ระหว่างบุคคล เช่น การให้คะแนนประเมินผลการปฏิบัติที่สูงเพื่อให้เกิดความเสมอภาคในกลุ่ม หรือความพยายามที่จะคงไว้ซึ่งระดับความสัมพันธ์ระหว่างบุคคล	1	1	1	1	1	1	

รายการ	ความคิดเห็นของ ผู้ทรงคุณวุฒิ						ข้อเสนอแนะ
	1	2	3	4	5	IOC	
2) เป้าหมายภายในบุคคล (internalized goals) หมายถึง ลักษณะความคาดหวังเกี่ยวกับการประเมิน อันเป็นผลมาจากความเชื่อและการให้คุณค่าภายในของ ผู้ประเมิน เช่น ผู้ประเมินที่มีความเชื่อมั่นต่อกระบวนการ ในการประเมิน หรือผู้ประเมินที่เชื่อว่าผลการประเมินที่ แม่นยำจะมีประโยชน์ต่อการนำไปใช้							
ตัวบ่งชี้ที่ 2 ในองค์ประกอบที่ 1 ความสามารถ (ability for rating) หมายถึง ทักษะ ของผู้ประเมินในการทำความเข้าใจคุณลักษณะที่ ต้องการประเมิน โดยความแตกต่างของทักษะนี้จะขึ้นกับ โอกาสของผู้ประเมินในการสังเกตพฤติกรรมของการ ผู้ถูกประเมิน และการได้รับการฝึก/อบรมการ ประเมินผลการปฏิบัติ	1	1	1	1	1	1	
ตัวบ่งชี้ที่ 3 ในองค์ประกอบที่ 1 บุคลิกภาพแบบมีจิตสำนึก (conscientiousness) หมายถึง คุณลักษณะของบุคคลที่แสดงออกเกี่ยวกับการ ตระหนักถึงการควบคุมหรือกระตุ้นตัวเองเพื่อให้เกิดการ ตอบสนองอย่างมีประสิทธิภาพ เช่น มีพลังในการทำงาน ต้องการที่จะประสบความสำเร็จ ทำงานเป็นระบบ ยึด มั่นในหน้าที่ความรับผิดชอบ	1	1	0	1	1	0.8	

รายการ	ความคิดเห็นของ ผู้ทรงคุณวุฒิ						ข้อเสนอแนะ
	1	2	3	4	5	IOC	
ตัวบ่งชี้ที่ 4 ในองค์ประกอบที่ 1							
แรงจูงใจ (motivation) หมายถึง สิ่งที่มีผลผลักดันให้ ผู้ประเมินเกิดความพยายามในการประเมินเพื่อให้บรรลุ เป้าหมายแห่งความถูกต้องของผลการประเมิน โดยคนที่มี แรงจูงใจสูงจะมีความพยายามอย่างไม่ลดละต่างกับคนที่ มีแรงจูงใจต่ำที่อาจล้มเลิกการกระทำก่อนที่จะบรรลุเป้าหมาย	1	1	0	1	1	0.8	
ตัวบ่งชี้ที่ 5 ในองค์ประกอบที่ 1							
ความรับผิดชอบที่ตรวจสอบได้ (accountability) หมายถึง ความพร้อมที่จะรายงานหรือรับการตรวจสอบ จากผู้ที่เกี่ยวข้องในผลจากการปฏิบัติหน้าที่เป็นผู้ประเมิน	1	1	1	1	1	1	
องค์ประกอบที่ 2							
การตัดสิน (judgment) หมายถึง รูปแบบในการ ตรวจสอบความสอดคล้องระหว่างข้อมูลจากการปฏิบัติ ที่เกิดขึ้นจริงกับเกณฑ์มาตรฐานในกระบวนการเรียนรู้ โดยใช้ปัญหาเป็นฐาน ซึ่งวัดได้จากตัวแปรสังเกตได้ 2 ตัวแปร คือ การรับรู้มาตรฐาน และกระบวนการ เปรียบเทียบ	1	1	0	1	1	0.8	ผู้ทรง คนที่ 5: การตัดสิน ต้อง พิจารณา กระบวนการ และการ ผลลัพธ์
ตัวบ่งชี้ที่ 1 ในองค์ประกอบที่ 2							
การรับรู้มาตรฐาน (standards) หมายถึง การตีความ ของผู้ประเมินเกี่ยวกับหลักเกณฑ์หรือแนวทางการปฏิบัติ ที่เหมาะสมสำหรับผู้เรียนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน	1	1	1	1	1	1	

รายการ	ความคิดเห็นของ ผู้ทรงคุณวุฒิ						ข้อเสนอแนะ
	1	2	3	4	5	IOC	
ตัวบ่งชี้ที่ 2 ในองค์ประกอบที่ 2 กระบวนการเปรียบเทียบ (comparison process) หมายถึง การใช้ความคิดอย่างเป็นลำดับของผู้ประเมิน เพื่อให้ได้มาซึ่งสารสนเทศสำหรับการประเมินผล การเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งประกอบด้วย กระบวนการคิด 2 ระยะ คือ 1) การรวบรวมข้อมูลตาม วัตถุประสงค์เฉพาะในการประเมินผล ซึ่งข้อมูลจะได้จาก การสังเกตพฤติกรรม หรือผลของการปฏิบัติ และ 2) การคัดเลือกและบูรณาการข้อมูล หรือการพิจารณา ตัดสินข้อมูล โดยผู้ประเมินจะพิจารณาคัดเลือก จำแนก และสรุปข้อมูลเฉพาะที่ตรงกับวัตถุประสงค์การประเมิน	1	1	0	1	1	0.8	
องค์ประกอบที่ 3 คุณภาพการให้คะแนนในวิชาการเรียนรู้โดยใช้ปัญหา เป็นฐาน (quality ratings in PBL) หมายถึง คุณลักษณะอันพึงประสงค์ที่สำคัญของการวัดและ ประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งวัดได้จาก ตัวแปรสังเกตได้ 2 ตัวแปร คือ ความแม่นยำ และ ความคลาดเคลื่อน	1	1	1	0	1	0.8	
ตัวบ่งชี้ที่ 1 ในองค์ประกอบที่ 3 ความแม่นยำ (accuracy) หมายถึง ผลการให้คะแนนตาม แบบประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานของผู้ประเมิน มีความถูกต้องใกล้เคียงตามความสามารถที่แท้จริงของผู้ถูก ประเมิน	1	1	1	1	1	1	

รายการ	ความคิดเห็นของ ผู้ทรงคุณวุฒิ						ข้อเสนอแนะ
	1	2	3	4	5	IOC	
ตัวบ่งชี้ที่ 2 ในองค์ประกอบที่ 3 ความคลาดเคลื่อน (error) หมายถึง ผลการให้คะแนนตามแบบประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานของผู้ประเมินมีความแตกต่างจากค่าความสามารถที่แท้จริงหรือค่าที่ยอมรับของผู้ถูกประเมิน	1	1	1	1	1	1	

ขั้นตอนที่ 8 ผู้วิจัยนำองค์ประกอบและตัวบ่งชี้ที่ผ่านการพิจารณาจากผู้ทรงคุณวุฒิเหล่านี้มาใช้สร้างโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และใช้กำหนดโครงสร้างเนื้อหาของแบบสอบถามที่ใช้วัดตัวบ่งชี้เหล่านี้

2. แบบสอบถามคุณลักษณะของผู้ประเมิน แบ่งออกเป็น 2 ตอน ได้แก่ ตอนที่ 1 ข้อมูลทั่วไปของผู้ตอบ จำนวน 6 ข้อ มีลักษณะเป็นคำถามปลายเปิด และแบบตรวจสอบรายการ (check list) เพื่อสำรวจข้อมูลภูมิหลัง ได้แก่ อายุ เพศ สังกัดคณะ สังกัดมหาวิทยาลัย สถานะผู้ประเมิน ตอนที่ 2 แบบสอบถามเกี่ยวกับตัวแปรต่างๆ ในงานวิจัย จำนวน 39 ข้อ ประกอบด้วยแบบสอบถามชนิดตัวเลือกคู่ (forced-choice items) คือ ใช่/ไม่ใช่ จำนวน 10 ข้อ ซึ่งวัดครอบคลุมตัวบ่งชี้ 2 ตัว ได้แก่ เป้าหมาย และความสามารถ และชนิดมาตราประมาณค่า 5 ระดับ (rating scale) คือ จริงที่สุด ค่อนข้างจริง จริงปานกลาง ค่อนข้างไม่จริง และไม่จริงเลย จำนวน 29 ข้อ ซึ่งวัดครอบคลุมตัวบ่งชี้ 5 ตัว ได้แก่ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ ความรับผิดชอบที่ตรวจสอบได้ การรับรู้มาตรฐานของการเรียนรู้โดยใช้ปัญหาเป็นฐาน และกระบวนการเปรียบเทียบ

ข้อคำถามเชิงบวกในแบบสอบถามชนิดตัวเลือกคู่ มีจำนวน 5 ข้อ ได้แก่ แบบสอบถามในตอนต้นที่ 1.2 ข้อที่ 2, 6, 7, 8 และ 10 มีเกณฑ์การให้คะแนนดังนี้

ใช่	ให้คะแนน	1 คะแนน
ไม่ใช่	ให้คะแนน	0 คะแนน

ข้อคำถามเชิงลบในแบบสอบถามชนิดตัวเลือกคู่ มีจำนวน 5 ข้อ ได้แก่ข้อ แบบสอบถามในตอน
1.2 ข้อที่ 1, 3, 4, 5 และ 9 มีเกณฑ์การให้คะแนนดังนี้

ใช่	ให้คะแนน	0 คะแนน
ไม่ใช่	ให้คะแนน	1 คะแนน

ข้อคำถามเชิงบวกในมาตรฐานค่า 5 ระดับ มีจำนวน 21 ข้อ ได้แก่ แบบสอบถามในตอน
1.3 ข้อที่ 1, 3, 5, 7, 9, 10, 11, 12, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 28 และ 29 มีเกณฑ์
การให้คะแนนดังนี้

จริงที่สุด	ให้คะแนน	5 คะแนน
ค่อนข้างจริง	ให้คะแนน	4 คะแนน
จริงปานกลาง	ให้คะแนน	3 คะแนน
ค่อนข้างไม่จริง	ให้คะแนน	2 คะแนน
ไม่จริงเลย	ให้คะแนน	1 คะแนน

ข้อคำถามเชิงลบในมาตรฐานค่า 5 ระดับ มีจำนวน 8 ข้อ ได้แก่ แบบสอบถามในตอน
1.3 ข้อที่ 2, 4, 6, 8, 13, 14, 18 และ 27 มีเกณฑ์การให้คะแนนดังนี้

จริงที่สุด	ให้คะแนน	1 คะแนน
ค่อนข้างจริง	ให้คะแนน	2 คะแนน
จริงปานกลาง	ให้คะแนน	3 คะแนน
ค่อนข้างไม่จริง	ให้คะแนน	4 คะแนน
ไม่จริงเลย	ให้คะแนน	5 คะแนน

โดยมีขั้นตอนการสร้างและตรวจสอบคุณภาพของเครื่องมือ ดังนี้

ขั้นตอนที่ 1 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการพัฒนาตัวบ่งชี้ และองค์ประกอบ
ของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และนำข้อเสนอแนะและ
ข้อสนับสนุนที่ได้รับจากการสัมภาษณ์ผู้ทรงคุณวุฒิมาปรับใช้

ขั้นตอนที่ 2 ผู้วิจัยสร้างตารางกำหนดจำนวนข้อคำถาม และพฤติกรรมที่ต้องการวัด (table of specification) เกี่ยวกับตัวบ่งชี้ของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (ดังตารางที่ 8) และดำเนินการสร้างข้อคำถาม จากนั้นนำแบบสอบถามที่สร้างขึ้นไปให้อาจารย์ที่ปรึกษาตรวจสอบความถูกต้องและความเหมาะสมของข้อคำถาม พร้อมทั้งขอคำแนะนำและข้อเสนอแนะเพื่อปรับปรุงแก้ไขแบบสอบถาม และนำมาปรับปรุงแก้ไขตามคำแนะนำของอาจารย์ที่ปรึกษา

ตารางที่ 8 แสดงการกำหนดจำนวนข้อคำถาม และพฤติกรรม/คุณลักษณะที่ต้องการวัด (table of specification)

หน่วย	พฤติกรรม/คุณลักษณะ ที่ต้องการวัด	ระดับ ความสำคัญ	จำนวนข้อ		ข้อที่
			ชนิด ตัวเลือกคู่	มาตรวัด ประมาณค่า	
1.	เป้าหมายของผู้ประเมิน 1.1 เป้าหมายระหว่างบุคคล 1.2 เป้าหมายภายในบุคคล	100% 60% 40%	5 3 2		ตอนที่ 1.2 ข้อที่ 1, 3, 4 ข้อที่ 2, 5
2.	ความสามารถในการให้คะแนน 2.1 ความรู้เกี่ยวกับการให้คะแนน 2.2 ประสบการณ์ในการให้คะแนน	100% 60% 40%	5 3 2		ตอนที่ 1.2 ข้อที่ 6, 7 ข้อที่ 8, 9, 10
3.	บุคลิกภาพแบบมีจิตสำนึก *แปลจากแบบวัด 50-Item Set of IPIP Big-Five Factor Markers ของ Goldberg โดยนำมาเฉพาะ construct ของ Conscientiousness จำนวน 10 ข้อ (Goldberg, 2001)	100%		10	ตอนที่ 1.3 ข้อ 1-10
4.	แรงจูงใจ	100%		5	ตอนที่ 1.3 ข้อ 11-15
5.	ความรับผิดชอบที่ตรวจสอบได้	100%		5	ตอนที่ 1.3 ข้อ 16-20
6.	การรับรู้มาตรฐาน	100%		5	ตอนที่ 1.3 ข้อ 21-25
7.	กระบวนการเปรียบเทียบ 7.1 การรวบรวมข้อมูล 7.2 การคัดเลือกและบูรณาการข้อมูล	100% 50% 50%		4 2 2	ตอนที่ 1.3 ข้อที่ 26, 29 ข้อที่ 27, 28

ขั้นตอนที่ 3 นำแบบสอบถามพร้อมทั้ง table of specification ไปให้ผู้เชี่ยวชาญทางด้านเนื้อหาและการวัดและประเมินผล จำนวน 5 คน ตรวจสอบคุณภาพของเครื่องมือในด้านความตรงเชิงเนื้อหา (content validity) จากนั้นนำมาหาค่าดัชนีความสอดคล้อง (index of congruency; IOC) ได้ค่าอยู่ระหว่าง 0.8 – 1 ซึ่งมีค่ามากกว่า 0.8 ขึ้นไป แสดงว่า ข้อคำถามทั้งหมดมีความเป็นตัวแทนของสิ่งที่ต้องการวัด (ศิริชัย กาญจนวาสี, 2552) โดยรายละเอียดดังแสดงตามตารางที่ 9

ตารางที่ 9 ค่าดัชนีของความสอดคล้องของแบบสอบถามคุณลักษณะของผู้ประเมิน

ข้อคำถาม	IOC	ข้อเสนอแนะ
1. เป้าหมาย		
1.1 ฉันคิดว่า การให้คะแนนงานกลุ่มแบบเท่ากันทุกคนจะช่วยให้เกิดความสามัคคีและลดการแข่งขันกันภายในกลุ่ม (-)	0.8	-
1.2 ฉันคิดว่า ผลการประเมินต้องมีการจำแนกจุดแข็งและจุดอ่อนของผู้ถูกประเมิน (+)	1	-
1.3 ฉันนิยมให้คะแนนประเมินบุคคลแบบภาพรวมมากกว่าการประเมินตามคุณลักษณะย่อย เพราะทำให้ผู้ถูกประเมินได้คะแนนมากขึ้น (-)	0.8	-
1.4 ฉันคิดว่า สัมพันธภาพระหว่างผู้ประเมินและผู้ถูกประเมินจะมีผลต่อการประเมินให้คะแนน (-)	0.8	ข้อความมีลักษณะเชิงลบมากเกินไป
1.5 ฉันมักจะให้คะแนนประเมินต่ำกว่าความเป็นจริง เพื่อให้ผู้ถูกประเมินเกิดการพัฒนาตนเองได้ดียิ่งขึ้น (-)	0.8	performance ดี อาจให้มากกว่าจริง ถ้าไม่ต้อาจให้น้อยกว่าที่ควรเป็น
2. ความสามารถ		
2.1 ฉันเคยมีประสบการณ์ในการให้คะแนนประเมินบุคคลหรือผลงาน (+)	1	-
2.2 ฉันเคยได้รับการอบรม/ได้รับข้อมูลคำแนะนำเกี่ยวกับวิธีการประเมินให้คะแนนบุคคลหรือผลงาน (+)	1	-
2.3 การประเมินผลการปฏิบัติจะให้ประโยชน์แก่ผู้ประเมินและผู้ถูกประเมิน (+)	1	-

ข้อคำถาม	IOC	ข้อเสนอแนะ
2.4 การประเมินเป็นการตัดสินตามความชอบหรือความพึงพอใจของผู้ประเมิน (-)	1	-
2.5 เป้าหมายของการประเมิน คือ สารสนเทศเพื่อการปรับปรุงพัฒนาหรือการตัดสินคุณค่าของสิ่งที่ประเมิน (+)	0.8	-
3. บุคลิกภาพแบบมีจิตสำนึก		
3.1 ฉันเตรียมพร้อมในทุกๆ เรื่อง (+)	1	-
3.2 ฉันมักจะวางสิ่งของไม่เป็นที่เป็นทาง (-)	0.8	-
3.3 ฉันใส่ใจรายละเอียดในการทำงาน (+)	1	-
3.4 ฉันมักจะทำสิ่งต่างๆ ให้เป็นเรื่องวุ่นวาย (-)	0.8	-
3.5 ฉันทำงานที่ได้รับมอบหมายในทันที (+)	1	-
3.6 ฉันมักลืมเก็บสิ่งของที่ใช้แล้วให้กลับเข้าที่ตามเดิม (-)	0.8	-
3.7 ฉันชอบปฏิบัติตามคำสั่ง (+)	0.8	-
3.8 ฉันเลี้ยงที่จะต้องทำงานที่ได้รับมอบหมาย (-)	1	-
3.9 ฉันปฏิบัติตามแผนกำหนดการของงาน (+)	1	-
3.10 ฉันทำงานอย่างละเอียดถี่ถ้วน (+)	1	-
4. แรงจูงใจ		
4.1 ฉันยินดีและเต็มใจทำหน้าที่เป็นผู้ประเมินในวิชา PBL (+)	0.8	-
4.2 เมื่อฉันทำหน้าที่เป็นผู้ประเมินในวิชา PBL ผลการประเมินของฉันจะต้องมีความถูกต้อง (+)	0.8	-
4.3 ถ้าฉันเป็นผู้ประเมินที่ยึดมั่นหลักการมากไป อาจทำให้เกิดผลเสียตามมาได้ (-)	0.8	-
4.4 ฉันคิดว่าเป็นเรื่องปกติ ถ้าจะประเมินคะแนนให้ตนเองหรือเพื่อนในกลุ่มแบบสูงๆ (-)	0.8	-
4.5 ฉันคิดว่า คะแนนจากการประเมินจะเป็นข้อมูลป้อนกลับเพื่อพัฒนาความสามารถของผู้ถูกประเมิน (+)	1	-
5. ความรับผิดชอบที่ตรวจสอบได้		
5.1 ฉันคิดว่า ควรมีระบบการตรวจสอบซ้ำในด้านความเหมาะสมและถูกต้องของคะแนนที่ได้จากผู้ประเมินทุกๆ คน (+)	0.8	-
5.2 ฉันยินดีที่จะได้รับการตรวจสอบในฐานะการทำหน้าที่เป็นผู้ประเมิน (+)	1	-

ข้อคำถาม	IOC	ข้อเสนอแนะ
5.3 ถ้าไม่มีกฎหรือข้อบังคับระบุไว้ ผู้ปฏิบัติงานไม่จำเป็นต้องรับผิดชอบ ผลจากการกระทำที่เกิดขึ้น (-)	0.8	-
5.4 ผู้ประเมินต้องแสดงความรับผิดชอบต่อผู้ถูกประเมินโดยให้การ ประเมินอย่างเที่ยงธรรม รับฟังความคิดเห็นและให้ข้อเท็จจริง (+)	1	-
5.5 ผู้ประเมินต้องแสดงความรับผิดชอบต่อผู้ที่มีบทบาทงานประเมินให้ โดยการทำงานอย่างเต็มความสามารถ ส่งผลประเมินทันตามกำหนดเวลา และให้ผลการประเมินที่ถูกต้อง (+)	0.8	-
6.การรับรู้มาตรฐาน		
6.1 ผู้เรียนในกลุ่ม PBL จะต้องมีความเข้าใจบทบาทเชิงรุกในการกำหนดปัญหา และวัตถุประสงค์การเรียนรู้ (+)	0.8	-
6.2 ในการเรียนการสอนแบบ PBL ผู้เรียนจะได้รับความรู้จากการไป ศึกษาค้นคว้าด้วยตนเอง (+)	0.8	-
6.3 ทักษะในการเรียนแบบ PBL ที่สำคัญ คือ การคิด วิเคราะห์ สังเคราะห์ ประเมิน และการถ่ายทอด (+)	0.8	-
6.4 การแสดงความคิดเห็นและเนื้อหาที่พูดของสมาชิกในกลุ่ม PBL จะบ่งบอก ถึงความรู้และความเข้าใจต่อเรื่องนั้นๆ (+)	0.8	-
6.5 กระบวนการกลุ่ม PBL จะสิ้นสุดลงเมื่อกลุ่มสามารถหาข้อมูลได้ ครบถ้วนต่อการพิสูจน์สมมติฐานทั้งหมด ตลอดจนสามารถสรุปเกี่ยวกับ หลักการและแนวทางในการนำความรู้ไปใช้ (+)	0.8	-
7.กระบวนการเปรียบเทียบ		
7.1 เมื่อเป็นผู้ประเมิน ฉันได้พิจารณาเลือกใช้ข้อมูลที่ตรงกับ วัตถุประสงค์หรือสิ่งต้องการจะประเมิน (+)	0.8	-
7.2 เมื่อเป็นผู้ประเมิน ฉันใช้ข้อมูลที่เป็นความคิดเห็นของตนเองหรือผู้ที่ เกี่ยวข้อง มากกว่าข้อมูลที่ได้จากการไปสังเกตโดยตรง (-)	0.8	-
7.3 ฉันใช้ข้อมูลหลายๆ ด้าน เพื่อประกอบการตัดสินใจให้คะแนนประเมิน (+)	0.8	-
7.4 ฉันได้พิจารณาซึ่งน้ำหนักความน่าเชื่อถือของข้อมูลก่อนการตัดสินใจ ให้คะแนนประเมิน (+)	0.8	-

ขั้นตอนที่ 4 นำแบบสอบถามที่ผ่านการพิจารณาความตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญและปรับปรุงแก้ไขตามคำแนะนำของผู้เชี่ยวชาญแล้วไปทดลองใช้กับนักศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏยะลา ชั้นปีที่ 1 ในวิชาเรียนที่ใช้ PBL จำนวน 30 คน เพื่อหาค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's alpha coefficient) ได้ค่าอยู่ระหว่าง 0.723 – 0.780 ซึ่งมีค่ามากกว่า 0.5 ขึ้นไป แสดงว่า แบบวัดมีความเที่ยงในระดับพอใช้จนถึงค่อนข้างดี (ศิริชัย กาญจนวาสี, 2552) โดยรายละเอียดดังแสดงตามตารางที่ 12

3. แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย เป็นการประเมินความสามารถหรือทักษะการปฏิบัติของผู้เรียนในรายวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งประกอบด้วยกระบวนการ 9 ขั้นตอน ได้แก่ 1) การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ 2) การกำหนดประเด็นปัญหา 3) การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้จากความรู้และประสบการณ์ที่มีอยู่ 4) การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล 5) การระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา 6) การศึกษาค้นคว้าด้วยตนเอง 7) การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้เรียนรู้มาต่อกลุ่ม 8) การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ และ 9) การสะท้อนความรู้ในบทเรียน โดยทำการสรุปทั้งในด้านประเด็นปัญหา ผลจากการไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน และกระบวนการในการแก้ปัญหา ร่วมกับการประเมินคุณลักษณะอันพึงประสงค์ที่ได้จากการเรียนรู้โดยใช้ปัญหาเป็นฐาน ได้แก่ ภาวะผู้นำ ความเคารพต่อสมาชิกในกลุ่ม และเจตคติต่อการเรียนรู้โดยใช้ปัญหาเป็นฐาน

โดยแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยมีลักษณะเป็นการให้คะแนนแบบรูบิกส์ (scoring rubric) จำนวน 19 ข้อ มีเกณฑ์การให้คะแนนดังนี้

ควรปรับปรุง (1 คะแนน) หมายถึง ผู้รับการประเมินแสดงออกว่า ไม่มี ตามรายการในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

ปานกลาง (2 คะแนน) หมายถึง ผู้รับการประเมินแสดงออกว่า มีน้อยครั้ง หรือ มีแต่ไม่ครบถ้วน ตามรายการในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

ดี (3 คะแนน) หมายถึง ผู้รับการประเมินแสดงออกว่า มีมากครั้ง และ มีครบถ้วน ตามรายการในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

ดีมาก (4 คะแนน) หมายถึง ผู้รับการประเมินแสดงออกว่า มีทุกครั้ง หรือ มีมากครั้ง มีครบถ้วน และเกิดประโยชน์ต่อกลุ่ม ตามรายการในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

ข้อมูลที่ได้จากส่วนนี้จะนำไปใช้วิเคราะห์ตัวบ่งชี้ 2 ตัว คือ ความแม่นยำ และความคลาดเคลื่อน โดยใช้โปรแกรม Facets ซึ่งมีเกณฑ์ในการพิจารณาดังนี้

ความแม่นยำ เป็นตัวบ่งชี้ที่มีลักษณะเป็นตัวแปรจัดกลุ่ม (categorical variable) หรือตัวแปรหุ่น (dummy variable) ซึ่งมีค่าเป็น 0 กับ 1 โดยค่าเท่ากับ 1 จะแสดงว่ามีความแม่นยำในการให้

คะแนน ซึ่งเกณฑ์ในการพิจารณา คือ ผลการประมาณค่าของผู้ประเมินจากโปรแกรม Facets ในค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐาน ที่มีค่าอยู่ระหว่าง 0.6 ถึง 1.4 (Wright et al., 1994) ส่วนค่าเท่ากับ 0 จะแสดงว่าไม่มีความแม่นยำในการให้คะแนน หรือมีค่า Infit Mean – Square น้อยกว่า 0.6 หรือมากกว่า 1.4

ความคลาดเคลื่อน เป็นตัวบ่งชี้ที่มีลักษณะเป็นตัวแปรต่อเนื่อง (continuous variable) ซึ่งมาจากการวิเคราะห์ความคลาดเคลื่อนชนิดกต/ปล่อยคะแนน โดยใช้โปรแกรม Facets จากผลการประมาณค่าของผู้ประเมิน ในค่า Measure ซึ่งมีหน่วยเป็นโลจิท (logits) ถ้ามีค่าเป็นบวกจะแสดงถึงความคลาดเคลื่อนชนิดกตคะแนน ค่าเป็นศูนย์จะแสดงถึงการไม่มีความคลาดเคลื่อน และค่าเป็นลบจะแสดงถึงความคลาดเคลื่อนชนิดปล่อยคะแนน (Linacre, 2012) ทั้งนี้ในการระบุขนาดของความคลาดเคลื่อนเพื่อใช้ในการเปรียบเทียบจะใช้ค่า Measure โดยไม่พิจารณาเครื่องหมาย +/- หรือทิศทางของความคลาดเคลื่อน

ขั้นตอนในการสร้างและตรวจสอบคุณภาพของเครื่องมือมีดังนี้

ขั้นตอนที่ 1 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการพัฒนาตัวบ่งชี้ของการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ในวิชาการเรียนรู้โดยใช้ปัญหาเป็นฐาน และนำข้อเสนอแนะและข้อสนับสนุนที่ได้รับจากการสัมภาษณ์ผู้ทรงคุณวุฒิมาปรับใช้

ขั้นตอนที่ 2 ผู้วิจัยสร้างตารางกำหนดจำนวนข้อคำถาม และพฤติกรรมที่ต้องการวัด (table of specification) เกี่ยวกับตัวบ่งชี้ของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (ดังตารางที่ 10) และดำเนินการสร้างข้อคำถาม จากนั้นนำแบบสอบถามที่สร้างขึ้นไปให้อาจารย์ที่ปรึกษาตรวจสอบความถูกต้องและความเหมาะสมของข้อคำถาม พร้อมทั้งขอคำแนะนำและข้อเสนอแนะเพื่อปรับปรุงแก้ไขแบบสอบถาม และนำมาปรับปรุงแก้ไขตามคำแนะนำของอาจารย์ที่ปรึกษา

ตารางที่ 10 แสดงการกำหนดจำนวนข้อคำถาม และพฤติกรรม/คุณลักษณะที่ต้องการวัด (table of specification)

พฤติกรรม/คุณลักษณะที่ต้องการวัด	ระดับ ความสำคัญ	จำนวน ข้อ	ข้อที่
1. การรวบรวม วิเคราะห์ ทำความเข้าใจในข้อมูลจากสถานการณ์ และการกำหนดประเด็นปัญหา	15.79%	3	1, 2, 4
2. การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้ จากความรู้และประสบการณ์ที่มีอยู่	10.53%	2	5, 16

ตารางที่ 10 (ต่อ)

พฤติกรรม/คุณลักษณะที่ต้องการวัด	ระดับ ความสำคัญ	จำนวน ข้อ	ข้อที่
3. การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล และการระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา	10.53%	2	6, 7
4. การศึกษาค้นคว้าด้วยตนเอง	10.53%	2	8, 9
5. การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้เรียนรู้มาต่อกลุ่ม	10.53%	2	10, 15
6. การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ	10.53%	2	3, 11
7. การสะท้อนความรู้ในบทเรียน โดยทำการสรุปเกี่ยวกับปัญหาในแต่ละประเด็น และผลการไปศึกษาเพิ่มเติม รวมถึงกระบวนการแก้ปัญหา	10.53%	2	17, 18
8. คุณลักษณะอันพึงประสงค์ที่ได้จากการเรียนรู้โดยใช้ปัญหาเป็นฐาน ได้แก่ ภาวะผู้นำ ความเคารพต่อสมาชิกในกลุ่ม และเจตคติต่อการเรียนรู้โดยใช้ปัญหาเป็นฐาน	21.05%	4	12, 13, 14, 19

ขั้นตอนที่ 3 นำแบบสอบถามพร้อมทั้งตารางแสดงการกำหนดจำนวนข้อคำถาม และพฤติกรรม/คุณลักษณะที่ต้องการวัด (table of specification) ไปให้ผู้เชี่ยวชาญทางด้านเนื้อหาและการวัดและประเมินผล จำนวน 5 คน ตรวจสอบคุณภาพของเครื่องมือในด้านความตรงเชิงเนื้อหา (content validity) จากนั้นนำมาหาค่าดัชนีความสอดคล้อง (index of congruency; IOC) ได้ค่าอยู่ระหว่าง 0.8 – 1 ซึ่งมีค่ามากกว่า 0.8 ขึ้นไป แสดงว่า ข้อคำถามทั้งหมดมีความเป็นตัวแทนของสิ่งที่ต้องการวัด (ศิริชัย กาญจนวาสี, 2552) โดยรายละเอียดดังแสดงตามตารางที่ 11

ตารางที่ 11 ค่าดัชนีของความสอดคล้องของแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย

ข้อความ	IOC	ข้อเสนอแนะ
1.ตั้งคำถามที่นำไปสู่การทำความเข้าใจในสถานการณ์ปัญหา	1	-
2.ชี้/ระบุข้อมูลที่สำคัญจากสถานการณ์ปัญหา	1	-
3.คิดวิเคราะห์ที่ไตร่ตรองอย่างเป็นระบบ	1	-
4.ประยุกต์เอาความรู้หรือประสบการณ์มาใช้กับสถานการณ์ใน PBL	1	-
5.ตั้งสมมติฐานที่เป็นไปได้สำหรับสาเหตุหรือแนวทางแก้ปัญหา	1	-
6.อธิบายเชื่อมโยงประเด็นต่างๆ อย่างเป็นเหตุ เป็นผล	1	-
7.ระบุประเด็นที่ต้องการเรียนรู้จากสถานการณ์ใน PBL ได้	1	-
8.ค้นหาข้อมูลความรู้ที่ทันสมัยและเป็นประโยชน์	1	-
9.ค้นคว้าหาข้อมูลได้อย่างครอบคลุมและครบถ้วน	1	-
10.ถ่ายทอดเนื้อหาความรู้ได้อย่างชัดเจน เข้าใจง่าย	1	-
11.นำความรู้ที่ได้จากการค้นคว้ามาใช้อธิบายตอบคำถามหรือสมมติฐานจากสถานการณ์ของปัญหา	1	-
12.ชี้แนะหรือนำพากลุ่มไปสู่แนวทางในการเรียนรู้/เป้าหมายความสำเร็จ	1	-
13.ให้เกียรติเพื่อนสมาชิกในกลุ่มอย่างเท่าเทียมกัน	1	-
14.ให้ความร่วมมือ/ ทำงานร่วมกับผู้อื่นได้	0.8	-
15.สนใจมีส่วนร่วม/ตั้งใจทำกิจกรรมกลุ่มอย่างสม่ำเสมอ	0.8	-
16.แสดงออกถึงการมีความคิดริเริ่ม/ สร้างสรรค์	1	-
17.สามารถให้ข้อมูลย้อนกลับ/ สะท้อนความคิดแก่กลุ่มได้	1	-
18.สามารถสรุปข้อมูล/เนื้อหาที่สำคัญของการเรียนรู้ในครั้งนี้	1	-
19.แสดงความพึงพอใจต่อการเรียนแบบ PBL ในครั้งนี้	0.8	ไม่แน่ใจว่าควรถามหรือไม่

ขั้นตอนที่ 4 นำแบบสอบถามที่ผ่านการพิจารณาความตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญและปรับปรุงแก้ไขตามคำแนะนำของผู้เชี่ยวชาญแล้วไปทดลองใช้กับนักศึกษาคณะครุศาสตร์มหาวิทยาลัยราชภัฏยะลา ชั้นปีที่ 1 ในวิชาเรียนที่ใช้ PBL จำนวน 30 คน โดยให้นักศึกษาทำการ

ประเมินตนเอง และนำมาหาค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's alpha coefficient) ได้ค่าเท่ากับ 0.893 ซึ่งมีค่ามากกว่า 0.5 ขึ้นไป แสดงว่า แบบวัดมีความเที่ยงในระดับดี (ศิริชัย กาญจนวาสี, 2552)

เพื่อให้มองเห็นภาพรวมของเครื่องมือที่ใช้ในการวิจัยนี้ ผู้วิจัยได้ทำตารางที่ 12 เพื่อสรุปคุณลักษณะของเครื่องมือทั้งหมดดังนี้

ตารางที่ 12 สรุปคุณลักษณะของเครื่องมือที่ใช้ในการวิจัย

เครื่องมือ	รูปแบบของเครื่องมือ	จำนวน ข้อ	ผู้ใช้	IOC	Cronbach's alpha
1. แบบสอบถาม คุณลักษณะ ของผู้ประเมิน	ชนิดตัวเลือกคู่ วัดใน 2 ตัว บ่งชี้ ได้แก่ เป้าหมาย และ ความสามารถ และชนิดมาตร ประมาณค่า 5 ระดับ วัดใน	39	กลุ่มผู้ประเมิน อาจารย์ เพื่อน และตนเอง ใน วิชา PBL	IOC \geq 0.80	0.780
1.1 บริบท	5 ตัวบ่งชี้ ได้แก่ บุคลิกภาพ	30			
1.2 การตัดสินใจ	แบบมีจิตสำนึก แรงจูงใจ ความรับผิดชอบที่ตรวจสอบ ได้ การรับรู้มาตรฐานของการ เรียนรู้โดยใช้ปัญหาเป็นฐาน และกระบวนการเปรียบเทียบ	9		0.723	
2. แบบประเมิน กระบวนการเรียนรู้ ในกลุ่มย่อย	ชนิดมาตรประมาณค่า 4 ระดับ	19	กลุ่มผู้ประเมิน อาจารย์ เพื่อน และตนเอง ใน วิชา PBL	IOC \geq 0.80	0.893

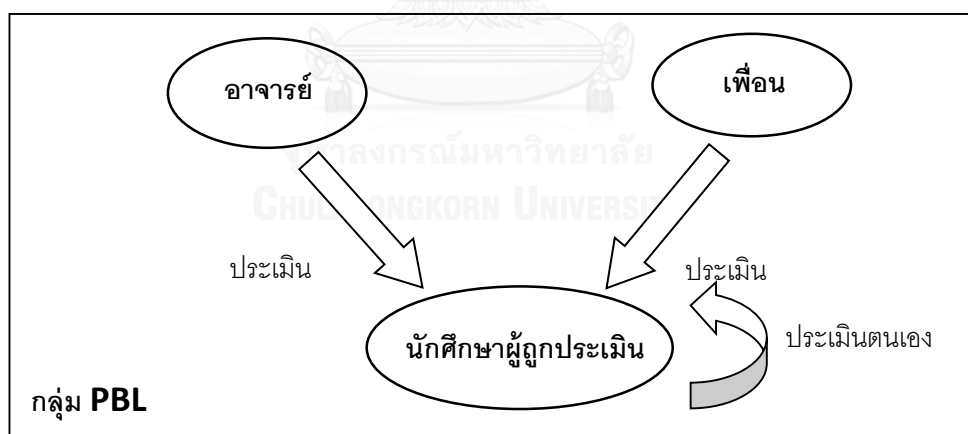
1.4 การเก็บรวบรวมข้อมูล

ขั้นตอนในการดำเนินการเก็บรวบรวมข้อมูลมีรายละเอียดดังนี้

- 1) ผู้วิจัยเสนอโครงการวิจัยเพื่อขอรับการพิจารณาจริยธรรมการวิจัยต่อคณะกรรมการพิจารณาจริยธรรมการวิจัยในคน กลุ่มสหสถาบัน ชุดที่ 1 (เลขที่ใบรับรองโครงการวิจัย COA No.070/2557)
- 2) ผู้วิจัยทำหนังสือขอความร่วมมือในการทำวิจัยจาก คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และส่งไปยังหน่วยงานที่เป็นต้นสังกัดของกลุ่มตัวอย่างเพื่อขอความอนุเคราะห์ในการเก็บรวบรวมข้อมูล
- 3) ผู้วิจัยจัดเตรียมและตรวจสอบความพร้อมของแบบสอบถามที่จะใช้ในการเก็บรวบรวมข้อมูล
- 4) ผู้วิจัยจัดเตรียมผู้ช่วยนักวิจัยจำนวน 3 คน เพื่อทำหน้าที่ช่วยเก็บข้อมูลแบบสอบถาม โดยผู้วิจัยได้ชี้แจงรายละเอียดเกี่ยวกับวัตถุประสงค์การวิจัย พร้อมอธิบายรายละเอียดเกี่ยวกับแบบสอบถามที่ใช้ในงานวิจัย วิธีการประสานงานกับผู้มีส่วนร่วมในการวิจัย และวิธีการเก็บข้อมูลแบบสอบถามแก่ผู้ช่วยนักวิจัย เพื่อให้มีการปฏิบัติที่เป็นมาตรฐานเดียวกัน และเปิดโอกาสให้ซักถามข้อสงสัยในประเด็นต่างๆ รวมถึงแจ้งให้ผู้ช่วยนักวิจัยทราบว่า หากมีปัญหาในระหว่างการประสานงานกับผู้มีส่วนร่วมในการวิจัย หรือระหว่างที่ทำการเก็บข้อมูลแบบสอบถามจะสามารถโทรศัพท์ติดต่อผู้วิจัยได้ตลอดเวลา และในระหว่างที่มีดำเนินการนั้นผู้วิจัยจะมีการติดตามผู้ช่วยนักวิจัยเป็นระยะเพื่อสอบถามปัญหาและอุปสรรคที่พบ พร้อมทั้งให้การแก้ไขปัญหาและอุปสรรคนั้น
- 5) ผู้วิจัยติดต่อประสานงานเพื่อนัดหมายวันและเวลาในการชี้แจงข้อมูลการวิจัยแก่อาจารย์และนักศึกษาในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของสถาบันอุดมศึกษาที่เป็นกลุ่มตัวอย่าง โดยวันและเวลาที่กำหนดจะต้องเป็นช่วงภายหลังเสร็จสิ้นกระบวนการกลุ่มในการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งเมื่อถึงกำหนดการที่ได้นัดหมายผู้วิจัยหรือผู้ช่วยนักวิจัยจะแนะนำตัว อธิบายวัตถุประสงค์ในการทำวิจัย ประโยชน์ของการวิจัย การรักษาความลับของข้อมูลส่วนตัวและจะไม่เปิดเผยต่อสาธารณะเป็นรายบุคคล แต่จะรายงานผลการวิจัยเป็นข้อมูลส่วนรวมซึ่งข้อมูลใดที่สามารถระบุถึงตัวอาสาสมัครได้จะไม่ปรากฏในรายงาน และขอความร่วมมือในการวิจัยตามความสมัครใจเท่านั้น ร่วมกับการชี้แจงว่าหากไม่ร่วมเข้าในการศึกษานี้จะไม่มีผลกระทบใดๆ ต่อการปฏิบัติงานหรือหน่วยงานในกรณีอาสาสมัครที่เป็นอาจารย์ หรือไม่มีผลกระทบใดๆ ต่อการเรียนและการให้คะแนนตัดเกรดในรายวิชาที่เรียนในกรณีอาสาสมัครที่เป็นนักศึกษา และสามารถขอยุติการเข้าร่วมงานวิจัยนี้

เมื่อใดก็ได้ที่ต้องการโดยไม่มีผลกระทบใดๆ เช่นกัน ซึ่งถ้าอาจารย์และนักศึกษามีปัญหาสงสัยใดๆ สามารถซักถามได้ทันที

6) ผู้วิจัยหรือผู้ช่วยนักวิจัยทำการสุ่มอย่างง่ายจากรายชื่อนักศึกษาในแต่ละกลุ่มย่อยและมีความสมัครใจเข้าร่วมการวิจัย เพื่อให้ได้นักศึกษาจำนวน 2 คน ซึ่งทำหน้าที่เป็นผู้ประเมินตนเอง 1 คน และผู้ประเมินเพื่อน 1 คน จากนั้นผู้วิจัยหรือผู้ช่วยนักวิจัยจะมอบเอกสารได้แก่ เอกสารแนะนำสำหรับอาสาสมัคร ใบยินยอมเข้าร่วมการวิจัยที่ให้ผู้เข้าร่วมการวิจัยลงนามเพื่อแสดงความยินยอมเป็นลายลักษณ์อักษร และแบบสอบถามคุณลักษณะของผู้ประเมินจำนวน 45 ข้อ และแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยจำนวน 19 ข้อ ให้แก่ผู้มีส่วนร่วมในการวิจัยที่สมัครใจเข้าร่วมการวิจัย ซึ่งในแต่ละกลุ่มย่อยประกอบด้วยอาจารย์ เพื่อน และตนเอง (ผู้เรียน) อย่างละ 1 คน และในการตอบแบบสอบถามจะใช้เวลารวมประมาณ 20 นาที จากนั้นเมื่อผู้มีส่วนร่วมในการวิจัยที่สมัครใจเข้าร่วมการวิจัยตอบแบบสอบถามแล้วเสร็จ ผู้วิจัยหรือผู้ช่วยนักวิจัยจะทำการเก็บรวบรวมกลับคืนโดยจะระมัดระวังเรื่องการเปิดเผยหรือการรั่วไหลของข้อมูลไปสู่บุคคลภายนอกและจะปกปิดข้อมูลเป็นความลับ โดยรายละเอียดของขั้นตอนการเก็บรวบรวมข้อมูลตามภาพประกอบที่ 10



ภาพที่ 10 ขั้นตอนการเก็บรวบรวมข้อมูลในระยะที่ 1

7) ผู้วิจัยนำแบบสอบถามและแบบประเมินมาตรวจสอบความสมบูรณ์ของข้อมูล และลงรหัสเลขที่ จากนั้นนำข้อมูลที่รวบรวมได้ไปวิเคราะห์ทางสถิติ ซึ่งในระหว่างการวิเคราะห์ข้อมูลผู้วิจัยจะระมัดระวังเรื่องการเปิดเผยหรือการรั่วไหลของข้อมูลไปสู่บุคคลภายนอก และมีการกำหนดรหัส

ของผู้เข้าร่วมการวิจัย ดังนั้นจึงไม่มีรายชื่อของผู้เข้าร่วมการวิจัยปรากฏอยู่ ซึ่งผู้ที่สามารถรับรู้ข้อมูลเหล่านี้นอกจากผู้วิจัย ผู้ช่วยนักวิจัย และผู้กำกับดูแลการวิจัยแล้ว จะต้องได้รับคำยินยอมจากผู้เข้าร่วมการวิจัยเป็นลายลักษณ์อักษรก่อน และผู้วิจัยจะเก็บรักษาข้อมูลเป็นความลับ และผลการวิจัยจะไม่ได้มีการนำเสนอผลเป็นรายบุคคล/สถาบันการศึกษา แต่จะนำเสนอเป็นภาพรวมของกลุ่มตัวอย่างทั้งหมด จึงจะไม่มีข้อมูลใดในการรายงานที่จะนำไปสู่การระบุตัวของผู้มีส่วนร่วมในการวิจัยได้

1.5 การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลผู้วิจัยใช้โปรแกรมสำเร็จรูป SPSS for Window (statistical package for social science) โปรแกรม Facets และโปรแกรม LISREL โดยมีขั้นตอนวิธีการวิเคราะห์ดังนี้

1) ข้อมูลเบื้องต้นเกี่ยวกับลักษณะของกลุ่มตัวอย่างและตัวแปร วิเคราะห์ข้อมูลโดยใช้สถิติบรรยาย โดยกรณีตัวแปรต่อเนื่อง ได้แก่ อายุ ระยะเวลาที่สอน PBL เกรดเฉลี่ยสะสม และจำนวนวิชาที่เคยเรียน PBL จะคำนวณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ส่วนกรณีตัวแปรไม่ต่อเนื่อง ได้แก่ เพศ วุฒิการศึกษา คณะ และชั้นปี จะคำนวณความถี่และร้อยละ ด้วยโปรแกรม SPSS

2) การวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ใช้โมเดลหลายองค์ประกอบของราสซ์ด้วยโปรแกรม Facets

3) ตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนน ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานกับข้อมูลเชิงประจักษ์ ด้วยโปรแกรม LISREL

4) ตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนน ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ด้วยโปรแกรม LISREL

ตอนที่ 2 การพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน

เมื่อเสร็จสิ้นการดำเนินการในระยะที่ 1 จะได้โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งมีความตรงกับข้อมูลเชิงประจักษ์ และสามารถอธิบายได้ว่า องค์ประกอบ ตัวบ่งชี้ และเส้นทางอิทธิพลที่สำคัญ รวมถึงมีพารามิเตอร์ใดที่แปรเปลี่ยนระหว่างกลุ่มผู้ประเมิน จากนั้นผู้วิจัยจะนำองค์ความรู้ที่ได้นี้มาดำเนินการในระยะที่ 2 คือ การพัฒนาโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ให้มีความเหมาะสมกับแต่ละ

กลุ่มผู้ประเมิน และทำการตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานนี้ โดยมีรายละเอียดของการดำเนินงานดังนี้

2.1 ประชากรและกลุ่มตัวอย่างในระยะที่ 2

ประชากร คือ อาจารย์และนักศึกษาของสถาบันอุดมศึกษาต่างๆ ในวิชาต่างๆ ที่จัดการสอนในรูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐาน

กลุ่มตัวอย่าง คือ อาจารย์และนักศึกษาของสถาบันอุดมศึกษาที่ได้จากวิธีการเลือกแบบเจาะจง (purposive sampling) จำนวน 4 กลุ่ม PBL และทำการสุ่มอย่างง่ายเข้ากลุ่มทดลองและกลุ่มควบคุม อย่างละ 2 กลุ่ม PBL (random assignment) โดยกำหนดเกณฑ์ในการคัดเลือกเพื่อความเท่าเทียมกันของกลุ่มทดลองและกลุ่มควบคุม คือ (1) ในกลุ่มทดลองและกลุ่มควบคุมจะต้องเป็นสถาบันอุดมศึกษาของรัฐที่มีการเรียนการสอนในวิชาเดียวกัน และใช้รูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐาน ครบทั้ง 9 ขั้นตอน ได้แก่ ก)การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ ข)การกำหนดประเด็นปัญหา ค)การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้ จากความรู้และประสบการณ์ที่มีอยู่ ง)การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล จ)การระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา ฉ)การศึกษาค้นคว้าด้วยตนเอง ช)การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้ไปเรียนรู้มาในกลุ่ม ซ)การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ และ ฅ)การสะท้อนความรู้ในบทเรียน โดยทำการสรุปทั้งในด้านประเด็นปัญหา ผลจากการไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน และกระบวนการในการแก้ปัญหา (2) ผู้บริหารสถาบันหรือผู้ที่มีอำนาจให้การอนุมัติในการศึกษาวิจัย (3) อาจารย์และนักศึกษามีครใจเข้าร่วมการวิจัย (4) อาจารย์จะต้องมีประสบการณ์ในการสอนวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานไม่ต่ำกว่า 1 ปี และ (5) ในแต่ละกลุ่ม PBL ต้องประกอบด้วยอาจารย์ 1 คน และกำหนดจำนวนนักศึกษาอยู่ระหว่าง 8 - 12 คน และเกณฑ์ในการคัดกลุ่ม PBL ออก คือ มีอาจารย์ประจำกลุ่มหรือนักศึกษาในกลุ่มแสดงความประสงค์ไม่เข้าร่วมการวิจัยตั้งแต่ 1 คนขึ้นไป

2.2 เครื่องมือที่ใช้ในการวิจัยในระยะที่ 2

เครื่องมือที่ใช้ในการวิจัยในระยะที่ 2 มี 2 ประเภท คือ

1. เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล เป็นแบบวัดเดียวกันกับที่ใช้ในขั้นตอนการวิจัยระยะที่ 1ซึ่งประกอบด้วย 1) แบบสอบถามคุณลักษณะของผู้ประเมิน จำนวน 1 ฉบับ แบ่งออกเป็น 2 ตอน ได้แก่ ตอนที่ 1 ข้อมูลทั่วไปของผู้ตอบ จำนวน 6 ข้อ มีลักษณะเป็นคำถามปลายเปิด และ

แบบตรวจสอบรายการ (check list) ตอนที่ 2 แบบสอบถามเกี่ยวกับตัวแปรต่างๆ ในงานวิจัย จำนวน 39 ข้อ ประกอบด้วยเป็นแบบสอบถามชนิดตัวเลือกคู่ (forced-choice items) และชนิดมาตราประมาณค่า 5 ระดับ (rating scale) และ 2) แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย จำนวน 1 ฉบับ มีจำนวน 19 ข้อ

2. เครื่องมือที่ใช้ในการทดลอง คือ คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา สำหรับอาจารย์ 1 ชุด และสำหรับนักศึกษา 1 ชุด โดยมีขั้นตอนของการสร้างคู่มือดังนี้

ขั้นตอนที่ 1 ศึกษาองค์ความรู้เกี่ยวกับโมเดลเชิงสาเหตุของคุณภาพการให้คะแนน ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งได้จากการวิจัยในระยะที่ 1 ว่ามีองค์ประกอบ ตัวบ่งชี้ หรือเส้นทางอิทธิพลใดที่สำคัญ และมีพารามิเตอร์ใดที่แปรเปลี่ยนหรือไม่แปรเปลี่ยนระหว่างกลุ่มผู้ประเมิน ร่วมกับศึกษาแนวคิดทฤษฎีที่เกี่ยวข้องกับการพัฒนาหรือสร้างให้เกิดตัวแปรสำคัญดังกล่าว เพื่อเป็นแนวทางในการพัฒนาคู่มือ

ขั้นตอนที่ 2 ผู้วิจัยพัฒนาโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน จากการศึกษารายละเอียดในขั้นตอนที่ 1 ซึ่งมีความเหมาะสมกับแต่ละกลุ่มผู้ประเมิน และสร้างคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา สำหรับอาจารย์ 1 ชุด และสำหรับนักศึกษา 1 ชุด โดยรายละเอียดของคู่มือสำหรับอาจารย์ ได้แก่ หลักการและวัตถุประสงค์ของคู่มือ แนวทางการปฏิบัติและเครื่องมือที่ใช้สำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย และแนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ ส่วนคู่มือสำหรับนักศึกษาจะมีรายละเอียดเฉพาะหัวข้อแนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ และแบบประเมินกระบวนการกลุ่ม ซึ่งทั้งสองหัวข้อนี้มีเนื้อหาเหมือนกับคู่มือสำหรับอาจารย์ (รายละเอียดของคู่มือดูที่ภาคผนวก ค)

ขั้นตอนที่ 3 ผู้วิจัยนำคู่มือที่สร้างขึ้นไปให้อาจารย์ที่ปรึกษาตรวจสอบความถูกต้องและความเหมาะสมของคู่มือ พร้อมทั้งขอคำแนะนำและข้อเสนอแนะเพื่อปรับปรุงแก้ไขคู่มือ และนำมาปรับปรุงแก้ไขตามคำแนะนำของอาจารย์ที่ปรึกษา

ขั้นตอนที่ 4 ผู้วิจัยนำคู่มือไปให้ผู้ทรงคุณวุฒิ ซึ่งเป็นผู้ที่มีความรู้และประสบการณ์ในด้านการเรียนรู้โดยใช้ปัญหาเป็นฐาน จำนวน 4 คน ตรวจสอบความถูกต้องและความเหมาะสมของคู่มือ พร้อมทั้งขอคำแนะนำและข้อเสนอแนะเพื่อปรับปรุงแก้ไขคู่มือ

2.3 การดำเนินการทดลองและการเก็บรวบรวมข้อมูลในระยะที่ 2

การวิจัยครั้งนี้ผู้วิจัยใช้ระเบียบวิธีวิจัยกึ่งทดลอง (quasi-experimental designs) แบบสุ่มตัวอย่างสองกลุ่มวัดสองครั้ง (pretest-posttest control group design with random assignment) เพื่อให้สามารถเปรียบเทียบผลการเปลี่ยนแปลงของตัวแปรตามอันเนื่องมาจากตัวแปรอิสระได้ชัดเจนมากยิ่งขึ้น โดยมีแบบแผนการทดลองดังนี้



- | | | |
|----|-----|---|
| E | คือ | กลุ่มทดลอง |
| C | คือ | กลุ่มควบคุม |
| O1 | คือ | การทดสอบก่อนการทดลอง |
| O2 | คือ | การทดสอบหลังการทดลอง |
| X | คือ | คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติ
สำหรับการเป็นผู้ประเมินที่มีคุณภาพ |

ผู้วิจัยเป็นผู้ดำเนินการรวบรวมข้อมูลกับกลุ่มตัวอย่าง ทั้งกลุ่มทดลองและกลุ่มควบคุมด้วยตนเอง โดยดำเนินการวิจัยตามลำดับขั้นตอนดังต่อไปนี้

1. ขั้นเตรียมก่อนดำเนินการทดลอง ผู้วิจัยติดต่อประสานงานกับต้นสังกัดของกลุ่มตัวอย่าง โดยขอรายชื่ออาจารย์และนักศึกษาเพื่อทำการสุ่มอย่างง่ายเข้ากลุ่มทดลองและกลุ่มควบคุมอย่างละ 2 กลุ่ม PBL และนัดหมายวันเวลาและสถานที่ในการชี้แจงข้อมูลการวิจัยแก่อาจารย์และนักศึกษาของกลุ่มทดลองและกลุ่มควบคุมแต่ละกลุ่ม รวมจำนวน 4 กลุ่ม โดยวันและเวลาที่กำหนดจะเป็นช่วง

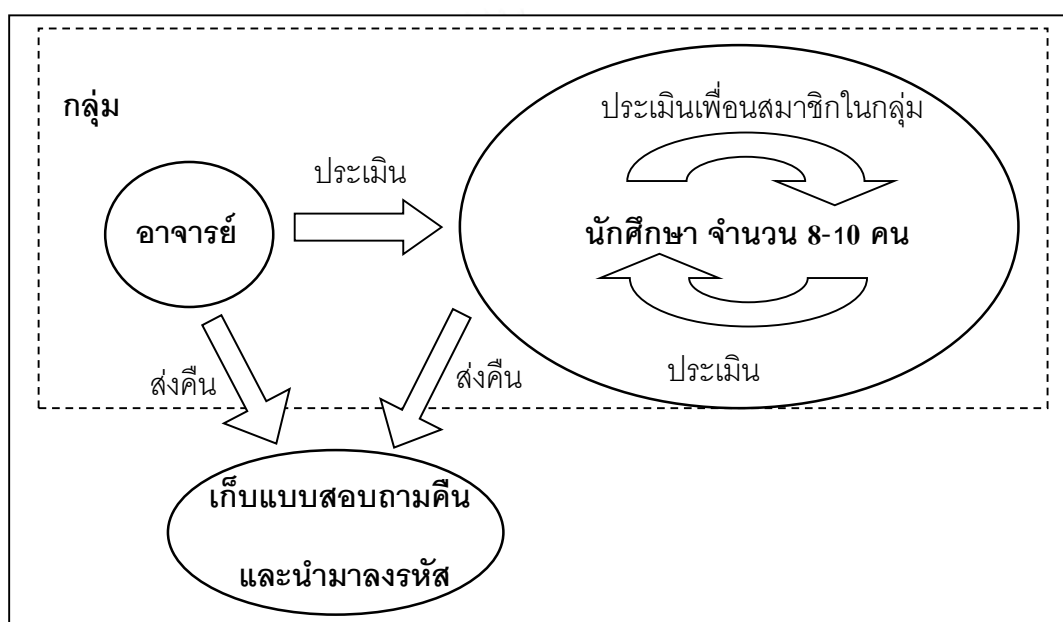
ภายหลังเสร็จสิ้นกระบวนการกลุ่มในวันแรกหรือวันเปิดโจทย์สถานการณ์ ซึ่งเมื่อถึงกำหนดการที่ได้ หนดหมาย ผู้วิจัยได้แนะนำตัวอธิบายวัตถุประสงค์ในการทำวิจัย ประโยชน์ของการวิจัย ขั้นตอนในการ ทำวิจัย การรักษาความลับของข้อมูลส่วนตัวซึ่งจะไม่เปิดเผยต่อสาธารณะเป็นรายบุคคล แต่จะ รายงานผลการวิจัยเป็นข้อมูลส่วนรวม และข้อมูลใดที่สามารถระบุถึงตัวอาสาสมัครได้จะไม่ปรากฏใน รายงาน และขอความร่วมมือในการวิจัยตามความสมัครใจเท่านั้น ร่วมกับการชี้แจงว่าหากไม่ร่วมเข้า ในการศึกษาจะไม่มีผลกระทบใดๆ ต่อการปฏิบัติงานหรือหน่วยงานในกรณีอาสาสมัครที่เป็นอาจารย์ หรือไม่มีผลกระทบใดๆ ต่อการเรียนและการให้คะแนน/ตัดเกรดในรายวิชาที่เรียนในกรณีอาสาสมัคร ที่เป็นนักศึกษา และสามารถขอยุติการเข้าร่วมงานวิจัยนี้เมื่อใดก็ได้ที่ต้องการโดยไม่มีผลกระทบใดๆ เช่นกัน ซึ่งถ้าอาจารย์และนักศึกษามีปัญหาสงสัยใดๆ สามารถซักถามได้ทันทีเมื่ออาจารย์และ นักศึกษาในกลุ่มย่อยทุกคนแสดงความยินดีให้ความร่วมมือในการวิจัย ผู้วิจัยจะดำเนินการในขั้นต่อไป แต่ถ้ามีอาจารย์ประจำกลุ่มหรือนักศึกษาในกลุ่มแสดงความประสงค์ไม่เข้าร่วมการวิจัยตั้งแต่ 1 คนขึ้นไป ผู้วิจัยจะติดต่อกลุ่ม PBL อื่นๆ แทนและดำเนินการในขั้นตอนนี้ใหม่

2. เมื่อเสร็จสิ้นการดำเนินการในข้อ 1 ผู้วิจัยจะมอบเอกสารได้แก่ เอกสารแนะนำสำหรับ อาสาสมัคร ใบยินยอมเข้าร่วมการวิจัยที่ให้ผู้เข้าร่วมการวิจัยลงนามเพื่อแสดงความยินยอม และ แบบสอบถามคุณลักษณะของผู้ประเมิน 1 ฉบับ จำนวน 45 ข้อ ซึ่งใช้เวลาทำประมาณ 10 นาที ให้แก่ ผู้มีส่วนร่วมในการวิจัยทุกคนในกลุ่ม PBL ที่สมัครใจเข้าร่วมการวิจัย ซึ่งในแต่ละกลุ่มจะมีสมาชิก คือ อาจารย์ 1 คน และนักศึกษาประมาณ 8 - 12 คน จากนั้นเมื่อผู้มีส่วนร่วมในการวิจัยตอบแบบสอบถาม แล้วเสร็จ ผู้วิจัยจะทำการเก็บรวบรวมกลับคืน

3. สำหรับกลุ่มทดลองจะได้รับการแจกคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและ แนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ในระดับอุดมศึกษา และผู้วิจัยจะอธิบายรายละเอียดเบื้องต้นในคู่มือแก่อาจารย์และผู้เรียนในกลุ่ม ทดลอง ส่วนกลุ่มควบคุมจะได้รับการแจกเฉพาะตัวอย่างแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย พร้อมคำอธิบายการใช้แบบประเมิน

4. ระยะเวลาหลังการทดลอง ผู้วิจัยจะติดต่อประสานงานเพื่อนัดหมายวันเวลาและสถานที่ในการ ตอบแบบสอบถามกับอาจารย์และนักศึกษาของแต่ละกลุ่มทั้งในกลุ่มควบคุมและกลุ่มทดลอง ซึ่งมี จำนวนกลุ่ม PBL รวม 4 กลุ่ม โดยวันและเวลาที่กำหนดเวลาจะเป็นช่วงภายหลังเสร็จสิ้นกระบวนการ กลุ่มในการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งเมื่อถึงกำหนดการที่ได้นัดหมายผู้วิจัยจะดำเนินการเก็บ

รวบรวมข้อมูลกับสมาชิกของแต่ละกลุ่ม PBL คือ อาจารย์ 1 คน และนักศึกษาประมาณ 8 - 12 คน โดยให้แต่ละคนตอบแบบสอบถามคุณลักษณะของผู้ประเมิน 1 ฉบับ จำนวน 45 ข้อ โดยใช้เวลาประมาณ 10 นาที (เป็นแบบสอบถามชุดเดียวกับที่ใช้ในวันเปิดโจทย์สถานการณ์) และแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย จำนวน 19 ข้อ ใช้เวลาประมาณ 20 นาที ซึ่งแต่ละฉบับมีรายละเอียดของการเก็บรวบรวมข้อมูลในแต่ละกลุ่ม PBL ดังนี้ คือ 1) อาจารย์ทำการประเมินนักศึกษาในกลุ่มทุกคน 2) นักศึกษาแต่ละคนทำการประเมินเพื่อนสมาชิกในกลุ่มทุกคน 3) นักศึกษาแต่ละคนทำการประเมินตนเอง (ภาพประกอบที่ 11) จากนั้นเมื่อผู้มีส่วนร่วมในการวิจัยตอบแบบสอบถามแล้วเสร็จ ผู้วิจัยจะทำการเก็บรวบรวมกลับคืน



ภาพที่ 11 ขั้นตอนการเก็บรวบรวมข้อมูลในระยะที่ 2

5. ผู้วิจัยนำแบบสอบถามและแบบประเมินที่ได้จากข้อ 2 และข้อ 4 มาจับคู่ข้อมูลตามรหัสผู้เข้าร่วมการวิจัย จากนั้นนำข้อมูลที่รวบรวมได้ไปวิเคราะห์ทางสถิติ

2.4 การวิเคราะห์ข้อมูล

ผู้วิจัยดำเนินการวิเคราะห์ข้อมูลตามขั้นตอนดังนี้

1) วิเคราะห์ความคลาดเคลื่อนในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง จากแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยในระยะหลังทดลอง โดยใช้โมเดลหลายองค์ประกอบของราสส์ ด้วยโปรแกรม Facets

2) วิเคราะห์ความแตกต่างของค่าเฉลี่ยของคะแนนหลังทดลองในแต่ละตัวแปรสาเหตุในโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา รวม 5 ตัวแปร ได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบ ที่ตรวจสอบได้ เปรียบเทียบระหว่างกลุ่มควบคุมและกลุ่มทดลอง โดยมีคะแนนก่อนทดลองเป็นตัวแปรร่วม ใช้สถิติการวิเคราะห์ความแปรปรวนร่วม (Analysis of covariance; ANCOVA)

3) วิเคราะห์ความแตกต่างของค่าเฉลี่ยของคะแนนที่ได้ในระยะหลังทดลองจากทั้ง 3 ตัวแปรตามในโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ได้แก่ การรับรู้มาตรฐาน กระบวนการเปรียบเทียบ และความคลาดเคลื่อน เปรียบเทียบระหว่างกลุ่มควบคุมและกลุ่มทดลอง โดยมีคะแนนที่ได้ในระยะก่อนทดลองเป็นตัวแปรร่วม ใช้สถิติการวิเคราะห์ความแปรปรวนตัวแปรร่วมพหุนาม (Multivariate analysis of covariance; MANCOVA)

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์การวิจัย 4 ประการ ได้แก่ 1) วิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา โดยใช้โมเดลหลายองค์ประกอบของราสซ์ 2) พัฒนาและตรวจสอบโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา 3) ตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง และ 4) พัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา โดยการนำเสนอผลการวิเคราะห์ข้อมูลจะแบ่งออกออกเป็น 5 ตอน ได้แก่ ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของกลุ่มตัวอย่าง ตอนที่ 2 ผลการวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ตอนที่ 3 ผลการตรวจสอบโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ตอนที่ 4 ผลการตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง และตอนที่ 5 ผลการพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

ในการวิเคราะห์ข้อมูลและผลการวิเคราะห์ข้อมูล ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้แทนค่าสถิติและตัวแปรต่างๆ ดังนี้

สัญลักษณ์ที่ใช้แทนค่าสถิติ

n	หมายถึง จำนวนตัวอย่าง (total number)
f	หมายถึง ความถี่ (frequency)
M	หมายถึง ค่าเฉลี่ย (mean)
Mdn	หมายถึง ค่ามัธยฐาน (median)
Mo	หมายถึง ค่าฐานนิยม (mode)
Min	หมายถึง ค่าต่ำสุด (minimum)

Max	หมายถึง ค่าสูงสุด (maximum)
SD	หมายถึง ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)
χ^2	หมายถึง ดัชนีตรวจสอบความกลมกลืนประเภทค่าสถิติไค-สแควร์ (Chi-square)
df	หมายถึง ค่าองศาอิสระ (degree of freedom)
p	หมายถึง ระดับนัยสำคัญทางสถิติ (statistic significant level)
CFI	หมายถึง ดัชนีความกลมกลืนเปรียบเทียบ (comparative fit index)
GFI	หมายถึง ดัชนีวัดระดับความกลมกลืน (goodness of fit index)
AGFI	หมายถึง ดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (adjusted goodness of fit index)
RMR	หมายถึง ดัชนีรากของค่าเฉลี่ยกำลังสองของส่วนที่เหลือ (root mean square residual)
RMSEA	หมายถึง ดัชนีรากของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่า (root mean square error of approximation)
R ²	หมายถึง สัมประสิทธิ์การพยากรณ์ (coefficient of determination)
TE	หมายถึง อิทธิพลรวม (total effect)
DE	หมายถึง อิทธิพลทางตรง (direct effect)
IE	หมายถึง อิทธิพลทางอ้อม (indirect effect)

อักษรย่อภาษาอังกฤษ

ตัวแปรแฝง

CONTEXT	หมายถึง บริบทของผู้ประเมิน
JUDGE	หมายถึง การตัดสิน
RATING	หมายถึง คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน

ตัวแปรสังเกตได้

GOAL	หมายถึง เป้าหมาย
ABILI	หมายถึง ความสามารถ
CONSCI	หมายถึง บุคลิกภาพแบบมีจิตสำนึก
MOTIV	หมายถึง แรงจูงใจ

ACCOU	หมายถึง ความรับผิดชอบที่ตรวจสอบได้
STAND	หมายถึง การรับรู้มาตรฐาน
COMPAR	หมายถึง กระบวนการเปรียบเทียบ
ERROR	หมายถึง ความคลาดเคลื่อน
ACCUR	หมายถึง ความแม่นยำ

ตอนที่ 1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของกลุ่มตัวอย่าง

กลุ่มตัวอย่างมาจากสถาบันอุดมศึกษาจำนวน 5 แห่ง ที่จัดการเรียนการสอนแบบ PBL ในรายวิชาต่างๆ เช่น Respiratory system, Cardiovascular system, Biomedical science, การพยาบาลเด็กและวัยรุ่น, การพยาบาลผู้ใหญ่และผู้สูงอายุ, การพยาบาลครอบครัว, การพยาบาลผู้ป่วยวิกฤต, การตรวจวิเคราะห์และคลินิกสัมพันธ์ เป็นต้น โดยมีจำนวนรวมทั้งสิ้น 120 กลุ่ม PBL ซึ่งประกอบด้วยกลุ่มผู้ประเมินที่เป็นอาจารย์จำนวน 120 คน กลุ่มนักศึกษาที่เป็นผู้ประเมินเพื่อนจำนวน 120 คน และกลุ่มนักศึกษาที่เป็นผู้ประเมินตนเองจำนวน 120 คน โดยผลการวิเคราะห์ข้อมูลเบื้องต้นของกลุ่มตัวอย่างประกอบด้วย ผลการวิเคราะห์ความถี่และร้อยละของกลุ่มตัวอย่างจำแนกตามตัวแปร เพศ คณะ และวุฒิการศึกษา ส่วนตัวแปร อายุ ระยะเวลาที่สอน PBL เกรดเฉลี่ยสะสม และจำนวนวิชา PBL ที่เคยเรียน ผลการวิเคราะห์จะนำเสนอเป็นค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

เมื่อพิจารณาตัวแปร เพศ คณะ วุฒิการศึกษา ชั้นปี อายุ ระยะเวลาที่สอน PBL เกรดเฉลี่ยสะสม และจำนวนวิชา PBL ที่เคยเรียน พบว่า กลุ่มตัวอย่างซึ่งเป็นอาจารย์ ส่วนใหญ่เป็นเพศหญิง มีจำนวน 88 คน คิดเป็นร้อยละ 73.33 อยู่ในคณะแพทยศาสตร์ ซึ่งมีจำนวน 52 คน คิดเป็นร้อยละ 43.33 มีวุฒิการศึกษาปริญญาเอก ซึ่งมีจำนวน 64 คน คิดเป็นร้อยละ 53.33 อายุเฉลี่ย 38.71 ปี ($SD=9.53$ ปี) และมีระยะเวลาที่สอน PBL เฉลี่ย 5.43 ปี ($SD=5.16$ ปี) ส่วนกลุ่มตัวอย่างซึ่งเป็นนักศึกษา ส่วนใหญ่เป็นเพศหญิง มีจำนวน 156 คน คิดเป็นร้อยละ 65 อยู่ในคณะแพทยศาสตร์ ซึ่งมีจำนวน 104 คน คิดเป็นร้อยละ 43.33 กำลังศึกษาอยู่ในชั้นปีที่ 3 ซึ่งมีจำนวน 138 คน คิดเป็นร้อยละ 57.50 อายุเฉลี่ย 20.20 ปี ($SD=1.54$ ปี) ค่าเฉลี่ยของเกรดเฉลี่ยสะสม 3.18 ($SD=0.43$) และมีจำนวนวิชา PBL ที่เคยเรียนเฉลี่ย 2.02 วิชา ($SD=2.66$ วิชา) ดังรายละเอียดแสดงในตารางที่ 13 และ 14

ตารางที่ 13 ความถี่และร้อยละของกลุ่มตัวอย่างจำแนกตามตัวแปร เพศ คณะ วุฒิการศึกษา และชั้นปี

ตัวแปร	อาจารย์ (n=120)		นักศึกษา (n=240)	
	จำนวน	ร้อยละ	จำนวน	ร้อยละ
เพศ				
ชาย	32	26.67	84	35.00
หญิง	88	73.33	156	65.00
รวม	120	100.00	240	100.00
คณะ				
แพทยศาสตร์	52	43.33	104	43.33
พยาบาลศาสตร์	21	17.50	42	17.50
ทันตแพทยศาสตร์	21	17.50	42	17.50
สหเวชศาสตร์	26	21.67	52	21.67
รวม	120	100.00	240	100.00
วุฒิการศึกษา				
ปริญญาตรี	23	19.17	-	-
ปริญญาโท	33	27.50	-	-
ปริญญาเอก	64	53.33	-	-
รวม	120	100.00	-	-
ชั้นปี				
1	-	-	42	17.50
2	-	-	24	10.00
3	-	-	138	57.50
4	-	-	32	13.30
5	-	-	4	1.70
รวม	-	-	240	100.00

ตารางที่ 14 ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง จำแนกตามตัวแปร อายุ ระยะเวลาที่สอน GPAX และจำนวนวิชา PBL ที่เคยเรียน

ตัวแปร	อาจารย์ (n=120)		นักศึกษา (n=240)	
	M	SD	M	SD
อายุ (range 22 ถึง 66 ปี)	38.71	9.53	20.20	1.54
ระยะเวลาที่สอน PBL (range 0 ถึง 25 ปี)	5.43	5.16	-	-
GPAX (missing = 42 ราย; range 1.93 ถึง 3.96)	-	-	3.18	0.43
จำนวนวิชา PBL ที่เคยเรียน (range 0 ถึง 12 วิชา)	-	-	2.02	2.66

ตอนที่ 2 ผลการวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมิน ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

ผลการวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยใช้โมเดลหลายองค์ประกอบของราสซ์ จากโปรแกรม Facets ในส่วนนี้จะนำเสนอภาพรวมของกลุ่มผู้ประเมินทั้ง 3 กลุ่ม คือ อาจารย์ เพื่อน และตนเอง ซึ่งทำการประเมินในผู้รับการประเมินจำนวน 120 คน โดยใช้แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยที่มีลักษณะเป็นมาตรวัดประมาณค่า 4 ระดับ (rating scale) จำนวน 19 ข้อ โดยมีรูปแบบการวิเคราะห์ 3 ฟาเซท (facets) คือ ผู้รับการประเมิน (ratee) * กลุ่มผู้ประเมิน (rater) * รายการประเมิน (items) ผลการวิเคราะห์ดังแสดงในตารางที่ 15 – 17

ผลการวิเคราะห์ค่าสถิติทดสอบความสอดคล้อง (fit) ระหว่างข้อมูล กับ Rasch model พบว่า ข้อมูลชุดนี้มีความสอดคล้องกับ Rasch model เป็นอย่างดี โดยพิจารณาจาก ค่าเฉลี่ยของ StRes (mean of the standardized residuals) = 0.0 และ SD (sample standard deviation) = 1.00 ซึ่งถ้าค่าเฉลี่ยของ StRes มีค่าเข้าใกล้ 0.0 และ SD มีค่าเข้าใกล้ 1.0 จะแสดงว่าข้อมูลมีความสอดคล้องกับ Rasch model (Linacre, 2012) ส่วนผลการวิเคราะห์ความคลาดเคลื่อนชนิดกด/ปล่อยคะแนนในตารางที่ 15 และ 16 เมื่อเปรียบเทียบระหว่างกลุ่มผู้ประเมิน พบว่า ผู้ประเมินกลุ่มตนเองมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด รองลงมาคือ กลุ่มอาจารย์ ส่วนกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน (Measure in Logits

ของผู้ประเมินกลุ่มตนเอง เพื่อน และอาจารย์ มีค่าเท่ากับ 0.47, -0.75, และ 0.28 ตามลำดับ; ซึ่งค่าเป็นบวกจะแสดงถึงการกดคะแนน และค่าเป็นลบจะแสดงถึงการปล่อยคะแนน) นอกจากนี้เมื่อพิจารณาพฤติกรรมการให้คะแนนของผู้ประเมินแต่ละกลุ่มจะพบว่า ผู้ประเมินกลุ่มเพื่อนให้คะแนนเฉลี่ยรายข้อมากกว่าผู้ประเมินกลุ่มอาจารย์และตนเอง (คะแนนเฉลี่ยรายข้อ = 3.21, 2.84 และ 2.77 ตามลำดับ) สอดคล้องกับผลการวิเคราะห์ในตารางที่ 17 นั่นคือ กลุ่มผู้ประเมินที่มีการใช้มาตรวัดคะแนนค่า 4 มากที่สุด คือ กลุ่มเพื่อน ซึ่งมีจำนวน 798 ครั้ง คิดเป็นร้อยละ 53.40 รองลงมาคือ กลุ่มอาจารย์ ซึ่งมีจำนวน 386 ครั้ง คิดเป็นร้อยละ 25.80 และกลุ่มตนเอง ซึ่งมีจำนวน 310 ครั้ง คิดเป็นร้อยละ 20.70 ตามลำดับ และเมื่อเปรียบเทียบเฉพาะในแต่ละกลุ่มผู้ประเมินพบว่า ผู้ประเมินทั้ง 3 กลุ่ม มีการใช้มาตรวัดคะแนนค่า 3 บ่อยที่สุด โดยกลุ่มตนเองใช้จำนวน 1,188 ครั้ง คิดเป็นร้อยละ 52.10 ส่วนกลุ่มอาจารย์ใช้จำนวน 1,202 ครั้ง คิดเป็นร้อยละ 52.70 และกลุ่มเพื่อนใช้จำนวน 1,182 ครั้ง คิดเป็นร้อยละ 51.80

ผลการวิเคราะห์ความแม่นยำในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน จากตารางที่ 16 เมื่อพิจารณาจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมินพบว่า ผู้ประเมินทั้ง 3 กลุ่ม คือ ตนเอง อาจารย์ และเพื่อน มีความแม่นยำในการให้คะแนนเท่ากับ 0.97, 0.98, และ 1.05 ตามลำดับ ซึ่งค่านี้อยู่ระหว่าง 0.6 ถึง 1.4 แสดงถึงการมีความแม่นยำในการให้คะแนน (Wright et al., 1994) ส่วนผลการวิเคราะห์อื่นๆ นอกจากเปรียบเทียบความคลาดเคลื่อนของผู้ประเมิน (คอลัมน์ Rater-type) และเปรียบเทียบการใช้มาตรวัดคะแนน (คอลัมน์ Scale) ที่แสดงในตารางที่ 15 มีดังนี้ 1) ความสามารถของผู้รับการประเมิน (คอลัมน์ Ratee) พบว่า ผู้รับการประเมินคนที่ 99 มีคะแนนสูงสุด รองลงมาคือ คนที่ 53 และ 101 ส่วนผู้ที่มีคะแนนต่ำสุด มี 3 คน คือ คนที่ 25, 38 และ 70 และ 2) รายการประเมินข้อที่มีการให้คะแนนสูงสุด (คอลัมน์ Item) คือ ข้อที่ 14 รองลงมาคือ ข้อที่ 13 ส่วนข้อที่มีการให้คะแนนต่ำสุด คือ ข้อที่ 1

จากผลการวิเคราะห์ผลการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของกลุ่มตัวอย่างสรุปได้ว่า ผู้ประเมินกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน ส่วนผู้ประเมินกลุ่มตนเองมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มอาจารย์ และเมื่อพิจารณาความแม่นยำในการให้คะแนนจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม มีการให้คะแนนที่มีความแม่นยำ

จากผลทั้งสองอย่างนี้แสดงว่า ในภาพรวมผู้ประเมินทั้งสามกลุ่มให้คะแนนที่มีความคลาดเคลื่อนเพียงเล็กน้อย และจัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้

ตารางที่ 15 แสดงการเปรียบเทียบระหว่างฟาเซท

Measr	+Ratee	-Rater-type	+Item	Scale
5	+	+	+	(4)
4	+ 99	+	+	+
3	+	+	+	+
	53 101 42 65 62 30 24 63 83			---
2	+ 7 12 69 28 50 57 68 74 75 84 85 95 98 100 105 34 46 56 67 4 16 27 35 45 60 61 64 102 103 120 18 51 55 66 71 87 89 94 111 88 108 113 115 5 20 21 32 37 41 49 86 104 112 116 73 79 96 117	+	+	+
1	+ 1 2 23 31 33 39 77 80 90 92 97 106 110 9 48 78 109 119 6 8 10 14 40 43 47 29 44 54 59 76 93 107 118 3 15 52 72 13 81 58 19 26 36 114	+	+	+
0	* 17 22 91 11 82 25 38 70	Self-rater Tutor-rater	14 13 15 19	3
		Peer-rater	8 18 6 10 11 3 7 16 2 4 5 9 12 17 1	---
-1	+	+	+	(1)
Measr	+Ratee	-Rater-type	+Item	Scale

ตารางที่ 16 ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน

กลุ่มผู้ประเมิน	Sum of Ratings	Count of Ratings	Obsvd Average	Measure in Logits	Infit MnSq
ตนเอง (most severe)	6,318	2,280	2.77	0.47	0.97
อาจารย์	6,483	2,280	2.84	0.28	0.98
เพื่อน (most lenient)	7,323	2,280	3.21	-0.75	1.05
<i>M</i>	6,708.00	2,280.00	2.94	0.00	1.00
<i>SD</i>	538.96	0.00	0.24	0.66	0.04

mean of the "StRes" (Standardized Residuals) = 0.0, *SD* (sample standard deviation) = 1.00

ตารางที่ 17 ความถี่ของการให้คะแนนในแต่ละมาตรวัด

กลุ่มผู้ประเมิน	มาตรวัด (score)											
	1			2			3			4		
	f	Used %	Rel. %	f	Used %	Rel. %	f	Used %	Rel. %	f	Used %	Rel. %
อาจารย์	51	2.2	44.0	641	28.1	38.7	1,202	52.7	33.7	386	16.9	25.8
เพื่อน	15	0.7	12.9	285	12.5	17.2	1,182	51.8	33.1	798	35.0	53.4
ตนเอง	50	2.2	43.1	732	32.1	44.1	1,188	52.1	33.3	310	13.6	20.7
รวม	116	5.1	100	1,658	72.7	100	3,572	156.7	100	1,494	65.5	100

ตอนที่ 3 ผลการตรวจสอบโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

การนำเสนอผลการตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานกับข้อมูลเชิงประจักษ์ แบ่งออกเป็น 3 ตอน ดังนี้

3.1 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวบ่งชี้ภายในโมเดล

ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรสังเกตได้จำนวน 9 ตัวแปร รวม 36 คู่ ในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยใช้ค่าสหสัมพันธ์แบบเพียร์สัน พบว่า มีจำนวน 23 คู่ ที่มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง -0.271 ถึง 0.512 โดยคู่ที่มีความสัมพันธ์กันมากที่สุดคือ ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) กับแรงจูงใจ (MOTIV) มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.512 รองลงมา คือ ตัวแปรการรับรู้มาตรฐาน (STAND) กับกระบวนการเปรียบเทียบ (COMPAR) มีค่าเท่ากับ 0.482 และตัวแปรที่มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีจำนวน 2 คู่ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -0.107 และ -0.134 ส่วนตัวแปรที่ไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ มีจำนวน 11 คู่ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง -0.073 ถึง 0.090 โดยมีรายละเอียดแสดงในตารางที่ 18

ตารางที่ 18 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่าสหสัมพันธ์ของตัวแปรสังเกตได้ในโมเดล (n= 360)

ตัวแปร	GOAL	ABILI	CONSCI	MOTIV	ACCOU	STAND	COMPAR	ERROR	ACCUR
GOAL	1.000								
ABILI	0.348**	1.000							
CONSCI	0.190**	0.292**	1.000						
MOTIV	0.338**	0.362**	0.404**	1.000					
ACCOUN	0.177**	0.167**	0.296**	0.431**	1.000				
STANDA	0.144**	0.157**	0.314**	0.418**	0.419**	1.000			
COMPAR	0.192**	0.35**	0.359**	0.512**	0.481**	0.482**	1.000		
ERROR	0.079	-0.031	-0.134*	-0.073	-0.073	-0.271**	-0.107*	1.000	
ACCUR	-0.058	0.029	0.060	-0.011	0.029	0.090	0.072	-0.185**	1.000
M	3.294	3.686	36.831	18.233	20.692	21.864	15.783	0.510	0.686
SD	1.111	0.952	4.255	2.57	2.688	2.364	1.877	1.619	0.465

หมายเหตุ * p<.05 และ ** p< .01

3.2 ผลการตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

การตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานกับข้อมูลเชิงประจักษ์ ซึ่งประกอบด้วยตัวแปรแฝงทั้งหมด 3 ตัวแปร คือ บริบทของผู้ประเมิน (CONTEXT) วัดจากตัวแปรสังเกตได้ 5 ตัวแปร ได้แก่ เป้าหมาย (GOAL) ความสามารถ (ABILI) บุคลิกภาพแบบมีจิตสำนึก (CONSCI) แรงจูงใจ (MOTIV) และความรับผิดชอบที่ตรวจสอบได้ (ACCOU) การตัดสิน (JUDGE) วัดจากตัวแปรสังเกตได้ 2 ตัวแปร ได้แก่ การรับรู้มาตรฐาน (STAND) และกระบวนการเปรียบเทียบ (COMPAR) และคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) วัดจากตัวแปรสังเกตได้ 2 ตัวแปร ได้แก่ ความคลาดเคลื่อน (ERROR) และความแม่นยำ (ACCUR) ผลการวิเคราะห์ด้วยโปรแกรม LISREL 8.52 พบว่า โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานมีความสอดคล้องกับข้อมูลเชิงประจักษ์ โดยพิจารณาจากค่าสถิติที่ใช้ตรวจสอบความตรงของโมเดล ได้แก่ ค่า $\chi^2 = 23.121$, $df = 23$, $p = .454$ และมีค่า $\chi^2/df = 1.005$ ซึ่งมีค่าน้อยกว่า 2 จึงไม่ปฏิเสธสมมติฐานหลักที่ว่า โมเดลสมมติฐานตามทฤษฎีที่พัฒนาขึ้นสอดคล้องกับข้อมูลเชิงประจักษ์ และสอดคล้องกับผลการวิเคราะห์ค่าดัชนี GFI= 0.990, AGFI= 0.980, RMR=0.04 และ RMSEA=0.0038 ดังรายละเอียดแสดงในตารางที่ 19 และ 20 และภาพที่ 12

จากตารางที่ 19 เมื่อพิจารณาตัวแปรสังเกตได้ในโมเดลเชิงสาเหตุ พบว่า เกือบทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นบวกและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ยกเว้นตัวแปรความคลาดเคลื่อน (ERROR) ที่มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นลบไม่แตกต่างจากศูนย์ และทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานอยู่ระหว่าง -0.940 ถึง 0.765 โดยในองค์ประกอบด้านบริบทของผู้ประเมิน (CONTEXT) มีตัวแปรแรงจูงใจ (MOTIV) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และบุคลิกภาพแบบมีจิตสำนึก (CONSCI) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.757, 0.605, และ 0.538 ตามลำดับ

เมื่อพิจารณาองค์ประกอบด้านการตัดสิน (JUDGE) พบว่า ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงกว่าตัวแปรการรับรู้มาตรฐาน (STAND) ซึ่งมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.765 และ 0.630 ตามลำดับ

เมื่อพิจารณาองค์ประกอบด้านคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า ตัวแปรความคลาดเคลื่อน (ERROR) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นลบ คือ -0.940 ส่วนตัวแปรความแม่นยำ (ACCUR) ซึ่งมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นบวก คือ 0.197 ตามลำดับ

สำหรับค่าความเที่ยงในการวัดตัวแปรสังเกตได้แต่ละตัวพบว่า มีค่าอยู่ระหว่าง 0.039 ถึง 0.884ซึ่งมีค่าค่อนข้างต่ำไปจนถึงสูง โดยตัวแปรที่มีค่าความเที่ยงในการวัดสูงที่สุด คือ ความคลาดเคลื่อน (ERROR) รองลงมาคือ กระบวนการเปรียบเทียบ (COMPAR) แรงจูงใจ (MOTIV) และการรับรู้มาตรฐาน (STAND) โดยมีค่าความเที่ยงในการวัด คือ 0.884, 0.585, 0.573 และ 0.397 ตามลำดับ

ตารางที่ 19 ค่าประมาณพารามิเตอร์ และค่าสถิติที่เกี่ยวข้องในการตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานตามกรอบแนวคิด

ตัวแปร	น.น.องค์ประกอบ		t	R ²	สปส.คะแนน องค์ประกอบ	
	b (SE)	β				
LX (น้ำหนักองค์ประกอบตัวแปรสังเกตได้ภายนอก)						
CONTEXT	GOAL	0.528(0.077)	0.331	6.872**	0.110	0.046
	ABILI	0.673(0.088)	0.423	7.669**	0.179	0.057
	CONSCI	0.858(0.100)	0.538	8.591**	0.290	0.092
	MOTIV	1.205(0.135)	0.757	8.960**	0.573	0.230
	ACCOU	0.964(0.113)	0.605	8.566**	0.366	0.124
LY(น้ำหนักองค์ประกอบตัวแปรสังเกตได้ภายใน)						
JUDGE	STAND	1.000	0.630	-	0.397	0.133
	COMPAR	1.213(0.110)	0.765	11.005**	0.585	0.242
RATING	ERROR	-4.778(3.252)	-0.940	-1.469	0.884	-0.192
	ACCUR	1.000	0.197	-	0.039	0.003

หมายเหตุ ** p < .01

จากตารางที่ 20 เมื่อพิจารณาค่าสัมประสิทธิ์การพยากรณ์ของสมการโครงสร้างตัวแปร (R^2) พบว่า ตัวแปรการตัดสิน (JUDGE) และคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) มีค่าเท่ากับ 0.841 และ 0.069 ตามลำดับ แสดงว่า ตัวแปรสาเหตุภายในโมเดล คือ บริบทของผู้ประเมิน (CONTEXT) สามารถอธิบายความแปรปรวนของการตัดสิน (JUDGE) และคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ได้ถึงร้อยละ 84.1 และ 6.9 ตามลำดับ

เมื่อพิจารณาผลการวิเคราะห์อิทธิพลทางตรงและอิทธิพลทางอ้อมของตัวแปรคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) โดยพิจารณาจากค่าอิทธิพลในรูปแบบคะแนนมาตรฐาน พบว่า ตัวแปรบริบทของผู้ประเมิน (CONTEXT) มีขนาดอิทธิพลทางตรงในทิศทางบวกมีค่าเท่ากับ 0.099 อย่างไม่มีนัยสำคัญทางสถิติ และมีอิทธิพลทางอ้อมในทิศทางบวกมีค่าเท่ากับ 0.155 อย่างไม่มีนัยสำคัญทางสถิติ ส่วนตัวแปรการตัดสิน (JUDGE) มีขนาดอิทธิพลทางตรงในทิศทางบวกมีค่าเท่ากับ 0.169 อย่างไม่มีนัยสำคัญทางสถิติ

เมื่อพิจารณาผลการวิเคราะห์อิทธิพลรวมของตัวแปรในโมเดลที่เป็นปัจจัยเชิงสาเหตุของการตัดสิน (JUDGE) โดยพิจารณาจากค่าอิทธิพลในรูปแบบคะแนนมาตรฐาน พบว่า ตัวแปรบริบทของผู้ประเมิน (CONTEXT) มีขนาดอิทธิพลรวมในทิศทางบวกมีค่าเท่ากับ 0.917 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า บริบทของผู้ประเมินจะส่งผลทางตรงในทิศทางบวกต่อการตัดสินของผู้ประเมิน

เมื่อพิจารณาผลการวิเคราะห์อิทธิพลรวมของตัวแปรในโมเดลที่เป็นปัจจัยเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) โดยพิจารณาจากค่าอิทธิพลในรูปแบบคะแนนมาตรฐาน พบว่า ตัวแปรบริบทของผู้ประเมิน (CONTEXT) มีค่าอิทธิพลรวมสูงสุดในทิศทางบวกมีค่าเท่ากับ 0.254 โดยไม่มีนัยสำคัญทางสถิติ ส่วนตัวแปรการตัดสิน (JUDGE) มีขนาดอิทธิพลรวมในทิศทางบวกมีค่าเท่ากับ 0.169 อย่างไม่มีนัยสำคัญทางสถิติ

ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรแฝงทั้ง 3 ตัวแปร พบว่า ทุกคู่มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยบริบทของผู้ประเมิน (CONTEXT) มีความสัมพันธ์ทางบวกกับการตัดสิน (JUDGE) ในระดับสูง ($r = 0.917$) บริบทของผู้ประเมิน (CONTEXT) มีความสัมพันธ์ทางบวกกับคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING)

ในระดับต่ำ ($r = 0.254$) และการตัดสิน (JUDGE) มีความสัมพันธ์ทางบวกกับคุณภาพการให้คะแนน
ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ในระดับต่ำ ($r = 0.260$)

ตารางที่ 20 ค่าสถิติผลการวิเคราะห์อิทธิพลของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่
มีการเรียนรู้โดยใช้ปัญหาเป็นฐานตามกรอบแนวคิด

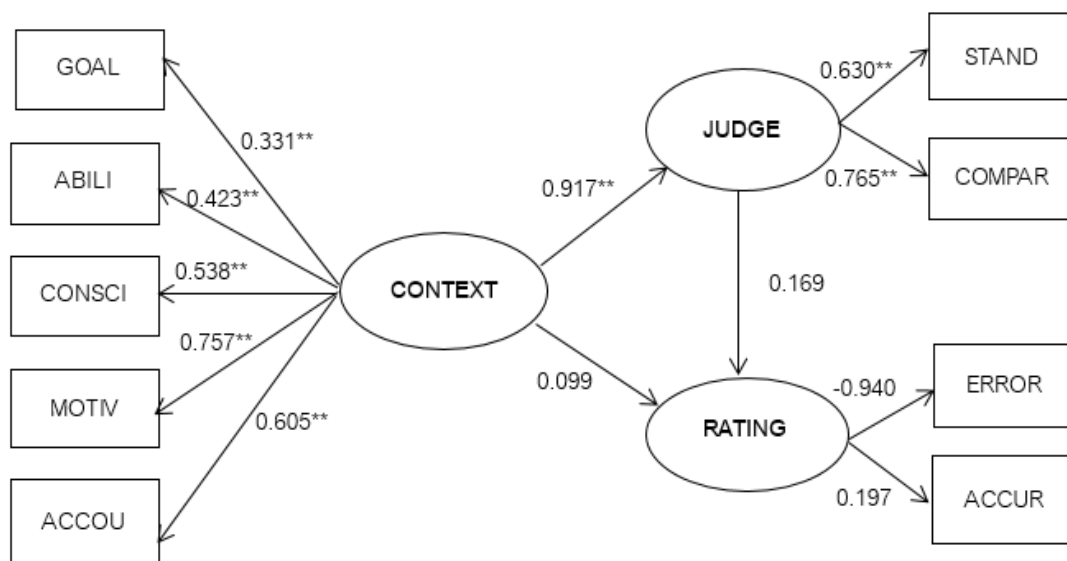
ตัวแปรสาเหตุ	สถิติ	ตัวแปรผล					
		การตัดสิน (JUDGE)			คุณภาพการให้คะแนนใน PBL (RATING)		
		DE	IE	TE	DE	IE	TE
บริบทของ ผู้ประเมิน (CONTEXT)	b	0.920	-	0.920	0.031	0.049	0.080
	β	0.917	-	0.917	0.099	0.155	0.254
	SE	0.106	-	0.106	0.132	0.135	0.045
	t	8.671**	-	8.671**	0.235	0.362	1.769
การตัดสิน (JUDGE)	b	-	-	-	0.053	-	0.053
	β	-	-	-	0.169	-	0.169
	SE	-	-	-	0.147	-	0.147
	t	-	-	-	0.359	-	0.359

Chi-square = 23.121; df =23; p = 0.454 ; GFI = 0.990; AGFI = 0.980; RMR =0.04; RMSEA = 0.0038

สมการโครงสร้างตัวแปร	CONTEXT	JUDGE	RATING
R^2	-	0.841	0.069

เมตริกซ์สหสัมพันธ์ระหว่างตัวแปรแฝง	CONTEXT	JUDGE	RATING
CONTEXT	1.000	-	-
JUDGE	0.917**	1.000	-
RATING	0.254**	0.260**	1.000

หมายเหตุ ** $p < .01$



ภาพที่ 12 โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

จากผลการตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานตามกรอบแนวคิดกับข้อมูลเชิงประจักษ์ สรุปได้ว่า โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ แต่เมื่อพิจารณาองค์ประกอบ คือ บริบทของผู้ประเมิน (CONTEXT) และการตัดสิน (JUDGE) พบว่า มีอิทธิพลต่อคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) อย่างไม่มีนัยสำคัญทางสถิติ ทั้งที่ค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานของตัวแปรสังเกตได้เกือบทั้งหมด มีค่าเป็นบวกและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ยกเว้นตัวแปรความคลาดเคลื่อน (ERROR) และความแม่นยำ (ACCUR) ที่มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานไม่แตกต่างจากศูนย์ ซึ่งไม่สอดคล้องกับกรอบแนวคิดในการวิจัยที่ได้นำเสนอในบทที่ 2 ดังนั้นผู้วิจัยจึงได้พัฒนาโมเดลทางเลือกดังมีรายละเอียดในหัวข้อถัดไป

3.3 โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา (โมเดลทางเลือก)

จากผลการตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานตามกรอบแนวคิดกับข้อมูลเชิงประจักษ์ในข้อ 3.2 ซึ่งไม่สอดคล้องกับกรอบแนวคิดในการวิจัย ดังนั้นผู้วิจัยจึงทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานอีกครั้ง พบว่า จากงานวิจัยของ Heidemeier & Moser (2009) ได้นำเสนอตัวแปรบริบทของผู้ประเมิน (CONTEXT) ที่ส่งผลต่อคุณภาพการให้คะแนน (RATING) โดยตรง และบริบทของผู้ประเมินจะวัดได้จากตัวแปรสังเกตได้ต่างๆ เช่น เป้าหมาย (GOAL) ความสามารถ (ABILI) บุคลิกภาพแบบมีจิตสำนึก (CONSCI) แรงจูงใจ (MOTIV) และความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และตัวแปรคุณภาพการให้คะแนน (RATING) จะวัดจากตัวแปรสังเกตได้ คือ การรับรู้มาตรฐาน (STAND) กระบวนการเปรียบเทียบ (COMPAR) และความคลาดเคลื่อน (ERROR) นอกจากนี้ Murphy & Cleveland (1995) ได้กล่าวว่า ความคลาดเคลื่อนของผู้ประเมินเป็นการตรวจสอบทางอ้อมของความแม่นยำในการให้คะแนน (indirect measures of rating accuracy หรือ inaccuracy) และการตรวจสอบความแม่นยำในการให้คะแนนเป็นการตรวจสอบทางตรง (direct measures of rating accuracy) หรือถ้าผลการวิเคราะห์ไม่พบความคลาดเคลื่อนของผู้ประเมินอาจกล่าวได้ว่าการให้คะแนนนั้นมีความแม่นยำ (rating accuracy) ซึ่งสอดคล้องกับหลักการคำนวณของโปรแกรม facets คือ ค่าของตัวแปรความแม่นยำ (ACCUR) มาจากการแปลงค่าของตัวแปรความคลาดเคลื่อน (ERROR) โดยโปรแกรมจะคำนวณค่าความคลาดเคลื่อนของผู้ประเมิน รวมถึงค่าความสามารถของผู้ถูกประเมิน และค่าความยากง่ายของข้อคำถามในหน่วยโลจิทก่อน จากนั้นจะมีการคำนวณค่ากำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากประมาณค่าของผู้ประเมิน ซึ่งเป็นค่า Infit Mean – Square หรือค่าความแม่นยำ (ACCUR) ดังนั้นจึงสามารถใช้ตัวแปรความคลาดเคลื่อน (ERROR) แทนตัวแปรความแม่นยำ (ACCUR) และในการปรับโมเดลทางเลือกจะใช้การจัดกลุ่มของตัวแปรแฝงภายในใหม่ ให้เหลือเพียงตัวแปรคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ซึ่งวัดจากตัวแปรสังเกตได้ 3 ตัวแปร ได้แก่ การรับรู้มาตรฐาน (STAND) กระบวนการเปรียบเทียบ (COMPAR) และความคลาดเคลื่อน (ERROR) ส่วนตัวแปรความแม่นยำ (ACCUR) ซึ่งเป็นตัวแปรสังเกตได้จะถูกลบออกจากโมเดล

การตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (โมเดลทางเลือก) ซึ่งประกอบด้วยตัวแปรแฝงทั้งหมด 2 ตัวแปร คือ บริบทของผู้ประเมิน (CONTEXT) วัดจากตัวแปรสังเกตได้ 5 ตัวแปร ได้แก่ เป้าหมาย (GOAL) ความสามารถ (ABILI) บุคลิกภาพแบบมีจิตสำนึก (CONSCI) แรงจูงใจ (MOTIV) และความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) วัดจากตัวแปรสังเกตได้ 3 ตัวแปร ได้แก่ การรับรู้มาตรฐาน (STAND) กระบวนการเปรียบเทียบ (COMPAR) และความคลาดเคลื่อน (ERROR) โดยใช้โปรแกรม LISREL 8.52 พบว่า โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ โดยพิจารณาจากค่าสถิติที่ใช้ตรวจสอบความตรงของโมเดล ได้แก่ ค่า $\chi^2 = 22.444$, $df = 19$, $p = .263$ และมีค่า $\chi^2/df = 1.181$ ซึ่งมีค่าน้อยกว่า 2 จึงไม่ปฏิเสธสมมติฐานหลักที่ว่า โมเดลสมมติฐานตามทฤษฎีที่พัฒนาขึ้นสอดคล้องกับข้อมูลเชิงประจักษ์ และสอดคล้องกับผลการวิเคราะห์ค่าดัชนี GFI= 0.985, AGFI= 0.971, RMR=0.035 และ RMSEA=0.0225 ดังรายละเอียดแสดงในตารางที่ 21 และ 22 และภาพที่ 13

จากตารางที่ 21 เมื่อพิจารณาตัวแปรสังเกตได้ในโมเดลเชิงสาเหตุ พบว่า ตัวแปรเกือบทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นบวกและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ยกเว้นตัวแปรความคลาดเคลื่อน (ERROR) ที่มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นลบและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานอยู่ระหว่าง -0.130 ถึง 0.773 โดยในองค์ประกอบด้านบริบทของผู้ประเมิน (CONTEXT) มีตัวแปรแรงจูงใจ (MOTIV) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และบุคลิกภาพแบบมีจิตสำนึก (CONSCI) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.759, 0.668, และ 0.516 ตามลำดับ

เมื่อพิจารณาองค์ประกอบด้านคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ การรับรู้มาตรฐาน (STAND) และความคลาดเคลื่อน (ERROR) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.773, 0.623, และ -0.130 ตามลำดับ

สำหรับค่าความเที่ยงในการวัดตัวแปรสังเกตได้แต่ละตัวพบว่า มีค่าอยู่ระหว่าง 0.017 ถึง 0.598 ซึ่งมีค่าค่อนข้างต่ำไปจนถึงปานกลาง โดยตัวแปรที่มีค่าความเที่ยงในการวัดสูงที่สุด คือ

กระบวนการเปรียบเทียบ (COMPAR) รองลงมาคือ แรงจูงใจ (MOTIV) และความรับผิดชอบที่ตรวจสอบได้ (ACCOU) โดยมีค่าความเที่ยงในการวัด คือ 0.598, 0.576, และ 0.447 ตามลำดับ

ตารางที่ 21 ค่าพารามิเตอร์ และค่าสถิติที่เกี่ยวข้องในการตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (โมเดลทางเลือก)

ตัวแปร	น.น.องค์ประกอบ		t	R ²	สปส.คะแนน องค์ประกอบ	
	b (SE)	β				
LX (น้ำหนักองค์ประกอบตัวแปรสังเกตได้ภายนอก)						
CONTEXT	GOAL	0.384(0.061)	0.351	6.293**	0.123	0.031
	ABILI	0.462(0.051)	0.494	8.999**	0.244	0.179
	CONSCI	2.197(0.229)	0.516	9.607**	0.267	0.028
	MOTIV	1.948(0.129)	0.759	15.085**	0.576	0.137
	ACCOU	1.794(0.142)	0.668	12.653**	0.447	0.110
LY (น้ำหนักองค์ประกอบตัวแปรสังเกตได้ภายใน)						
RATING	STAND	1.478	0.628	-	0.394	0.095
	COMPAR	1.451(0.142)	0.773	10.193**	0.598	0.208
	ERROR	-0.210(0.092)	-0.130	-2.275**	0.017	0.009

หมายเหตุ * $p < .05$ และ ** $p < .01$

จากตารางที่ 22 เมื่อพิจารณาค่าสัมประสิทธิ์การพยากรณ์ของสมการโครงสร้าง (R²) พบว่าตัวแปรคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) มีค่าเท่ากับ 0.809 แสดงว่า ตัวแปรสาเหตุภายในโมเดล คือ บริบทของผู้ประเมิน (CONTEXT) สามารถอธิบายความแปรปรวนของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ได้ถึงร้อยละ 80.9

เมื่อพิจารณาค่าอิทธิพลรวมของตัวแปรคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า ตัวแปรบริบทของผู้ประเมิน (CONTEXT) มีค่าอิทธิพลรวมเท่ากับ

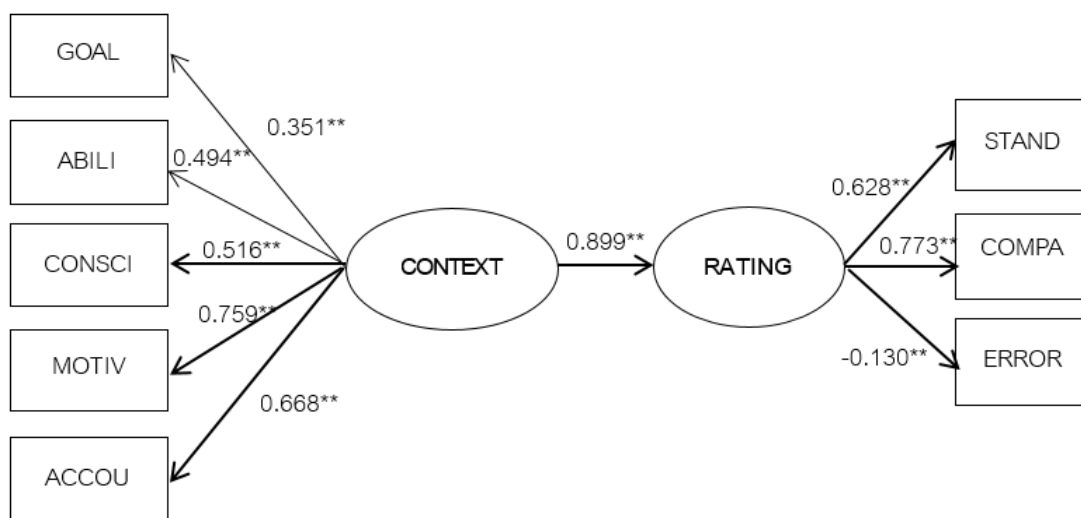
0.899 อย่างมีนัยสำคัญทางสถิติ ที่ระดับ .01 แสดงว่า บริบทของผู้ประเมินจะส่งผลทางตรงในทิศทางบวกต่อคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน

ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรแฝง 2 ตัวแปร คือ บริบทของผู้ประเมิน (CONTEXT) และคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่ามีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยมีค่าความสัมพันธ์เท่ากับ 0.899 แสดงว่า บริบทของผู้ประเมิน (CONTEXT) มีความสัมพันธ์ทางบวกกับคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ในระดับสูง ($r = 0.899, p = .000$)

ตารางที่ 22 ค่าสถิติผลการวิเคราะห์อิทธิพลของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (โมเดลทางเลือก)

ตัวแปรสาเหตุ	สถิติ	ตัวแปรผล		
		คุณภาพการให้คะแนนใน PBL (RATING)		
		DE	IE	TE
บริบทของผู้ประเมิน (CONTEXT)	b	0.899	-	0.899
	β	0.899	-	0.899
	SE	0.086	-	0.086
	t	10.439**	-	10.439**
Chi-square =22.444; df =19; p =.263; GFI =0.985; AGFI =0.971; RMR=0.035; RMSEA =0.0225				
สมการโครงสร้างตัวแปร		CONTEXT	RATING	
R ²		-	0.809	
เมตริกซ์สหสัมพันธ์ระหว่างตัวแปรแฝง		CONTEXT	RATING	
CONTEXT		1.000		
RATING		0.899**	1.000	

หมายเหตุ ** $p < .01$



ภาพที่ 13 โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา (โมเดลทางเลือก)

ตอนที่ 4 ผลการตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง

การวิเคราะห์ในส่วนนี้เป็นการวิเคราะห์เพื่อทดสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาที่พัฒนาและตรวจสอบความตรงของโมเดลจนมีความสอดคล้องกับข้อมูลเชิงประจักษ์ดีแล้ว ($\chi^2 = 22.444$, $df = 19$, $p = .263$, $GFI = 0.985$, $AGFI = 0.971$, $RMR = 0.035$ และ $RMSEA = 0.0225$) ทดสอบระหว่างกลุ่มผู้ประเมิน 3 กลุ่ม คือ อาจารย์ เพื่อน และตนเอง โดยนำเสนอแบ่งเป็น 2 ตอนย่อย ดังนี้

4.1 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวบ่งชี้ภายในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาจำแนกตามกลุ่มผู้ประเมิน

การนำเสนอผลการวิเคราะห์ในตอนนี้เป็นการศึกษาความสัมพันธ์สหสัมพันธ์ระหว่างตัวแปรในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานจำแนกตามกลุ่มผู้ประเมิน เพื่อให้ทราบลักษณะความสัมพันธ์ของตัวแปรในการพิจารณาความเหมาะสมของเมทริกซ์สหสัมพันธ์ โดยมีรายละเอียดดังนี้

4.1.1 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวบ่งชี้ภายในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาของผู้ประเมินกลุ่มอาจารย์

ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรสังเกตได้จำนวน 9 ตัวแปร รวม 28 คู่ ในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยใช้ค่าสหสัมพันธ์แบบเพียร์สัน พบว่า มีจำนวน 17 คู่ ที่มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง -0.255 ถึง 0.545 โดยคู่ที่มีความสัมพันธ์กันมากที่สุด คือ ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) กับ การรับรู้มาตรฐาน (STAND) มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.545 รองลงมา คือ กระบวนการเปรียบเทียบ (COMPAR) กับ ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.529 และตัวแปรที่มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีจำนวน 1 คู่ คือ เป้าหมาย (GOAL) กับ ความสามารถ (ABILI) ซึ่งมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.219 ส่วนตัวแปรที่ไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ มีจำนวน 10 คู่ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง 0.003 ถึง 0.168 โดยมีรายละเอียดแสดงในตารางที่ 23

ตารางที่ 23 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่าสหสัมพันธ์ของตัวแปรในผู้ประเมินกลุ่มอาจารย์ (n= 120)

ตัวแปร	GOAL	ABILI	CONSCI	MOTIV	ACCOU	STAND	COMPAR	ERROR
GOAL	1.000							
ABILI	0.219*	1.000						
CONSCI	0.003	0.360**	1.000					
MOTIV	0.256**	0.422**	0.375**	1.000				
ACCOU	0.094	0.258**	0.363**	0.462**	1.000			
STANDA	0.025	0.168	0.410**	0.474**	0.448**	1.000		
COMPAR	0.023	0.371**	0.422**	0.493**	0.529**	0.545**	1.000	
ERROR	0.091	-0.076	-0.255**	-0.008	0.018	-0.256**	-0.042	1.000
<i>M</i>	3.983	4.275	38.700	19.200	21.192	22.233	16.508	0.382
<i>SD</i>	0.917	0.860	4.283	2.630	2.677	2.340	1.983	1.583

หมายเหตุ * $p < .05$ และ ** $p < .01$

4.1.2 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวบ่งชี้ภายในโมเดลเชิงสาเหตุของ คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาของ ผู้ประเมินกลุ่มเพื่อน

ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรสังเกตได้จำนวน 9 ตัวแปร รวม 28 คู่ ในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยใช้ค่าสหสัมพันธ์แบบเพียร์สัน พบว่า มีจำนวน 11 คู่ ที่มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง 0.263 ถึง 0.523 โดยคู่ที่มีความสัมพันธ์กันมากที่สุด คือ ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) กับ การรับรู้มาตรฐาน (STAND) มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.523 รองลงมา คือ กระบวนการเปรียบเทียบ (COMPAR) กับ ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.475 และตัวแปรที่มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีจำนวน 4 คู่ และมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง 0.181 ถึง 0.234 ส่วนตัวแปรที่ไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ มีจำนวน 13 คู่ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง -0.002 ถึง 0.171 โดยมีรายละเอียดแสดงในตารางที่ 24

ตารางที่ 24 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่าสหสัมพันธ์ของตัวแปรในผู้ประเมินกลุ่มเพื่อน (n= 120)

ตัวแปร	GOAL	ABILI	CONSCI	MOTIV	ACCOU	STAND	COMPAR	ERROR
GOAL	1.000							
ABILI	0.124	1.000						
CONSCI	-0.002	0.181*	1.000					
MOTIV	0.263**	0.270**	0.350**	1.000				
ACCOUN	0.144	0.127	0.274**	0.417**	1.000			
STANDA	0.122	0.171	0.138	0.408**	0.451**	1.000		
COMPAR	0.124	0.157	0.234*	0.456**	0.475**	0.523**	1.000	
ERROR	0.025	-0.227*	-0.075	-0.085	-0.078	-0.234*	-0.272**	1.000
M	2.908	3.392	35.908	18.008	20.567	21.917	15.558	1.048
SD	1.037	0.802	3.658	2.310	2.595	2.266	1.846	1.408

หมายเหตุ * p<.05 และ ** p< .01

4.1.3 ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรชี้ภายในโมเดลเชิงสาเหตุของ คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาของ ผู้ประเมินกลุ่มตนเอง

ผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรสังเกตได้จำนวน 9 ตัวแปร รวม 28 คู่ ในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยใช้ค่าสหสัมพันธ์แบบเพียร์สัน พบว่า มีจำนวน 12 คู่ ที่มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง 0.263 ถึง 0.467 โดยคู่ที่มีความสัมพันธ์กันมากที่สุด คือ ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) กับ แรงจูงใจ (MOTIV) มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.467 รองลงมา คือ ความคลาดเคลื่อน (ERROR) กับ การรับรู้มาตรฐาน (STAND) มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -0.387 และตัวแปรที่มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 มีจำนวน 4 คู่ และมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง 0.220 ถึง 0.234 ส่วนตัวแปรที่ไม่มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ มีจำนวน 12 คู่ โดยมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง -0.006 ถึง 0.177 โดยมีรายละเอียดแสดงในตารางที่ 25

ตารางที่ 25 ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และค่าสหสัมพันธ์ของตัวแปรในผู้ประเมินกลุ่มตนเอง (n= 120)

ตัวแปร	GOAL	ABILI	CONSCI	MOTIV	ACCOU	STAND	COMPAR	ERROR
GOAL	1.000							
ABILI	0.234*	1.000						
CONSCI	0.173	0.012	1.000					
MOTIV	0.263**	0.160	0.327**	1.000				
ACCOUN	0.163	-0.006	0.177	0.360**	1.000			
STANDA	0.171	0.039	0.317**	0.322**	0.334**	1.000		
COMPAR	0.115	0.263**	0.220*	0.467**	0.386**	0.332**	1.000	
ERROR	-0.154	-0.103	-0.315**	-0.220*	-0.231*	-0.387**	-0.160	1.000
M	2.992	3.392	35.883	17.492	20.317	21.442	15.283	0.665
SD	1.041	0.910	4.197	2.473	2.738	2.435	1.567	1.312

หมายเหตุ * p<.05 และ ** p< .01

4.2 ผลการตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง

การวิเคราะห์ในตอนนี้มีจุดประสงค์เพื่อทดสอบสมมติฐานเกี่ยวกับความไม่แปรเปลี่ยนของโมเดลสมการโครงสร้างกลุ่มพหุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาตามตัวแปรกลุ่มผู้ประเมิน 3 กลุ่ม คือ ผู้ประเมินกลุ่มอาจารย์ จำนวน 120 คน ผู้ประเมินกลุ่มเพื่อน จำนวน 120 คน และผู้ประเมินกลุ่มตนเอง จำนวน 120 คน ในการทดสอบประกอบด้วยสมมติฐานเกี่ยวกับความไม่แปรเปลี่ยนของรูปแบบโมเดล และความไม่แปรเปลี่ยนของค่าพารามิเตอร์ในโมเดล ซึ่งในครั้งนี้มีสมมติฐานที่ทดสอบรวมทั้งสิ้น 2 สมมติฐาน ได้แก่

1. ความไม่แปรเปลี่ยนของรูปแบบโมเดล

$$H_0: \text{Form 1} = \text{Form 2} = \text{Form 3}$$

2. ความไม่แปรเปลี่ยนของสัมประสิทธิ์ถดถอยของตัวแปรสังเกตได้บนตัวแปรแฝงภายนอก (ΛX) และสมมติฐานที่ 1

$$H_0: \Lambda X (1) = \Lambda X (2) = \Lambda X (3)$$

ผลการวิเคราะห์เพื่อทดสอบความไม่แปรเปลี่ยนของโมเดลสมการโครงสร้างกลุ่มพหุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาระหว่างกลุ่มผู้ประเมิน 3 กลุ่ม ตามสมมติฐานดังกล่าวข้างต้น แสดงรายละเอียดในตารางที่ 26

ตารางที่ 26 ผลการทดสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมิน

สมมติฐาน	χ^2	df	χ^2/df	p	GFI	RMSEA	CFI	Diff. χ^2	Diff. df	p
1. รูปแบบ	34.804	54	0.645	.980	0.976	0.000	1.000	-	-	
2. LX = IN	61.441	64	0.960	.568	0.963	0.000	0.987	26.637	10	.003
3. LX, LY = IN	หยุดการทดสอบ									

จากตารางที่ 26 เมื่อพิจารณาผลการวิเคราะห์โมเดลสมการโครงสร้างกลุ่มพหุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาระหว่างกลุ่มผู้ประเมินในสมมติฐานที่ 1 ซึ่งเป็นการทดสอบความไม่แปรเปลี่ยนของรูปแบบโมเดล โดยไม่มีการกำหนดให้

ค่าพารามิเตอร์ระหว่างกลุ่มผู้ประเมินที่ต่างกันมีค่าเท่ากัน หรือก็คือการทดสอบความสอดคล้องของโมเดลกับข้อมูลเชิงประจักษ์ในแต่ละกลุ่มผู้ประเมิน ผลการทดสอบพบว่า ไม่ปฏิเสธสมมติฐานที่ 1 โดยพิจารณาจากค่า $\chi^2 = 34.804$, $df = 54$, $\chi^2/df = 0.645$, $p = .980$, $GFI = 0.976$, $RMSEA = 0.000$, และ $CFI = 1.000$ จากข้อมูลจะเห็นว่า ค่าไค-สแควร์ไม่มีนัยสำคัญ ค่าไค-สแควร์สัมพัทธ์ (χ^2/df) มีค่าน้อยกว่า 2 ค่า RMSEA เข้าใกล้ 0 และ ค่า CFI อยู่ในช่วง 0 ถึง 1 ซึ่งทุกค่าให้ผลที่สอดคล้องกัน จึงยอมรับสมมติฐานที่ 1 คือ โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ แสดงว่า รูปแบบของโมเดลมีความไม่แปรเปลี่ยนระหว่างกลุ่มผู้ประเมิน

ผลการทดสอบโมเดลตามสมมติฐานที่ 2 ซึ่งเป็นการทดสอบความไม่แปรเปลี่ยนของค่าพารามิเตอร์สัมประสิทธิ์ถดถอยของตัวแปรสังเกตได้บนตัวแปรแฝงภายนอก (LX) และสมมติฐานที่ 1 โดยกำหนดให้เมทริกซ์ LX มีค่าเท่ากันทั้ง 3 กลุ่มผู้ประเมิน ผลการทดสอบพบว่า ไม่ปฏิเสธสมมติฐานที่ 2 โดยพิจารณาจากค่า $\chi^2 = 61.441$, $df = 64$, $\chi^2/df = 0.960$, $p = .568$, $GFI = 0.963$, $RMSEA = 0.000$, และ $CFI = 0.987$ แสดงว่า การกำหนดเงื่อนไขโดยให้ค่าพารามิเตอร์สัมประสิทธิ์ถดถอยของตัวแปรสังเกตได้บนตัวแปรแฝงภายนอก (LX) มีค่าเท่ากันทั้ง 3 กลุ่มผู้ประเมินทำให้โมเดลยังคงมีความสอดคล้องกับข้อมูลเชิงประจักษ์ แต่เมื่อพิจารณาผลการทดสอบความแตกต่างของค่า χ^2 ระหว่างสมมติฐานที่ 2 กับสมมติฐานที่ 1 ซึ่งมีค่าเท่ากับ 26.637 ที่ $df = 10$ $p < 0.05$ พบว่า มีนัยสำคัญทางสถิติ และสอดคล้องกับผลต่างของค่า CFI ระหว่างสมมติฐานที่ 2 กับสมมติฐานที่ 1 โดยมีค่าเท่ากับ 0.013 ซึ่งมากกว่า 0.01 (Cheung & Rensvold, 2002) ดังนั้นจึงสรุปได้ว่า สัมประสิทธิ์ถดถอยของตัวแปรสังเกตได้บนตัวแปรแฝงภายนอก มีความแปรเปลี่ยนระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง จึงหยุดการทดสอบสมมติฐาน

ผลการทดสอบสมมติฐานเกี่ยวกับความไม่แปรเปลี่ยนของโมเดลสมการโครงสร้างกลุ่มพหุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาตามตัวแปรกลุ่มผู้ประเมิน 3 กลุ่ม คือ ผู้ประเมินกลุ่มอาจารย์ เพื่อน และตนเอง สรุปได้ว่า มีความไม่แปรเปลี่ยนของรูปแบบโมเดลระหว่างกลุ่มผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง แต่มีความแปรเปลี่ยนของค่าน้ำหนักองค์ประกอบของแต่ละตัวบ่งชี้ ดังนั้นในการนำเสนอผลการวิเคราะห์เพื่อทดสอบความไม่แปรเปลี่ยนของโมเดล ผู้วิจัยนำเสนอเฉพาะโมเดลที่มีความไม่แปรเปลี่ยนของรูปแบบโมเดล ซึ่งเป็นโมเดลที่ไม่มีเงื่อนไขกำหนดให้ค่าพารามิเตอร์ของโมเดลมีค่าเท่ากันระหว่าง 3 กลุ่มผู้ประเมิน

จากตารางที่ 27 เมื่อพิจารณาตัวแปรสังเกตได้ในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาจำแนกตามผู้ประเมินกลุ่มอาจารย์พบว่า ตัวแปรเกือบทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นบวกและแตกต่างกันจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ .01 ยกเว้นตัวแปรความคลาดเคลื่อน (ERROR) ที่มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นลบและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานอยู่ระหว่าง 0.151 ถึง 0.815 โดยองค์ประกอบด้านบริบทของผู้ประเมิน (CONTEXT) มีตัวแปรแรงจูงใจ (MOTIV) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และบุคลิกภาพแบบมีจิตสำนึก (CONSCI) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.721, 0.651 และ 0.604 ตามลำดับ เมื่อพิจารณาองค์ประกอบด้านคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ การรับรู้มาตรฐาน (STAND) และความคลาดเคลื่อน (ERROR) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.815, 0.705 และ -0.196 ตามลำดับ สำหรับค่าความเที่ยงในการวัดตัวแปรสังเกตได้แต่ละตัวพบว่ามีค่าอยู่ระหว่าง 0.023 ถึง 0.664 โดยตัวแปรที่มีค่าความเที่ยงในการวัดสูงที่สุด คือ กระบวนการเปรียบเทียบ (COMPAR) รองลงมาคือ แรงจูงใจ (MOTIV) และการรับรู้มาตรฐาน (STAND) โดยมีค่าความเที่ยงในการวัด คือ 0.664, 0.520, และ 0.497 ตามลำดับ

ผู้ประเมินกลุ่มเพื่อน พบว่า ตัวแปรเกือบทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นบวกและแตกต่างกันจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 01 ยกเว้นตัวแปรความคลาดเคลื่อน (ERROR) ที่มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นลบและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานอยู่ระหว่าง 0.234 ถึง 0.836 โดยในองค์ประกอบด้านบริบทของผู้ประเมิน (CONTEXT) มีตัวแปรแรงจูงใจ (MOTIV) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และบุคลิกภาพแบบมีจิตสำนึก (CONSCI) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.726, 0.655, และ 0.392 ตามลำดับ เมื่อพิจารณาองค์ประกอบด้านคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ การรับรู้มาตรฐาน (STAND) และความคลาดเคลื่อน (ERROR) โดยมีค่าน้ำหนักองค์ประกอบ

ในรูปคะแนนมาตรฐาน คือ 0.836, 0.690 และ -0.339 ตามลำดับ สำหรับค่าความเที่ยงในการวัดตัวแปรสังเกตได้แต่ละตัวพบว่า มีค่าอยู่ระหว่าง 0.055 ถึง 0.699 โดยตัวแปรที่มีค่าความเที่ยงในการวัดสูงที่สุด คือ กระบวนการเปรียบเทียบ (COMPAR) รองลงมาคือ แรงจูงใจ (MOTIV) และการรับรู้มาตรฐาน (STAND) โดยมีค่าความเที่ยงในการวัด คือ 0.699, 0.527 และ 0.476 ตามลำดับ

ผู้ประเมินกลุ่มตนเอง พบว่า ตัวแปรเกือบทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นบวกและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ยกเว้นตัวแปรความคลาดเคลื่อน (ERROR) ที่มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นลบและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานอยู่ระหว่าง 0.212 ถึง 0.656 โดยในองค์ประกอบด้านบริบทของผู้ประเมิน (CONTEXT) มีตัวแปรแรงจูงใจ (MOTIV) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และบุคลิกภาพแบบมีจิตสำนึก (CONSCI) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.656, 0.524 และ 0.465 ตามลำดับ เมื่อพิจารณาองค์ประกอบด้านคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ การรับรู้มาตรฐาน (STAND) และความคลาดเคลื่อน (ERROR) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.633, 0.592 และ -0.497 ตามลำดับ สำหรับค่าความเที่ยงในการวัดตัวแปรสังเกตได้แต่ละตัว พบว่า มีค่าอยู่ระหว่าง 0.110 ถึง 0.430 โดยตัวแปรที่มีค่าความเที่ยงในการวัดสูงที่สุด คือ แรงจูงใจ (MOTIV) รองลงมาคือ กระบวนการเปรียบเทียบ (COMPAR) และการรับรู้มาตรฐาน (STAND) โดยมีค่าความเที่ยงในการวัด คือ 0.430, 0.401 และ 0.351 ตามลำดับ

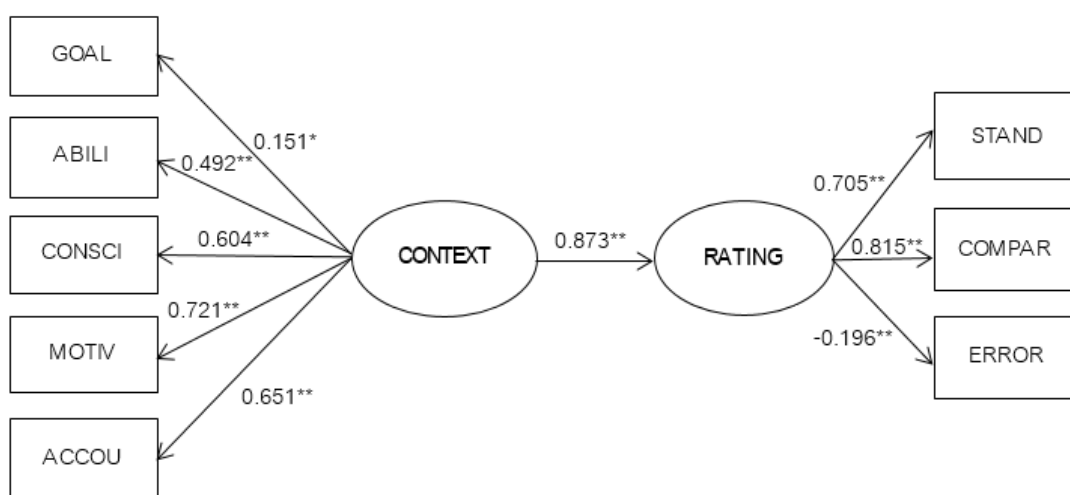
เมื่อพิจารณาค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานของตัวบ่งชี้โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาโดยเปรียบเทียบระหว่างกลุ่มผู้ประเมิน สำหรับองค์ประกอบที่ 1 ด้านบริบทของผู้ประเมิน (CONTEXT) ตัวบ่งชี้ที่มีน้ำหนักมากที่สุดของทั้ง 3 กลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง คือ ตัวบ่งชี้แรงจูงใจ (MOTIV) มีค่าน้ำหนักเท่ากับ 0.721, 0.726 และ 0.656 ตามลำดับ สำหรับตัวบ่งชี้ที่มีน้ำหนักน้อยที่สุดของผู้ประเมินกลุ่มอาจารย์และเพื่อน คือ ตัวบ่งชี้เป้าหมาย (GOAL) มีค่าน้ำหนักเท่ากับ 0.151 และ 0.234 ตามลำดับ และผู้ประเมินกลุ่มตนเอง คือ ตัวบ่งชี้ความสามารถ (ABILI) มีค่าน้ำหนักเท่ากับ 0.212 ส่วนองค์ประกอบที่ 2 ด้านคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหา

เป็นฐาน (RATING) เมื่อเปรียบเทียบระหว่าง 3 กลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง พบว่า ตัวบ่งชี้ที่มีน้ำหนักมากที่สุด คือ ตัวบ่งชี้กระบวนการเปรียบเทียบ (COMPAR) มีค่าน้ำหนักเท่ากับ 0.815, 0.836 และ 0.633 ตามลำดับ สำหรับตัวบ่งชี้ที่มีน้ำหนักน้อยที่สุด คือ ตัวบ่งชี้ความคลาดเคลื่อน (ERROR) มีค่าน้ำหนักเท่ากับ -0.196, -0.339 และ -0.497 ตามลำดับ

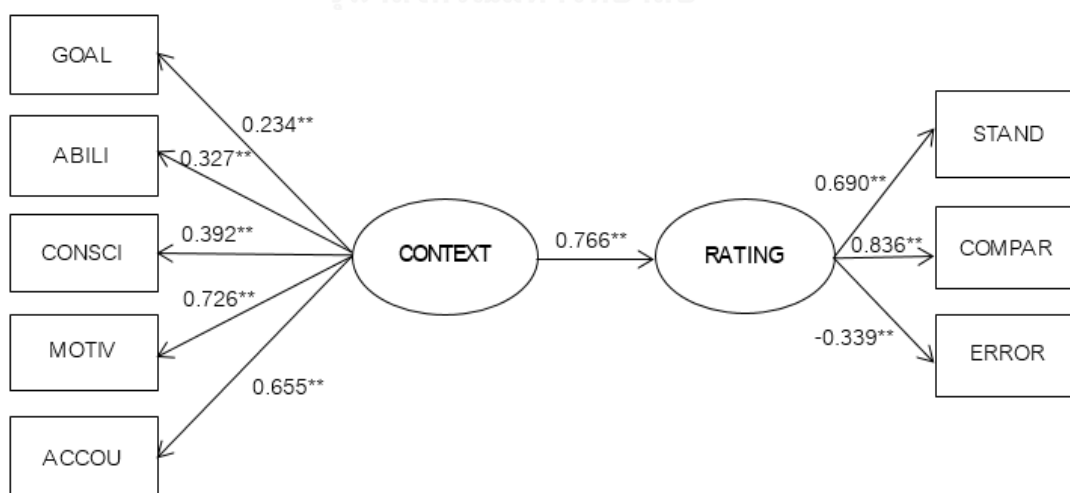
เมื่อพิจารณาค่าสัมประสิทธิ์การพยากรณ์ของสมการโครงสร้าง (R^2) ของตัวแปรแฝงภายใน คือ คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ของทั้ง 3 กลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง ซึ่งมีค่าเท่ากับ 0.761, 0.586 และ 0.939 ตามลำดับ แสดงว่า ตัวแปรสาเหตุภายในโมเดล คือ บริบทของผู้ประเมิน (CONTEXT) สามารถอธิบายความแปรปรวนของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ในกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ได้ร้อยละ 76.1, 58.6 และ 93.9 ตามลำดับ

จากผลการวิเคราะห์ค่าน้ำหนักองค์ประกอบของตัวบ่งชี้ในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาแต่ละกลุ่มผู้ประเมิน แสดงให้เห็นว่า ในองค์ประกอบด้านบริบทของผู้ประเมิน (CONTEXT) และด้านคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า มีความไม่แปรเปลี่ยนของค่าน้ำหนักองค์ประกอบของตัวบ่งชี้ในแต่ละกลุ่มผู้ประเมิน คือ องค์ประกอบด้านบริบทของผู้ประเมิน (CONTEXT) พบว่า ตัวบ่งชี้ที่มีน้ำหนักมากที่สุดของทั้ง 3 กลุ่มผู้ประเมิน คือ ตัวบ่งชี้แรงจูงใจ (MOTIV) ส่วนองค์ประกอบด้านคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า ตัวบ่งชี้ที่มีน้ำหนักมากที่สุด คือ ตัวบ่งชี้กระบวนการเปรียบเทียบ (COMPAR) และเมื่อพิจารณาค่าอิทธิพลทางตรงที่ส่งผลต่อคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ของแต่ละกลุ่มผู้ประเมิน พบว่า อิทธิพลจากบริบทของผู้ประเมิน (CONTEXT) มีผลต่อคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานค่อนข้างสูงทั้ง 3 กลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง โดยมีค่าอิทธิพลเท่ากับ 0.873, 0.766 และ 0.969 ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และจากผลการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรแฝง 2 ตัวแปร คือ บริบทของผู้ประเมิน (CONTEXT) และคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ของทั้ง 3 กลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง พบว่า มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยมีค่าความสัมพันธ์เท่ากับ 0.873, 0.766 และ 0.969 ตามลำดับ สรุปว่า ถ้าบริบทของผู้ประเมิน (CONTEXT) มีค่าสูง คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานจะมีค่าสูงขึ้นด้วย รายละเอียดแสดงในภาพที่ 14 - 16 และ

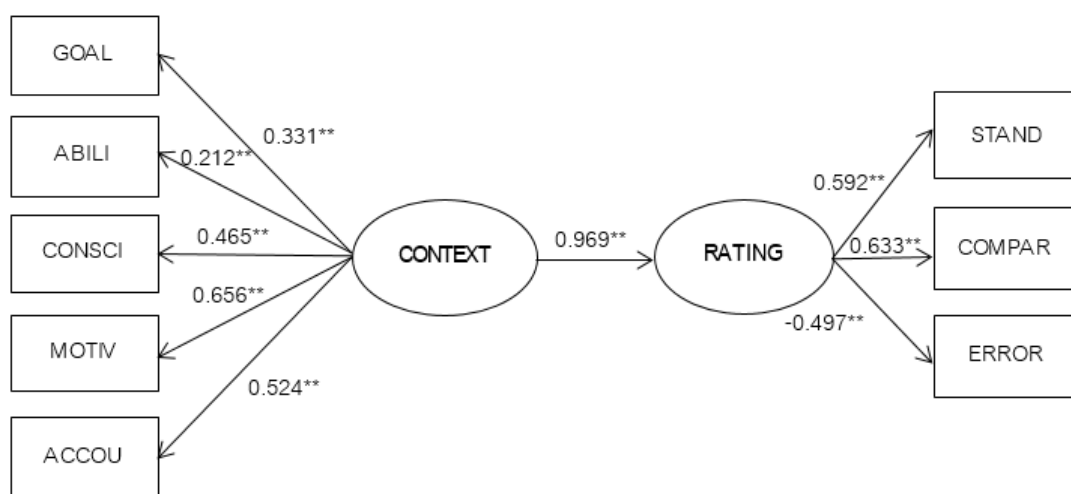
ตารางที่ 27 ซึ่งผลการศึกษาในส่วนนี้จะนำไปประยุกต์ใช้ในการพัฒนาโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ให้มีความเหมาะสมกับแต่ละกลุ่มผู้ประเมิน และสร้างเป็นคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา ซึ่งเป็นการศึกษาในระยะที่ 2 ต่อไป



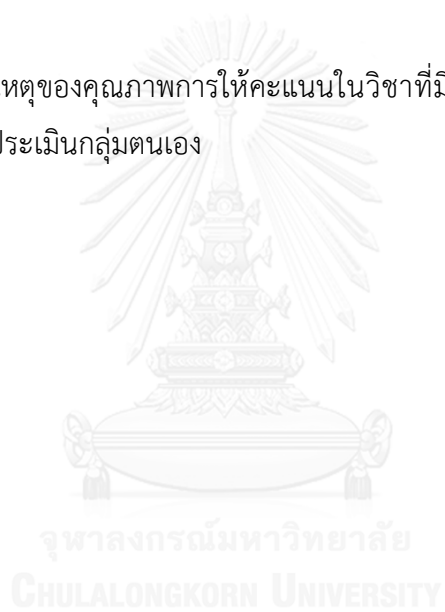
ภาพที่ 14 โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาในผู้ประเมินกลุ่มอาจารย์



ภาพที่ 15 โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาในผู้ประเมินกลุ่มเพื่อน



ภาพที่ 16 โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน
ในระดับอุดมศึกษาในผู้ประเมินกลุ่มตนเอง



ตารางที่ 27 ผลการวิเคราะห์ความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานใน ระดับอุดมศึกษา ระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อนและตนเอง

ค่าสถิติ $\chi^2 = 34.804$, $df = 54$, $\chi^2/df = 0.645$, $p = .980$, $GFI = 0.976$, และ $RMSEA = 0.000$

ค่าน้ำหนักองค์ประกอบของตัวแปรสังเกตได้

ตัวแปร	กลุ่มอาจารย์				กลุ่มเพื่อน				กลุ่มตนเอง				
	b (SE)	β	t	R ² สปส.	b (SE)	β	t	R ² สปส.	b (SE)	β	t	R ² สปส.	
LX (น้ำหนักองค์ประกอบตัวแปรสังเกตได้ภายนอก)													
CONTEXT	0.151 (0.053)		2.820*	0.023	0.030	0.234 (0.056)	4.153**	0.055	0.061	0.331 (0.063)	5.219**	0.110	0.092
ABILI	0.492 (0.054)	0.492	9.146**	0.242	0.127	0.327 (0.056)	5.847**	0.107	0.090	0.212 (0.062)	3.401**	0.045	0.055
CONSCI	0.604 (0.056)	0.604	10.814**	0.364	0.186	0.392 (0.062)	6.360**	0.154	0.113	0.465 (0.061)	7.597**	0.216	0.146
MOTIV	0.721 (0.055)	0.721	13.152**	0.520	0.293	0.726 (0.073)	9.974**	0.527	0.375	0.656 (0.073)	9.006**	0.430	0.284
ACCOU	0.651 (0.052)	0.651	12.469**	0.424	0.221	0.655 (0.060)	11.004**	0.430	0.281	0.524 (0.062)	8.508**	0.274	0.178

ตารางที่ 27 (ต่อ)

ตัวแปร	กลุ่มอาจารย์				กลุ่มเพื่อน				กลุ่มตนเอง			
	b (SE)	β	t	R ² <u>ลบ</u>	b (SE)	β	t	R ² <u>ลบ</u>	b (SE)	β	t	R ² <u>ลบ</u>
LY (น้ำหนักองค์ประกอบตัวแปรสังเกตได้ภายใน)												
RATING STAND	0.700 (0.050)	0.705	13.898**	0.497	0.904 (0.085)	0.690	10.607**	0.476	0.576 (0.066)	0.592	8.717**	0.351
COMPAR	0.810 (0.058)	0.815	13.860**	0.664	1.095 (0.092)	0.836	11.925**	0.699	0.616 (0.065)	0.633	9.480**	0.401
ERROR	-0.195 (0.054)	-0.196	-3.613**	0.038	-0.444 (0.076)	-0.339	-5.799**	0.115	-0.483 (0.062)	-0.497	-7.781**	0.247
ขนาดอิทธิพลจากตัวแปรสาเหตุไปตัวแปรผล												
RATING												
ตัวแปรสาเหตุ	กลุ่มอาจารย์				กลุ่มเพื่อน				กลุ่มตนเอง			
CONTEXT	DE	IE	TE	TE	DE	IE	TE	TE	DE	IE	TE	TE
	0.878	-	0.878	0.878	0.585	-	0.585	0.585	0.997	-	0.997	0.997
	0.873	-	0.873	0.873	0.766	-	0.766	0.766	0.969	-	0.969	0.969
SE	0.082	-	0.082	0.082	0.062	-	0.062	0.062	0.067	-	0.067	0.067
t	10.652**	-	10.652**	10.652**	9.409**	-	9.409**	9.409**	14.934**	-	14.934**	14.934**
ผู้ประเมิน	กลุ่มอาจารย์				กลุ่มเพื่อน				กลุ่มตนเอง			
ผลการโครงสร้างตัวแปร	RATING				RATING				RATING			
R ²	0.761				0.586				0.939			
เมทริกซ์สหสัมพันธ์	CONTEXT				CONTEXT				CONTEXT			
CONTEXT	1	-	-	-	1	-	-	-	1	-	-	-
RATING	0.873**	1	1	1	0.766**	1	1	1	0.969**	1	1	1

ตอนที่ 5 ผลการพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

การนำเสนอผลการพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา แบ่งออกเป็น 2 ตอนย่อย ดังนี้

5.1 ผลการพัฒนาโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

การพัฒนาโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาให้มีความเหมาะสมกับแต่ละกลุ่มผู้ประเมิน ผู้วิจัยได้ดำเนินการโดยนำผลการวิเคราะห์ข้อมูลในตอนต้นที่ 4 คือ โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ที่มีความไม่แปรเปลี่ยนของรูปแบบโมเดลระหว่างกลุ่มผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง แต่มีความแปรเปลี่ยนของค่าน้ำหนักองค์ประกอบของแต่ละตัวบ่งชี้

ค่าพารามิเตอร์ของตัวแปรสังเกตได้บนตัวแปรแฝงภายนอกระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง มาใช้ในการพัฒนาโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา และได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเพิ่มเติมในการพัฒนาต่อยอดในรายละเอียดของการใช้งานโมเดล เพื่อทำให้เกิดความชัดเจน เป็นรูปธรรม และได้ประโยชน์อย่างสูงสุดจากการใช้งานโมเดล โดยได้จัดทำเป็นคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา สำหรับอาจารย์ 1 ชุด และสำหรับนักศึกษา 1 ชุด

รายละเอียดของคู่มือสำหรับอาจารย์ ได้แก่ หลักการและวัตถุประสงค์ของคู่มือ แนวทางการปฏิบัติและเครื่องมือที่ใช้สำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย และแนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ ส่วนคู่มือสำหรับนักศึกษาจะมีรายละเอียดเฉพาะหัวข้อแนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ และแบบประเมินกระบวนการกลุ่ม ซึ่งทั้งสองหัวข้อนี้มีเนื้อหาเหมือนกับคู่มือสำหรับอาจารย์ ซึ่งมีรายละเอียดโดยสังเขปดังนี้ (รายละเอียดของคู่มือฉบับสมบูรณ์ดูที่ภาคผนวก ค)

แนวทางการปฏิบัติในการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานตามแนวคิดของ Moore & Poikela (2011) เช่น การประเมินไม่ควรจะมุ่งเน้นที่การได้ทำหรือไม่ได้ทำเนื่องจากไม่ช่วยให้เกิดการเสริมพลังแก่ผู้เรียน แต่ควรจะให้ความสำคัญในการประเมินเกี่ยวกับกับวิธีการและผลลัพธ์

ที่เกิดขึ้น ควรให้ผู้ที่มีส่วนได้ส่วนเสียมีส่วนร่วมในการประเมินอย่างครอบคลุมและมีการแลกเปลี่ยนความรู้ที่เกิดขึ้นจากการประเมิน การประเมินเพื่อปรับปรุงควรให้ความสำคัญกับกระบวนการในการเรียนและการประเมินเพื่อตัดสินควรให้ความสำคัญกับผลกระทบและประสิทธิภาพที่เกิดขึ้นจากการเรียน การประเมินควรเป็นไปอย่างเปิดเผยและซื่อสัตย์ซึ่งเกิดขึ้นได้เมื่อทุกคนในกลุ่มมีการปฏิบัติตามกฎ การประเมินที่มีประสิทธิภาพจะให้ข้อมูลป้อนกลับที่ต่อเนื่องเป็นประโยชน์ต่อการพัฒนา ปรับปรุง และช่วยในการตัดสินใจ การประเมินควรจะแทรกสอดและรวมอยู่ในทุกๆ กระบวนการของการเรียนรู้ โดยใช้ปัญหาเป็นฐาน และสัดส่วนของประเด็นที่ใช้ในการประเมินควรสอดคล้องกับกิจกรรมที่เกิดขึ้นจริง เป็นต้น

แนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ เช่น การพัฒนาเป้าหมายของผู้ประเมิน ซึ่งเกี่ยวข้องกับความคาดหวังของผู้ประเมินที่มีต่อกระบวนการและผลลัพธ์ของการประเมินที่มีคุณภาพ โดยมีแนวทางการปฏิบัติดังนี้ 1) ทำความเข้าใจเกี่ยวกับผลลัพธ์ของการประเมินที่มีคุณภาพ ได้แก่ ช่วยอธิบายจุดแข็งและโอกาสสำหรับการพัฒนาทั้งตัวผู้เรียนเองและระบบการเรียนการสอน และ 2) ทำความเข้าใจถึงบทบาทของการเป็นผู้ประเมินที่ดี ได้แก่ เข้าใจในวัตถุประสงค์และวิธีการประเมิน จัดหาข้อมูลที่ทันสมัย น่าเชื่อถือ และมีประโยชน์ เพื่อใช้ประกอบการประเมิน ยึดมั่นในหลักความยุติธรรมโดยแสดงตนเป็นกลางหรือเป็นเอกเทศต่อการประเมิน และหมั่นสังเกต ศึกษาเพิ่มเติม ตลอดจนแลกเปลี่ยนความรู้และมุมมองประสบการณ์ระหว่างผู้ประเมิน

คู่มือที่ได้พัฒนาขึ้นนี้ ผู้วิจัยได้นำไปให้ผู้ทรงคุณวุฒิ ซึ่งเป็นผู้ที่มีความรู้และประสบการณ์ในด้านการเรียนรู้โดยใช้ปัญหาเป็นฐาน จำนวน 4 คน ตรวจสอบความถูกต้องและความเหมาะสมของคู่มือสำหรับอาจารย์ และคู่มือสำหรับนักศึกษา โดยผลการตรวจสอบคู่มือสำหรับอาจารย์ พบว่ามีผู้เชี่ยวชาญที่ประเมินว่าคู่มือมีความถูกต้องและเหมาะสมทั้งในด้านโครงสร้างและเนื้อหาสาระของคู่มือจำนวน 3/4 คน และคู่มือสำหรับนักศึกษา พบว่า มีผู้เชี่ยวชาญที่ประเมินว่าคู่มือมีความถูกต้องและเหมาะสมทั้งในด้านโครงสร้างและเนื้อหาสาระของคู่มือ จำนวน 4 คน ดังนั้นแสดงว่ามีผู้เชี่ยวชาญมากกว่าครึ่งหนึ่งที่แสดงความเห็นด้วยกับคู่มือทั้ง 2 ชุดนี้ ส่วนผลการวิเคราะห์เนื้อหา (content analysis) เกี่ยวกับความคิดเห็น/ ข้อเสนอแนะอื่นๆ จะแสดงรายละเอียดไว้ในตารางที่ 28 ซึ่งผู้วิจัยได้นำข้อเสนอแนะเหล่านี้มาพิจารณาปรับใช้ในคู่มือ เพื่อการนำไปใช้ปฏิบัติจริงได้อย่างมีประสิทธิภาพ

ตารางที่ 28 แสดงความคิดเห็นเพิ่มเติมของผู้ทรงคุณวุฒิ (จำนวน 4 ท่าน)

ประเด็น ที่	ความคิดเห็นของ ผู้ทรงคุณวุฒิ	จำนวนคนที่ เห็นด้วย (n = 4)	เหตุผล	การปรับใช้ใน คู่มือ
1	ภาพรวมในด้านความถูกต้อง และความเหมาะสมของคู่มือ ● คู่มือสำหรับอาจารย์	3	ผู้ทรงที่ไม่เห็นด้วย ได้ให้ข้อมูลว่าควร ใช้คู่มือเดียวกันทั้ง อาจารย์และ นักศึกษา	-
	● คู่มือสำหรับนักศึกษา	4	-	-
2	ควรลดเนื้อหาของคู่มือทั้ง สองชุดลง	2	ไม่ใช่สิ่งที่อาจารย์ ผู้สอนซึ่งไม่ใช่ นักวิชาการทางการ ศึกษาจะสนใจ	พัฒนาเป็น คู่มือทั้งฉบับ สั้นและฉบับ สมบูรณ์
3	ผู้ประเมินต้องทำความเข้าใจ หรือศึกษาคู่มือก่อน	2	เพื่อสื่อสารให้ เข้าใจตรงกัน	เพิ่มใน คำแนะนำการ ใช้งานของคู่มือ

5.2 ผลการตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็น ฐานในระดับอุดมศึกษา

ผลการวิจัยในตอนนี้เป็นที่น่าเสนอผลการทดลองใช้คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่ม
ย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็น
ฐานระดับอุดมศึกษา (ฉบับสั้น) ซึ่งประกอบด้วย ชุดสำหรับอาจารย์ และชุดสำหรับนักศึกษา เนื่องจาก
ในขั้นเตรียมก่อนดำเนินการทดลอง ผู้วิจัยได้ติดต่อประสานงานกับต้นสังกัดของกลุ่มตัวอย่างเพื่อขอ
อนุมัติดำเนินการวิจัย โดยมีการอธิบายรายละเอียดเกี่ยวกับขั้นตอนการทำวิจัยพร้อมทั้งนำเครื่องมือ
ที่ใช้ในการวิจัยรวมถึงคู่มือทั้งฉบับสั้นและฉบับสมบูรณ์ประกอบการพิจารณา ซึ่งทางต้นสังกัดได้
อนุมัติให้ดำเนินการวิจัยโดยใช้คู่มือฉบับสั้นโดยให้เหตุผล คือ ผู้เรียนมีเวลาค่อนข้างจำกัด โดยผลการ
วิเคราะห์ในส่วนนี้จะแบ่งการนำเสนอออกเป็น 4 ตอนย่อย ดังนี้

5.2.1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของกลุ่มตัวอย่าง

จำนวนกลุ่มตัวอย่างซึ่งประกอบด้วยอาจารย์และนักศึกษา (ชั้นปี 3) ของสถาบันอุดมศึกษาแห่งหนึ่งที่จัดการสอนในรูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐาน แบ่งเป็นกลุ่มทดลองจำนวน 2 กลุ่ม PBL และกลุ่มควบคุมจำนวน 2 กลุ่ม PBL ซึ่งในแต่ละกลุ่มจะมีสมาชิก คือ อาจารย์ 1 คน และนักศึกษาประมาณ 11 - 12 คน (รายละเอียดดังตารางที่ 29) โดยมีช่วงเวลาในการเก็บข้อมูล 2 ครั้ง คือ 1) ภายหลังเสร็จสิ้นกระบวนการกลุ่มในวันแรกหรือวันเปิดโจทยสถานการณ์ ทั้งกลุ่มทดลองและกลุ่มควบคุมทำแบบสอบถามคุณลักษณะของผู้ประเมิน 1 ฉบับ จำนวน 45 ข้อ ใช้เวลาทำประมาณ 10 นาที จากนั้นเฉพาะกลุ่มทดลองจะได้รับการแจกคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย และแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา และผู้วิจัยจะอธิบายรายละเอียดเบื้องต้นในคู่มือแก่อาจารย์และผู้เรียนในกลุ่มทดลอง ส่วนกลุ่มควบคุมจะได้รับการแจกเฉพาะตัวอย่างแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย พร้อมคำอธิบายการใช้แบบประเมิน และ 2) ภายหลังเสร็จสิ้นกระบวนการกลุ่มในวันสุดท้ายหรือวันปิดโจทยสถานการณ์ ซึ่งมีระยะเวลาห่างจากครั้งที่ 1 เป็นเวลา 13 วัน ทั้งกลุ่มทดลองและกลุ่มควบคุมตอบแบบสอบถามคุณลักษณะของผู้ประเมิน 1 ฉบับ จำนวน 45 ข้อ โดยใช้เวลาประมาณ 10 นาที (เป็นแบบสอบถามชุดเดียวกับที่ใช้ในครั้งที่ 1) และแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย จำนวน 19 ข้อ ใช้เวลาประมาณ 20 นาที

ตารางที่ 29 จำนวนตัวอย่าง อัตราการตอบกลับและอัตราความครบถ้วนของแบบสอบถามที่ได้รับจากกลุ่มตัวอย่าง (ในแต่ละกลุ่มประกอบด้วยอาจารย์ 1 คน และนักศึกษา 11-12 คน)

กลุ่ม PBL	จำนวน (คน)			อัตราการตอบกลับ				อัตราความครบถ้วน			
	น.ศ.	อ.	รวม	น.ศ.	อ.	รวม	%	น.ศ.	อ.	รวม	%
1 (ทดลอง)	12	1	13	12	1	13	100	7	1	8	61.54
2 (ทดลอง)	11	1	12	11	1	12	100	7	1	8	66.67
3 (ควบคุม)	11	1	12	11	1	12	100	7	1	8	66.67
4 (ควบคุม)	11	1	12	10	1	11	91.67	7	1	8	66.67
รวม	45	4	49	44	4	48	97.96	28	4	32	65.31

จากตารางที่ 29 ในการเก็บรวบรวมข้อมูลผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างจำนวน 4 กลุ่ม PBL รวมจำนวน 49 คน และได้รับข้อมูลกลับคืนมาทั้งสิ้น 48 คน คิดเป็น

ร้อยละ 97.96 ถือได้ว่าเป็นอัตราการตอบกลับในระดับสูง เนื่องจากทั้งหมดเป็นกลุ่มตัวอย่างที่ให้ ความยินยอมในการเข้าร่วมการวิจัย ส่วนกลุ่มตัวอย่างที่ไม่อนุญาตให้ทำวิจัยจะไม่ได้ระบุไว้ และ แบบสอบถามที่ได้รับกลับคืนมาจากกลุ่มตัวอย่าง พบว่ามีความครบถ้วนของข้อมูลจำนวน 32 คน คิดเป็นร้อยละ 65.31

เมื่อพิจารณาตัวแปร เพศ วุฒิการศึกษา อายุ ระยะเวลาที่สอน PBL และเกรดเฉลี่ยสะสม พบว่า กลุ่มตัวอย่างซึ่งเป็นอาจารย์ ทั้งหมดเป็นเพศหญิง มีจำนวน 4 คน คิดเป็นร้อยละ 100.00 มีวุฒิ การศึกษาปริญญาโทและเอกอย่างละเท่าๆ กัน คือ กลุ่มละ 2 คน มีอายุเฉลี่ย 37.50 ปี ($SD=8.58$) และมีระยะเวลาที่สอน PBL เฉลี่ย 4 ปี ($SD=2.94$) ส่วนกลุ่มตัวอย่างซึ่งเป็นนักศึกษา ส่วนใหญ่เป็น เพศหญิง มีจำนวน 26 คน คิดเป็นร้อยละ 92.86 มีอายุเฉลี่ย 20.93 ปี ($SD=0.54$) และมีค่าเฉลี่ยของ เกรดเฉลี่ยสะสม 2.96 ($SD=0.33$) ดังรายละเอียดแสดงในตารางที่ 30 และ 31

ตารางที่ 30 ผลการวิเคราะห์ความถี่และร้อยละของกลุ่มตัวอย่างจำแนกตามตัวแปร เพศ และวุฒิ การศึกษา

ตัวแปร	อาจารย์ (n=4)		นักศึกษา (n=28)	
	จำนวน	ร้อยละ	จำนวน	ร้อยละ
เพศ				
ชาย	0	0	2	7.14
หญิง	4	100.00	26	92.86
รวม	4	100.00	28	100.00
วุฒิการศึกษา				
ปริญญาตรี	0	0	-	-
ปริญญาโท	2	50.00	-	-
ปริญญาเอก	2	50.00	-	-
รวม	4	100.00	-	-

ตารางที่ 31 ผลการวิเคราะห์ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของกลุ่มตัวอย่าง จำแนกตามตัวแปร อายุ ระยะเวลาที่สอน และGPAX

ตัวแปร	อาจารย์ (n=4)		นักศึกษา (n=28)	
	M	SD	M	SD
อายุ	37.50	8.58	20.93	0.54
ระยะเวลาที่สอน PBL (ปี)	4.00	2.94	-	-
GPAX (missing = 2 ราย)	-	-	2.96	0.33

ผลการวิเคราะห์ข้อมูลพื้นฐาน ได้แก่ คะแนนทั้งก่อนและหลังทดลองเปรียบเทียบระหว่างกลุ่มทดลองและกลุ่มควบคุม โดยใช้สถิติ Independent t-test พบว่า (1) ตัวแปรทั้ง 7 ตัวแปร ได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ ความรับผิดชอบที่ตรวจสอบได้ การรับรู้มาตรฐาน และกระบวนการเปรียบเทียบ มีคะแนนเฉลี่ยก่อนทดลองไม่แตกต่างกันระหว่างกลุ่มทดลองกับกลุ่มควบคุม ($p > .05$) (2) ตัวแปรทั้ง 7 ตัวแปร ได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ ความรับผิดชอบที่ตรวจสอบได้ การรับรู้มาตรฐาน และกระบวนการเปรียบเทียบ มีคะแนนเฉลี่ยหลังทดลองไม่แตกต่างกันระหว่างกลุ่มทดลองกับกลุ่มควบคุม ($p > .05$) และมีค่าขนาดอิทธิพล (effect size) อยู่ระหว่าง 0.048 ถึง 0.515 ซึ่งอยู่ในระดับน้อยจนถึงปานกลาง (Sullivan & Feinn, 2012) และ (3) ตัวแปรความคลาดเคลื่อนที่วัดเฉพาะหลังทดลอง พบว่า กลุ่มทดลองมีความคลาดเคลื่อนน้อยกว่ากลุ่มควบคุมอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (กลุ่มทดลอง $M = 1.43$, $SD = 1.18$ และกลุ่มควบคุม $M = 3.80$, $SD = 2.54$) และมีค่าขนาดอิทธิพลเท่ากับ 1.199 ซึ่งมีค่ามากกว่า 0.8 แสดงว่ามีขนาดอิทธิพลมาก (Sullivan & Feinn, 2012) ดังรายละเอียดแสดงในตารางที่ 32

ตารางที่ 32 ค่าสถิติจากการวิเคราะห์ข้อมูลพื้นฐานเปรียบเทียบระหว่างกลุ่มทดลองและกลุ่มควบคุม

ตัวแปร	กลุ่มทดลอง (n = 16)	กลุ่มควบคุม (n = 16)	t	df	Sig. (2-tailed)	d
คะแนนเป้าหมาย (ก่อนทดลอง)						
<i>M (SD)</i>	2.75 (1.00)	3.00 (1.15)	0.655	30	.518	-
Range	0 - 5	0 - 5				
คะแนนเป้าหมาย (หลังทดลอง)						
<i>M (SD)</i>	3.19 (1.38)	3.00 (1.10)	-0.426	30	.673	0.152
Range	0 - 5	0 - 5				
คะแนนความสามารถ (ก่อนทดลอง)						
<i>M (SD)</i>	4.06 (0.93)	3.75 (0.93)	-0.951	30	.349	-
Range	0 - 5	0 - 5				
คะแนนความสามารถ (หลังทดลอง)						
<i>M (SD)</i>	3.50 (1.1)	3.69 (0.79)	0.555	30	.583	0.198
Range	0 - 5	0 - 5				
คะแนนบุคลิกภาพแบบมีจิตสำนึก (ก่อนทดลอง)						
<i>M (SD)</i>	36.13 (5.46)	35.00 (5.34)	-0.589	30	.560	-
Range	10 - 50	10 - 50				
คะแนนบุคลิกภาพแบบมีจิตสำนึก (หลังทดลอง)						
<i>M (SD)</i>	35.13 (5.38)	33.88 (5.34)	-0.660	30	.514	0.233
Range	10 - 50	10 - 50				
คะแนนแรงจูงใจ (ก่อนทดลอง)						
<i>M (SD)</i>	18 (2.39)	16.13 (5.10)	-1.332	30	.193	-
Range	5 - 25	5 - 25				
คะแนนแรงจูงใจ (หลังทดลอง)						
<i>M (SD)</i>	18.44 (2.39)	17.19 (2.51)	-1.442	30	.160	0.510
Range	5 - 25	5 - 25				

ตารางที่ 32 (ต่อ)

ตัวแปร	กลุ่มทดลอง (n = 16)	กลุ่มควบคุม (n = 16)	t	df	Sig. (2-tailed)	d
คะแนนความรับผิดชอบที่ตรวจสอบได้ (ก่อนทดลอง)						
<i>M (SD)</i>	20.00 (2.22)	18.81 (5.60)	-0.788	30	.437	-
Range	5 - 25	5 - 25				
คะแนนความรับผิดชอบที่ตรวจสอบได้ (หลังทดลอง)						
<i>M (SD)</i>	20.06 (2.41)	20.19 (2.93)	0.132	30	.896	0.048
Range	5 - 25	5 - 25				
คะแนนการรับรู้มาตรฐาน (ก่อนทดลอง)						
<i>M (SD)</i>	21.94 (2.49)	20.13 (6.17)	-1.089	30	.285	-
Range	5 - 25	5 - 25				
คะแนนการรับรู้มาตรฐาน (หลังทดลอง)						
<i>M (SD)</i>	21.94 (1.95)	20.88 (2.16)	-1.462	30	.154	0.515
Range	5 - 25	5 - 25				
คะแนนกระบวนการเปรียบเทียบ (ก่อนทดลอง)						
<i>M (SD)</i>	15.25 (1.65)	14.75 (4.45)	-0.421	30	.677	-
Range	4 - 20	4 - 20				
คะแนนกระบวนการเปรียบเทียบ (หลังทดลอง)						
<i>M (SD)</i>	15.25 (2.08)	14.31 (1.89)	-1.335	30	.192	0.473
Range	4 - 20	4 - 20				
คะแนนความคลาดเคลื่อน (หลังทดลอง)						
<i>M (SD)</i>	0.59 (0.49)	1.58 (1.06)	3.386**	30	.002	1.199
(expected value of 0.00)						

หมายเหตุ ** p < .01

5.2.2 ผลการวิเคราะห์ความคลาดเคลื่อนในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง

ผลการวิเคราะห์ความคลาดเคลื่อนในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยใช้โมเดลหลายองค์ประกอบของราสช์ จากโปรแกรม Facets ในส่วนนี้จะนำเสนอแยกเป็น 4 กลุ่ม PBL โดยในแต่ละกลุ่ม PBL ประกอบด้วยกลุ่มผู้ประเมิน 3 กลุ่ม คือ อาจารย์ เพื่อน และตนเอง ซึ่งทำการประเมินในผู้รับการประเมินจำนวน 7 คน โดยใช้แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยที่มีลักษณะเป็นมาตรวัดประมาณค่า 4 ระดับ (rating scale) จำนวน 19 ข้อ โดยมีรูปแบบการวิเคราะห์ 4 ฟาเซต (facets) คือ ผู้ประเมิน (rater) * สถานะผู้ประเมิน (rater-type) * ผู้รับการประเมิน (ratee) * รายการประเมิน (items) ผลการวิเคราะห์ของแต่ละกลุ่ม PBL มีรายละเอียดดังนี้

5.2.2.1 กลุ่มที่ 1 (กลุ่มทดลอง)

ผลการวิเคราะห์ค่าสถิติทดสอบความสอดคล้อง (fit) ระหว่างข้อมูล กับ Rasch model พบว่า ข้อมูลชุดนี้มีความสอดคล้องกับ Rasch model เป็นอย่างดี โดยพิจารณาจาก ค่าเฉลี่ยของ StRes (mean of the standardized residuals) = -0.05 และ SD (sample standard deviation) = 1.16 ซึ่งถ้าค่าเฉลี่ยของ StRes มีค่าเข้าใกล้ 0.0 และ SD มีค่าเข้าใกล้ 1.0 จะแสดงว่าข้อมูลมีความสอดคล้องกับ Rasch model (Linacre, 2012) ส่วนผลการวิเคราะห์ความคลาดเคลื่อนชนิดกด/ปล่อยคะแนนในตารางที่ 33 และ 34 เมื่อเปรียบเทียบระหว่างกลุ่มผู้ประเมิน พบว่า ผู้ประเมินกลุ่มตนเองมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด รองลงมาคือ กลุ่มเพื่อน ส่วนกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน (Measure in Logits ของผู้ประเมินกลุ่มตนเอง เพื่อน และอาจารย์ มีค่าเท่ากับ 0.49, 0.16, และ -0.65 ตามลำดับ; ซึ่งค่าเป็นบวกจะแสดงถึงการกดคะแนน และค่าเป็นลบจะแสดงถึงการปล่อยคะแนน) นอกจากนี้เมื่อพิจารณาพฤติกรรมการให้คะแนนของผู้ประเมินแต่ละกลุ่มจะพบว่า ผู้ประเมินกลุ่มอาจารย์ให้คะแนนเฉลี่ยรายข้อมากกว่าผู้ประเมินกลุ่มเพื่อนและตนเอง (คะแนนเฉลี่ยรายข้อ = 3.43, 3.20 และ 3.14 ตามลำดับ) สอดคล้องกับผลการวิเคราะห์ในตารางที่ 35 นั่นคือ กลุ่มผู้ประเมินที่มีการใช้มาตรวัดคะแนนค่า 4 มากที่สุด คือ กลุ่มอาจารย์ คิดเป็นร้อยละ 42.86 รองลงมาคือ กลุ่มเพื่อน คิดเป็นร้อยละ 25.19 และกลุ่มตนเอง คิดเป็นร้อยละ 24.06 และเมื่อเปรียบเทียบเฉพาะในแต่ละกลุ่มผู้ประเมินพบว่า ผู้ประเมินทั้ง 3 กลุ่ม มีการใช้มาตรวัดคะแนนค่า 3 บ่อยที่สุด โดยกลุ่มตนเอง คิดเป็นร้อยละ 66.17 ส่วนกลุ่มอาจารย์ใช้จำนวน คิดเป็นร้อยละ 57.14 และกลุ่มเพื่อน คิดเป็นร้อยละ 69.30

ผลการวิเคราะห์ความแม่นยำในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน จากตารางที่ 34 เมื่อพิจารณาจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม คือ ตนเอง อาจารย์ และเพื่อน มีความแม่นยำในการให้คะแนนเท่ากับ 1.00, 1.00, และ 0.94 ตามลำดับ ซึ่งค่านี้อยู่ระหว่าง 0.6 ถึง 1.4 แสดงถึงการมีความแม่นยำในการให้คะแนน (Wright et al., 1994) ส่วนผลการวิเคราะห์อื่นๆ นอกจากเปรียบเทียบความคลาดเคลื่อนของผู้ประเมิน (คอลลัมน์ Rater-type) และเปรียบเทียบการใช้มาตรวัดคะแนน (คอลลัมน์ Scale) ที่แสดงในตารางที่ 33 มีดังนี้ 1) ความสามารถของผู้รับการประเมิน (คอลลัมน์ Ratee) พบว่า ผู้รับการประเมิน คนที่ 7 มีคะแนนสูงสุด รองลงมาคือ คนที่ 6 ส่วนผู้ที่มีคะแนนต่ำสุด มี 3 คน คือ คนที่ 2, 3, และ 4 และ 2) รายการประเมินข้อที่มีการให้คะแนนสูงสุด (คอลลัมน์ Item) คือ ข้อที่ 13 รองลงมาคือ ข้อที่ 14 ส่วนข้อที่มีการให้คะแนนต่ำสุด คือ ข้อที่ 17

จากผลการวิเคราะห์ผลการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของกลุ่มที่ 1 (ทดลอง) สรุปได้ว่า ผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน ส่วนผู้ประเมินกลุ่มตนเองมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มเพื่อน และเมื่อพิจารณาความแม่นยำในการให้คะแนนจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม มีการให้คะแนนที่มีความแม่นยำ จากผลทั้งสองอย่างนี้แสดงว่า ในภาพรวมผู้ประเมินทั้งสามกลุ่มให้คะแนนที่มีความคลาดเคลื่อนเพียงเล็กน้อย (Measure in Logits อยู่ระหว่าง -0.65 ถึง 0.49) และจัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้ (Infit Mean – Square อยู่ระหว่าง 0.94 ถึง 1.00)

ตารางที่ 33 แสดงการเปรียบเทียบระหว่างฟาเซท (กลุ่มที่ 1: กลุ่มทดลอง)

Measr	+Ratee	-Rater	-Rater-type	+Item	Scale
5				13	(4)
4				14	
3				15	---
2		104			
1				19	
	107 106		Self-rater Peer-rater	8	
* 0 *	101 105	*	*	11	* 3 *
	102 103 104		Tutor-rater	5 16 12	
-1		108 106		2 1 3 6 7 9 18	
-2		105		10 4	
		101 107		17	
-3		103 102			---
-4					(2)
Measr	+Ratee	-Rater	-Rater-type	+Item	Scale

ตารางที่ 34 ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน (กลุ่มที่ 1: กลุ่มทดลอง)

ผู้ประเมินคนที่	Sum of Ratings	Count of Ratings	Obsvd Average	Measure in Logits	Infit MnSq
ตนเอง (most severe)	418	133	3.14	0.49	1.00
เพื่อน	2551	798	3.20	0.16	1.00
อาจารย์ (most lenient)	456	133	3.43	-0.65	0.94
<i>M</i>	1141.67	354.67	3.26	0.00	0.98
<i>SD</i>	1220.67	383.94	0.15	0.59	0.03

mean of the "StRes" (Standardized Residuals) = -0.05, *SD* (sample standard deviation) = 1.16

ตารางที่ 35 ความถี่ของการให้คะแนนในแต่ละมาตรวัด (กลุ่มที่ 1: กลุ่มทดลอง)

กลุ่มผู้ประเมิน	มาตรวัด (score)							
	1		2		3		4	
	f	Used %	f	Used %	f	Used %	f	Used %
อาจารย์	0	0	0	0	76	57.1	57	42.9
เพื่อน	0	0	44	5.5	553	69.3	201	25.2
ตนเอง	0	0	13	9.8	88	66.2	32	24.1
รวม	0	0	57	15.3	717	192.6	290	92.1

5.2.2.2 กลุ่มที่ 2 (กลุ่มทดลอง)

ผลการวิเคราะห์ค่าสถิติทดสอบความสอดคล้อง (fit) ระหว่างข้อมูล กับ Rasch model พบว่า ข้อมูลชุดนี้มีความสอดคล้องกับ Rasch model เป็นอย่างดี โดยพิจารณาจาก ค่าเฉลี่ยของ StRes (mean of the standardized residuals) = 0.0 และ *SD* (sample standard deviation) = 1.01 ซึ่งถ้าค่าเฉลี่ยของ StRes มีค่าเข้าใกล้ 0.0 และ *SD* มีค่าเข้าใกล้ 1.0 จะแสดงว่าข้อมูลมีความสอดคล้องกับ Rasch model (Linacre, 2012) ส่วนผลการวิเคราะห์ความคลาดเคลื่อนชนิดกด/ปล่อยคะแนนในตารางที่ 36 และ 37 เมื่อเปรียบเทียบระหว่างกลุ่มผู้ประเมิน พบว่าผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด รองลงมา

คือ กลุ่มตนเอง ส่วนกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน (Measure in Logits ของผู้ประเมินกลุ่มอาจารย์ เพื่อน และตนเองมีค่าเท่ากับ 0.28, 0.04 และ -0.32 ตามลำดับ; ซึ่งค่าเป็นบวกจะแสดงถึงการกดคะแนน และค่าเป็นลบจะแสดงถึงการปล่อยคะแนน) นอกจากนี้เมื่อพิจารณาพฤติกรรมการให้คะแนนของผู้ประเมินแต่ละกลุ่มจะพบว่า ผู้ประเมินกลุ่มเพื่อนให้คะแนนเฉลี่ยรายข้อมากกว่าผู้ประเมินกลุ่มตนเองและอาจารย์ (คะแนนเฉลี่ยรายข้อ = 3.18, 3.10 และ 2.91 ตามลำดับ) สอดคล้องกับผลการวิเคราะห์ในตารางที่ 38 นั่นคือ กลุ่มผู้ประเมินที่มีการใช้มาตรวัดคะแนนค่า 4 มากที่สุด คือ กลุ่มเพื่อน คิดเป็นร้อยละ 39.97 รองลงมาคือ กลุ่มตนเอง คิดเป็นร้อยละ 33.83 และกลุ่มอาจารย์ คิดเป็นร้อยละ 19.55 และเมื่อเปรียบเทียบเฉพาะในแต่ละกลุ่มผู้ประเมินพบว่า ผู้ประเมินกลุ่มอาจารย์และกลุ่มตนเอง มีการใช้มาตรวัดคะแนนค่า 3 บ่อยที่สุด คิดเป็นร้อยละ 51.88 และ 42.11 ส่วนกลุ่มเพื่อนใช้มาตรวัดคะแนนค่า 4 บ่อยที่สุด คิดเป็นร้อยละ 39.97

ผลการวิเคราะห์ความแม่นยำในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน จากตารางที่ 37 เมื่อพิจารณาจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม คือ ตนเอง เพื่อน และอาจารย์ มีความแม่นยำในการให้คะแนนเท่ากับ 1.16, 1.03, และ 0.68 ตามลำดับ ซึ่งค่านี้อยู่ระหว่าง 0.6 ถึง 1.4 แสดงถึงการมีความแม่นยำในการให้คะแนน (Wright et al., 1994) ส่วนผลการวิเคราะห์อื่นๆ นอกจากเปรียบเทียบความคลาดเคลื่อนของผู้ประเมิน (คอลลัมภ์ Rater-type) และเปรียบเทียบการใช้มาตรวัดคะแนน (คอลลัมภ์ Scale) ที่แสดงในตารางที่ 36 มีดังนี้ 1) ความสามารถของผู้รับการประเมิน (คอลลัมภ์ Ratee) พบว่า ผู้รับการประเมินคนที่ 2 และคนที่ 6 มีคะแนนสูงสุด รองลงมาคือ คนที่ 4, 5, และ 7 ส่วนผู้ที่มีคะแนนต่ำสุด คือ คนที่ 1 และ 2) รายการประเมินข้อที่มีการให้คะแนนสูงสุด (คอลลัมภ์ Item) คือ ข้อที่ 13 และข้อที่ 14 รองลงมาคือ ข้อที่ 15 ส่วนข้อที่มีการให้คะแนนต่ำสุด คือ ข้อที่ 6

จากผลการวิเคราะห์ผลการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของกลุ่มที่ 2 (ทดลอง) สรุปได้ว่า ผู้ประเมินกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน ส่วนผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มตนเอง และเมื่อพิจารณาความแม่นยำในการให้คะแนนจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม มีการให้คะแนนที่มีความแม่นยำ จากผลทั้งสองอย่างนี้แสดงว่า ในภาพรวมผู้ประเมินทั้งสามกลุ่มให้คะแนนที่มี

ความคลาดเคลื่อนเพียงเล็กน้อย (Measure in Logits อยู่ระหว่าง -0.32 ถึง 0.28) และจัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้ (Infit Mean – Square อยู่ระหว่าง 0.68 ถึง 1.16)

ตารางที่ 36 แสดงการเปรียบเทียบระหว่างฟาเซท (กลุ่มที่ 2: กลุ่มทดลอง)

Measr	+Ratee	-Rater	-Rater-type	+Item	Scale
7				13 14	(4)
				15	
6					
5					
4					
3					
		205			
2		208			

1					
				16 18 19	
				12	
*	202 206	201	Tutor-rater		
0	204 205 207	206	Self-rater	2	3
	203		Peer-rater	1 5 10	
	201	207			
		202		11 17	
-1				3 7 8	
				4 9	
				6	---
-2		203			
		204			
-3					(2)
Measr	+Ratee	-Rater	-Rater-type	+Item	Scale

ตารางที่ 37 ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน (กลุ่มที่ 2: กลุ่มทดลอง)

ผู้ประเมินคนที่	Sum of Ratings	Count of Ratings	Obsvd Average	Measure in Logits	Infit MnSq
อาจารย์ (most severe)	387	133	2.91	0.28	0.68
ตนเอง	412	133	3.10	0.04	1.16
เพื่อน (most lenient)	2,534	798	3.18	-0.32	1.03
<i>M</i>	1,111.00	354.67	3.06	0.00	0.96
<i>SD</i>	1,232.42	383.94	0.14	0.30	0.25

mean of the "StRes" (Standardized Residuals) = 0.0, *SD* (sample standard deviation) = 1.01

ตารางที่ 38 ความถี่ของการให้คะแนนในแต่ละมาตรวัด (กลุ่มที่ 2: กลุ่มทดลอง)

กลุ่มผู้ประเมิน	มาตรวัด (score)							
	1		2		3		4	
	f	Used %	f	Used %	f	Used %	f	Used %
อาจารย์	0	0	38	28.57	69	51.88	26	19.55
เพื่อน	0	0	179	22.43	300	37.59	319	39.97
ตนเอง	0	0	32	24.06	56	42.11	45	33.83
รวม	0	0	249	75.06	425	131.58	390	93.36

5.2.2.3 กลุ่มที่ 3 (กลุ่มควบคุม)

ผลการวิเคราะห์ค่าสถิติทดสอบความสอดคล้อง (fit) ระหว่างข้อมูล กับ Rasch model พบว่า ข้อมูลชุดนี้มีความสอดคล้องกับ Rasch model เป็นอย่างดี โดยพิจารณาจาก ค่าเฉลี่ยของ StRes (mean of the standardized residuals) = -0.02 และ *SD* (sample standard deviation) = 1.07 ซึ่งถ้าค่าเฉลี่ยของ StRes มีค่าเข้าใกล้ 0.0 และ *SD* มีค่าเข้าใกล้ 1.0 จะแสดงว่าข้อมูลมีความสอดคล้องกับ Rasch model (Linacre, 2012) ส่วนผลการวิเคราะห์ความคลาดเคลื่อน

ชนิดกต/ปล่อยคะแนนในตารางที่ 39 และ 40 เมื่อเปรียบเทียบระหว่างกลุ่มผู้ประเมิน พบว่า ผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกตคะแนนมากที่สุด ส่วนกลุ่มตนเองมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนนมากที่สุด รองลงมาคือ กลุ่มเพื่อน (Measure in Logits ของผู้ประเมินกลุ่มอาจารย์ เพื่อน และตนเอง มีค่าเท่ากับ 1.31, -0.44 และ -0.87 ตามลำดับ; ซึ่งค่าเป็นบวกจะแสดงถึงการกตคะแนน และค่าเป็นลบจะแสดงถึงการปล่อยคะแนน) นอกจากนี้เมื่อพิจารณาพฤติกรรมการให้คะแนนของผู้ประเมินแต่ละกลุ่มจะพบว่า ผู้ประเมินกลุ่มตนเองให้คะแนนเฉลี่ยรายข้อมากกว่าผู้ประเมินกลุ่มเพื่อนและอาจารย์ (คะแนนเฉลี่ยรายข้อ = 3.62, 3.59 และ 2.58 ตามลำดับ) สอดคล้องกับผลการวิเคราะห์ในตารางที่ 41 นั่นคือ กลุ่มผู้ประเมินที่มีการใช้มาตรวัดคะแนนค่า 4 มากที่สุด คือ กลุ่มตนเอง คิดเป็นร้อยละ 69.17 รองลงมาคือ กลุ่มเพื่อน คิดเป็นร้อยละ 65.41 ส่วนกลุ่มอาจารย์ไม่มีการใช้มาตรวัดคะแนน 4 หรือคิดเป็นร้อยละ 0 และเมื่อเปรียบเทียบเฉพาะในแต่ละกลุ่มผู้ประเมินพบว่า ผู้ประเมินกลุ่มตนเองและกลุ่มเพื่อน มีการใช้มาตรวัดคะแนนค่า 4 บ่อยที่สุด คิดเป็นร้อยละ 69.17 และ 65.41 ตามลำดับ ส่วนกลุ่มอาจารย์ใช้มาตรวัดคะแนนค่า 3 บ่อยที่สุด คิดเป็นร้อยละ 63.91

ผลการวิเคราะห์ความแม่นยำในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน จากตารางที่ 40 เมื่อพิจารณาจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากค่าประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม คือ ตนเอง เพื่อน และอาจารย์ มีความแม่นยำในการให้คะแนนเท่ากับ 0.94, 0.98, และ 0.96 ตามลำดับ ซึ่งค่านี้อยู่ระหว่าง 0.6 ถึง 1.4 แสดงถึงการมีความแม่นยำในการให้คะแนน (Wright et al., 1994) ส่วนผลการวิเคราะห์อื่นๆ นอกจากเปรียบเทียบความคลาดเคลื่อนของผู้ประเมิน (คอลัมน์ Rater-type) และเปรียบเทียบการใช้มาตรวัดคะแนน (คอลัมน์ Scale) ที่แสดงในตารางที่ 39 มีดังนี้ 1) ความสามารถของผู้รับการประเมิน (คอลัมน์ Ratee) พบว่า ผู้รับการประเมินคนที่ 1 และคนที่ 6 มีคะแนนสูงสุด รองลงมาคือ คนที่ 4 และคนที่ 7 ส่วนผู้ที่มีคะแนนต่ำสุด คือ คนที่ 3 และ 2) รายการประเมินข้อที่มีการให้คะแนนสูงสุด (คอลัมน์ Item) คือ ข้อที่ 15 รองลงมาคือ ข้อที่ 13 และ 14 ส่วนข้อที่มีการให้คะแนนต่ำสุด คือ ข้อที่ 1

จากผลการวิเคราะห์ผลการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของกลุ่มที่ 3 (ควบคุม) สรุปได้ว่า ผู้ประเมินกลุ่มตนเองมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มเพื่อน ส่วนผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนชนิดกตคะแนน และเมื่อพิจารณาความแม่นยำในการให้คะแนนจากค่า Infit Mean – Square หรือค่ายก

กำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม มีการให้คะแนนที่มีความแม่นยำ จากผลทั้งสองอย่างนี้แสดงว่า ในภาพรวมผู้ประเมินทั้งสามกลุ่มให้คะแนนที่มีความคลาดเคลื่อนเพียงเล็กน้อย (Measure in Logits อยู่ระหว่าง -0.87 ถึง 1.31) และจัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้ (Infit Mean - Square อยู่ระหว่าง 0.94 ถึง 0.98)

ตารางที่ 39 แสดงการเปรียบเทียบระหว่างฟาเซท (กลุ่มที่ 3: กลุ่มควบคุม)

Measr	+Ratee	-Rater	-Rater-type	+Item	Scale
4	+	+	+	+	(4)

3	+	+	+	15	+
2	+	+	+	13 14	+
			Tutor-rater	16	3
1	+	+	+	17	+
	301 306			19	
	304 307				
* 0 *	* 302 *	*	*	* 8 9 10 *	* --- *
		306	Peer-rater	3	
	305		Self-rater	4 5 18	
-1	+ 303	+	+	+ 2 6 7 11 12 +	+
		301 305		1	2
-2	+	+	+	+	+
-3	+	+	+	+	---
-4	+	303	+	+	+
-5	+	302	+	+	+
		304	+	+	+
-6	+	+	+	+	+
		308	+	+	+
-7	+	307	+	+	(1)
Measr	+Ratee	-Rater	-Rater-type	+Item	Scale

ตารางที่ 40 ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน (กลุ่มที่ 3: กลุ่มควบคุม)

ผู้ประเมินคนที่	Sum of Ratings	Count of Ratings	Obsvd Average	Measure in Logits	Infit MnSq
อาจารย์ (most severe)	343	133	2.58	1.31	0.96
เพื่อน	2861	798	3.59	-0.44	0.98
ตนเอง (most lenient)	482	133	3.62	-0.87	0.94
<i>M</i>	1228.67	354.67	3.26	0.00	0.96
<i>SD</i>	1415.35	383.94	0.59	1.15	0.02

mean of the "StRes" (Standardized Residuals) = -0.02, *SD* (sample standard deviation) = 1.07

ตารางที่ 41 ความถี่ของการให้คะแนนในแต่ละมาตรวัด (กลุ่มที่ 3: กลุ่มควบคุม)

กลุ่มผู้ประเมิน	มาตรวัด (score)							
	1		2		3		4	
	f	Used %	f	Used %	f	Used %	f	Used %
อาจารย์	8	6.02	40	30.08	85	63.91	0	0.00
เพื่อน	0	0.00	55	6.89	221	27.69	522	65.41
ตนเอง	0	0.00	9	6.77	32	24.06	92	69.17
รวม	8	6.02	104	43.73	338	115.66	614	134.59

5.2.2.4 กลุ่มที่ 4 (กลุ่มควบคุม)

ผลการวิเคราะห์ค่าสถิติทดสอบความสอดคล้อง (fit) ระหว่างข้อมูล กับ Rasch model พบว่า ข้อมูลชุดนี้มีความสอดคล้องกับ Rasch model เป็นอย่างดี โดยพิจารณาจาก ค่าเฉลี่ยของ StRes (mean of the standardized residuals) = -0.02 และ *SD* (sample standard deviation) = 1.07 ซึ่งถ้าค่าเฉลี่ยของ StRes มีค่าเข้าใกล้ 0.0 และ *SD* มีค่าเข้าใกล้ 1.0 จะแสดงว่า ข้อมูลมีความสอดคล้องกับ Rasch model (Linacre, 2012) ส่วนผลการวิเคราะห์ความคลาดเคลื่อนชนิดกด/ปล่อยคะแนนในตารางที่ 42 และ 43 เมื่อเปรียบเทียบระหว่างกลุ่มผู้ประเมิน พบว่า ผู้ประเมิน

กลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด ส่วนกลุ่มตนเองและกลุ่มเพื่อนมีการให้คะแนนที่ใกล้เคียงกันและมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน (Measure in Logits ของผู้ประเมินกลุ่มอาจารย์ ตนเอง และเพื่อน มีค่าเท่ากับ 1.34, -0.64 และ -0.70 ตามลำดับ; ซึ่งค่าเป็นบวกจะแสดงถึงการกดคะแนน และค่าเป็นลบจะแสดงถึงการปล่อยคะแนน) นอกจากนี้เมื่อพิจารณาพฤติกรรมการให้คะแนนของผู้ประเมินแต่ละกลุ่มจะพบว่า ผู้ประเมินกลุ่มเพื่อนและกลุ่มตนเองให้คะแนนเฉลี่ยรายข้อมากกว่าผู้ประเมินกลุ่มตนเองและอาจารย์ (คะแนนเฉลี่ยรายข้อ = 3.71, 3.71 และ 2.23 ตามลำดับ) สอดคล้องกับผลการวิเคราะห์ในตารางที่ 44 นั่นคือกลุ่มผู้ประเมินที่มีการใช้มาตรวัดคะแนนค่า 4 มากที่สุด คือ กลุ่มตนเอง คิดเป็นร้อยละ 72.93 รองลงมาคือ กลุ่มเพื่อน คิดเป็นร้อยละ 72.18 และกลุ่มอาจารย์ คิดเป็นร้อยละ 12.78 ตามลำดับ และเมื่อเปรียบเทียบเฉพาะในแต่ละกลุ่มผู้ประเมินพบว่า ผู้ประเมินกลุ่มตนเองและกลุ่มเพื่อน มีการใช้มาตรวัดคะแนนค่า 4 บ่อยที่สุด คิดเป็นร้อยละ 72.93 และ 72.18 ส่วนกลุ่มอาจารย์ใช้มาตรวัดคะแนนค่า 2 บ่อยที่สุด คิดเป็นร้อยละ 34.59

ผลการวิเคราะห์ความแม่นยำในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน จากตารางที่ 43 เมื่อพิจารณาจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐาน ซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม คือ ตนเอง เพื่อน และอาจารย์ มีความแม่นยำในการให้คะแนนเท่ากับ 0.77, 0.97, และ 0.98 ตามลำดับ ซึ่งค่านี้อยู่ระหว่าง 0.6 ถึง 1.4 แสดงถึงการมีความแม่นยำในการให้คะแนน (Wright et al., 1994) ส่วนผลการวิเคราะห์อื่นๆ นอกจากเปรียบเทียบความคลาดเคลื่อนของผู้ประเมิน (คอลัมน์ Rater-type) และเปรียบเทียบการใช้มาตรวัดคะแนน (คอลัมน์ Scale) ที่แสดงในตารางที่ 42 มีดังนี้ 1) ความสามารถของผู้รับการประเมิน (คอลัมน์ Ratee) พบว่า ผู้รับการประเมินคนที่ 5 มีคะแนนสูงสุด รองลงมาคือ คนที่ 2 ส่วนผู้ที่มีคะแนนต่ำสุด คนที่ 3 และคนที่ 7 และ 2) รายการประเมินข้อที่มีการให้คะแนนสูงสุด (คอลัมน์ Item) คือ ข้อที่ 14 รองลงมาคือ ข้อที่ 13 และ 15 ส่วนข้อที่มีการให้คะแนนต่ำสุด คือ ข้อที่ 1

จากผลการวิเคราะห์ผลการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของกลุ่มที่ 4 (ควบคุม) สรุปได้ว่า ผู้ประเมินกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มตนเอง ส่วนผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด และเมื่อพิจารณาความแม่นยำในการให้คะแนนจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนน

จากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม มีการให้คะแนนที่มีความแม่นยำ จากผลทั้งสองอย่างนี้แสดงว่า ในภาพรวมผู้ประเมินทั้งสามกลุ่มให้คะแนนที่มีความคลาดเคลื่อนเพียงเล็กน้อย (Measure in Logits อยู่ระหว่าง -0.70 ถึง 1.34) และจัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้ (Infit Mean – Square อยู่ระหว่าง 0.77 ถึง 0.98)

ตารางที่ 42 แสดงการเปรียบเทียบระหว่างฟาเซท (กลุ่มที่ 4: กลุ่มควบคุม)

Measr	+Ratee	-Rater	-Rater-type	+Item	Scale
5					(4)
4				14	
3					---
2				13 15	
1	405		Tutor-rater	19	3
0	402 404			10 9 8	
	401 406 403 407	401	Peer-rater Self-rater	7 18 4 6 16	*
-1		402		17 2	
-2		405		11 12 3	2
-3		403		1	---
-4		404			
-5		406 407 408			(1)
Measr	+Ratee	-Rater	-Rater-type	+Item	Scale

ตารางที่ 43 ความคลาดเคลื่อนและความแม่นยำของกลุ่มผู้ประเมิน (กลุ่มที่ 4: กลุ่มควบคุม)

ผู้ประเมินคนที่	Sum of Ratings	Count of Ratings	Obsvd Average	Measure in Logits	Infit MnSq
อาจารย์ (most severe)	296	133	2.23	1.34	0.98
ตนเอง	494	133	3.71	-0.64	0.77
เพื่อน (most lenient)	2957	798	3.71	-0.70	0.97
<i>M</i>	1249.00	354.67	3.22	0.00	0.91
<i>SD</i>	1482.48	383.94	0.85	1.16	0.12

mean of the "StRes" (Standardized Residuals) = -0.02, *SD* (sample standard deviation) = 1.07

ตารางที่ 44 ความถี่ของการให้คะแนนในแต่ละมาตรวัด (กลุ่มที่ 4: กลุ่มควบคุม)

กลุ่มผู้ประเมิน	มาตรวัด (score)							
	1		2		3		4	
	f	Used %	f	Used %	f	Used %	f	Used %
อาจารย์	37	27.82	46	34.59	33	24.81	17	12.78
เพื่อน	0	0.00	13	1.63	209	26.19	576	72.18
ตนเอง	0	0.00	2	1.50	34	25.56	97	72.93
รวม	37	27.82	61	37.72	276	76.57	690	157.89

5.2.3 ผลการเปรียบเทียบคะแนนหลังทดลองของแต่ละตัวแปรสาเหตุในโมเดล คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา (จำนวน 5 ตัวแปร) ระหว่างกลุ่มควบคุมและกลุ่มทดลองโดยมีคะแนนก่อนทดลองเป็นตัวแปรร่วม

ผลการตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนร่วม (Analysis of covariance; ANCOVA) ได้แก่ (1) การแจกแจงของคะแนนก่อนทดลองและคะแนนหลังทดลองของตัวแปร ได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบ ที่ตรวจสอบได้ ทดสอบด้วย Kolmogorov-Smirnov Test พบว่า ทุกตัวแปรมีการแจกแจงเป็นแบบปกติ ($p > .05$) (2) ความแปรปรวนของคะแนนหลังทดลองในแต่ละกลุ่มเท่ากัน (homogeneity of variance) ทดสอบด้วย Levene's test พบว่า เกือบทุกตัวแปรมีความแปรปรวนในแต่ละกลุ่มเท่ากัน

($p > .05$) ยกเว้นตัวแปรความสามารถ แต่เมื่อพิจารณาสัดส่วนความแปรปรวนพบว่า มีค่าไม่เกิน 2.00 ดังนั้นจึงสรุปได้ว่า มีความแปรปรวนไม่แตกต่างกัน (Field, 2013) และ (3) ความสัมพันธ์ระหว่าง covariate กับตัวแปรตามต้องเหมือนกันทุกกลุ่ม (homogeneity of regression slopes) ทดสอบด้วยการวิเคราะห์ interaction ระหว่าง covariate กับตัวแปรอิสระ โดยผลต้องไม่ต่างกัน ($p > .05$) ซึ่งผลการทดสอบนี้พบว่า ความสัมพันธ์ระหว่าง covariate กับตัวแปรตามมีรูปแบบเดียวกันในทุกกลุ่มของตัวแปรอิสระ ($p > .05$) ยกเว้นตัวแปรแรงจูงใจ ($p = 0.028$) ดังนั้นสรุปได้ว่า ตัวแปรทั้ง 4 ตัวแปรได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก และความรับผิดชอบที่ตรวจสอบได้ สามารถวิเคราะห์โดยใช้สถิติ ANCOVA ได้ ยกเว้นตัวแปรแรงจูงใจ ซึ่งต้องใช้สถิติ Independent t-test ในการเปรียบเทียบคะแนนก่อนและหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลอง และใช้สถิติ Paired sample t-test เพื่อเปรียบเทียบคะแนนก่อนและหลังทดลองภายในกลุ่มควบคุมและกลุ่มทดลอง ดังมีรายละเอียดของผลการทดสอบดังนี้

ตารางที่ 45 ผลการวิเคราะห์ความแปรปรวนร่วมของคะแนนหลังทดลอง ระหว่างกลุ่มควบคุมและกลุ่มทดลองโดยมีคะแนนก่อนทดลองเป็นตัวแปรร่วม

แหล่งของความแปรปรวน	Sum of squares	df	Mean Square	F	Sig.
คะแนนเป้าหมายก่อนทดลอง	14.788	1	14.788	13.549	.001
กลุ่ม (ควบคุม/ทดลอง)	0.966	1	0.966	0.885	.355
Error	31.650	29	1.091		
Total	353.000	32			
R Squared = 0.323 (Adjusted R Squared = 0.276)					
คะแนนความสามารถก่อนทดลอง	12.841	1	12.841	25.512	.000
กลุ่ม (ควบคุม/ทดลอง)	1.289	1	1.289	2.561	.120
Error	14.597	29	0.503		
Total	441.000	32			
R Squared = 0.473 (Adjusted R Squared = 0.437)					

ตารางที่ 45 (ต่อ)

แหล่งของความแปรปรวน	Sum of Squares	df	Mean Square	F	Sig.
คะแนนบุคลิกภาพแบบมีจิตสำนึกก่อนทดลอง	745.030	1	745.030	185.506	.000
กลุ่ม (ควบคุม/ทดลอง)	0.357	1	0.357	0.089	.768
Error	116.470	29	4.016		
Total	38962.000	32			
R Squared = 0.867 (Adjusted R Squared = 0.858)					
คะแนนความรับผิดชอบที่ตรวจสอบได้ก่อนทดลอง	79.896	1	79.896	17.102	.000
กลุ่ม (ควบคุม/ทดลอง)	2.636	1	2.636	0.564	.459
Error	135.479	29	4.672		
Total	13176.000	32			
R Squared = 0.371 (Adjusted R Squared = 0.328)					

จากตารางที่ 45 ผลการวิเคราะห์ความแปรปรวนร่วมของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองโดยมีคะแนนก่อนทดลองเป็นตัวแปรร่วม พบว่า (1) ตัวแปรเป้าหมายมีค่าเฉลี่ยของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองไม่แตกต่างกัน ภายหลังจากมีการควบคุมคะแนนก่อนทดลองแล้ว $F(1,29) = 0.885$, $p = .355$ ส่วนคะแนนก่อนทดลองมีความสัมพันธ์กับคะแนนหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนหลังทดลองได้ร้อยละ 27.6 (2) ตัวแปรความสามารถ มีค่าเฉลี่ยของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองไม่แตกต่างกัน ภายหลังจากมีการควบคุมคะแนนก่อนทดลองแล้ว $F(1,29) = 2.561$, $p = .120$ ส่วนคะแนนก่อนทดลองมีความสัมพันธ์กับคะแนนหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนหลังทดลองได้ร้อยละ 43.7 (3) ตัวแปรบุคลิกภาพแบบมีจิตสำนึก มีค่าเฉลี่ยของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองไม่แตกต่างกัน ภายหลังจากมีการควบคุมคะแนนก่อนทดลองแล้ว $F(1,29) = 0.089$, $p = .768$ ส่วนคะแนนก่อนทดลองมีความสัมพันธ์กับคะแนนหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนหลังทดลองได้ร้อยละ 85.8 และ (4) ตัวแปรความรับผิดชอบที่ตรวจสอบได้ มีค่าเฉลี่ยของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองไม่แตกต่างกัน ภายหลังจากมีการควบคุมคะแนนก่อนทดลองแล้ว $F(1,29) = 0.564$, $p =$

.459 ส่วนคะแนนก่อนทดลองมีความสัมพันธ์กับคะแนนหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนหลังทดลองได้ร้อยละ 32.8

จากที่กล่าวมาข้างต้นว่า ตัวแปรแรงจูงใจก่อนทดลองไม่เป็นไปตามข้อตกลงเบื้องต้นสำหรับการใช้เป็นตัวแปรร่วม (covariate) ในการวิเคราะห์สถิติ ANCOVA ที่ว่าความสัมพันธ์ระหว่างตัวแปรร่วมกับตัวแปรตามต้องเหมือนกันทุกกลุ่ม (homogeneity of regression slopes) เนื่องจากผลการวิเคราะห์ interaction ระหว่างตัวแปรร่วมกับตัวแปรตามมีความแตกต่างกันระหว่างกลุ่ม ($p = 0.028$) ดังนั้นจึงใช้สถิติ Paired sample t-test เพื่อเปรียบเทียบคะแนนก่อนและหลังทดลองภายในกลุ่มควบคุมและกลุ่มทดลอง ดังมีรายละเอียดของผลการทดสอบดังนี้

ตารางที่ 46 ผลการทดสอบความแตกต่างของค่าเฉลี่ยคะแนนแรงจูงใจ

ตัวแปร/คู่ที่เปรียบเทียบ	n	M	SD	t	df	Sig. (2-tailed)
กลุ่มทดลอง (Paired sample t-test)						
คะแนนฯ ก่อนทดลอง	16	18.00	2.394	-1.163	15	.263
คะแนนฯ หลังทดลอง	16	18.44	2.394			
กลุ่มควบคุม (Paired sample t-test)						
คะแนนฯ ก่อนทดลอง	16	16.13	5.097	-1.078	15	.298
คะแนนฯ หลังทดลอง	16	17.19	2.509			

จากตารางที่ 46 แสดงให้เห็นว่า (1) ภายในกลุ่มทดลองมีค่าเฉลี่ยคะแนนแรงจูงใจก่อนทดลอง ($M = 18.00, SD = 2.39$) ไม่แตกต่างจากหลังทดลอง ($M = 18.44, SD = 2.39$) โดยมีค่า $t(15) = -1.163, p = .263$ และ (2) ภายในกลุ่มควบคุมมีค่าเฉลี่ยคะแนนแรงจูงใจก่อนทดลอง ($M = 16.13, SD = 5.097$) ไม่แตกต่างจากหลังทดลอง ($M = 17.19, SD = 2.509$) โดยมีค่า $t(15) = -1.078, p = .298$

5.2.4 ผลการเปรียบเทียบคะแนนหลังทดลองของทั้ง 3 ตัวแปรตามในโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ระหว่างกลุ่มควบคุมและกลุ่มทดลอง โดยมีคะแนนที่ได้ในระยะก่อนทดลองเป็นตัวแปรร่วม

ผลการตรวจสอบข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนตัวแปรร่วมพหุนาม (Multivariate analysis of covariance; MANCOVA) ได้แก่ (1) ความเป็นเอกพันธ์ของเมทริกซ์ความ

แปรปรวน-ความแปรปรวนร่วมของประชากรทดสอบด้วย Box's M พบว่า การรับรู้มาตรฐานหลังทดลอง กระบวนการเปรียบเทียบหลังทดลอง และความคลาดเคลื่อนหลังทดลอง มีค่า Box's M = 21.393, sig. = .004 ซึ่งน้อยกว่าระดับนัยสำคัญทางสถิติที่ .05 แสดงว่าตัวแปรตามทั้งสามตัวมีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของประชากรไม่เป็นเอกพันธ์กัน แต่เนื่องจากกลุ่มตัวอย่างทั้งสองกลุ่มมีขนาดเท่ากัน จึงสามารถใช้ Pillai's trace ที่มีความแกร่ง (robustness) สำหรับการเพิกเฉยต่อข้อตกลงเบื้องต้นในข้อนี้ได้ (Field, 2013) (2) ความสัมพันธ์ระหว่างตัวแปรตามทดสอบด้วย Bartlett's Test of Sphericity พบว่ามีค่า $\chi^2 = 35.217$ (df= 5, p= .000) แสดงว่า ตัวแปรตามมีความสัมพันธ์กัน (3) ความแปรปรวนของตัวแปรตามทั้งสามตัวแปรทดสอบด้วย Levene's test พบว่า การรับรู้มาตรฐานหลังทดลอง sig. = .992 กระบวนการเปรียบเทียบหลังทดลอง sig. = .683 และความคลาดเคลื่อนหลังทดลอง sig. = .091 ซึ่งมากกว่าระดับนัยสำคัญทางสถิติที่ .05 แสดงว่า ค่าความแปรปรวนของตัวแปรตามแต่ละตัวไม่แตกต่างกันระหว่างกลุ่มทดลองและกลุ่มควบคุม และ (4) ความสัมพันธ์ระหว่าง covariate กับตัวแปรตามต้องเหมือนกันทุกกลุ่ม (homogeneity of regression slopes) ทดสอบด้วยการวิเคราะห์ interaction ระหว่าง covariate กับตัวแปรอิสระ โดยผลต้องไม่ต่างกัน (p > .05) ซึ่งผลการทดสอบนี้พบว่า ความสัมพันธ์ระหว่าง covariate กับตัวแปรตามมีรูปแบบเดียวกันในทุกกลุ่มของตัวแปรอิสระ (p > .05) ยกเว้นตัวแปรคะแนนกระบวนการเปรียบเทียบก่อนทดลอง (p = .046) ดังนั้นสรุปได้ว่า ตัวแปรตามทั้ง 3 ตัวแปร ได้แก่ คะแนนการรับรู้มาตรฐานหลังทดลอง คะแนนกระบวนการเปรียบเทียบหลังทดลอง และความคลาดเคลื่อน เป็นไปตามข้อตกลงเบื้องต้นและสามารถวิเคราะห์สถิติ MANCOVA โดยมีเฉพาะคะแนนการรับรู้มาตรฐานก่อนทดลองที่สามารถใช้เป็นตัวแปรร่วมได้

ตารางที่ 47 ผลการวิเคราะห์ความแปรปรวนตัวแปรร่วมพหุนามคะแนนหลังทดลองของตัวแปรการรับรู้มาตรฐาน กระบวนการเปรียบเทียบ และความคลาดเคลื่อน ระหว่างกลุ่มควบคุมและกลุ่มทดลอง

แหล่งความแปรปรวน	Multivariate effects		Univariate effects		
	Pillai's Trace	F (3,27) values	F (1,29) values		
			คะแนนการรับรู้มาตรฐานหลังทดลอง	คะแนนกระบวนการเปรียบเทียบหลังทดลอง	คะแนนความคลาดเคลื่อน
Covariate: คะแนนการรับรู้มาตรฐานก่อนทดลอง (R Squared)	0.289	3.666*	10.805** (0.320)	5.689* (0.211)	1.814 (0.319)
Factor: กลุ่ม (ทดลอง vs. ควบคุม) กลุ่มทดลอง <i>M (SD)</i> กลุ่มควบคุม <i>M (SD)</i>	0.263	3.220*	1.022 21.94 (1.95) 20.88 (2.16)	0.888 15.25 (2.08) 14.31 (1.89)	9.627** 0.59 (0.49) 1.58 (1.06)

Box's M = 21.393, F = 3.176, sig. = .004

Bartlett's Test of Sphericity = 35.217, df= 5, p= .000

Levene's test: การรับรู้มาตรฐานหลังทดลอง F=.000, sig. = .992; กระบวนการเปรียบเทียบหลังทดลอง F=.170, sig. = .683; และความคลาดเคลื่อนหลังทดลอง F=3.054, sig. = .091

หมายเหตุ ** p < .01, * p < .05

จากตารางที่ 47 ผลการวิเคราะห์ความแปรปรวนตัวแปรร่วมพหุนาม (MANCOVA) พบว่า กลุ่ม(ควบคุม/ทดลอง) มีผลต่อตัวแปรตามแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อพิจารณาตัวแปรตามพร้อมกันทั้ง 3 ตัวแปร ได้แก่ คะแนนการรับรู้มาตรฐานหลังทดลอง คะแนนกระบวนการเปรียบเทียบหลังทดลอง และความคลาดเคลื่อน (Pillai's Trace = 0.263, F(3,27) = 3.220, p < .05) และเมื่อแยกการทดสอบอิทธิพลของตัวแปรการรับรู้มาตรฐานก่อนทดลอง (covariate) และกลุ่ม(ควบคุม/ทดลอง) ในตัวแปรตามทีละตัว ผลพบว่า (1) กลุ่ม (ควบคุม/ทดลอง) ไม่มีอิทธิพล

ต่อคะแนนการรับรู้มาตรฐานหลังทดลอง และคะแนนกระบวนการเปรียบเทียบหลังทดลอง หลังจากมีการควบคุมตัวแปรการรับรู้มาตรฐานก่อนทดลอง แต่มีอิทธิพลต่อความคลาดเคลื่อนอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 หลังจากมีการควบคุมตัวแปรการรับรู้มาตรฐานก่อนทดลอง (ค่า $F(1,29) = 9.627, p < .05$) โดยกลุ่มทดลองมีค่าเฉลี่ยความคลาดเคลื่อนน้อยกว่ากลุ่มควบคุม (กลุ่มทดลอง $M = 0.59, SD = 0.49$ และกลุ่มควบคุม $M = 1.58, SD = 1.06$) (2) คะแนนการรับรู้มาตรฐานก่อนทดลองมีความสัมพันธ์กับคะแนนการรับรู้มาตรฐานหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนการรับรู้มาตรฐานหลังทดลองได้ร้อยละ 32.0 และ (3) คะแนนการรับรู้มาตรฐานก่อนทดลองมีความสัมพันธ์กับคะแนนกระบวนการเปรียบเทียบหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนกระบวนการเปรียบเทียบหลังทดลองได้ร้อยละ 21.1



บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยเรื่อง การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสส์ครั้งนี้ เป็นการวิจัยและพัฒนา (research and development) มีวัตถุประสงค์เพื่อ 1) วิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานโดยใช้โมเดลหลายองค์ประกอบของราสส์ 2) พัฒนาและตรวจสอบโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน 3) ตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง และ 4) พัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยมีขั้นตอนในการวิจัยแบ่งเป็น 2 ระยะ ได้แก่

ระยะที่ 1 การพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง มีวิธีดำเนินการวิจัยสรุปได้เป็น 2 ขั้นตอน ดังนี้ (1) การสำรวจและคัดเลือกองค์ประกอบและตัวบ่งชี้ของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา โดยผู้วิจัยได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง จากนั้นได้สัมภาษณ์/สอบถามจากผู้ทรงคุณวุฒิที่มีความรู้และความเชี่ยวชาญในด้านการเรียนรู้โดยใช้ปัญหาเป็นฐาน จำนวน 5 คน และนำข้อมูลที่ได้มาวิเคราะห์เนื้อหา (content analysis) และคำนวณค่าดัชนีความสอดคล้อง (IOC) และ (2) การตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ผู้วิจัยได้ทำการศึกษาโดยเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่าง คือ อาจารย์และนักศึกษาของสถาบันอุดมศึกษาต่างๆ ในวิชาต่างๆ ที่จัดการสอนในรูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐานตรงตามเกณฑ์ในการคัดเลือก จำนวน 120 กลุ่ม PBL ซึ่งในการเก็บข้อมูลของงานวิจัยนี้ผู้วิจัยได้ขอการรับรองโครงการวิจัยจากคณะกรรมการพิจารณาจริยธรรมการวิจัยในคน กลุ่มสหสถาบัน ชุดที่ 1 จุฬาลงกรณ์มหาวิทยาลัย (รหัสชุดโครงการ COA No. 070/2557) เครื่องมือที่ใช้ในการวิจัยมี 2 ชุด ได้แก่ แบบสอบถามคุณลักษณะของผู้ประเมิน และแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย โดยเครื่องมือทุกฉบับมีการตรวจสอบความตรงเชิงเนื้อหาจากการตัดสินของผู้ทรงคุณวุฒิ ได้ค่าดัชนีความสอดคล้อง (IOC) อยู่ระหว่าง 0.80 ถึง 1.00 และมีค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's alpha coefficient) อยู่ระหว่าง 0.723 – 0.893 ทำการวิเคราะห์ข้อมูลด้วยสถิติ

บรรยาย MFRM ตรวจสอบความตรงของโมเดลฯ และตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง

ระยะที่ 2 การพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน มีวิธีดำเนินการวิจัยสรุปได้เป็น 2 ขั้นตอน ดังนี้ (1) การพัฒนาโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาให้มีความเหมาะสมกับแต่ละกลุ่มผู้ประเมิน ผู้วิจัยได้ดำเนินการโดยนำผลการวิเคราะห์ข้อมูลในตอนต้นที่ 4 คือ โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาที่มีความไม่แปรเปลี่ยนของรูปแบบโมเดลระหว่างกลุ่มผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง แต่มีความแปรเปลี่ยนของค่าน้ำหนักองค์ประกอบของแต่ละตัวบ่งชี้ ร่วมกับการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเพิ่มเติม เพื่อการพัฒนาต่อยอดในรายละเอียดของการใช้งานโมเดล โดยได้จัดทำเป็นคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา สำหรับอาจารย์ 1 ชุด และสำหรับนักศึกษา 1 ชุด จากนั้นผู้วิจัยได้นำคู่มือไปให้ผู้ทรงคุณวุฒิ ซึ่งเป็นผู้ที่มีความรู้และประสบการณ์ในด้านการเรียนรู้โดยใช้ปัญหาเป็นฐาน จำนวน 4 คน ตรวจสอบความถูกต้องและความเหมาะสมของคู่มือ และ (2) การตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ผู้วิจัยได้ดำเนินการโดยนำคู่มือสำหรับอาจารย์ และนักศึกษาไปทดลองใช้กับกลุ่มตัวอย่าง คือ อาจารย์และนักศึกษา (ชั้นปี 3) ของสถาบันอุดมศึกษาแห่งหนึ่งที่จัดการสอนในรูปแบบการเรียนรู้โดยใช้ปัญหาเป็นฐาน แบ่งเป็นกลุ่มทดลองจำนวน 2 กลุ่ม PBL และกลุ่มควบคุมจำนวน 2 กลุ่ม PBL เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล เป็นแบบวัดเดียวกันกับที่ใช้ในขั้นตอนการวิจัยระยะที่ 1 ซึ่งประกอบด้วยแบบสอบถามคุณลักษณะของผู้ประเมิน และแบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ทำการวิเคราะห์ข้อมูลด้วยสถิติบรรยาย, MFRM, ANCOVA และ MANCOVA

สรุปผลการวิจัย

ระยะที่ 1 การพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ผลการวิจัยสรุปได้ดังนี้

1. โมเดลที่พัฒนาขึ้นประกอบด้วยตัวแปรแฝงภายนอก 2 ตัวแปร คือ 1) บริบทของผู้ประเมิน ซึ่งวัดได้จากตัวแปรสังเกตได้ 5 ตัวแปร ได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้ และ 2) การตัดสินใจ ซึ่งวัดได้จากตัวแปรสังเกตได้ 2 ตัวแปร

ได้แก่ การรับรู้มาตรฐาน และกระบวนการเปรียบเทียบ โดยตัวแปรทั้งหมดนี้จะส่งผลต่อตัวแปรแฝงภายใน คือ คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งวัดได้จากตัวแปรสังเกตได้ 2 ตัวแปร ได้แก่ ความแม่นยำ และความคลาดเคลื่อน นอกจากนี้โมเดลฯ ยังมีตัวแปรกำกับ (moderator variable) คือ กลุ่มผู้ประเมิน ซึ่งแบ่งเป็น อาจารย์ เพื่อน และตนเอง และค่า IOC ขององค์ประกอบและตัวบ่งชี้ภายในโมเดลฯ ที่พัฒนาขึ้นมีค่าอยู่ระหว่าง 0.8 – 1

2. ผลการวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานโดยใช้โมเดลหลายองค์ประกอบของราสส์ พบว่าผู้ประเมินกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน ส่วนผู้ประเมินกลุ่มตนเองมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มอาจารย์ และเมื่อพิจารณาความแม่นยำในการให้คะแนนจากค่า Infit Mean – Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากค่าประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม มีการให้คะแนนที่มีความแม่นยำ จากผลทั้งสองอย่างนี้แสดงว่าในภาพรวมผู้ประเมินทั้งสามกลุ่มให้คะแนนที่มีความคลาดเคลื่อนเพียงเล็กน้อย และจัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้

3. ผลการตรวจสอบความตรงของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา พบว่า โมเดลฯ มีความสอดคล้องกับข้อมูลเชิงประจักษ์ จึงไม่ปฏิเสธสมมติฐานหลักที่ว่า โมเดลสมมติฐานตามทฤษฎีที่พัฒนาขึ้นสอดคล้องกับข้อมูลเชิงประจักษ์ แต่เมื่อพิจารณาองค์ประกอบบริบทของผู้ประเมิน (CONTEXT) และการตัดสิน (JUDGE) พบว่า มีอิทธิพลต่อคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) อย่างไม่มีนัยสำคัญทางสถิติ ทั้งที่ค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานของตัวแปรสังเกตได้เกือบทั้งหมด มีค่าเป็นบวกและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ยกเว้นตัวแปรความคลาดเคลื่อน (ERROR) และความแม่นยำ (ACCUR) ที่มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานไม่แตกต่างจากศูนย์ ดังนั้นผลการวิเคราะห์นี้จึงไม่สอดคล้องกับกรอบแนวคิดในการวิจัย ผู้วิจัยจึงได้พัฒนาโมเดลทางเลือกจากการทบทวนเอกสารและงานวิจัย โดยโมเดลทางเลือกประกอบด้วยตัวแปรแฝงทั้งหมด 2 ตัวแปร คือ บริบทของผู้ประเมิน (CONTEXT) วัดจากตัวแปรสังเกตได้ 5 ตัวแปร ได้แก่ เป้าหมาย (GOAL) ความสามารถ (ABILI) บุคลิกภาพแบบมีจิตสำนึก (CONSCI) แรงจูงใจ (MOTIV) และความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) วัดจากตัวแปรสังเกตได้ 3 ตัวแปร ได้แก่

การรับรู้มาตรฐาน (STAND) กระบวนการเปรียบเทียบ (COMPAR) และความคลาดเคลื่อน (ERROR) ซึ่งผลการตรวจสอบความตรงของโมเดลทางเลือกโดยใช้โปรแกรม LISREL 8.52 พบว่า โมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์ จึงไม่ปฏิเสธสมมติฐานหลักที่ว่า โมเดลสมมติฐานตามทฤษฎีที่พัฒนาขึ้นสอดคล้องกับข้อมูลเชิงประจักษ์ เมื่อพิจารณาค่าอิทธิพลรวมของตัวแปรคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า ตัวแปรบริบทของผู้ประเมิน (CONTEXT) มีค่าอิทธิพลรวมเท่ากับ 0.899 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า บริบทของผู้ประเมินจะส่งผลทางตรงในทิศทางบวกต่อคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยมีค่าสัมประสิทธิ์การพยากรณ์ของสมการโครงสร้าง (R^2) เท่ากับ 0.809 แสดงว่า บริบทของผู้ประเมิน (CONTEXT) สามารถอธิบายความแปรปรวนของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ได้ถึงร้อยละ 80.9 และในองค์ประกอบบริบทของผู้ประเมิน (CONTEXT) ตัวแปรสังเกตได้ทั้งหมดมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานเป็นบวกและแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งตัวแปรแรงจูงใจ (MOTIV) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) บุคลิกภาพแบบมีจิตสำนึก (CONSCI) ความสามารถ (ABILI) และเป้าหมาย (GOAL) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.759, 0.668, 0.516, 0.494 และ 0.351 ตามลำดับ ส่วนองค์ประกอบคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) พบว่า ตัวแปรกระบวนการเปรียบเทียบ (COMPAR) มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานสูงที่สุด รองลงมาคือ การรับรู้มาตรฐาน (STAND) และความคลาดเคลื่อน (ERROR) โดยมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐาน คือ 0.773, 0.623, และ -0.130 ตามลำดับ

4. ผลการตรวจสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง พบว่า มีความไม่แปรเปลี่ยนของรูปแบบโมเดลระหว่างกลุ่มผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง แต่มีความแปรเปลี่ยนของค่าน้ำหนักองค์ประกอบของแต่ละตัวบ่งชี้ ซึ่งเมื่อพิจารณาค่าอิทธิพลทางตรงที่ส่งผลต่อคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ของแต่ละกลุ่มผู้ประเมิน พบว่า อิทธิพลจากบริบทของผู้ประเมิน (CONTEXT) มีผลต่อคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานค่อนข้างสูงทั้ง 3 กลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง โดยมีค่าอิทธิพลเท่ากับ 0.873, 0.766 และ 0.969 ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ บริบทของผู้ประเมิน (CONTEXT) สามารถอธิบายความ

แปรปรวนของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ในกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ได้ร้อยละ 76.1, 58.6 และ 93.9 ตามลำดับ ซึ่งในองค์ประกอบบริบทของผู้ประเมิน (CONTEXT) ตัวบ่งชี้ที่มีน้ำหนักมากที่สุดของทั้ง 3 กลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง คือ ตัวบ่งชี้แรงจูงใจ (MOTIV) รองลงมาคือ ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และบุคลิกภาพแบบมีจิตสำนึก (CONSCI) ส่วนตัวบ่งชี้ที่มีน้ำหนักน้อยที่สุดของผู้ประเมินกลุ่มอาจารย์และเพื่อน คือ ตัวบ่งชี้เป้าหมาย (GOAL) และในผู้ประเมินกลุ่มตนเอง คือ ตัวบ่งชี้ความสามารถ (ABILI) ส่วนองค์ประกอบคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) เมื่อเปรียบเทียบระหว่าง 3 กลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง พบว่าตัวบ่งชี้ที่มีน้ำหนักมากที่สุด คือ ตัวบ่งชี้กระบวนการเปรียบเทียบ (COMPAR) รองลงมาคือ การรับรู้มาตรฐาน (STAND) และความคลาดเคลื่อน (ERROR)

ระยะที่ 2 การพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ผลการวิจัยสรุปได้ดังนี้

1. คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา ประกอบด้วยชุดสำหรับอาจารย์ 1 ชุด และสำหรับนักศึกษา 1 ชุด โดยรายละเอียดของคู่มือสำหรับอาจารย์ ได้แก่ หลักการและวัตถุประสงค์ของคู่มือ แนวทางการปฏิบัติและเครื่องมือที่ใช้สำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย และแนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ ส่วนคู่มือสำหรับนักศึกษาจะมีรายละเอียดเฉพาะหัวข้อแนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ และแบบประเมินกระบวนการกลุ่ม ซึ่งทั้งสองหัวข้อนี้มีเนื้อหาเหมือนกับคู่มือสำหรับอาจารย์ และแบบประเมินกระบวนการกลุ่ม เมื่อนำคู่มือทั้ง 2 ชุดนี้ไปให้ผู้ทรงคุณวุฒิ จำนวน 4 คน ตรวจสอบความถูกต้องและความเหมาะสมของคู่มือ ผลพบว่า คู่มือทั้ง 2 ชุด มีความถูกต้องและความเหมาะสม แต่ควรพัฒนาให้มีคู่มือฉบับสั้นเพิ่มขึ้น ซึ่งผู้วิจัยได้นำข้อเสนอแนะเหล่านั้นมาปรับใช้และได้มีการพัฒนาคู่มือฉบับสั้นเพิ่มอีก 2 ชุด คือ สำหรับอาจารย์ 1 ชุด และสำหรับนักศึกษา 1 ชุด ซึ่งเนื้อหาของคู่มือฉบับสั้นนำมาจากการสรุปประเด็นสำคัญของคู่มือฉบับสมบูรณ์

2. ผลการวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานโดยใช้โมเดลหลายองค์ประกอบของราสซ์ พบว่า (1) ในกลุ่มที่ 1 (ทดลอง) ผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน ส่วนผู้ประเมินกลุ่มตนเองมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มเพื่อน แต่จัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้ (2) ในกลุ่มที่ 2 (ทดลอง) ผู้ประเมินกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนน ส่วนผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มตนเอง แต่จัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้ (3) ในกลุ่มที่ 3 (ควบคุม) ผู้ประเมินกลุ่มตนเองมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มเพื่อน ส่วนผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนน แต่จัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้ และ (4) ในกลุ่มที่ 4 (ควบคุม) ผู้ประเมินกลุ่มเพื่อนมีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดปล่อยคะแนนมากที่สุด รองลงมาคือ ผู้ประเมินกลุ่มตนเอง ส่วนผู้ประเมินกลุ่มอาจารย์มีแนวโน้มให้คะแนนที่มีความคลาดเคลื่อนชนิดกดคะแนนมากที่สุด แต่จัดว่าเป็นการให้คะแนนที่มีความแม่นยำหรืออยู่ในเกณฑ์ที่ยอมรับได้

3. ผลการวิเคราะห์ความแปรปรวนร่วมของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองโดยมีคะแนนก่อนทดลองเป็นตัวแปรร่วม พบว่า (1) ตัวแปรเป้าหมาย มีค่าเฉลี่ยของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองไม่แตกต่างกัน ภายหลังจากมีการควบคุมคะแนนก่อนทดลองแล้ว (2) ตัวแปรความสามารถ มีค่าเฉลี่ยของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองไม่แตกต่างกัน ภายหลังจากมีการควบคุมคะแนนก่อนทดลองแล้ว (3) ตัวแปรบุคลิกภาพแบบมีจิตสำนึก มีค่าเฉลี่ยของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองไม่แตกต่างกัน ภายหลังจากมีการควบคุมคะแนนก่อนทดลองแล้ว และ (4) ตัวแปรความรับผิดชอบที่ตรวจสอบได้ มีค่าเฉลี่ยของคะแนนหลังทดลองระหว่างกลุ่มควบคุมและกลุ่มทดลองไม่แตกต่างกัน ภายหลังจากมีการควบคุมคะแนนก่อนทดลองแล้ว และคะแนนก่อนทดลองมีความสัมพันธ์กับคะแนนหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนหลังทดลองได้ร้อยละ 27.6 – 85.8 ส่วนผลการเปรียบเทียบค่าเฉลี่ยคะแนนแรงจูงใจภายในกลุ่มทดลองพบว่า ค่าเฉลี่ยคะแนนแรงจูงใจก่อนทดลองไม่แตกต่างจากหลังทดลอง เช่นเดียวกับการเปรียบเทียบภายในกลุ่มควบคุมที่พบว่าค่าเฉลี่ยคะแนนแรงจูงใจก่อนทดลองไม่แตกต่างจากหลังทดลอง

4. ผลการวิเคราะห์ความแปรปรวนตัวแปรร่วมพหุนาม (MANCOVA) พบว่า กลุ่ม(ควบคุม/ทดลอง) มีผลต่อตัวแปรตามแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อพิจารณาตัวแปรตามพร้อมกันทั้ง 3 ตัวแปร ได้แก่ คะแนนการรับรู้มาตรฐานหลังทดลอง คะแนนกระบวนการเปรียบเทียบหลังทดลอง และความคลาดเคลื่อน โดยเมื่อแยกการทดสอบอิทธิพลของตัวแปรการรับรู้มาตรฐานก่อนทดลอง(covariate) และกลุ่ม(ควบคุม/ทดลอง) ในตัวแปรตามทีละตัว ผลพบว่า กลุ่ม(ควบคุม/ทดลอง) ไม่มีอิทธิพลต่อคะแนนการรับรู้มาตรฐานหลังทดลองและคะแนนกระบวนการเปรียบเทียบหลังทดลอง แต่มีอิทธิพลต่อความคลาดเคลื่อน โดยกลุ่มทดลองมีความคลาดเคลื่อนน้อยกว่ากลุ่มควบคุม และคะแนนการรับรู้มาตรฐานก่อนทดลองมีความสัมพันธ์กับคะแนนการรับรู้มาตรฐานหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนการรับรู้มาตรฐานหลังทดลองได้ร้อยละ 32.0 และคะแนนการรับรู้มาตรฐานก่อนทดลองมีความสัมพันธ์กับคะแนนกระบวนการเปรียบเทียบหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนกระบวนการเปรียบเทียบหลังทดลองได้ร้อยละ 21.1 คะแนนการรับรู้มาตรฐานก่อนทดลองมีความสัมพันธ์กับคะแนนความคลาดเคลื่อนหลังทดลอง

อภิปรายผล

ระยะที่ 1 การพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรต่างๆ ที่ส่งผลต่อความแม่นยำในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานของกลุ่มผู้ประเมิน ซึ่งได้แก่ อาจารย์ เพื่อน และตนเอง โดยผลจากการวิเคราะห์ครั้งแรก พบว่า โมเดลมีสอดคล้องกับข้อมูลเชิงประจักษ์ แต่ไม่สอดคล้องกับกรอบแนวคิดในการวิจัย ดังนั้นผู้วิจัยจึงได้พัฒนาโมเดลทางเลือกจากการทบทวนเอกสารและงานวิจัยอีกครั้ง ซึ่งได้โมเดลที่มีการจัดกลุ่มของตัวแปรสังเกตได้ใหม่ กล่าวคือ จากเดิมองค์ประกอบการตัดสินใจ วัดจากตัวแปรสังเกตได้คือ การรับรู้มาตรฐาน และกระบวนการเปรียบเทียบ ได้ปรับเป็นรวมอยู่ในองค์ประกอบคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งเดิมวัดจากตัวแปรสังเกตได้ คือ ความคลาดเคลื่อน และความแม่นยำ โดยปรับใหม่เป็นการวัดจากตัวแปรสังเกตได้ 3 ตัวแปร คือ การรับรู้มาตรฐาน กระบวนการเปรียบเทียบ และความคลาดเคลื่อน ส่วนตัวแปรความแม่นยำ ได้นำออกจากโมเดลฯ และผลการตรวจสอบโมเดลใหม่ พบว่า มีความสอดคล้องกับกลมกลืนกับข้อมูลเชิงประจักษ์ และเป็นไปตามกรอบแนวคิดในการวิจัย จากผลการวิจัยนี้พบว่า มีประเด็นที่น่าสนใจ คือ ในองค์ประกอบคุณภาพการให้คะแนนในวิชาที่มี

การเรียนรู้โดยใช้ปัญหาเป็นฐาน ไม่จำเป็นต้องวัดจากตัวแปรสังเกตได้ทั้งสองตัวแปร คือ ความแม่นยำ และความคลาดเคลื่อน แต่ควรเลือกวัดเพียงตัวแปรใดตัวแปรหนึ่งเท่านั้น เนื่องจาก Murphy & Cleveland (1995) ได้กล่าวว่า ความคลาดเคลื่อนของผู้ประเมินเป็นการตรวจสอบทางอ้อมของความแม่นยำในการให้คะแนน (indirect measures of rating accuracy หรือ inaccuracy) หรือถ้าผลการวิเคราะห์ไม่พบความคลาดเคลื่อนของผู้ประเมินอาจกล่าวได้ว่าการให้คะแนนนั้นมีความแม่นยำ (rating accuracy) ซึ่งสอดคล้องกับหลักการคำนวณของโปรแกรม facets คือ ค่าของตัวแปรความแม่นยำ (ACCUR) มาจากการแปลงค่าของตัวแปรความคลาดเคลื่อน (ERROR) โดยโปรแกรมจะคำนวณค่าความคลาดเคลื่อนของผู้ประเมิน รวมถึงค่าความสามารถของผู้ถูกประเมิน และค่าความยากง่ายของข้อคำถามในหน่วยโลจิทก่อน จากนั้นจะมีการคำนวณค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน ซึ่งเป็นค่า Infit Mean - Square หรือค่าความแม่นยำ (ACCUR)

จากผลการวิเคราะห์ความแม่นยำและความคลาดเคลื่อนในการให้คะแนนของกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานโดยใช้โมเดลหลายองค์ประกอบของราสส์ พบว่าผู้ประเมินทั้ง 3 กลุ่ม คือ อาจารย์ เพื่อน และตนเอง ทุกกลุ่มต่างก็มีความคลาดเคลื่อนในการให้คะแนนแตกต่างกันเพียงเล็กน้อย แต่เมื่อพิจารณาความแม่นยำในการให้คะแนนจากค่า Infit Mean - Square หรือค่ายกกำลังสองของความคลาดเคลื่อนมาตรฐานซึ่งเป็นผลต่างของคะแนนจากการประเมินกับคะแนนจากการประมาณค่าของผู้ประเมิน พบว่า ผู้ประเมินทั้ง 3 กลุ่ม มีการให้คะแนนที่มีความแม่นยำ จึงอาจอธิบายได้ว่าเนื่องจากกลุ่มตัวอย่างทั้งหมดเป็นคนที่เกี่ยวข้องกับการประกอบวิชาชีพทางสุขภาพ ซึ่งเน้นการมีคุณธรรมและจริยธรรมมาก ดังนั้นจึงมีการให้คะแนนที่มีความคลาดเคลื่อนเพียงเล็กน้อยและมีการให้คะแนนที่มีความแม่นยำ ส่วนประเด็นที่น่าสนใจ คือ ในการวัดตัวแปรความคลาดเคลื่อน จะพบว่ามีหลากหลาย (variation) ของค่าคะแนนมากกว่าตัวแปรความแม่นยำ ดังนั้นจึงอาจเป็นเหตุผลสำหรับการพิจารณาเลือกตัวแปรมาใช้ในการวิจัย

จากผลการทดสอบความไม่แปรเปลี่ยนของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ที่พบว่า มีความไม่แปรเปลี่ยนของรูปแบบโมเดลระหว่างกลุ่มผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง แต่มีความแปรเปลี่ยนของค่าน้ำหนักองค์ประกอบของแต่ละตัวบ่งชี้ โดยตัวบ่งชี้ในบริบทของผู้ประเมินที่มีความสำคัญมากที่สุดของทั้ง 3 กลุ่มผู้ประเมิน คือ ตัวบ่งชี้แรงจูงใจ โดยผลนี้สอดคล้องกับงานของ Murphy, Cleveland, Skattebo, & Kinney (2004); Roch (2007) และ

Salvemini, Reilly, & Smither (1993) ที่พบว่า แรงจูงใจของผู้ประเมินเป็นตัวทำนายที่ดีที่สุดสำหรับคุณภาพการให้คะแนน และตัวบ่งชี้ที่มีความสำคัญรองลงมา คือ ความรับผิดชอบที่ตรวจสอบได้ และบุคลิกภาพแบบมีจิตสำนึก ซึ่งสอดคล้องกับผลการศึกษาของ Bernardin, Tyler, & Villanova (2009) และ Curtis, Harvey, & Ravden (2005) ที่พบว่า ความรับผิดชอบที่ตรวจสอบได้ และบุคลิกภาพแบบมีจิตสำนึก สามารถทำนายคุณภาพการให้คะแนนได้ ส่วนตัวบ่งชี้ที่มีความสำคัญน้อยที่สุดของผู้ประเมินกลุ่มอาจารย์และเพื่อน คือ เป้าหมาย ซึ่งสาเหตุที่ตัวบ่งชี้เป้าหมายมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานน้อย อาจเนื่องมาจากตัวแปรเป้าหมายและแรงจูงใจมีความสัมพันธ์กันในทางทฤษฎี กล่าวคือ เป้าหมายจะเป็นตัวกระตุ้นให้เกิดแรงจูงใจในการแสดงพฤติกรรมต่างๆ (Locke, 1996) ดังนั้นผลการวิจัยจึงพบว่า ตัวบ่งชี้แรงจูงใจมีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานมากกว่าตัวบ่งชี้เป้าหมาย ส่วนตัวบ่งชี้ที่มีความสำคัญน้อยที่สุดของผู้ประเมินกลุ่มตนเอง คือ ความสามารถ และยังเป็นตัวบ่งชี้ที่มีน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานต่ำสุดอันดับสองของผู้ประเมินกลุ่มอาจารย์ และกลุ่มเพื่อน ซึ่งการที่ตัวบ่งชี้ความสามารถนี้มีความสำคัญน้อย จะมีความสอดคล้องกับ Eva (2001) ที่ได้กล่าวไว้ว่า การฝึกอบรมพัฒนาความสามารถเกี่ยวกับการใช้เกณฑ์การประเมิน และการสร้างแบบประเมินที่ตรงตามคุณลักษณะที่จะใช้ในประเมิน ยังไม่ใช่วิธีการที่สามารถแก้ไขปัญหาการขาดความแม่นยำในการประเมินได้ทั้งหมด ส่วนองค์ประกอบคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษามีตัวบ่งชี้ที่มีความสำคัญมากที่สุดของทั้ง 3 กลุ่มผู้ประเมิน คือ กระบวนการเปรียบเทียบ รองลงมาคือ การรับรู้มาตรฐาน และความคลาดเคลื่อน ซึ่งสอดคล้องกับงานของ Landy & Farr, (1980) และ Murphy & Cleveland (1995) ที่กล่าวว่า คุณภาพการให้คะแนนจะต่ำถ้ามีความคลาดเคลื่อนของผู้ประเมินที่ให้คะแนนแบบกดหรือปล่อยคะแนน (McManus et al., 2006; Murphy & Cleveland, 1995) ดังนั้นจากผลการวิจัยส่วนนี้พบว่ามีประเด็นที่น่าสนใจ คือ การพัฒนาคุณภาพการให้คะแนนสำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (tutorial-based assessment) ในการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยมีผู้ประเมินที่เป็น อาจารย์ เพื่อน และตนเอง ควรให้ความสำคัญและพัฒนาปัจจัยด้านบริบทของผู้ประเมิน ซึ่งได้แก่ แรงจูงใจ ความรับผิดชอบที่ตรวจสอบได้ บุคลิกภาพแบบมีจิตสำนึก และเป้าหมาย ควบคู่ไปกับการฝึกอบรมความรู้ด้านการประเมิน (rater training) ซึ่งทั่วไปไม่มีเป้าหมายเฉพาะเพียงการฝึกให้ผู้ประเมินมีความเข้าใจตรงกันโดยฝึกให้เกิดความคุ้นเคยกับแบบประเมินและเกณฑ์การประเมิน

ระยะที่ 2 การพัฒนาและตรวจสอบโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน เป็นการนำผลการพัฒนาโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการตรวจสอบความไม่แปรเปลี่ยนของโมเดลระหว่างกลุ่มผู้ประเมิน อาจารย์ เพื่อน และตนเอง ในตอนที่ 1 ร่วมกับการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเพิ่มเติม เพื่อมาใช้ในการพัฒนาโมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ซึ่งเป็นการพัฒนาต่อยอดในรายละเอียดของการใช้งานโมเดล โดยการจัดทำเป็นคู่มือ การประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา สำหรับอาจารย์ 1 ชุด และสำหรับนักศึกษา 1 ชุด อันจะทำให้เกิดความชัดเจน เป็นรูปธรรม และได้ประโยชน์อย่างสูงสุดจากการใช้งานโมเดล โดยมีเป้าหมายสูงสุด คือ ช่วยในการปรับปรุงหรือพัฒนาให้เกิดคุณภาพในการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งจากผลการวิจัยนี้พบว่า มีประเด็นที่น่าสนใจซึ่งเป็นความคิดเห็นของผู้ทรงคุณวุฒิที่มีความรู้และประสบการณ์ในด้านการเรียนรู้โดยใช้ปัญหาเป็นฐาน เกี่ยวกับการใช้งานคู่มือ คือ ผู้ประเมินต้องทำความเข้าใจและศึกษาคู่มือก่อนทุกครั้ง ควรมีการประชุมปรึกษาหารือถึงแนวทางการปฏิบัติในคู่มืออย่างต่อเนื่อง และผู้ประเมินต้องมีการทำความเข้าใจและสื่อสารเป็นภาษาเดียวกัน แต่อีกมุมมองหนึ่งของผู้ทรงคุณวุฒิที่น่าสนใจซึ่งได้กล่าวไว้ว่า คู่มือสำหรับอาจารย์และนักศึกษาควรเหมือนกัน โดยลดความละเอียดของตัวแปรต่างๆ ลง เนื่องจากไม่ใช่สิ่งที่อาจารย์ผู้สอนซึ่งไม่ใช่นักวิชาการทางการศึกษาจะสนใจ ทั้งนี้ผู้วิจัยขออธิบายเกี่ยวกับประเด็นนี้ว่า การเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ส่วนใหญ่มีในสาขาทางสุขภาพศาสตร์ เช่น แพทย์ ทันตแพทย์ พยาบาล เภสัช ซึ่งหลักสูตรเหล่านี้จะประกอบด้วยจำนวนวิชาและหน่วยกิตจำนวนมากเพราะมีวิชาปฏิบัติร่วมด้วย ดังนั้นจึงนิยมใช้การเรียนรู้โดยใช้ปัญหาเป็นฐานซึ่งมีข้อดี คือ ผู้เรียนได้เรียนรู้ตัวอย่างจากสถานการณ์ปัญหาซึ่งมีความใกล้เคียงกับสถานการณ์จริง มีฐานความรู้ที่กว้าง พัฒนาทักษะการแก้ปัญหา และทักษะการเรียนรู้แบบนำตนเอง แต่เนื่องจากอาจารย์และนักศึกษาค่อนข้างมีเวลาจำกัด เห็นได้จากการเปิดปิดโจทย์สถานการณ์ส่วนใหญ่จะใช้เวลาห่างกันประมาณ 1 สัปดาห์ จึงต้องการคู่มือที่สั้นโดยสามารถสรุปหลักการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพได้ เพื่อที่จะใช้เวลาที่เหลือในการศึกษาค้นคว้าด้วยตนเองหรือการเรียนรู้แบบนำตนเอง (self-directed learning) ตามกระบวนการของการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งจากประเด็นนี้จึงสรุปได้ว่าผู้สอนและผู้เรียนมีลักษณะที่แตกต่างกัน 2 กลุ่ม คือ กลุ่มที่ต้องการทราบเฉพาะหลักหรือประเด็นสำคัญสำหรับการเป็นผู้ประเมินที่มีคุณภาพกับกลุ่มที่ให้ความสำคัญกับรายละเอียดของการปฏิบัติเพื่อที่จะเกิดความเข้าใจและสื่อสารเป็นภาษาเดียวกัน ดังนั้นผู้วิจัยจึงพัฒนาคู่มือทั้งฉบับสั้นและฉบับสมบูรณ์ตามความต้องการของผู้ใช้งานเพื่อนำไปใช้ปฏิบัติจริงได้อย่างมีประสิทธิภาพ

ผลจากการทดลองใช้คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติ สำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา (ฉบับสั้น) ซึ่งประกอบด้วย ชุดสำหรับอาจารย์ และชุดสำหรับนักศึกษา มีประเด็นที่น่าสนใจ คือ ตัวแปรเชิงสาเหตุ ในโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานทั้ง 5 ตัวแปร คือ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้ มีค่าเฉลี่ยของคะแนนหลังการใช้คู่มือไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติระหว่างกลุ่มทดลอง และกลุ่มควบคุม แต่พบว่าคะแนนก่อนทดลองของตัวแปร เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก และความรับผิดชอบที่ตรวจสอบได้ มีความสัมพันธ์กับคะแนนหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนหลังทดลองได้ร้อยละ 27.6 – 85.8 ดังนั้นอาจมีสาเหตุที่เป็นไปได้คือ ตัวแปรเหล่านี้ส่วนใหญ่เป็นตัวแปรทางจิตวิทยาจึงต้องการเวลาในการพัฒนา ดังนั้นจึงไม่พบความแตกต่างของคะแนนหลังทดลองระหว่างกลุ่มทดลองและกลุ่มควบคุม แต่ทั้งนี้ผู้ประเมินก็ควรที่จะพัฒนาตัวแปรเหล่านี้ต่อไปเนื่องจากคะแนนหลังทดลองมีความสัมพันธ์กับคะแนนก่อนทดลอง ส่วนตัวแปรความสามารถถึงแม้จะไม่ใชตัวแปรทางจิตวิทยา แต่เป็นตัวแปรที่มีค่าน้ำหนักองค์ประกอบในรูปคะแนนมาตรฐานน้อยที่สุดในโมเดลฯ และในงานวิจัยส่วนใหญ่มักจะพบว่า การจัดอบรมให้ความรู้เพื่อพัฒนาความสามารถในการให้คะแนน ไม่ค่อยส่งผลทำให้เกิดคุณภาพและความแม่นยำในการให้คะแนน (Eva, 2001) ดังนั้นจึงอาจต้องการเวลาในการพัฒนาเช่นกัน

ส่วนผลการวิเคราะห์ความแปรปรวนตัวแปรร่วมพหุนาม พบว่า กลุ่ม(ควบคุม/ทดลอง) ไม่มีอิทธิพลต่อคะแนนการรับรู้มาตรฐานหลังทดลองและคะแนนกระบวนการเปรียบเทียบหลังทดลอง แต่มีอิทธิพลต่อความคลาดเคลื่อน โดยกลุ่มทดลองมีความคลาดเคลื่อนน้อยกว่ากลุ่มควบคุม และคะแนนการรับรู้มาตรฐานก่อนทดลองมีความสัมพันธ์กับคะแนนการรับรู้มาตรฐานหลังทดลอง และคะแนนกระบวนการเปรียบเทียบหลังทดลอง โดยสามารถอธิบายความผันแปรของคะแนนหลังทดลองได้ร้อยละ 32.0 และ 21.1 ตามลำดับ ทั้งนี้อาจอธิบายได้ว่า การพัฒนาให้เกิดคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาซึ่งวัดได้จากทั้ง 3 ตัวแปร คือ การรับรู้มาตรฐาน กระบวนการเปรียบเทียบ และความคลาดเคลื่อน อาจต้องการเวลาในการพัฒนา ดังนั้นจึงไม่พบความแตกต่างของคะแนนหลังทดลองระหว่างกลุ่มทดลองและกลุ่มควบคุม แต่ทั้งนี้ผู้ประเมินก็ควรที่จะพัฒนาตัวแปรเหล่านี้ต่อไป เนื่องจากคะแนนการรับรู้มาตรฐานหลังทดลอง คะแนนกระบวนการเปรียบเทียบหลังทดลอง และคะแนนความคลาดเคลื่อนหลังทดลอง มีความสัมพันธ์กับคะแนนการรับรู้มาตรฐานก่อนทดลอง และมีประเด็นที่น่าสนใจ คือ ในการวัดคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา อาจมุ่งเน้นที่การวัดตัวแปรความคลาดเคลื่อน เนื่องจากเป็นตัวบ่งชี้ที่มีความไวมากที่สุด

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลวิจัยไปใช้

1. การพัฒนาคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ซึ่งเป็นการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (tutorial-based assessment) ในการเรียนรู้โดยใช้ปัญหาเป็นฐาน โดยมีผู้ประเมินที่เป็น อาจารย์ เพื่อน และตนเอง ควรให้ความสำคัญและพัฒนาปัจจัยด้านบริบทของผู้ประเมิน ซึ่งได้แก่ แรงจูงใจ ความรับผิดชอบที่ตรวจสอบได้ บุคลิกภาพแบบมีจิตสำนึก เป้าหมาย และความสามารถ ซึ่งจะส่งผลคุณภาพการให้คะแนนมากน้อยต่างกันตามลำดับความสำคัญ

2. การพัฒนาตัวแปรเชิงสาเหตุภายในโมเดลหรือการใช้คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษานั้น ควรที่จะต้องคำนึงถึงปัจจัยอื่นๆ ที่อาจส่งผลได้ เช่น ระยะเวลาความพร้อมหรือความแตกต่างระหว่างบุคคล เนื่องจากตัวแปรเชิงสาเหตุภายในโมเดลฯ ส่วนใหญ่เป็นตัวแปรทางด้านจิตวิทยา

3. คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษาที่พัฒนาขึ้นในงานวิจัยนี้ ประกอบด้วยคู่มือของอาจารย์ และนักศึกษา ทั้งฉบับสั้น และฉบับสมบูรณ์ รวมจำนวน 4 ชุด ซึ่งผู้ใช้สามารถเลือกได้ตามความเหมาะสม โดยคู่มือที่สั้นเหมาะกับผู้สอนและผู้เรียนที่มีเวลาไม่มาก ซึ่งจะเป็นการสรุปหลักการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพ ส่วนคู่มือฉบับสมบูรณ์จะประกอบด้วยแนวทางการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพ นอกจากนี้ในทางปฏิบัติจึงเพื่อให้เกิดผลที่ชัดเจนขึ้น ควรจัดให้มีการฝึกอบรม (training) เกี่ยวกับแนวปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพ โดยเน้นการพัฒนาปัจจัยด้านบริบทของผู้ประเมิน ซึ่งได้แก่ แรงจูงใจ ความรับผิดชอบที่ตรวจสอบได้ บุคลิกภาพแบบมีจิตสำนึก และเป้าหมาย ควบคู่ไปกับการฝึกอบรมความรู้ด้านการประเมิน การใช้แบบประเมิน และการใช้เกณฑ์การประเมิน ซึ่งเนื้อหาหรือรายละเอียดทั้งหมดนี้ได้มีอยู่ในคู่มือฉบับสมบูรณ์ และระยะเวลาในการอบรมอาจใช้เวลาประมาณ 3-4 ชั่วโมง เพื่อให้กลุ่มผู้ประเมินที่เป็น อาจารย์ เพื่อน และตนเอง มีโอกาสทำความเข้าใจประเด็นสำคัญต่างๆ ได้ตรงกัน และยังสามารถรับรู้ถึงปัญหาและอุปสรรคของผู้ประเมินได้

4. ภายหลังจากใช้คู่มือถ้าต้องการวัดคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา อาจมุ่งเน้นที่การวัดตัวแปรความคลาดเคลื่อนเพียงตัวแปรเดียว เนื่องจากเป็นตัวบ่งชี้ที่มีความไวมากที่สุด

5. การวิจัยนี้ทำการศึกษาในกลุ่มตัวอย่างที่อยู่ในระดับอุดมศึกษา และส่วนใหญ่อยู่ในสาขาทางสุขศาสตร์ เช่น แพทย์ ทันตแพทย์ พยาบาล เกษัช ซึ่งกลุ่มตัวอย่างนี้อาจมีคุณลักษณะหรือบุคลิกภาพที่มีความเฉพาะแตกต่างจากสาขาอื่นๆ ทำให้อาจส่งผลกระทบต่อลำดับความสำคัญของตัวแปรเชิงสาเหตุที่ส่งผลกระทบต่อคุณภาพการให้คะแนนมากน้อยต่างกัน ดังนั้นในการนำผลการวิจัยไปใช้อาจต้องคำนึงถึงเหตุผลข้อนี้ด้วย

ข้อเสนอแนะในการวิจัยต่อไป

1. ควรมีการวิจัยเกี่ยวกับปัจจัยที่ส่งผลกระทบต่อคุณภาพการให้คะแนนอย่างเป็นองค์รวมมากยิ่งขึ้น โดยนำตัวแปรที่ต่างระดับกันและมีความเกี่ยวข้องกับคุณภาพการให้คะแนน เช่น คุณลักษณะหรือประเภทของแบบประเมินต่างๆ เทคนิคของการประเมินแบบต่างๆ ตลอดจนเกณฑ์การประเมินต่างๆ เป็นต้น มาใช้ศึกษาร่วมกันภายในโมเดลและใช้การวิเคราะห์แบบพหุระดับ (multilevel analysis) เพื่อช่วยในการทำความเข้าใจถึงตัวแปรในเชิงสาเหตุได้อย่างครอบคลุมและเป็นระบบ

2. ควรมีการวิจัยเพื่อเน้นการพัฒนาแบบวัดของตัวแปรทางจิตวิทยา ซึ่งได้แก่ เป้าหมาย บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้ เนื่องจากตัวแปรเหล่านี้มีความสำคัญและส่งผลกระทบต่อพฤติกรรมการให้คะแนนของผู้ประเมิน โดยพัฒนาแบบวัดเหล่านี้ให้มีคุณภาพระดับมาตรฐานสากล ซึ่งต้องมีระบบการสร้างและการตรวจสอบคุณภาพของเครื่องมือที่ครอบคลุม เช่น การตรวจสอบในเชิงจิตมิติของโครงสร้างองค์ประกอบและความตรงเชิงเนื้อหา การตรวจสอบความตรงเชิงทฤษฎี การวิเคราะห์ความสัมพันธ์เปรียบเทียบระหว่างแบบวัดที่มีความใกล้เคียงและแบบวัดที่แตกต่างกัน การออกแบบสถานการณ์เพื่อตรวจสอบผลที่เกิดขึ้นโดยธรรมชาติ การทดลองโดยการจัดการกระทำให้เกิดตัวแปรเหล่านี้ขึ้น รวมถึงการวิจัยตรวจสอบในทางคลินิก เป็นต้น

3. ควรมีการวิจัยเชิงคุณภาพเกี่ยวกับตัวแปรต่างๆ ภายในโมเดล โดยการใช้การสัมภาษณ์จากกลุ่มตัวอย่างทั้งอาจารย์และนักศึกษาที่มีการเรียนการสอนโดยใช้ PBL เพื่อให้สามารถเข้าใจปัจจัยเชิงสาเหตุที่ส่งผลกระทบต่อคุณภาพการให้คะแนนได้มากขึ้น

4. ควรศึกษาผลของการใช้โมเดลคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในช่วงเวลาที่ยาวขึ้น โดยมีการเก็บข้อมูลซ้ำหลายครั้ง (repeated measure) เพื่อตรวจสอบว่าระยะเวลาที่มีผลในลักษณะของการเป็นตัวแปรแทรกซ้อนระหว่างตัวแปรเชิงสาเหตุกับตัวแปรตามหรือไม่

5. การวิจัยเชิงทดลองเพื่อให้เห็นผลที่ชัดเจนของตัวแทรกแซง (intervention) โดยเฉพาะในการจัดการกระทำเพื่อพัฒนาตัวแปรทางจิตวิทยา เช่น เป้าหมาย บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้ อาจต้องใช้ระยะเวลาในการศึกษาที่ยาวขึ้น มีตัวแทรกแซงที่มีประสิทธิภาพ

ตลอดจนมีการควบคุมตัวแปรแทรกซ้อนเป็นอย่างดีหรืออาจทำการตรวจสอบสภาพการณ์ (manipulation check) เพื่อเป็นข้อมูลในการตัดสินใจว่าตัวแทรกแซงที่ต้องการศึกษาเกิดขึ้นในกลุ่มตัวอย่างจริงหรือไม่ และอาจใช้เป็นหลักฐานสำหรับความตรงเชิงโครงสร้างของตัวแทรกแซง (construct validity) (Wheelan, 2005)



รายการอ้างอิง

ภาษาไทย

- นงลักษณ์ วิรัชชัย. (2542). โมเดลลิสรล: สถิติวิเคราะห์สำหรับการวิจัย กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- น้ำผึ้ง อินทะเนตร. (2554). การศึกษาคุณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์เมื่อจำนวนผู้ตรวจให้คะแนนต่างกัน โดยใช้โมเดลการสรุปอ้างอิงและโมเดลหลายองค์ประกอบของราส์ซ. มหาวิทยาลัยศรีนครินทรวิโรฒ.
- นุชนารถ วงศ์จำปา. (2553). การพัฒนาตัวบ่งชี้การจัดการเรียนรู้แบบใช้ปัญหาเป็นฐานในระดับอุดมศึกษา. จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2550). ทฤษฎีการทดสอบแนวใหม่ (3 ed.). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการทดสอบแบบดั้งเดิม (6 ed.). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สำนักงานคณะกรรมการกฤษฎีกา. (2542). พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542. ราชกิจจานุเบกษา ฉบับกฤษฎีกา เล่มที่ 116 ตอนที่ 74 ก 19 สิงหาคม 2542, 1-23.
- สำนักงานคณะกรรมการกฤษฎีกา. (2545). พระราชบัญญัติการศึกษาแห่งชาติ (ฉบับที่ 2) พ.ศ. 2545. ราชกิจจานุเบกษา ฉบับกฤษฎีกา เล่มที่ 119 ตอนที่ 123 ก 19 ธันวาคม 2545, 16-21.
- สำนักงานคณะกรรมการกฤษฎีกา. (2553). พระราชบัญญัติการศึกษาแห่งชาติ (ฉบับที่ 3) พ.ศ. 2553. ราชกิจจานุเบกษา ฉบับกฤษฎีกา เล่มที่ 127 ตอนที่ 45 ก 22 กรกฎาคม 2553, 1-3.
- สำนักงานคณะกรรมการอุดมศึกษา. (2554). คู่มือการประกันคุณภาพการศึกษาภายในสถานศึกษา ระดับอุดมศึกษา พ.ศ. 2553 (1 ed.). กรุงเทพฯ: ห้างหุ้นส่วนจำกัด ภาพพิมพ์.
- สุภมาส อังศุโชติ, สมถวิล วิจิตรวรรณ, & รัชนิกุล ภิญโญภาณุวัฒน์. (2554). สถิติวิเคราะห์สำหรับการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์: เทคนิคการใช้โปรแกรม LISREL (3 ed.). กรุงเทพฯ: เจริญดี มั่นคงการพิมพ์.

ภาษาอังกฤษ

- Ajpru, H., Pasiphol, S., & Wongwanich, S. (2011). Development of an instructional quality assurance model in nursing science. *Research in Higher Education Journal*, 13, 1-14.

- Barell, J. (2010). *Problem-based learning: the foundation for 21 st century skills*.
Bloomington: Solution Tree Press.
- Barrett, S. (2005). *Raters and examination*. The Netherlands: Springer.
- Barrows, H. S., & Tamblyn, R. (1980). *Problem-Based Learning: An Approach to Medical Education*. New York: 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.
- Bernardin, H. J., Tyler, C. L., & Villanova, P. (2009). Rating level and accuracy as a function of rater personality. *International Journal of Selection and Assessment*, 17(3), 300-310.
- Bridges, E. M., & Hallinger, P. (1992). Eugene: . (1992). *Problem-based learning for administrators*. Eugene, Oregon: ERIC Clearinghouse on Educational Management.
- Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor model. *Structural Equation Modeling*, 12(3), 471-492.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Classen, S., Wen, P. S., Velozo, C. A., Bedard, M., Winter, S. M., Brumback, B. A., et al. (2012). Rater reliability and rater effects of the Safe Driving Behavior Measure. *Am J Occup Ther*, 66(1), 69-77.
- Congdon, P. J., & McQueen, J. (2000). The stability of rater severity in large-scale assessment Programs. *Journal of Educational Measurement*, 37(2), 163-178.
- Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity". *Psychological Bulletin*, 52(3), 177-193.
- Curtis, A. B., Harvey, R. D., & Ravden, D. (2005). Sources of political distortions in performance appraisals. *Group and Organization Management*, 30(1), 42-60.
- DeCarlo, L. T., Kim, T. K., & Johnson, M. S. (2011). A hierarchical rater model for constructed responses, with a signal detection rater model. *Journal of Educational Measurement*, 48(3), 333-356.

- Eckes, T. (2009). *Many-facet Rasch measurement*. Strasbourg, France: Council of Europe/Language Policy Division.
- Engelhard, G. (1994). Examining rater error in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement*, 31(2), 93-112.
- Engelhard, G. (1996). Evaluating rater accuracy in performance assessments. *Journal of Educational Measurement*, 33(1), 56-70.
- Eva, K. W. (2001). Assessing tutorial-based assessment. *Adv Health Sci Educ Theory Pract*, 6(3), 243-257.
- 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.
- Farrokhi, F., Esfandiari, R., & Dalili, M. V. (2011). Applying the many-facet Rasch model to detect centrality in self-assessment, peer-assessment and teacher assessment. *World Applied Sciences Journal*, 15, 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.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock 'n' roll* (4 ed.). London: Sage publications.
- Gorman, C. A., & Rentsch, J. R. (2009). Evaluating frame-of-reference rater training effectiveness using performance schema accuracy. *J Appl Psychol*, 94(5), 1336-1344.
- Haiyang, S. (2010). An application of classical Test theory and manyfacet-Rasch measurement in analyzing the reliability of an english test for non-English major graduates. *Chinese Journal of Applied Linguistics*, 33(2), 87-102.
- Hedge, J. W., & Teachout, M. S. (2000). Exploring the concept of acceptability as a criterion for evaluating performance measures. *Group & Organization Management*, 25(1), 22-44.
- Heidemeier, H., & Moser, K. (2009). Self-other agreement in job performance ratings: a meta-analytic test of a process model. *J Appl Psychol*, 94(2), 353-370.

- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, 16(3), 235-266.
- Iramaneerat, C., & Yudkowsky, R. (2007). Rater errors in a clinical skills assessment of medical students. *Eval Health Prof*, 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. *Adv Health Sci Educ Theory Pract*, 13(4), 479-493.
- Jie, Z., & Lianzhen, H. (2008). Study of sources of score variability in performance assessment using MFRM: A case of speaking test in PET Band3. *CELEA Journal*, 31(4), 40-49.
- Kassebaum, D. G. (1989). Change in medical education: the courage and will to be different. *Academic Medicine*, 64(8), 446-447.
- 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.
- Kasten, R., & Weintraub, Z. (1999). Rating errors and rating accuracy: A field experiment. *Human Performance*, 12(2), 137-153.
- Khoury, G. C., & Analoui, F. (2004). Innovative management model for performance appraisal: The case of the Palestinian Public Universities. *Management Research News*, 27(1/2), 56-73.
- Kim, H. J. (2011). *Investigating raters' development of rating ability on a second language speaking assessment*. Columbia University.
- Landy, F. J., & Farr, J. L. (1980). Performance rating. *Psychological Bulletin*, 87(1), 72-107.
- Lee, R. M. K. W., & Kwan, C. Y. (1997). The use of problem-based learning in medical education. *Journal of Medical Education*, 1(2), 149-157.
- Linacre, J. M. (1996). *Generalizability theory and many-facet Rasch measurement* (3 ed.). Norwood, NJ: Ablex.
- Linacre, J. M. (2012). *A user's guide to FACETS Rasch-model computer programs. Program Manual 3.70.0*. Beaverton, Oregon: Winsteps.com.

- Locke, E. A. (1996). Motivation through conscious goal setting. *Applied & Preventive Psychology, 5*, 117-124.
- MacMillan, P. D. (2000). Classical, generalizability, and multifaceted Rasch detection of interrater variability in Large, Sparse Data Sets. *The Journal of Experimental Education, 68*(2), 167-190.
- McManus, I. C., Thompson, M., & Mollon, J. (2006). Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Med Educ, 6*(42), 1-22.
- Moore, I., & Poikela, S. (2011). *Evaluating problem-based learning initiatives*. New York: Taylor & Francis.
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational and goal-based perspectives*. Thousand Oaks, CA: Sage.
- Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *J Appl Psychol, 89*(1), 158-164.
- Papinczak, T., Young, L., Groves, M., & Haynes, M. (2007). An analysis of peer, self, and tutor assessment in problem-based learning tutorials. *Med Teach, 29*(5), e122-132.
- Park, H. (2006). The effect of interpersonal affect on performance ratings: Is it real or spurious? *Seoul Journal of Business, 12*(2), 37-56.
- Payne, S. C., Horner, M. T., Boswell, W. R., Schroeder, A. N., & Stine-Cheyne, K. J. (2009). Comparison of online and traditional performance appraisal systems. *Journal of Managerial Psychology 24*(6), 526-544.
- Roch, S. G. (2007). Why convene rater teams: An investigation of the benefits of anticipated discussion, consensus, and rater motivation. *Organizational Behavior and Human Decision Processes, 104*, 14-29.
- Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings : Assessing the psychometric quality of rating data. *Psychological Bulletin, 88*(2), 413-428.
- Salvemini, N. J., Reilly, R. R., & Smither, J. W. (1993). The influence of rater motivation on assimilation effects and accuracy in performance ratings. *Organizational Behavior and Human Decision Processes, 55*, 41-60.

- Schmitt, N., & G., K. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*, 210–222.
- Simsek, B., Pakdil, F., Dengiz, B., & Testik, M. C. (2013). Driver performance appraisal using GPS terminal measurements: A conceptual framework. *Transportation Research Part C, 26*, 49–60.
- Sullivan, G. M., & Feinn, R. (2012). Using effect size – or why the p value is not enough. *Journal of Graduate Medical Education, 4*, 279-282.
- Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology, 73*(3), 497-506.
- Tangdhanakanond, K., Pitiyanuwat, S., & Archwamety, T. (2006). Assessment of achievement and personal qualities under constructionist learning environment. *Education, 126*(3), 495–503.
- Tziner, A., Murphy, K., Cleveland, J. N., Yavo, A., & Hayoon, E. (2008). A New old question: Do contextual factors relate to rating behavior: An investigation with peer evaluations. *International Journal of Selection and Assessment, 16*(1), 59-67.
- Vecchio, R. P., & Anderson, R. J. (2009). Agreement in self–other ratings of leader effectiveness: The role of demographics and personality. *International Journal of Selection and Assessment, 17*(2), 165-179.
- Walsh, A. (2005). *The Tutor in Problem Based Learning: A Novice's Guide*. Hamilton ON Canada: Faculty of Health Sciences, McMaster University.
- Wang, X. M., Wong, K. F., & Kwong, J. Y. (2010). The roles of rater goals and ratee performance levels in the distortion of performance ratings. *J Appl Psychol, 95*(3), 546-561.
- Wheelan, S. A. (2005). *The handbook of group research and practice*. Thousand Oaks, CA: SAGE Publications.
- Wolfe, E. W., Chiu, C. W. T., & M., M. C. (2000). *Detecting rater effects in simulated data with a multifaceted Rash rating scale model* (Vol. 5). Norwood, NJ: Ablex.

- Wong, K. F., & Kwong, J. Y. (2007). Effects of rater goals on rating patterns: evidence from an experimental field study. *J Appl Psychol*, *92*(2), 577-585.
- Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Löf, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370.





ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ก

รายนามผู้ทรงคุณวุฒิในการวิจัย

การสำรวจและคัดเลือกองค์ประกอบและตัวบ่งชี้ของโมเดลเชิงสาเหตุของคุณภาพการให้คะแนน
ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

- | | |
|--------------------------------|--|
| 1. รศ. ดร.พิมพ์พันธ์ เดชะคุปต์ | คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย |
| 2. รศ. พญ.วัลลีย์ สัตยาศัย | คณะแพทยศาสตร์ มหาวิทยาลัยธรรมศาสตร์ |
| 3. ผศ. ดร.อัญชลี ชยานุวัชร | สำนักวิชาศิลปะศาสตร์ มหาวิทยาลัยวลัยลักษณ์ |
| 4. ผศ. ดร.ปิยะพงศ์ โชติพันธุ์ | สำนักวิชาเทคโนโลยีการเกษตร มหาวิทยาลัยวลัยลักษณ์ |
| 5. ผศ. ดร.เกียรติกำจร กุศล | สำนักวิชาพยาบาลศาสตร์ มหาวิทยาลัยวลัยลักษณ์ |

การตรวจสอบความตรงเชิงเนื้อหาของเครื่องมือที่ใช้ในการวิจัย

- | | |
|-------------------------------|--|
| 1. รศ. ดร.ศิริเดช สุชีวะ | คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย |
| 2. ผศ. ดร.ณัฐภรณ์ หลาวทอง | คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย |
| 3. ผศ. ดร.อัญชลี ชยานุวัชร | สำนักวิชาศิลปะศาสตร์ มหาวิทยาลัยวลัยลักษณ์ |
| 4. ผศ. ดร.ปิยะพงศ์ โชติพันธุ์ | สำนักวิชาเทคโนโลยีการเกษตร มหาวิทยาลัยวลัยลักษณ์ |
| 5. ผศ. ดร.เกียรติกำจร กุศล | สำนักวิชาพยาบาลศาสตร์ มหาวิทยาลัยวลัยลักษณ์ |

การตรวจสอบความถูกต้องและความเหมาะสมของคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่ม
ย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหา
เป็นฐานระดับอุดมศึกษา

- | | |
|-------------------------------|--|
| 1. อ.ดร. นพ.दनัย ว่างสุตรค | คณะแพทยศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย |
| 2. ผศ. ดร.อัญชลี ชยานุวัชร | สำนักวิชาศิลปะศาสตร์ มหาวิทยาลัยวลัยลักษณ์ |
| 3. ผศ. ดร.ปิยะพงศ์ โชติพันธุ์ | สำนักวิชาเทคโนโลยีการเกษตร มหาวิทยาลัยวลัยลักษณ์ |
| 4. ผศ. ดร.เกียรติกำจร กุศล | สำนักวิชาพยาบาลศาสตร์ มหาวิทยาลัยวลัยลักษณ์ |

ภาคผนวก ข
เครื่องมือที่ใช้ในการวิจัย
แบบสอบถามคุณลักษณะของผู้ประเมิน

ส่วนที่ 1 แบบสอบถามสำหรับผู้ประเมิน

1.1 คำชี้แจง โปรดเติมข้อมูลลงในช่องว่าง หรือ ทำเครื่องหมาย ✓ ลงในช่องว่างให้ตรงกับความเป็นจริงของท่าน

1. อายุ.....ปี
2. เพศ 1) ชาย 2) หญิง
3. สังกัดคณะ..... สังกัดมหาวิทยาลัย.....
4. สถานะภาพ 1) อาจารย์ประจำกลุ่ม PBL 2) นิสิต/นักศึกษาในกลุ่ม PBL
5. ตอบเฉพาะนิสิต/นักศึกษา
 - 5.1 กำลังศึกษาอยู่ชั้นปีที่.....
 - 5.2 ผลการเรียนเฉลี่ย (GPA).....
 - 5.3 เคยเรียนในรายวิชาที่ใช้ PBL จำนวน.....วิชา
6. ตอบเฉพาะอาจารย์
 - 6.1 ระดับการศึกษาสูงสุด 1) ปริญญาตรี 2) ปริญญาโท 3) ปริญญาเอก
 - 6.2 ประสบการณ์เกี่ยวกับการสอนโดยใช้ PBL ระยะเวลา.....ปี

1.2 คำชี้แจง โปรดเติมเครื่องหมาย ✓ ลงในช่องว่างที่ตรงกับความคิดเห็นของท่านมากที่สุด ข้อละ 1 คำตอบ

คำถาม		ใช่	ไม่ใช่
1	ฉันคิดว่า การให้คะแนนงานกลุ่มแบบเท่ากันทุกคนจะช่วยให้เกิดความสามัคคีและลดการแข่งขันกันภายในกลุ่ม		
2	ฉันคิดว่า ผลการประเมินต้องมีการจำแนกจุดแข็งและจุดอ่อนของผู้ถูกประเมิน		
3	ฉันนิยมให้คะแนนประเมินบุคคลแบบภาพรวมมากกว่าการประเมินตามคุณลักษณะย่อย เพราะทำให้ผู้ถูกประเมินได้คะแนนมากขึ้น		
4	ฉันคิดว่า สัมพันธภาพระหว่างผู้ประเมินและผู้ถูกประเมินจะมีผลต่อการประเมินให้คะแนน		
5	ฉันมักจะให้คะแนนประเมินต่ำกว่าความเป็นจริง เพื่อให้ผู้ถูกประเมินเกิดการพัฒนาตนเองได้ดียิ่งขึ้น		
6	ฉันเคยมีประสบการณ์ในการให้คะแนนประเมินผลบุคคลหรือผลงาน		
7	ฉันเคยได้รับการอบรม/ได้รับข้อมูลคำแนะนำเกี่ยวกับวิธีการประเมินให้คะแนนบุคคลหรือผลงาน		
8	การประเมินผลการปฏิบัติจะให้ประโยชน์แก่ผู้ประเมินและผู้ถูกประเมิน		
9	การประเมินเป็นการตัดสินตามความชอบหรือความพึงพอใจของผู้ประเมิน		
10	เป้าหมายของการประเมิน คือ สารสนเทศเพื่อการปรับปรุงพัฒนา หรือการตัดสินคุณค่าของสิ่งที่ประเมิน		

1.3 คำชี้แจง โปรดเติมเครื่องหมาย ✓ ลงในช่องว่างที่ตรงกับระดับความคิดเห็นของท่านมากที่สุด ข้อละ 1 คำตอบ

ประเด็นคำถาม	ระดับความคิดเห็น				
	จริงที่สุด	ค่อนข้างจริง	จริงปานกลาง	ค่อนข้างไม่จริง	ไม่จริงเลย
1	ฉันเตรียมพร้อมในทุกๆ เรื่อง				
2	ฉันมักจะวางสิ่งของไม่เป็นที่เป็นทาง				
3	ฉันใส่ใจรายละเอียดในการทำงาน				
4	ฉันมักจะทำสิ่งต่างๆ ให้เป็นเรื่องวุ่นวาย				
5	ฉันทำงานที่ได้รับมอบหมายในทันที				

ประเด็นคำถาม		ระดับความคิดเห็น				
		จริง ที่สุด	ค่อนข้าง จริง	จริงปน กลาง	ค่อนข้าง ไม่จริง	ไม่จริง เลย
6	ฉันมักลืมเก็บสิ่งของที่ใส่แล้วให้กลับเข้าที่ตามเดิม					
7	ฉันชอบปฏิบัติตามคำสั่ง					
8	ฉันเลี้ยงที่จะต้องทำงานที่ได้รับมอบหมาย					
9	ฉันปฏิบัติตามแผนกำหนดการของงาน					
10	ฉันทำงานอย่างละเอียดถี่ถ้วน					
11	ฉันยินดีและเต็มใจทำหน้าที่เป็นผู้ประเมินในวิชา PBL					
12	เมื่อฉันทำหน้าที่เป็นผู้ประเมินในวิชา PBL ผลการประเมินของฉันจะต้องมีความถูกต้อง					
13	ถ้าฉันเป็นผู้ประเมินที่ยึดมั่นหลักการมากไป อาจทำให้เกิดผลเสียตามมาได้					
14	ฉันคิดว่าเป็นเรื่องปกติ ถ้าจะประเมินคะแนนให้ตนเองหรือเพื่อนในกลุ่มแบบสูงๆ					
15	ฉันคิดว่า คะแนนจากการประเมินจะเป็นข้อมูลป้อนกลับเพื่อพัฒนาความสามารถของผู้ถูกประเมิน					
16	ฉันคิดว่า ควรมีระบบการตรวจสอบซ้ำในด้านความเหมาะสมและถูกต้องของคะแนนที่ได้จากผู้ประเมินทุกๆ คน					
17	ฉันยินดีที่จะได้รับการตรวจสอบในฐานะการทำหน้าที่เป็นผู้ประเมิน					
18	ถ้าไม่มีกฎหรือข้อบังคับระบุไว้ ผู้ปฏิบัติงานไม่จำเป็นต้องรับผิดชอบผลจากการกระทำที่เกิดขึ้น					
19	ผู้ประเมินต้องแสดงความรับผิดชอบต่อผู้ถูกประเมินโดยให้การประเมินอย่างเที่ยงธรรม รับฟังความคิดเห็น และให้ข้อเท็จจริง					

ประเด็นคำถาม	ระดับความคิดเห็น					
	จริง ที่สุด	ค่อนข้าง จริง	จริงปน กลาง	ค่อนข้าง ไม่จริง	ไม่จริง เลย	
20	ผู้ประเมินต้องแสดงความรับผิดชอบต่อผู้ที่มอบหมายงานประเมินให้ โดยการทำงานอย่างเต็มความสามารถ ส่งผลประเมินทันตามกำหนดเวลา และให้ผลการประเมินที่ถูกต้อง					
21	ผู้เรียนในกลุ่ม PBL จะต้องมีความท้าทายในการกำหนดปัญหา และวัตถุประสงค์การเรียนรู้					
22	ในการเรียนการสอนแบบ PBL ผู้เรียนจะได้รับความรู้จากการไปศึกษาค้นคว้าด้วยตนเอง					
23	ทักษะในการเรียนแบบ PBL ที่สำคัญ คือ การคิดวิเคราะห์ สังเคราะห์ ประเมิน และการถ่ายทอด					
24	การแสดงความคิดเห็นและเนื้อหาที่พูดของสมาชิกในกลุ่ม PBL จะบ่งบอกถึงความรู้และความเข้าใจต่อเรื่องนั้นๆ					
25	กระบวนการกลุ่ม PBL จะสิ้นสุดลงเมื่อกลุ่มสามารถหาข้อมูลได้ครบถ้วนต่อการพิสูจน์สมมติฐานทั้งหมด ตลอดจนสามารถสรุปเกี่ยวกับหลักการและแนวทางในการนำความรู้ไปใช้					
26	เมื่อเป็นผู้ประเมิน ฉันได้พิจารณาเลือกใช้ข้อมูลที่ตรงกับวัตถุประสงค์หรือสิ่งต้องการจะประเมิน					
27	เมื่อเป็นผู้ประเมิน ฉันใช้ข้อมูลที่เป็นความคิดเห็นของตนเองหรือผู้ที่เกี่ยวข้อง มากกว่าข้อมูลที่ได้จากการไปสังเกตโดยตรง					
28	ฉันใช้ข้อมูลหลายๆ ด้าน เพื่อประกอบการตัดสินใจให้คะแนนประเมิน					
29	ฉันได้พิจารณาซึ่งน้ำหนักความน่าเชื่อถือของข้อมูลก่อนการตัดสินใจให้คะแนนประเมิน					

แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (กลุ่มที่__)

คำชี้แจง โปรดกรอคะแนนที่ตรงมากที่สุด กับพฤติกรรมของนิสิต/ นักศึกษาภายในกลุ่ม โดยเกณฑ์การประเมินมีดังนี้

- 4 = ดีมาก หมายถึง มีทุกครั้ง หรือ มีมากกว่าครึ่ง มีความถี่ และเกิดประโยชน์ต่อกลุ่ม ตามรายการประเมินพฤติกรรมในข้อนี้ๆ ตลอดจนกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 3 = ดี หมายถึง มีมากกว่าครึ่ง และ มีครบถ้วน ตามรายการประเมินพฤติกรรมในข้อนี้ๆ ตลอดจนกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 2 = ปานกลาง หมายถึง มีน้อยครั้ง หรือ มีไม่ครบถ้วน ตามรายการประเมินพฤติกรรมในข้อนี้ๆ ตลอดจนกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 1 = ควรปรับปรุง หมายถึง ไม่มี พฤติกรรมตามรายการประเมินในข้อนี้ๆ ตลอดจนกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

สถานะของผู้ประเมิน 1) อาจารย์ 2) นักศึกษาชื่อ.....

	ชื่อนักศึกษา/ ระดับคะแนน								
	1	2	3	4	5	6	7	8	9
1	ตั้งคำถามที่นำไปสู่การทำความเข้าใจในสถานการณ์ปัญหา								
2	ชี้/ระบุข้อมูลที่สำคัญจากสถานการณ์ปัญหา								
3	คิดวิเคราะห์โดยตรงอย่างเป็นระบบ								
4	ประยุกต์เอาความรู้หรือประสบการณ์มาใช้กับสถานการณ์ใน PBL								
5	ตั้งสมมติฐานที่เป็นไปได้สำหรับสาเหตุหรือแนวทางการแก้ปัญหา								
6	อธิบายเชื่อมโยงประเด็นต่างๆ อย่างเป็นเหตุ เป็นผล								

	รายการประเมิน	ชั้นศึกษา/ ระดับคะแนน											
		1	2	3	4	5	6	7	8	9			
7	ระบุประเด็นที่ต้องการเรียนรู้จากสถานการณ์ใน PBL ได้												
8	ค้นหาข้อมูลความรู้ที่ทันสมัยและเป็นประโยชน์												
9	ค้นคว้าหาข้อมูลได้อย่างครอบคลุมและครบถ้วน												
10	ถ่ายทอดเนื้อหาความรู้ได้อย่างชัดเจน เข้าใจง่าย												
11	นำความรู้ที่ได้จากการค้นคว้ามาใช้อธิบายตอบคำถามหรือ สมมติฐานจากสถานการณ์ของปัญหา												
12	ชี้แนะหรือนำพากลุ่มไปสู่แนวทางในการเรียนรู้/เป้าหมายความสำเร็จ												
13	ให้เกิดมิตรเพื่อนสมาชิกในกลุ่มอย่างเท่าเทียมกัน												
14	ให้ความร่วมมือ/ ทำงานร่วมกับผู้อื่นได้												
15	สนใจมีส่วนร่วม/ตั้งใจทำกิจกรรมกลุ่มอย่างสม่ำเสมอ												
16	แสดงออกถึงการมีความคิดริเริ่ม/ สร้างสรรค์												
17	สามารถให้ข้อมูลย้อนกลับ/ สะท้อนความคิดเห็นแก่กลุ่มได้												
18	สามารถสรุปเนื้อหาที่สำคัญของการเรียนรู้ในครั้งนี้												
19	แสดงความพึงพอใจต่อการเรียนแบบ PBL ในครั้งนี้												





คณะกรรมการพิจารณาจริยธรรมการวิจัยในคน กลุ่มสหสถาบัน ชุดที่ 1 จุฬาลงกรณ์มหาวิทยาลัย
อาคารสถาบัน 2 ชั้น 4 ซอยจุฬาลงกรณ์ 62 ถนนพญาไท เขตปทุมวัน กรุงเทพฯ 10330
โทรศัพท์: 0-2218-8147 โทรสาร: 0-2218-8147 E-mail: eccu@chula.ac.th

COA No. 070/2557

ใบรับรองโครงการวิจัย

โครงการวิจัยที่ 040.1/57 : การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสซ์

ผู้วิจัยหลัก : นางสาวศุภรัตน์ อิงชาติเจริญ

หน่วยงาน : คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

คณะกรรมการพิจารณาจริยธรรมการวิจัยในคน กลุ่มสหสถาบัน ชุดที่ 1 จุฬาลงกรณ์มหาวิทยาลัย ได้พิจารณา โดยใช้หลัก ของ The International Conference on Harmonization – Good Clinical Practice (ICH-GCP) อนุมัติให้ดำเนินการศึกษาวิจัยเรื่องดังกล่าวได้

ลงนาม.....
(รองศาสตราจารย์ นายแพทย์ปริดา ทิสนประดิษฐ) (ผู้ช่วยศาสตราจารย์ ดร.นันทรี ชัยชนะวงศาโรจน์)
ประธาน กรรมการและเลขานุการ

วันที่รับรอง : 21 เมษายน 2557

วันหมดอายุ : 20 เมษายน 2558

เอกสารที่คณะกรรมการรับรอง

- 1) โครงการวิจัย
- 2) ข้อมูลสำหรับกลุ่มประชากรหรือผู้มีส่วนร่วมในการวิจัยและใบยินยอมของกลุ่มประชากรหรือผู้มีส่วนร่วมในการวิจัย
- 3) ผู้วิจัย
- 4) แบบสอบถาม



เลขที่โครงการวิจัย 040.1/57
วันที่รับรอง 21 เม.ย. 2557
วันหมดอายุ 20 เม.ย. 2558

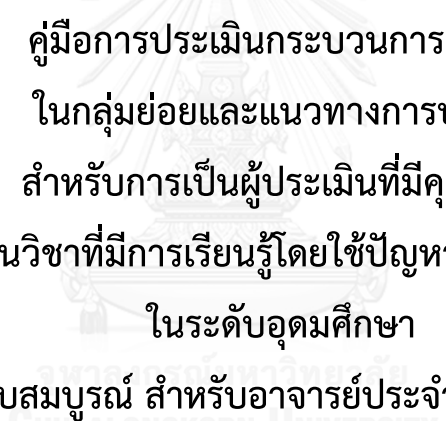
เงื่อนไข

1. ข้าพเจ้ารับทราบว่าเป็นการพิจารณาจริยธรรม หากดำเนินการเก็บข้อมูลการวิจัยก่อนได้รับการอนุมัติจากคณะกรรมการพิจารณาจริยธรรมการวิจัยฯ
2. หากใบรับรองโครงการวิจัยหมดอายุ การดำเนินการวิจัยต้องยุติ เมื่อต้องการต่ออายุต้องขออนุมัติใหม่ล่วงหน้าไม่ต่ำกว่า 1 เดือน พร้อมส่งรายงานความก้าวหน้าการวิจัย
3. ต้องดำเนินการวิจัยตามที่ระบุไว้ในโครงการวิจัยอย่างเคร่งครัด
4. ใช้เอกสารข้อมูลสำหรับกลุ่มประชากรหรือผู้มีส่วนร่วมในการวิจัย ใบยินยอมของกลุ่มประชากรหรือผู้มีส่วนร่วมในการวิจัย และเอกสารเชิญเข้าร่วมวิจัย (ถ้ามี) เฉพาะที่ประทับตราคณะกรรมการเท่านั้น
5. หากเกิดเหตุการณ์ไม่พึงประสงค์ร้ายแรงในสถานที่เก็บข้อมูลที่ขออนุมัติจากคณะกรรมการ ต้องรายงานคณะกรรมการภายใน 5 วันทำการ
6. หากมีการเปลี่ยนแปลงการดำเนินการวิจัย ให้ส่งคณะกรรมการพิจารณารับรองก่อนดำเนินการ
7. โครงการวิจัยไม่เกิน 1 ปี ส่งแบบรายงานสิ้นสุดโครงการวิจัย (AF 03-12) และบทคัดย่อผลการวิจัยภายใน 30 วัน เมื่อโครงการวิจัยเสร็จสิ้น สำหรับโครงการวิจัยที่เป็นวิทยานิพนธ์ให้ส่งบทคัดย่อผลการวิจัย ภายใน 30 วัน เมื่อโครงการวิจัยเสร็จสิ้น

ภาคผนวก ง

คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



คู่มือการประเมินกระบวนการเรียนรู้
ในกลุ่มย่อยและแนวทางการปฏิบัติ
สำหรับการเป็นผู้ประเมินที่มีคุณภาพ
ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน
ในระดับอุดมศึกษา
(ฉบับสมบูรณ์ สำหรับอาจารย์ประจำกลุ่ม PBL)

คำนำ

การพัฒนาคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพ ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาในครั้งนี้ เป็นส่วนหนึ่งของงานวิทยานิพนธ์เรื่อง “การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมิน ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสซ์” ของ ศุภรัตน์ อิงชาติเจริญ นิสิตหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา จุฬาลงกรณ์มหาวิทยาลัย โดยมี ผู้ช่วยศาสตราจารย์ ดร.กมลวรรณ ตังธนานนท์ และ รองศาสตราจารย์ ดร.โชติกา ภาษีผล เป็นอาจารย์ที่ปรึกษา ซึ่งมีเป้าหมายสำคัญในการจัดทำคู่มือ คือ เพื่อเสนอแนะแนวทางการปฏิบัติพร้อมทั้งเครื่องมือที่ใช้สำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย เพื่อให้เกิดการประเมินผลการเรียนรู้ได้อย่างมีคุณภาพทั้งในด้านมีความแม่นยำ และลดความคลาดเคลื่อนในการให้คะแนนการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

ในโอกาสนี้ผู้วิจัยขอขอบพระคุณท่านอาจารย์ที่ปรึกษา ท่านอาจารย์ผู้ทรงคุณวุฒิ ตลอดจนท่านคณาจารย์ และนิสิต/นักศึกษาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ซึ่งได้ให้ความอนุเคราะห์ช่วยเหลือเป็นอย่างดีทั้งในขั้นตอนของการพัฒนาเครื่องมือ การเก็บรวบรวมข้อมูล และการพัฒนาเป็นคู่มือนี้

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ศุภรัตน์ อิงชาติเจริญ
ผู้จัดทำคู่มือ

สารบัญ

บทที่

- 1 หลักการและวัตถุประสงค์ของคู่มือ
 - 1.1 หลักการและแนวคิด
 - 1.2 วัตถุประสงค์ของคู่มือ
 - 1.3 ประโยชน์ที่คาดว่าจะได้รับจากการใช้คู่มือ
 - 1.4 คำจำกัดความที่ใช้ในคู่มือ
- 2 แนวทางการปฏิบัติและเครื่องมือที่ใช้สำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย
 - 2.1 สรุปแนวคิด ทฤษฎี และผลการวิจัยที่เกี่ยวข้องกับการเรียนรู้โดยใช้ปัญหาเป็นฐาน
 - 2.2 แนวทางการปฏิบัติสำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา
 - 2.3 แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา
- 3 แนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ
 - 3.1 สรุปแนวคิด ทฤษฎี และผลการวิจัยที่เกี่ยวข้องกับปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ
 - 3.2 แนวทางการปฏิบัติสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ

รายการอ้างอิง

ภาคผนวก

แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย

บทที่ 1

หลักการและวัตถุประสงค์ของคู่มือ

1.1 หลักการและแนวคิด

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 แก้ไขเพิ่มเติม พ.ศ. 2545 และ พ.ศ. 2553 กล่าวว่า การศึกษาเป็นกระบวนการเรียนรู้เพื่อความเจริญงอกงามของบุคคลและสังคม โดยมีหลักสำคัญ คือ ผู้เรียนทุกคนมีความสามารถเรียนรู้และพัฒนาตนเองได้ และผู้เรียนมีความสำคัญที่สุด ซึ่งกระบวนการจัดการศึกษาต้องส่งเสริมให้ผู้เรียนสามารถพัฒนาตามธรรมชาติและเต็มตามศักยภาพ ส่วนการประเมินผู้เรียนควรใช้วิธีการที่หลากหลาย เช่น ประเมินพัฒนาการของผู้เรียน สังเกตพฤติกรรมการเรียน การร่วมกิจกรรม และการทดสอบ ตามความเหมาะสมของแต่ละระดับและรูปแบบการศึกษา นอกจากนี้ยังกำหนดให้มีระบบประกันคุณภาพการศึกษาเพื่อพัฒนาคุณภาพและมาตรฐานการศึกษาทุกระดับ (สำนักงานคณะกรรมการการศึกษา, 2542, 2545, 2553) การเรียนรู้โดยใช้ปัญหาเป็นฐาน (Problem Based Learning: PBL) เป็นการเรียนรู้ซึ่งมาจากการขับเคลื่อนของปัญหาที่เป็นบทเรียน โดยผู้เรียนต้องทำการระบุดูวัตถุประสงค์ที่จำเป็นสำหรับการเรียนรู้จากนั้นจะนำความรู้ที่ได้จากการไปศึกษาค้นคว้าด้วยตนเองมาใช้ในการแก้ไขปัญหา ซึ่งเป้าหมายสำคัญจะอยู่ที่การใช้ปัญหาเป็นจุดเริ่มต้นสำหรับการเรียนรู้ และประโยชน์ที่ได้จากความรู้ใหม่ คือ การเสริมสร้างความเข้าใจและเกิดเป็นความรู้ที่ติดตัวไป (Walsh, 2005) นอกจากนี้การประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานจะ ใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (tutorial-based assessment) หรือประเมินผลจากการปฏิบัติที่เกิดขึ้นจริงในบริบทของการเรียนรู้ ซึ่งเป็นรูปแบบหนึ่งของการประเมินทางเลือก (alternative assessment) ดังนั้นการเรียนรู้โดยใช้ปัญหาเป็นฐานจึงประกอบด้วยวิธีการเรียนรู้และการประเมินผลการเรียนรู้ที่สอดคล้องกับพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 นอกจากนี้ยังเหมาะสมกับการจัดการเรียนการสอนในศตวรรษที่ 21 ที่ต้องการให้มนุษย์ชาติกลายเป็นนักสร้างสรรค์ อนุรักษ์วัฒนธรรม สร้างนวัตกรรมทางเทคโนโลยี และออกแบบวิถีชีวิตและการปกครอง ซึ่งทักษะเหล่านี้ล้วนได้แก่ การคิดเชิงวิพากษ์ การแก้ไขปัญหา ความร่วมมือ ความสร้างสรรค์ การชี้แนะตนเอง ความเป็นผู้นำ การปรับตัวความรับผิดชอบ จิตสำนึกต่อโลก และทักษะการสืบค้น (Barell, 2010)

การประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานซึ่งใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อยจากผู้ประเมินที่มีสถานะต่างกัน ได้แก่ อาจารย์ เพื่อน และตนเอง (ผู้เรียน) เพื่อที่จะได้ข้อมูลสำคัญในการใช้แนะแนวทางการพัฒนาความสามารถ หรือใช้ตัดสินความสามารถของผู้เรียน แต่จุดอ่อนของวิธีการประเมินนี้ คือ ปัญหาความแม่นยำในการให้คะแนนระหว่างการประเมินโดยอาจารย์ เพื่อน และตนเอง (Eva, 2001, Papinczak et al., 2007) ซึ่งผลการศึกษาเปรียบเทียบความสัมพันธ์ระหว่างการให้คะแนนโดยเพื่อนและโดยอาจารย์ พบว่ามีความสัมพันธ์กันตั้งแต่ระดับต่ำจนถึงระดับปานกลาง

นอกจากนี้การให้คะแนนตนเองมีแนวโน้มที่จะมีความสัมพันธ์ระดับต่ำกับการให้คะแนนโดยอาจารย์หรือเพื่อน (Papinczak et al., 2007; Eva, 2001) และในการแก้ไขปัญหาการขาดความมั่นใจในการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานจากผู้ประเมินอาจารย์ เพื่อน และตนเอง ได้มีหลายงานวิจัยที่แนะนำให้ใช้เทคนิคต่างๆ เช่น ฝึกรอบมการใช้เกณฑ์การประเมิน การสร้างแบบประเมินให้ตรงตามคุณลักษณะที่จะประเมิน เป็นต้น แต่พบว่า วิธีการเหล่านี้ยังไม่สามารถแก้ไขปัญหาให้หมดไปได้ (Eva, 2001)

จากแนวคิดและสภาพปัญหาที่กล่าวมาในข้างต้น ผู้จัดทำจึงพัฒนาคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา เพื่อให้เกิดการประเมินผลการเรียนรู้ได้อย่างมีคุณภาพทั้งในด้านความแม่นยำ ลดความคลาดเคลื่อน มีความเที่ยงตรง และยุติธรรม โดยนำแนวคิดมาจากตัวแปรในงานวิทยานิพนธ์ของผู้จัดทำ ซึ่งมีชื่อเรื่อง “การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสส์” มาประยุกต์ใช้ในการพัฒนาคู่มือนี้

1.2 วัตถุประสงค์ของคู่มือ

1. ให้แนวทางการปฏิบัติและเครื่องมือที่ใช้สำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยระหว่างกลุ่มผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา
2. ให้แนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพทั้งในด้านความแม่นยำและลดความคลาดเคลื่อนในการให้คะแนนการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

1.3 ประโยชน์ที่คาดว่าจะได้รับจากการใช้คู่มือ

1. ได้แนวทางการปฏิบัติและเครื่องมือที่ใช้สำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยสำหรับกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา
2. ได้แนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพทั้งในด้านความแม่นยำและลดความคลาดเคลื่อนในการให้คะแนนการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

3. สามารถใช้เป็นแนวทางและเครื่องมือสำหรับการทำวิจัยด้านการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

1.4 คำจำกัดความที่ใช้ในคู่มือ

การเรียนรู้โดยใช้ปัญหาเป็นฐาน หมายถึง รูปแบบของการเรียนการสอนที่ใช้การขับเคลื่อนของปัญหาเป็นจุดเริ่มต้นสำหรับการเรียนรู้ เพื่อที่จะเสริมสร้างความเข้าใจและพัฒนาเป็นความรู้ที่ติดตัวไป โดยมีกระบวนการเรียนรู้ใน 9 ขั้นตอน ได้แก่ 1) การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ 2) การกำหนดประเด็นปัญหา 3) การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหานั้นเป็นไปได้จากความรู้และประสบการณ์ที่มีอยู่ 4) การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล 5) การระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา 6) การศึกษาค้นคว้าด้วยตนเอง 7) การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้เรียนรู้มาต่อกลุ่ม 8) การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ และ 9) การสะท้อนความรู้ในบทเรียน โดยทำการสรุปทั้งในด้านประเด็นปัญหา ผลจากการไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน ตลอดจนกระบวนการในการแก้ปัญหา ส่วนการวัดและประเมินผลการเรียนรู้จะใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อยจากกลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง

การประเมินกระบวนการเรียนรู้ในกลุ่มย่อย หมายถึง การพิจารณาตัดสินเกี่ยวกับความสามารถหรือทักษะการปฏิบัติของผู้เรียนในรายวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ตามกระบวนการเรียนรู้ทั้ง 9 ขั้นตอน ร่วมกับการประเมินคุณลักษณะอันพึงประสงค์ที่ได้จากการเรียนรู้โดยใช้ปัญหาเป็นฐาน ได้แก่ ภาวะผู้นำ ความเคารพต่อสมาชิกในกลุ่ม และเจตคติต่อการเรียนรู้โดยใช้ปัญหาเป็นฐาน

กลุ่มผู้ประเมิน หมายถึง บุคคลผู้ทำหน้าที่ในการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ซึ่งได้แก่ อาจารย์ เพื่อนนักศึกษา และตัวนักศึกษาเอง ซึ่งบุคคลทั้งสามฝ่ายนี้จะมีความแตกต่างกันในด้านของระดับความสัมพันธ์และบทบาทหน้าที่ต่อผู้ถูกประเมิน

อาจารย์ หมายถึง บุคลากรวิชาชีพที่ทำหน้าที่หลักทางด้านการเรียนการสอน และการส่งเสริมให้เกิดการเรียนรู้ผ่านวงจรของกระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐาน ในสถาบันระดับอุดมศึกษา

เพื่อน หมายถึง นักศึกษาระดับอุดมศึกษาในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งจะต้องทำหน้าที่ในการประเมินนักศึกษาที่อยู่ภายในกลุ่มเดียวกัน

ตนเอง หมายถึง ผู้เรียนซึ่งเป็นนักศึกษาระดับอุดมศึกษาในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งจะต้องทำหน้าที่ในการประเมินตนเอง

บทที่ 2

แนวทางการปฏิบัติและเครื่องมือที่ใช้สำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย

เนื้อหาสาระในบทนี้ได้แบ่งการนำเสนอออกเป็น 3 ตอน ได้แก่ 1) สรุปแนวคิด ทฤษฎี และผลการวิจัยที่เกี่ยวข้องกับการเรียนรู้โดยใช้ปัญหาเป็นฐาน 2) แนวทางการปฏิบัติสำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา และ 3) แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ซึ่งรายละเอียดของแต่ละตอนมีดังนี้

2.1 สรุปแนวคิด ทฤษฎี และผลการวิจัยที่เกี่ยวข้องกับการเรียนรู้โดยใช้ปัญหาเป็นฐาน

การเรียนรู้จากปัญหาเป็นแนวทางพื้นฐานอย่างหนึ่งสำหรับการดำรงชีวิตของมนุษย์ ซึ่งความพยายามในการแก้ไขปัญหาเป็นเรื่องที่พบได้ในชีวิตประจำวัน โดยประสบการณ์ที่ได้จากการเรียนรู้ในสถานการณ์ปัญหาจะก่อให้เกิดความรู้และข้อมูลต่างๆ ที่สามารถนำไปประยุกต์ใช้กับปัญหาอื่นๆ ได้ในอนาคต และการได้มีโอกาสใช้ความรู้ต่างๆ นี้บ่อยครั้งยิ่งจะทำให้เกิดเป็นความจำระยะยาว ซึ่งจะต่างจากความรู้ที่ได้จากการอ่านที่มักเป็นความจำระยะสั้น ดังนั้นการเรียนรู้โดยใช้ปัญหาเป็นฐานจึงเป็นการเรียนรู้ที่เกิดจากกระบวนการทำความเข้าใจหรือแก้ไขปัญหาในสถานการณ์ต่างๆ ซึ่งไม่จัดเป็นวิธีการเรียนรู้อันแปลกใหม่จากเดิม (Barrows & Tamblyn, 1980) การเรียนรู้โดยใช้ปัญหาเป็นฐานได้มีการริเริ่มใช้ในหลักสูตรแพทยศาสตร์ของ McMaster University ประเทศแคนาดา ในปี 1960 ซึ่งในช่วงแรกยังไม่เป็นที่ยอมรับมากนักแต่ภายหลังได้มีผลการศึกษาเปรียบเทียบ พบว่า ประสิทธิภาพของการเรียนและความพึงพอใจของผู้เรียนและผู้สอนจากการเรียนรู้โดยใช้ปัญหาเป็นฐานมีมากกว่าระบบการเรียนการสอนแบบดั้งเดิม (Kassebaum, 1989) นอกจากนี้ยังมีการนำมาใช้ในการเรียนการสอนของสาขาต่างๆ ตั้งแต่ระดับโรงเรียนจนถึงมหาวิทยาลัย ในส่วนของประโยชน์จากการเรียนรู้โดยใช้ปัญหาเป็นฐานที่ผู้เรียนจะได้รับ คือ ได้เรียนรู้ตัวอย่างจากสถานการณ์ปัญหาซึ่งมีความใกล้เคียงกับสถานการณ์จริง ฝึกการทำงานร่วมกันอย่างมีประสิทธิภาพ พัฒนาทักษะการแก้ปัญหาและเกิดความรู้ที่ได้จากการลงมือกระทำ มีฐานความรู้ที่กว้าง ทราบวิธีการประยุกต์ใช้ความรู้ในสถานการณ์ต่างๆ เกิดความรู้สึกตื่นตัว มีแรงจูงใจ เกิดการเรียนรู้ด้วยตนเอง (self-directed learning) มีพัฒนาการของการกำหนด/ตัดสินใจได้โดยตนเอง (self-determination) มีความมั่นใจในทักษะการเรียนรู้และการให้เหตุผลของตน เกิดความประทับใจต่อองค์ความรู้ และจะช่วยส่งเสริมให้เกิดเป็นการเรียนรู้ตลอดชีวิต (lifelong learning) ส่วนข้อจำกัดของการเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ การจะประสบผลสำเร็จในการเรียนจะขึ้นอยู่กับวินัยของผู้เรียนที่จะเผชิญกับสิ่งที่ไม่รู้หรือปัญหาที่ยังไม่เข้าใจ และอาจารย์

จำเป็นต้องมีทักษะในการแนะแนวหรือให้คำแนะนำ สามารถสร้างหรือเตรียมสถานการณ์โจทย์ปัญหา สำหรับการเรียนรู้ ส่วนปัญหาอื่นที่อาจพบได้ในระยะแรกของการเรียน คือ อาจรู้สึกเครียดหรือกดดัน ซึ่งสามารถที่จะแก้ไขได้โดยการปฏิบัติตามขั้นตอนของกระบวนการเรียนรู้อย่างถูกต้อง (Barrows & Tamblyn, 1980; Hmelo-Silver, 2004; Lee & Kwan, 1997)

กระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐานประกอบด้วยรายละเอียดของกิจกรรม คือ ขั้นตอนแรกกลุ่มผู้เรียนจะได้รับสถานการณ์ปัญหาที่มีความซับซ้อนโดยมีข้อมูลให้เพียงเล็กน้อย ซึ่งกลุ่มผู้เรียนต้องใช้การคิดวิเคราะห์ ไตร่ตรอง จำแนกข้อเท็จจริงต่างๆ ที่เกี่ยวข้องในสถานการณ์ ตลอดจนการตั้งคำถามต่ออาจารย์ประจำกลุ่มเพื่อที่จะได้รับข้อมูลเพิ่มเติมและทำความเข้าใจสถานการณ์ปัญหานั้นได้ดียิ่งขึ้น ต่อมากลุ่มผู้เรียนจะต้องสร้างสมมติฐานที่เป็นแนวทางการแก้ไขสถานการณ์ปัญหาที่เป็นไปได้ เพื่อที่จะแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูลที่พวกเขาได้รวบรวมมา จากนั้นสิ่งที่สำคัญของกระบวนการเรียนรู้ก็คือ การที่กลุ่มผู้เรียนได้ช่วยกันระบุประเด็นสำหรับการเรียนรู้เพิ่มเติมเพื่อใช้ในการแก้ไขสถานการณ์ปัญหาที่พวกเขายังขาดอยู่ และกลุ่มจะต้องมีการแบ่งประเด็นกันสำหรับการไปศึกษาค้นคว้าด้วยตนเองต่อ (self-directed learning) ซึ่งจะทำให้เกิดเป็นความเชี่ยวชาญในประเด็นนั้นๆ โดยภายหลังจากการศึกษาค้นคว้าด้วยตนเองแล้วกลุ่มผู้เรียนจะกลับมารวมกันอีกครั้งเพื่อแลกเปลี่ยนสิ่งที่ได้ไปเรียนรู้มาและนำความรู้ใหม่ที่ได้มาทดสอบสมมติฐานที่กำหนดไว้หรือทำการปรับเปลี่ยนสมมติฐานอีกครั้งตามความรู้ใหม่ที่ได้รับ และในขั้นตอนสุดท้ายกลุ่มผู้เรียนจะมีการสะท้อนถึงบทเรียนที่ได้รับโดยมีการสรุปเกี่ยวกับปัญหาในแต่ละประเด็นและผลจากการไปศึกษาเพิ่มเติมด้วยตนเอง รวมถึงกระบวนการแก้ไขปัญหาของกลุ่ม ขณะที่บทบาทของกลุ่มถือเป็นส่วนหนึ่งที่มีความสำคัญต่อการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งข้อตกลงเบื้องต้นข้อหนึ่งในการเรียนรู้โดยใช้ปัญหาเป็นฐานคือ โครงสร้างของกลุ่มต้องมีขนาดเล็กเพื่อที่จะช่วยให้เกิดการกระจายของความคิดและช่วยเหลือกันในการแก้ไขสถานการณ์ปัญหาอย่างทั่วถึง นอกจากนี้ ยังมีผลการวิจัยที่พบว่า การอภิปรายโต้ตอบกันในกลุ่มขนาดเล็กจะเพิ่มการแก้ไขปัญหาและเกิดทักษะการคิดขั้นสูง (higher order thinking) ตลอดจนเป็นการส่งเสริมการแลกเปลี่ยนความรู้ ในการอภิปรายร่วมกันของสมาชิกกลุ่มจะต้องมีผู้ประสานงาน (facilitator) หรืออาจใช้วิธีการอื่น เช่น การแบ่งหน้าที่ การร่วมช่วยกันทุกฝ่าย เป็นต้น ซึ่งจะทำให้เกิดการจัดสรรการมีส่วนร่วมของสมาชิกทุกคนและเกิดความครอบคลุมของทุกประเด็น การศึกษา กระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐานจากที่ได้กล่าวมาสามารถสรุปได้เป็น 9 ขั้นตอน ดังนี้ 1) การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ 2) การกำหนดประเด็นปัญหา 3) การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้ จากความรู้และประสบการณ์ที่มีอยู่ 4) การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล 5) การระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา 6) การศึกษาค้นคว้าด้วยตนเอง 7) การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้ไปเรียนรู้มาในกลุ่ม 8) การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการ

ปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ และ 9) การสะท้อนความรู้ในบทเรียน โดยทำการสรุป ทั้งในด้านประเด็นปัญหา ผลจากการไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน และกระบวนการในการ แก้ปัญหา

บทบาทของอาจารย์ในการเรียนรู้โดยใช้ปัญหาเป็นฐานแบ่งเป็น 3 ระยะ ได้แก่ (Bridges, 1992) 1) ก่อนเริ่มสถานการณ์ปัญหา (prior) อาจารย์มีกิจกรรมที่สำคัญ คือ การทบทวนกระบวนการ สำหรับการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งในกรณีที่ผู้เรียนยังไม่คุ้นเคย อาจารย์ควรมีตัวอย่างของ กิจกรรมกลุ่มให้เห็นหรือให้ผู้เรียนทดลองทำกลุ่ม การรวบรวมแหล่งข้อมูลตลอดจนเครื่องมือและวัสดุ อุปกรณ์ที่จำเป็นต้องใช้สำหรับการเรียนรู้ในแต่ละโจทย์สถานการณ์ และการช่วยกำหนดบทบาท หน้าที่ของผู้เรียนภายในกลุ่ม เช่น ผู้นำกลุ่ม ผู้ประสานงาน และผู้จดบันทึก 2) ระหว่างการทำกิจกรรม (during) อาจารย์จะเป็นผู้สังเกตรูปแบบการมีส่วนร่วมของสมาชิกในกลุ่มโดยพิจารณาจาก 3 ประเด็น ได้แก่ ประเด็นแรก คือ เนื้อหาที่พูด ซึ่งจะแสดงให้เห็นว่าผู้เรียนมีความเข้าใจหรือมีความรู้จากการที่ได้ ไปศึกษาค้นคว้าด้วยตนเองมา ประเด็นที่สอง คือ กระบวนการ ซึ่งจะพิจารณาว่าใครเป็นผู้พูด มีการ แสดงความคิดเห็นอย่างไร และมีความเกี่ยวข้องกับเนื้อหาหรือไม่ และประเด็นที่สาม คือ ความถี่ ซึ่งจะ พิจารณาว่าการมีส่วนร่วมในกลุ่มของผู้เรียนมีมากน้อยเพียงไร ทั้งนี้อาจารย์อาจต้องมีการแทรกแซง กลุ่มถ้าผู้เรียนนำเสนอข้อมูลที่ไม่ตรงกับเนื้อหา หรือในกรณีที่ต้องการตั้งคำถามให้เกิดการคิด ในประเด็นสำคัญหรือเพื่อตรวจสอบการรับรู้ของผู้เรียน ตลอดจนการให้คำแนะนำเพื่อปรับปรุง กระบวนการกลุ่ม และในระหว่างที่ทำกลุ่มอาจารย์อาจให้ความรู้แก่ผู้เรียนในกรณีที่ผู้เรียนต้องการให้ อาจารย์ช่วยอธิบายในประเด็นที่ยังไม่เข้าใจ หรืออาจให้ช่วยเป็นแหล่งข้อมูลเหมือนกับการเป็น ผู้เชี่ยวชาญ นอกจากนี้อาจารย์ต้องมีการให้กำลังใจผู้เรียนเพื่อเป็นการสนับสนุนเขาหรือในกรณีที่ผู้เรียน บางคนที่ขาดความมั่นใจในตนเองอาจารย์จะต้องไม่กดดัน แต่ควรสร้างบรรยากาศการช่วยเหลือภายใน กลุ่มเพื่อให้ผู้เรียนรู้สึกผ่อนคลายในการพูดปัญหาที่เกิดขึ้นของเขาและอาจชี้ว่าความล้มเหลวที่เกิดจะ เป็นโอกาสที่ดีสำหรับการเรียนรู้ และ 3) ช่วงท้าย (following) ในช่วงสิ้นสุดของการทำกิจกรรม อาจารย์ต้องทำหน้าที่สำคัญ ได้แก่ การชักชวนให้ผู้เรียนพูดสะท้อนความคิดเห็นเกี่ยวกับสิ่งที่ตนเองได้ ทำและแนวทางในการพัฒนาตนเองต่อไป การให้ผู้เรียนเขียนแสดงความคิดเห็นเพื่อเป็นข้อมูลสะท้อน กลับเกี่ยวกับการกระทำและผลที่เกิดขึ้นทั้งของตนเองและเพื่อนในกลุ่ม และการตรวจสอบข้อมูล สะท้อนกลับของผู้เรียนและใช้ข้อมูลเหล่านี้เพื่อการแก้ไขและปรับปรุงให้ดีขึ้น

2.2 แนวทางการปฏิบัติสำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยระหว่างกลุ่มผู้ประเมิน อาจารย์ เพื่อน และตนเอง ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

การวัดและประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานในทางปฏิบัติสามารถประเมินได้หลาย รูปแบบโดยรวมอยู่ในการประเมินเพื่อปรับปรุง (formative evaluation) และการประเมินเพื่อตัดสิน

(summative evaluation) และสถานะของผู้ประเมินอาจเป็นการประเมินตนเองของผู้เรียน (self-assessment) การประเมินโดยเพื่อน (peer assessment) และการประเมินโดยอาจารย์ (instructor assessment) รูปแบบการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานที่ผ่านมามักจะเน้นประเมินในด้านสภาพแวดล้อมของการทำงานร่วมกัน (cooperative environment) สภาพแวดล้อมที่เอื้อให้เกิดการเรียนรู้ (supportive environment) กระบวนการเรียนรู้ (learning process) ลักษณะบุคลิกภาพ (personal characteristics) และกลไกการทำงานเป็นกลุ่ม (group dynamics) เพื่อหลีกเลี่ยงปัญหาการแข่งขันชิงเกรด ซึ่ง Norman (Norman, 1991 cite in Eva, 2001) ได้กล่าวไว้ว่า การประเมินโดยใช้วัตถุประสงค์เชิงพฤติกรรมเป็นหลักถือเป็นแนวคิดที่ขัดแย้งกับการเรียนรู้โดยใช้ปัญหาเป็นฐาน และการเรียนรู้ด้วยตนเอง (self-directed learning) เนื่องจากการประเมินรูปแบบนี้จะสนับสนุนให้เกิดการทดสอบซึ่งไม่ได้เป็นการสนองตอบต่อวัตถุประสงค์ในการเรียนรู้โดยใช้ปัญหาเป็นฐาน แต่รูปแบบการประเมินที่น่าจะเหมาะสมกับการเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ การประเมินภายในบริบทของการปฏิบัติ เนื่องจากการเรียนรู้โดยใช้ปัญหาเป็นฐานได้ถูกออกแบบมาเพื่อเน้นในเรื่องของการวิเคราะห์และแก้ไขปัญหานั้น เครื่องมือสำหรับใช้ในการประเมินควรจะมุ่งเน้นในด้านทักษะกระบวนการ และทัศนคติของผู้เรียนแต่ละคนภายในบริบทของบทเรียนหรือสถานการณ์นั้นๆ เช่น การให้เหตุผล การศึกษาค้นคว้าด้วยตนเอง (self-directed learning) การสื่อสาร แรงจูงใจภายใน การคิดอย่างมีวิจารณญาณ และความสามารถทำงานร่วมกันเป็นทีม ซึ่งรูปแบบการประเมินดังกล่าวนี้ คือ การประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (tutorial-based assessment) แต่จุดอ่อนของวิธีนี้คือ อาจพิสูจน์ได้ยากกว่า มีการบรรลุตามเป้าหมายของการเรียนหรือไม่ ซึ่งต่างจากรูปแบบของการทดสอบที่แสดงให้เห็นถึงการวัดในเรื่องความรู้ที่ชัดเจน นอกจากนี้การประเมินตนเองและการประเมินโดยเพื่อน ยังเป็นเรื่องที่โต้แย้งกันอยู่มากเนื่องจากการประเมินตนเองของผู้เรียนอาจไม่ได้บ่งชี้ถึงความสามารถที่แท้จริง และเป็นเรื่องปกติที่ความสัมพันธ์ของบุคคลจะทำให้เกิดมุมมองด้านบวกซึ่งจะทำให้เกิดการลังเลในการให้คะแนนตามวัตถุประสงค์การประเมิน แต่ในทางตรงข้ามข้อมูลเหล่านี้จะมีประโยชน์ในด้านการใช้เป็นข้อมูลสะท้อนกลับเพื่อให้เกิดการพัฒนาขึ้น (Eva, 2001: 244; Walsh, 2005)

หลักหรือองค์ประกอบที่สำคัญในการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานตามแนวคิดของ Moore & Poikela (2011) มี 9 ข้อ ดังนี้ 1) การประเมินควรสอบถามข้อมูลจากอาจารย์ นักศึกษา และผู้ที่เกี่ยวข้องเพื่อให้ได้ข้อมูลเกี่ยวกับประสิทธิภาพของการเรียน โดยเฉพาะนักศึกษาควรจะมีบทบาทเชิงรุกในการประเมินเกี่ยวกับพฤติกรรม ทักษะ ตลอดจนการเรียนรู้ที่เกิดขึ้นกับตนเอง 2) ทักษะสำคัญที่ทำให้เกิดการเรียนรู้อย่างมีประสิทธิภาพ คือ การประเมินตนเอง (self-evaluation) และการสะท้อนความคิด (critical reflection) ซึ่งทั้งสองทักษะนี้เกิดขึ้นในการเรียนรู้โดยใช้ปัญหาเป็นฐาน และในการประเมินด้านกระบวนการเรียนรู้โดยใช้ปัญหาเป็นฐานไม่ควรจะมุ่งเน้นที่การได้ทำหรือไม่ได้ทำ เนื่องจากไม่ช่วยให้เกิดการเสริมพลังแก่ผู้เรียน แต่ควรจะให้

ความสำคัญในการประเมินเกี่ยวกับวิธีการและผลลัพธ์ที่เกิดขึ้น 3) การประเมินควรจะให้ผู้ที่มีส่วนได้ส่วนเสียได้มีส่วนร่วมในการประเมินอย่างครอบคลุมและควรมีการแลกเปลี่ยนความรู้ที่เกิดขึ้นจากการประเมิน 4) การประเมินควรแยกแยะระหว่างการประเมินเพื่อปรับปรุง (formative evaluation) และการประเมินเพื่อตัดสิน (summative evaluation) โดยการประเมินเพื่อปรับปรุงควรให้ความสำคัญกับกระบวนการในการเรียน และการประเมินเพื่อตัดสินควรให้ความสำคัญกับผลกระทบและประสิทธิภาพที่เกิดขึ้นจากการเรียน 5) การประเมินโดยมุ่งเน้นผลลัพธ์ควรแยกจากการติดตามพฤติกรรมในการเรียน ความก้าวหน้าของกระบวนการ และบทบาทหน้าที่ในการเรียน ซึ่งการประเมินโดยมุ่งเน้นผลลัพธ์จะช่วยในการจำแนกลักษณะของการเปลี่ยนแปลงที่เกิดขึ้นจริงกับสิ่งที่เป็นมุมมองของผู้เรียนได้ 6) การประเมินควรเป็นไปอย่างเปิดเผยและซื่อสัตย์ซึ่งทำได้โดยทุกคนในกลุ่มมีการปฏิบัติตามกฎหรือมีความคุ้นเคยในการปฏิบัติตามระเบียบข้อตกลงร่วมกัน อันจะทำให้การประเมินได้ข้อมูลจริงที่เป็นประโยชน์ต่อการปรับปรุง 7) การประเมินที่มีประสิทธิภาพจะให้ข้อมูลย้อนกลับที่ต่อเนื่องเป็นประโยชน์ต่อการพัฒนา ปรับปรุง และช่วยในการตัดสินใจ นอกจากนี้การประเมินควรจะแทรกสอดและรวมอยู่ในทุกๆ กระบวนการของการเรียนรู้โดยใช้ปัญหาเป็นฐาน 8) สัดส่วนของประเด็นที่ใช้ในการประเมินควรสอดคล้องกับกิจกรรมที่เกิดขึ้นจริง และ 9) การประเมินควรเป็นไปอย่างมีจรรยาบรรณและสุภาพ

2.3 แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยระหว่างกลุ่มผู้ประเมินอาจารย์ เพื่อน และตนเอง ในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

การประเมินกระบวนการเรียนรู้ในกลุ่มย่อยเป็นการประเมินความสามารถหรือทักษะการปฏิบัติของผู้เรียนในรายวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งประกอบด้วยกระบวนการ 9 ขั้นตอน ได้แก่ 1) การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ 2) การกำหนดประเด็นปัญหา 3) การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้จากความรู้และประสบการณ์ที่มีอยู่ 4) การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล 5) การระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา 6) การศึกษาค้นคว้าด้วยตนเอง 7) การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้เรียนรู้มาต่อกลุ่ม 8) การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ และ 9) การสะท้อนความรู้ในบทเรียน โดยทำการสรุปทั้งในด้านประเด็นปัญหา ผลจากการไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน และกระบวนการในการแก้ปัญหา ร่วมกับการประเมินคุณลักษณะอันพึงประสงค์ที่ได้จากการเรียนรู้โดยใช้ปัญหาเป็นฐาน ได้แก่ ภาวะผู้นำ ความเคารพต่อสมาชิกในกลุ่ม และเจตคติต่อการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งช่วงเวลาที่เหมาะสมสำหรับการประเมินนี้ คือ ภายหลังจากเสร็จสิ้นกระบวนการกลุ่มในการเรียนรู้โดยใช้ปัญหาเป็นฐานภายในสัปดาห์แรก กลุ่มผู้ประเมินประกอบด้วย การประเมินตนเองของผู้เรียน (self-

assessment) การประเมินโดยเพื่อน (peer assessment) และการประเมินโดยอาจารย์ (instructor assessment)

แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยของคู่มือเล่มนี้ นำมาจากงานวิทยานิพนธ์เรื่อง “การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสซ์” ของผู้จัดทำคู่มือ ซึ่งประกอบด้วยรายการประเมินจำนวน 19 ข้อ ดังมีรายละเอียดแสดงในตารางที่ 1 ใช้เวลาทำแบบประเมินประมาณ 10 นาที ในแต่ละข้อเป็นมาตรวัดประมาณค่า 4 ระดับ (rating scale) ได้แก่

ดีมาก หมายถึง ผู้รับการประเมินแสดงออกว่า มีทุกครั้ง หรือ มีบางครั้ง มีครบถ้วน และเกิดประโยชน์ต่อกลุ่ม ตามรายการในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

ดี หมายถึง ผู้รับการประเมินแสดงออกว่า มีบางครั้ง และ มีครบถ้วน ตามรายการในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

ปานกลาง หมายถึง ผู้รับการประเมินแสดงออกว่า มีน้อยครั้ง หรือ มีแต่ไม่ครบถ้วน ตามรายการในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

ควรปรับปรุง หมายถึง ผู้รับการประเมินแสดงออกว่า ไม่มี ตามรายการในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

ตารางที่ 1 แสดงการกำหนดจำนวนข้อคำถาม และพฤติกรรม/คุณลักษณะที่ต้องการวัด (table of specification)

พฤติกรรม/คุณลักษณะที่ต้องการวัด	ระดับ ความสำคัญ	จำนวน ข้อ	ข้อที่
1. การรวบรวม วิเคราะห์ ทำความเข้าใจในข้อมูลจากสถานการณ์ และการกำหนดประเด็นปัญหา	15.79%	3	1, 2, 4
2. การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้ จากความรู้และประสบการณ์ที่มีอยู่	10.53%	2	5, 16
3. การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล และการระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา	10.53%	2	6, 7
4. การศึกษาค้นคว้าด้วยตนเอง	10.53%	2	8, 9

พฤติกรรม/คุณลักษณะที่ต้องการวัด	ระดับ ความสำคัญ	จำนวน ข้อ	ข้อที่
5. การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้เรียนรู้มาต่อ กลุ่ม	10.53%	2	10, 15
6. การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการ ปรับเปลี่ยน สมมติฐานตามความรู้ใหม่ที่ได้รับ	10.53%	2	3, 11
7. การสะท้อนความรู้ในบทเรียน โดยทำการสรุปเกี่ยวกับ ปัญหาในแต่ละประเด็น และผลการไปศึกษาเพิ่มเติมรวมถึง กระบวนการแก้ปัญหา	10.53%	2	17, 18
8. คุณลักษณะอันพึงประสงค์ที่ได้จากการเรียนรู้โดยใช้ ปัญหาเป็นฐาน ได้แก่ ภาวะผู้นำ ความเคารพต่อสมาชิกใน กลุ่ม และเจตคติต่อการเรียนรู้โดยใช้ปัญหาเป็นฐาน	21.05%	4	12, 13, 14, 19

ผลการตรวจสอบคุณภาพของเครื่องมือในด้านความตรงเชิงเนื้อหา (content validity) จากผู้เชี่ยวชาญทางด้านเนื้อหาและการวัดผล จำนวน 5 ท่าน ได้ค่าดัชนีความสอดคล้อง (index of congruency; IOC) อยู่ระหว่าง 0.8 – 1 ซึ่งมีค่ามากกว่า 0.5 แสดงว่า ข้อคำถามทั้งหมดมีความเป็นตัวแทนของสิ่งที่ต้องการวัด และผลการตรวจสอบด้านความเที่ยงโดยนำไปทดลองใช้กับนักศึกษามหาวิทยาลัยแห่งหนึ่งจำนวน 30 คน โดยให้นักศึกษาทำการประเมินตนเอง และนำมาหาค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's alpha coefficient) ได้ค่าเท่ากับ 0.893

(แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อยแสดงไว้ในภาคผนวก)

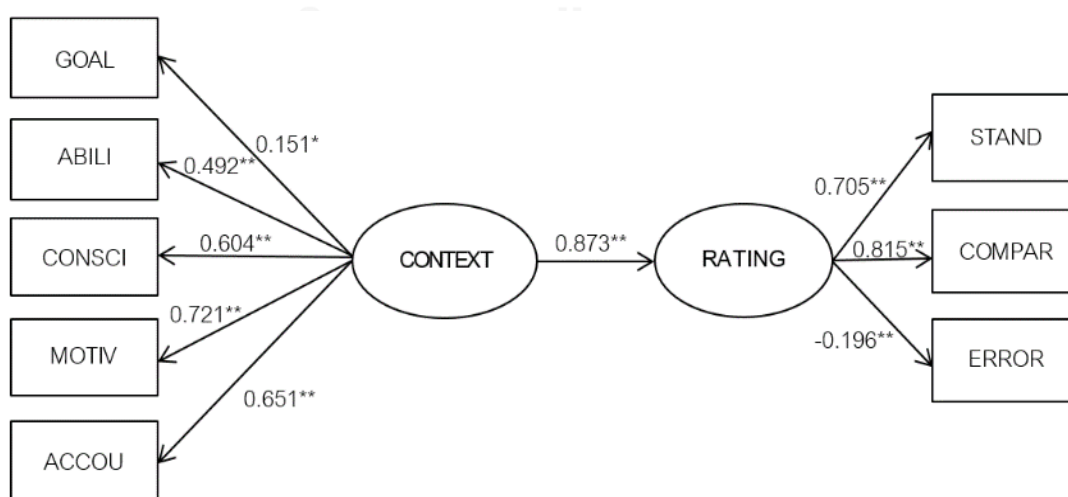
บทที่ 3

แนวทางสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ

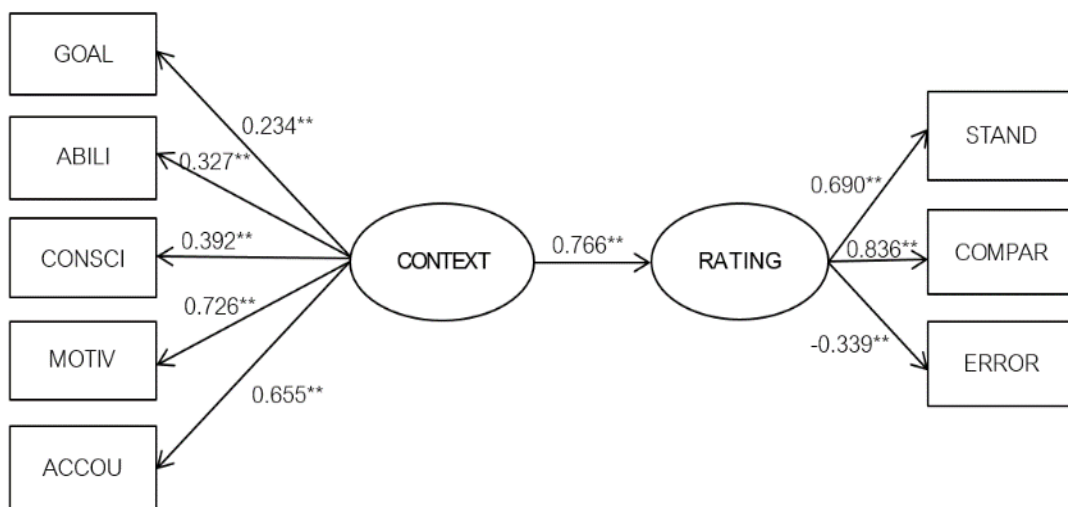
ผู้ประเมินควรศึกษาหรือทำความเข้าใจในคู่มือนี้ก่อนทำการประเมินจริง ซึ่งเนื้อหาสาระในบทนี้ ได้แบ่งการนำเสนอออกเป็น 2 ตอน ได้แก่ 1) ปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ และ 2) แนวทางการปฏิบัติสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ ซึ่งรายละเอียดของแต่ละตอนมีดังนี้

3.1 ปัจจัยเชิงสาเหตุที่ส่งผลต่อคุณภาพการให้คะแนนการปฏิบัติ

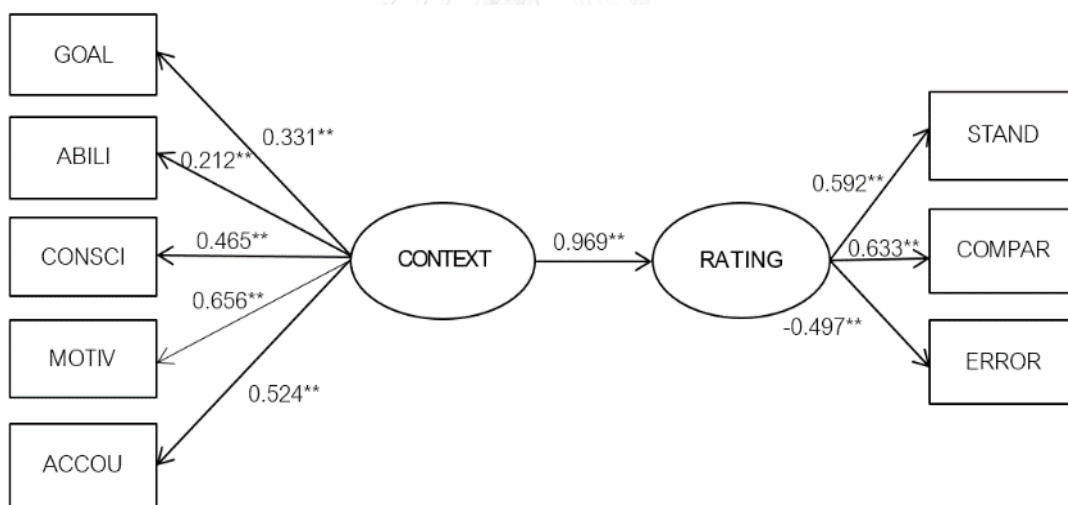
จากผลการศึกษาของงานวิทยานิพนธ์เรื่อง “การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสส์” ของผู้จัดทำคู่มือ พบว่า โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา ซึ่งประกอบด้วยตัวแปรแฝงทั้งหมด 2 ตัวแปร คือ 1) บริบทของผู้ประเมิน (CONTEXT) วัดจากตัวแปรสังเกตได้ 5 ตัวแปร ได้แก่ เป้าหมาย (GOAL) ความสามารถ (ABILI) บุคลิกภาพแบบมีจิตสำนึก (CONSCI) แรงจูงใจ (MOTIV) และความรับผิดชอบที่ตรวจสอบได้ (ACCOU) และ 2) คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) วัดจากตัวแปรสังเกตได้ 3 ตัวแปร ได้แก่ การรับรู้มาตรฐาน (STAND) กระบวนการเปรียบเทียบ (COMPAR) และความคลาดเคลื่อน (ERROR) มีรูปแบบไม่แตกต่างระหว่างกลุ่มผู้ประเมินทั้ง 3 กลุ่ม คือ อาจารย์ เพื่อน และตนเอง(รายละเอียดแสดงในภาพที่ 1-3)



ภาพที่ 1 โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาในผู้ประเมินกลุ่มอาจารย์



ภาพที่ 2 โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาในผู้ประเมินกลุ่มเพื่อน



ภาพที่ 3 โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษาในผู้ประเมินตนเอง

จากผลการวิเคราะห์อธิบายได้ว่า บริบทของผู้ประเมิน ซึ่งประกอบด้วย เป้าหมาย (GOAL) ความสามารถ (ABILI) บุคลิกภาพแบบมีจิตสำนึก (CONSCI) แรงจูงใจ (MOTIV) และความรับผิดชอบที่ตรวจสอบได้ (ACCOU) จะส่งผลทางตรงในทิศทางบวกต่อคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานไม่แตกต่างกันใน 3 กลุ่มผู้ประเมิน

3.2 แนวทางการปฏิบัติสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ

การพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ ตามโมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา (รายละเอียดกล่าวไว้ในข้อ 3.1) มีแนวทางดังต่อไปนี้

3.2.1 บริบทของผู้ประเมิน หมายถึง สถานการณ์หรือเงื่อนไขที่เป็นตัวกำหนดให้เกิดความแตกต่างของพฤติกรรมการให้คะแนนระหว่างผู้ประเมิน ซึ่งประกอบด้วย 5 ตัวแปรที่เกี่ยวข้อง ได้แก่ เป้าหมาย ความสามารถ บุคลิกภาพแบบมีจิตสำนึก แรงจูงใจ และความรับผิดชอบที่ตรวจสอบได้ ซึ่งแต่ละตัวแปรมีรายละเอียดดังนี้

1) เป้าหมาย (goals) หมายถึง ลักษณะความคาดหวังของผู้ประเมินที่มีต่อกระบวนการและผลลัพธ์ของการประเมิน โดยลักษณะของเป้าหมายอาจมีความซับซ้อน หรือเป็นเป้าหมายในเชิงตรงกันข้าม (conflicting goals) เป้าหมายที่สำคัญของผู้ประเมินภายในบริบทของการศึกษามี 2 ประเภท คือ

เป้าหมายระหว่างบุคคล (interpersonal goals) หมายถึง ลักษณะความคาดหวังของผู้ประเมินที่เกี่ยวข้องกับการใช้ผลการประเมินเพื่อพัฒนาความสัมพันธ์ระหว่างบุคคล เช่น การให้คะแนนประเมินผลการปฏิบัติที่สูงเพื่อให้เกิดความเสมอภาคในกลุ่ม หรือความพยายามที่จะคงไว้ซึ่งระดับความสัมพันธ์ระหว่างบุคคล

เป้าหมายภายในบุคคล (internalized goals) หมายถึง ลักษณะความคาดหวังเกี่ยวกับการประเมินอันเป็นผลมาจากความเชื่อและการให้คุณค่าภายในของผู้ประเมิน เช่น ผู้ประเมินที่มีความเชื่อมั่นต่อกระบวนการในการประเมิน หรือผู้ประเมินที่เชื่อว่าผลการประเมินที่แม่นยำจะมีประโยชน์ต่อการนำไปใช้

แนวทางการพัฒนาเป้าหมายของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. ทำความเข้าใจเกี่ยวกับผลลัพธ์ของการประเมินที่มีคุณภาพ เช่น - ทราบว่า ผลการประเมินจะช่วยอธิบายเกี่ยวกับจุดแข็งและโอกาสสำหรับการพัฒนาทั้งตัวผู้เรียนเอง และระบบการเรียนการสอน	นายแก้วทราบว่า เมื่อตนเองทำการประเมินเพื่อนในกลุ่ม คือ นางสาวปาน ได้อย่างถูกต้องแม่นยำ จะทำให้ผู้ที่เกี่ยวข้อง เช่น - นางสาวปาน รับรู้ว่าตนมีจุดแข็ง คือ มีความสามารถด้านการพูด แต่มีจุดอ่อนด้านขาดทักษะการคิดวิเคราะห์แก้ปัญหา ดังนั้น

แนวทาง	ตัวอย่าง
<ul style="list-style-type: none"> - ทราบว่า ผลการประเมินจะช่วยจุดประกายความคิดของผู้ที่เกี่ยวข้อง และจะนำไปสู่การพัฒนาอย่างต่อเนื่อง 	<p>นางสาวปานจึงควรเน้นพัฒนาตนเองด้านการคิดวิเคราะห์แก้ปัญหาให้มากขึ้น</p> <ul style="list-style-type: none"> - อาจารย์ทราบว่า นางสาวปานมีจุดอ่อนด้านการคิดวิเคราะห์แก้ปัญหา จึงช่วยตั้งคำถามฝึกให้นางสาวปานได้ใช้ความคิดวิเคราะห์แก้ปัญหา
<p>2. เข้าใจถึงบทบาทของการเป็นผู้ประเมินที่ดี คือ</p> <ul style="list-style-type: none"> - ทำความเข้าใจในวัตถุประสงค์และวิธีการประเมิน - จัดหาข้อมูลที่ทันสมัย น่าเชื่อถือ และมีประโยชน์ เพื่อใช้ประกอบการประเมิน - ยึดมั่นในหลักความยุติธรรมโดยแสดงตนเป็นกลางหรือเป็นเอกเทศต่อการประเมิน - หมั่นสังเกต ศึกษาเพิ่มเติม ตลอดจนแลกเปลี่ยนความรู้และมุมมองประสบการณ์ระหว่างผู้ประเมิน 	<p>นายแก้วทราบว่า ก่อนทำหน้าที่เป็นผู้ประเมินตนเองจะต้องศึกษาวัตถุประสงค์และวิธีการประเมินจากเอกสารประกอบการประเมินนั้นๆ จนเข้าใจถ่องถ้วนเป็นอย่างดี และจะต้องจัดหาข้อมูลที่ถูกต้อง และสอดคล้องกับวัตถุประสงค์การประเมิน เพื่อใช้ประกอบการตัดสินใจ และขณะที่นายแก้วทำหน้าที่เป็นผู้ประเมิน ก็ต้องแสดงความเป็นกลาง และมี ความยุติธรรมในการประเมิน</p>

2) ความสามารถ (ability for ratings) หมายถึง ศักยภาพของผู้ประเมินในการทำความเข้าใจคุณลักษณะที่ต้องการประเมิน โดยความแตกต่างของศักยภาพนี้จะขึ้นกับการได้รับความรู้หรือประสบการณ์จากการฝึกอบรมเกี่ยวกับการประเมินผลการปฏิบัติ หรือการประเมินผลบุคคล ตลอดจนโอกาสของผู้ประเมินในการสังเกตพฤติกรรมหรือคุณลักษณะที่ต้องการประเมินจากผู้ถูกประเมิน

แนวทางการพัฒนาความสามารถของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
<p>1. ศึกษาเพิ่มเติมเกี่ยวกับหลักการ และเทคนิคการประเมิน ดังนี้</p> <ul style="list-style-type: none"> - การประเมิน เป็นวิธีการให้ข้อเท็จจริงหรือเสนอสารสนเทศที่เป็นประโยชน์ เพื่อตอบสนองต่อความต้องการใช้สารสนเทศของผู้เกี่ยวข้อง เช่น ต้องการทราบปัญหาที่ควรแก้ไข ความก้าวหน้าหรือพัฒนาการ ตลอดจนผลสำเร็จที่เกิดขึ้น เป็นต้น 	<p>นายแก้วได้ไปศึกษาความรู้เพิ่มเกี่ยวกับหลักการ และเทคนิคการประเมินจากห้องสมุด</p>

แนวทาง	ตัวอย่าง
<p>- การประเมิน เป็นกระบวนการที่เกี่ยวข้องกับการกำหนดคุณค่าของสิ่งที่สนใจ ด้วยการสรุปอ้างอิงจากสิ่งที่สังเกตได้โดยตรง ซึ่งเป็นตัวบ่งชี้คุณค่านั้น เพื่อนำมาเปรียบเทียบกับเกณฑ์มาตรฐานที่ใช้ในการตัดสินคุณค่า หรือการประเมิน (evaluation) = การวัด (measurement) + การตัดสิน (judgment)</p> <p>- นักประเมิน ควรมีความสามารถในการตัดสินคุณค่าให้กับสิ่งที่ประเมินได้อย่างยุติธรรม มีความอิสระจากสิ่งที่ประเมิน เข้าใจและรอบรู้ในเรื่องที่ประเมิน สามารถวินิจฉัยคุณค่าของความสำเร็จ เช่น ประสิทธิภาพ ประสิทธิภาพ และสามารถให้ข้อเสนอแนะเพื่อการปรับปรุงพัฒนาคุณค่าของสิ่งนั้นได้อย่างเหมาะสม</p> <p>- ผลการประเมิน ควรมีความยุติธรรม น่าเชื่อถือ สมเหตุสมผล และเป็นประโยชน์</p>	
<p>2. ฝึกทักษะในการคิด วิเคราะห์ สังเคราะห์ และประเมินค่าซึ่งเป็นทักษะที่จำเป็นสำหรับผู้ประเมิน</p>	<p>- นายแก้วมักฝึกการตั้งคำถามและหาคำตอบด้วยตนเอง รวมถึงฝึกการคิดสรุปความสัมพันธ์ของสิ่งต่างๆ หรือข้อมูลที่มีความเกี่ยวข้องกัน</p>
<p>3. วางแผนและบริหารเวลาที่ใช้ในการประเมินอย่างมีประสิทธิภาพ</p>	<p>- เมื่อได้รับมอบหมายให้ทำการประเมิน นายแก้วจึงกำหนดเวลาที่จะใช้ในการประเมินให้ชัดเจน และปฏิบัติตามกำหนดเวลานั้นอย่างเคร่งครัด</p>
<p>4. หมั่นทบทวนข้อผิดพลาดตลอดจนความคงเส้นคงวาจากผลการประเมินของตน</p>	<p>- เมื่อนายแก้วทำการประเมินเสร็จได้ทำการตรวจทานผลการประเมินว่ามีข้อผิดพลาดตรงจุดไหนบ้าง และในประเด็นการประเมิน</p>

แนวทาง	ตัวอย่าง
	ที่ใกล้เคียงกันนั้นมีการประเมินที่เหมือนหรือแตกต่างกันหรือไม่

3) บุคลิกภาพแบบมีจิตสำนึก (conscientiousness) หมายถึง ลักษณะของบุคคลที่แสดงถึงการตระหนักในการควบคุมหรือกระตุ้นตนเองเพื่อให้เกิดการตอบสนองอย่างมีประสิทธิภาพ เช่น มีพลังในการทำงาน ต้องการที่จะประสบความสำเร็จ ทำงานเป็นระบบ ยึดมั่นในหน้าที่ความรับผิดชอบ แนวทางการพัฒนาบุคลิกภาพแบบมีจิตสำนึกของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. มีความซื่อสัตย์ต่อตนเอง และผู้อื่น	- นายแก้วประเมินให้คะแนนตนเองและเพื่อนในกลุ่มอย่างตรงตามความสามารถหรือการปฏิบัติที่เกิดขึ้นจริง
2. มีวินัยในตนเอง และมีความรับผิดชอบต่อหน้าที่	- นายแก้วแสดงความรับผิดชอบต่อหน้าที่ในการเป็นผู้ประเมินโดยให้ผลการประเมินที่ถูกต้องแม่นยำ
3. ได้รับความสามารถของตน และมีความพยายามพัฒนาตนให้ประสบความสำเร็จ	- นายแก้วทราบว่าตนยังไม่มีความรู้เกี่ยวกับหลักการประเมิน จึงได้ไปพัฒนาความรู้โดยการศึกษเพิ่มเติมในเรื่องนั้น
4. ฝึกความสามารถในการเริ่มงานและปฏิบัติงานให้สำเร็จ แม้ว่าจะมีอุปสรรค	- เมื่อนายแก้วยังไม่ทราบวิธีการในการประเมิน ก็ตั้งใจศึกษาเรียนรู้ และถึงแม้จะเจออุปสรรค ก็สามารถผ่านพ้นไปจนกระทั่งเกิดความเข้าใจแจ่มแจ้ง
5. ฝึกความเป็นระเบียบ รอบคอบ ตลอดจนมีแบบแผนในการทำงาน เพื่อให้งานออกมาสมบูรณ์แบบ	- นายแก้วฝึกคิดทบทวน ไตร่ตรองในการทำงานต่างๆ เพื่อป้องกันความผิดพลาด เช่น การตรวจทานผลการประเมินทุกๆ ครั้ง

4) แรงจูงใจ (motivation) หมายถึง สิ่งที่มีผลผลักดันให้ผู้ประเมินเกิดความพยายามในการประเมินเพื่อให้บรรลุเป้าหมายแห่งความถูกต้องของผลการประเมิน โดยคนที่มีแรงจูงใจสูงจะมีความพยายามอย่างไม่ลดละ ต่างกับคนที่มีแรงจูงใจต่ำที่อาจล้มเลิกการกระทำก่อนที่จะบรรลุเป้าหมาย

แนวทางการพัฒนาแรงจูงใจของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. รู้จักเป้าหมายของการประเมิน คือ สารสนเทศหรือข้อมูลที่มีน้ำหนักเชื่อถือ สมเหตุสมผล และสามารถนำไปใช้ประโยชน์ได้ต่อทั้งในแง่การพัฒนาผู้เรียน ตลอดจนการเรียนการสอน	- นายแก้วตั้งเป้าหมายว่า ผลการประเมินต้องเป็นข้อมูลที่ถูกต้อง น่าเชื่อถือ และสามารถนำไปใช้ประโยชน์ เช่น ใช้เป็นข้อมูลย้อนกลับเพื่อแสดงถึงความสามารถของผู้ถูกประเมิน
2. ต้องการเป็นผู้ประเมินที่น่าเชื่อถือ มีคุณธรรมและจริยธรรม	- นายแก้วเห็นว่าผู้ประเมินที่ทำหน้าที่ได้อย่างมีคุณธรรมและจริยธรรมจะช่วยพัฒนาสังคมให้น่าอยู่ จึงต้องการเป็นผู้ประเมินเช่นนั้น
3. รับรู้ว่าผลการประเมินที่แม่นยำ เป็นเรื่องที่ทำหายความสามารถของผู้ประเมิน	- นายแก้วทราบว่า การประเมินที่ให้ผลประเมินได้อย่างถูกต้อง แม่นยำ เป็นเรื่องที่ไม่ง่าย และทำหายความสามารถของผู้ประเมิน
4. เชื่อมั่นในความสามารถของตนเองด้านการประเมิน	- นายแก้วมั่นใจในความสามารถของตนในการทำหน้าที่ผู้ประเมินว่าจะทำได้อย่างถูกต้อง แม่นยำ

5) ความรับผิดชอบที่ตรวจสอบได้ (accountability) หมายถึง ความพร้อมที่จะรายงานหรือรับการตรวจสอบจากผู้ที่เกี่ยวข้องในผลจากการปฏิบัติหน้าที่เป็นผู้ประเมิน

แนวทางการพัฒนาความรับผิดชอบที่ตรวจสอบได้ของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. ทำงานอย่างใส่ใจ เป็นระบบ มีระเบียบ	- นายแก้วปฏิบัติหน้าที่เป็นผู้ประเมินอย่างใส่ใจ โดยปฏิบัติตามวิธีการประเมินอย่างถูกต้อง
2. ย้ำเตือนตนเองว่าการประเมินแต่ละรายการจะต้องมีความถูกต้อง และมีข้อมูลยืนยัน/สนับสนุนอย่างเพียงพอ	- นายแก้วหมั่นคิดทบทวนก่อนตัดสินผลการประเมินในแต่ละรายการ ว่ามีความถูกต้องหรือยัง และมีข้อมูลเพียงพอสำหรับการตัดสินหรือไม่

แนวทาง	ตัวอย่าง
3. สามารถแสดงหลักฐาน/ ข้อมูลสนับสนุนในผลการประเมินของตนได้ เช่น กรณีที่พบปัญหาความไม่สอดคล้องกันของผลประเมินจากผู้ประเมินอื่นๆ	- เมื่อผลการประเมินของตนไม่ตรงกับผู้ประเมินคนอื่นๆ นายแก้วก็สามารถให้เหตุผลหรือหลักฐานข้อมูลยืนยันในผลการประเมินของตนเองได้
4. ตรวจสอบซ้ำในผลการประเมิน เพื่อป้องกันความผิดพลาดหรือความประมาทเลินเล่อ โดยไม่ได้ตั้งใจ	- เมื่อนายแก้วทำการประเมินเสร็จทุกครั้งจะทำการตรวจสอบซ้ำในขั้นตอนและผลของการประเมิน
5. รับผิดชอบส่งผลประเมินให้ทันตามกำหนดเวลา	- เมื่อนายแก้วได้รับมอบหมายให้ทำการประเมินก็สามารถส่งผลการประเมินได้ตามกำหนดเวลา

3.2.2 คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (quality ratings in PBL) หมายถึง คุณลักษณะเกี่ยวกับความแม่นยำและความคลาดเคลื่อนของการให้คะแนนซึ่งเป็นการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งประกอบด้วย 3 ตัวแปรที่เกี่ยวข้อง ได้แก่ การรับรู้มาตรฐาน (standards) กระบวนการเปรียบเทียบ (comparison process) และความคลาดเคลื่อน (error) ซึ่งแต่ละตัวแปรมีรายละเอียดดังนี้

1) การรับรู้มาตรฐาน (standards) หมายถึง การตีความของผู้ประเมินเกี่ยวกับหลักเกณฑ์หรือแนวทางการปฏิบัติที่เหมาะสมสำหรับผู้เรียนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน

แนวทางการพัฒนาการรับรู้มาตรฐานของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
- ศึกษาหลักการและแนวทางการปฏิบัติที่เหมาะสมสำหรับผู้เรียนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน ดังนี้ <u>หลักการ:</u> การเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ รูปแบบของการเรียนการสอนที่ใช้การขับเคลื่อนของปัญหาที่มีความใกล้เคียงกับสถานการณ์จริงเป็นจุดเริ่มต้นสำหรับการเรียนรู้ และประโยชน์ที่ได้คือ เสริมสร้างความเข้าใจและพัฒนาเป็นความรู้ที่ติดตัวไป ฝึกการทำงานร่วมกันอย่างมีประสิทธิภาพ	นายแก้วได้ไปศึกษาความรู้เพิ่มเติมเกี่ยวกับหลักหรือแนวทางการปฏิบัติที่เหมาะสมสำหรับผู้เรียนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน

แนวทาง	ตัวอย่าง
<p>พัฒนาทักษะการแก้ปัญหาและเกิดความรู้ที่ได้จากการลงมือกระทำ มีฐานความรู้ที่กว้าง ทราบวิธีการประยุกต์ใช้ความรู้ในสถานการณ์ต่างๆ เกิดความรู้สึกรู้ตัว มีแรงจูงใจ มีพัฒนาการของการกำหนด/ตัดสินใจได้โดยตนเอง มีความมั่นใจในทักษะการเรียนรู้และการให้เหตุผลของตน เกิดความประทับใจต่อองค์ความรู้ และในการวัดและประเมินผลการเรียนรู้จะใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อยจากกลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง เนื่องจากเป็นผู้ที่ทราบข้อมูลของสมาชิกในกลุ่มเป็นอย่างดี จึงสามารถที่จะสะท้อนข้อมูลได้อย่างถูกต้องและเป็นประโยชน์</p> <p><u>กระบวนการ:</u> เน้นการอภิปรายโต้ตอบร่วมกันของสมาชิกในกลุ่มขนาดเล็ก ซึ่งประกอบด้วย 9 ขั้นตอนสำคัญ ดังนี้ 1) การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ 2) การกำหนดประเด็นปัญหา 3) การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้ จากความรู้และประสบการณ์ที่มีอยู่ 4) การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล 5) การระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา 6) การศึกษาค้นคว้าด้วยตนเอง 7) การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้ไปเรียนรู้มาในกลุ่ม 8) การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ และ 9) การสะท้อนความรู้ในบทเรียน โดยทำการสรุปทั้งในด้านประเด็นปัญหา ผลจาก</p>	

แนวทาง	ตัวอย่าง
การไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน และ กระบวนการในการแก้ปัญหา	

2) กระบวนการเปรียบเทียบ (comparison process) หมายถึง การใช้ความคิดอย่างเป็นลำดับของผู้ประเมินเพื่อให้ได้มาซึ่งสารสนเทศสำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ซึ่งประกอบด้วย 2 ขั้นตอน คือ 1) การรวบรวมข้อมูลเพื่อใช้ในการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย โดยข้อมูลจะได้รับการสังเกตพฤติกรรม หรือผลของการปฏิบัติ และ 2) การคัดเลือกและบูรณาการข้อมูล โดยผู้ประเมินจะพิจารณาคัดเลือก จำแนก และสรุปข้อมูลที่ตรงกับวัตถุประสงค์การประเมิน

แนวทางการพัฒนากระบวนการเปรียบเทียบของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. มีการรวบรวมหาข้อมูลทั้งเชิงปริมาณและเชิงคุณภาพ ที่ครอบคลุมครบถ้วน และเหมาะสมตรงกับรายการที่ต้องการประเมิน โดยข้อมูลอาจมาจากการสังเกตพฤติกรรม หรือผลที่เกิดขึ้นจากการปฏิบัติ	- นายแก้วรวบรวมข้อมูลทั้งความถี่ และ ลักษณะของคำถามที่นางสาวปาน (เพื่อนในกลุ่ม) ได้แสดงออกระหว่างการทำกิจกรรมกลุ่ม เพื่อใช้เป็นข้อมูลสำหรับการประเมินทักษะการตั้งคำถามของนางสาวปาน
2. พิจารณาคัดเลือก จำแนก และตรวจสอบความถูกต้องของข้อมูล	- นายแก้วพิจารณาเลือกใช้ข้อมูลการมีส่วนร่วมทำกิจกรรมกลุ่มของนางสาวปาน (เพื่อนในกลุ่ม) เพื่อให้ตรงกับรายการประเมิน คือ ความสนใจทำกิจกรรมกลุ่มของนางสาวปาน
3. วิเคราะห์ สรุป และแปลผลการประเมิน ให้ตรงกับวัตถุประสงค์การประเมิน	- นายแก้ววิเคราะห์ข้อมูลที่ได้จากการสังเกตเพื่อน คือ นางสาวปาน ว่าสามารถถ่ายทอดเนื้อหาความรู้ให้กับสมาชิกในกลุ่มได้อย่างชัดเจน เข้าใจง่าย จึงทำการสรุป และแปลผลการประเมินให้ นางสาวปานมีความสามารถอยู่ในระดับดีมาก

3) ความคลาดเคลื่อน (error) หมายถึง ผลการให้คะแนนตามแบบประเมินผลกระบวนการเรียนรู้ในกลุ่มย่อยของผู้ประเมินมีความแตกต่างจากค่าความสามารถที่แท้จริงหรือค่าที่ยอมรับของผู้ถูกประเมิน

แนวทางการพัฒนาการป้องกันความคลาดเคลื่อนของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. เข้าใจสาเหตุของการประเมินที่คลาดเคลื่อน ไม่ตรงกับความจริง	<p>นายแก้วเข้าใจว่าผลการประเมินที่คลาดเคลื่อนเกิดจากสาเหตุ เช่น</p> <ul style="list-style-type: none"> - การเข้าใจผิดในวัตถุประสงค์การประเมิน เช่น แทนที่จะเข้าใจให้ถูกต้องว่าวัตถุประสงค์การประเมินก็เพื่อให้ข้อมูลย้อนกลับในการพัฒนาความสามารถ แต่กลับคิดว่าควรให้คะแนนประเมินเพื่อนมากๆ จะได้เป็นการช่วยเหลือกัน - ไม่เข้าใจเกณฑ์ในการประเมินว่าแต่ละระดับคะแนนมีความแตกต่างในการตัดสินอย่างไร - มีผลประโยชน์เข้ามาเกี่ยวข้อง หรือเป็นผู้ประเมินที่มีความลำเอียง เช่น ให้คะแนนตามความประทับใจโดยรวม ชอบบกดหรือปล่อยคะแนน นิยมให้ค่าคะแนนกลางๆ เป็นต้น
2. เข้าใจผลที่เกิดจากการประเมินที่คลาดเคลื่อน	<p>นายแก้วเข้าใจผลที่เกิดจากการประเมินที่คลาดเคลื่อน คือ ทำให้เกิดการบิดเบือนในการวัดผลและส่งผลกระทบต่อความตรงหรือความยุติธรรมในการประเมิน ตลอดจนไม่สามารถนำผลคะแนนเหล่านั้นไปใช้ประโยชน์ได้</p>
3. ทำความเข้าใจอย่างถ่องแท้ในวิธีการรายการ/ตัวบ่งชี้ของแบบประเมิน และเกณฑ์การประเมิน	<p>นายแก้วอ่านเอกสารหรือคำอธิบายในแบบประเมินเพื่อทำความเข้าใจเกี่ยวกับวิธีการรายการที่ต้องประเมิน และเกณฑ์ในการประเมิน ว่าประกอบด้วยอะไรบ้างและทำอะไรจึงจะถูกต้อง</p>
4. กำจัด หรือควบคุมปัจจัยที่ส่งผลกระทบต่อความคลาดเคลื่อนในการประเมิน	<p>นายแก้ว เตือนตนเองอยู่เสมอว่าต้องไม่ลำเอียงในการประเมิน และต้องพยายามทำ</p>

แนวทาง	ตัวอย่าง
	หน้าที่ผู้ประเมินอย่างเป็นกลางและมีความยุติธรรมในการประเมิน
5. หมั่นตรวจสอบความถูกต้องของผลการประเมิน	นายแก้ว พยายามคิดทบทวนก่อนตัดสินผลการประเมิน และเมื่อประเมินเสร็จทุกครั้งก็ทำการตรวจสอบผลของการประเมินซ้ำ

รายการอ้างอิง

ภาษาไทย

ศิริชัย กาญจนวาสี. (2552). ทฤษฎีการประเมิน. (พิมพ์ครั้งที่ 7). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สำนักงานคณะกรรมการกฤษฎีกา. (2542). พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542.

ราชกิจจานุเบกษา ฉบับกฤษฎีกา เล่มที่ 116 ตอนที่ 74 ก 19 สิงหาคม 2542: 1-23.

สำนักงานคณะกรรมการกฤษฎีกา. (2545). พระราชบัญญัติการศึกษาแห่งชาติ (ฉบับที่ 2) พ.ศ. 2545. ราชกิจจานุเบกษา ฉบับกฤษฎีกา เล่มที่ 119 ตอนที่ 123 ก 19 ธันวาคม 2545: 16-21.

สำนักงานคณะกรรมการกฤษฎีกา. (2553). พระราชบัญญัติการศึกษาแห่งชาติ (ฉบับที่ 3) พ.ศ. 2553.

ราชกิจจานุเบกษา ฉบับกฤษฎีกา เล่มที่ 127 ตอนที่ 45 ก 22 กรกฎาคม 2553: 1-3.

สำนักงานคณะกรรมการอุดมศึกษา. (2554). คู่มือการประกันคุณภาพการศึกษาภายในสถานศึกษาระดับอุดมศึกษา พ.ศ. 2553 (พิมพ์ครั้งที่ 1). กรุงเทพฯ: ห้างหุ้นส่วนจำกัด ภาพพิมพ์.

ภาษาอังกฤษ

Barell, J. (2010). Problem-based learning: the foundation for 21 st century skills. In J. Bellanca & R. Brandt (Eds.), 21 st century skills: Rethinking how students learn. Bloomington, IN: Solution Tree Press.

Barrows, H. S., & Tamblyn, R. (1980). *Problem-Based Learning: An Approach to Medical Education*. New York: Springer.

Bridges, E. M., & Hallinger, P. (1992). *Problem-based learning for administrators*. Eugene: ERIC Clearinghouse on Educational Management.

- Eva, K. W. (2001). Assessing tutorial-based assessment. *Advances in Health Sciences Education, 6*: 243–257.
- Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn?. *Educational Psychology Review, 16*(3), 235-266.
- Kassebaum, D. G. (1989). Change in medical education: the courage and will to be different. *Academic Medicine, 64*(8), 446-447.
- Lee, R. M. K. W., & Kwan, C. Y. (1997). The use of problem-based learning in medical education. *Journal of Medical Education, 1*(2), 149-157.
- Moore, I., & Poikela, S. (2011). Evaluating problem-based learning initiatives. In T. Barrett, & S. Moore (Eds.), *New approaches to problem-based learning revitalising your practice in higher education* (pp.100-111). New York: Taylor & Francis.
- Papinczak et al., (2007). An analysis of peer, self, and tutor assessment in problem-based learning tutorials. *Medical Teacher, 29*, 122-132.
- Walsh, A. (2005). *The Tutor in Problem Based Learning: A Novice's Guide*. McMaster University, Faculty of Health Sciences, Hamilton ON Canada. Retrieved from <http://fhs.mcmaster.ca/facdev/documents/tutorPBL.pdf>

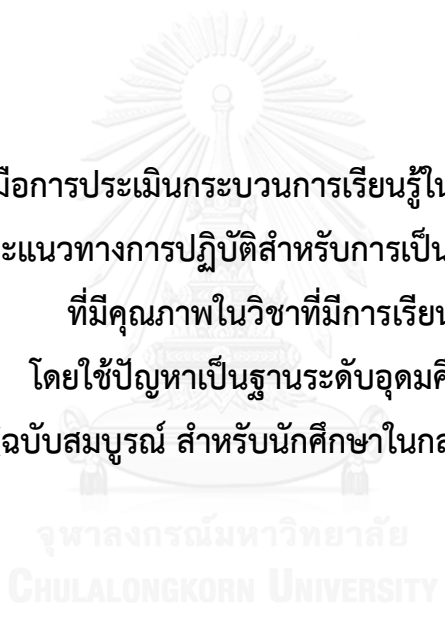
ภาคผนวก: แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (กลุ่มที่__)
 คำชี้แจง โปรดกรอคะแนนที่ตรงมากที่สุด กับพฤติกรรมของนิสิต/ นักศึกษาภายในกลุ่ม โดยเกณฑ์การประเมินมีดังนี้

- 4 = ดีมาก หมายถึง มีทุกครั้ง หรือ มีมากครั้ง และเกิดประโยชน์ต่อกลุ่ม ตามรายการประเมินพฤติกรรมในข้อนี้ๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 3 = ดี หมายถึง มีมากครั้ง และ มีครบถ้วน ตามรายการประเมินพฤติกรรมในข้อนี้ๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 2 = ปานกลาง หมายถึง มีบ่อยครั้ง หรือ มีไม่ครบถ้วน ตามรายการประเมินพฤติกรรมในข้อนี้ๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 1 = ควรปรับปรุง หมายถึง ไม่มี พฤติกรรมตามรายการประเมินในข้อนี้ๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

สถานะของผู้ประเมิน 1) อาจารย์ 2) นักศึกษาชื่อ.....

	ชื่อนักศึกษา/ ระดับคะแนน									
	1	2	3	4	5	6	7	8	9	
รายการประเมิน										
1										
2										
3										
4										
5										
6										

	ชื่อนักศึกษา/ ระดับคะแนน								
	1	2	3	4	5	6	7	8	9
	รายการประเมิน								
7									
	ระบุประเด็นที่ต้องการเรียนรู้จากสถานการณ์ใน PBL ได้								
8									
	ค้นหาข้อมูลความรู้ที่ทันสมัยและเป็นประโยชน์								
9									
	ค้นคว้าหาข้อมูลได้อย่างครอบคลุมและครบถ้วน								
10									
	ถ่ายทอดเนื้อหาความรู้ได้อย่างชัดเจน เข้าใจง่าย								
11									
	นำความรู้ที่ได้จากการค้นคว้ามาใช้อธิบายตอบคำถามหรือ สมมติฐานจากสถานการณ์ของปัญหา								
12									
	ชี้แนะหรือนำพากลุ่มไปสู่แนวทางในการเรียนรู้/เป้าหมายความสำเร็จ								
13									
	ให้เกิดยติเพื่อนสมาชิกในกลุ่มอย่างเท่าเทียมกัน								
14									
	ให้ความร่วมมือ/ ทำงานร่วมกับผู้อื่นได้								
15									
	สนใจมีส่วนร่วม/ตั้งใจทำกิจกรรมกลุ่มอย่างสม่ำเสมอ								
16									
	แสดงออกถึงการมีความคิดริเริ่ม/ สร้างสรรค์								
17									
	สามารถให้ข้อมูลย้อนกลับ/ สะท้อนความคิดเห็นแก่กลุ่มได้								
18									
	สามารถสรุปเนื้อหาที่สำคัญของการเรียนรู้ในครั้งนี้								
19									
	แสดงความพึงพอใจต่อการเรียนแบบ PBL ในครั้งนี้								



คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย
และแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมิน
ที่มีคุณภาพในวิชาที่มีการเรียนรู้
โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา
(ฉบับสมบูรณ์ สำหรับนักศึกษาในกลุ่ม PBL)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

คำนำ

การประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานที่ใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (tutotal-based assessment) จากผู้ประเมินที่มีสถานะต่างกัน ได้แก่ อาจารย์ เพื่อน และตนเอง (ผู้เรียน) เพื่อที่จะได้ข้อมูลสำคัญในการใช้แนะแนวทางการพัฒนาความสามารถ หรือใช้ตัดสินความสามารถของผู้เรียน แต่การประเมินด้วยวิธีนี้ก็มีจุดอ่อน คือ ปัญหาการขาดความแม่นยำในการให้คะแนนระหว่างการประเมินโดยอาจารย์ เพื่อน และตนเอง ซึ่งผลการศึกษาเปรียบเทียบความสัมพันธ์ระหว่างการให้คะแนนจากอาจารย์และจากเพื่อน พบว่า มีความสัมพันธ์กันในระดับต่ำจนถึงระดับปานกลาง นอกจากนี้การให้คะแนนตนเองมีแนวโน้มที่จะมีความสัมพันธ์ระดับต่ำกับการให้คะแนนจากอาจารย์หรือจากเพื่อน และในการแก้ไขปัญหาการขาดความแม่นยำในการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานจากผู้ประเมินอาจารย์ เพื่อน และตนเอง ได้มีหลายงานวิจัยที่แนะนำให้ใช้เทคนิคต่างๆ เช่น ฝึกอบรมการใช้เกณฑ์การประเมิน การสร้างแบบประเมินให้ตรงตามคุณลักษณะที่จะประเมิน เป็นต้น แต่พบว่า วิธีการเหล่านี้ยังไม่สามารถแก้ไขปัญหาคือให้หมดไปได้

จากแนวคิดและสภาพปัญหาที่กล่าวมาในข้างต้น ผู้จัดทำจึงพัฒนาคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา (ฉบับสั้น สำหรับนักศึกษาในกลุ่ม PBL) เพื่อให้เกิดการประเมินผลการเรียนรู้ได้อย่างมีคุณภาพทั้งในด้านความแม่นยำ ความเที่ยงตรง และยุติธรรม โดยนำแนวคิดมาจากตัวแปรในงานวิทยานิพนธ์ของผู้จัดทำ ซึ่งมีชื่อเรื่อง “การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสส์” มาประยุกต์ใช้ในการพัฒนาคู่มือนี้

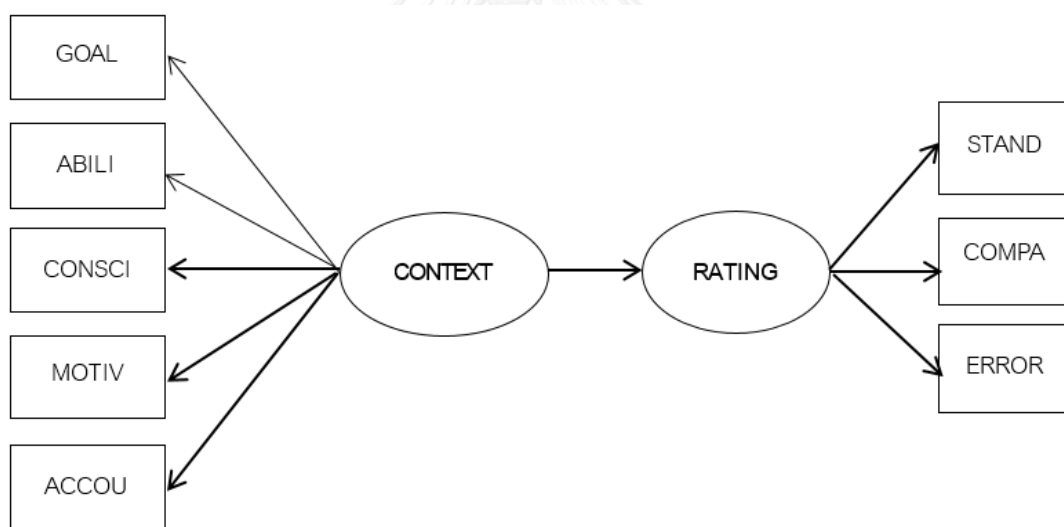
ในโอกาสนี้ผู้วิจัยขอขอบพระคุณท่านอาจารย์ที่ปรึกษา ท่านอาจารย์ผู้ทรงคุณวุฒิ ตลอดจนท่านคณาจารย์ และนิสิต/นักศึกษาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งได้ให้ความอนุเคราะห์ช่วยเหลือเป็นอย่างดีทั้งในขั้นตอนของการพัฒนาเครื่องมือ การเก็บรวบรวมข้อมูล และการพัฒนาเป็นคู่มือนี้

ศุภรัตน์ อิงชาติเจริญ

ผู้จัดทำคู่มือ

แนวทางการปฏิบัติสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ

จากงานวิทยานิพนธ์เรื่อง “การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสซ์” ของผู้จัดทำคู่มือ พบว่า โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา (ดังภาพที่ 1) ประกอบด้วย องค์ประกอบของตัวแปรต้น คือ บริบทของผู้ประเมิน (CONTEXT) ซึ่งวัดจากตัวแปรสังเกตได้ 5 ตัวแปร ได้แก่ 1)เป้าหมาย (GOAL) 2)ความสามารถ (ABILI) 3)บุคลิกภาพแบบมีจิตสำนึก (CONSCI) 4)แรงจูงใจ (MOTIV) และ 5)ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) ซึ่งตัวแปรที่กล่าวมาทั้งหมดนี้จะมีอิทธิพลทางบวกต่อองค์ประกอบของตัวแปรตาม คือ คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ซึ่งวัดจากตัวแปรสังเกตได้ 3 ตัวแปร ได้แก่ 6)การรับรู้มาตรฐาน (STAND) 7)กระบวนการเปรียบเทียบ (COMPAR) และ 8)ความคลาดเคลื่อน (ERROR)



ภาพที่ 1: โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

ดังนั้นแนวทางการปฏิบัติสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ เพื่อเป็นการแก้ไขปัญหาคาดความแม่นยำในการให้คะแนนระหว่างกลุ่มผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง (ผู้เรียน) จึงควรมุ่งเน้นการพัฒนาคุณลักษณะหรือ

พฤติกรรมที่เป็นตัวแปรต่างๆ ภายในโมเดลนี้ ดังนั้นผู้ประเมินควรศึกษาหรือทำความเข้าใจในคู่มือนี้ ก่อนทำการประเมินจริง ซึ่งรายละเอียดของแนวทางการปฏิบัติสำหรับแต่ละตัวแปรมีดังนี้

1) เป้าหมาย (goals) หมายถึง ลักษณะความคาดหวังของผู้ประเมินที่มีต่อกระบวนการและผลลัพธ์ของการประเมิน โดยลักษณะของเป้าหมายอาจมีความซับซ้อน หรือเป็นเป้าหมายในเชิงตรงกันข้าม (conflicting goals) เป้าหมายที่สำคัญของผู้ประเมินภายในบริบทของการศึกษามี 2 ประเภท คือ

- เป้าหมายระหว่างบุคคล (interpersonal goals) หมายถึง ลักษณะความคาดหวังของผู้ประเมินที่เกี่ยวข้องกับการใช้ผลการประเมินเพื่อพัฒนาความสัมพันธ์ระหว่างบุคคล เช่น การให้คะแนนประเมินผลการปฏิบัติที่สูงเพื่อให้เกิดความเสมอภาคในกลุ่ม หรือความพยายามที่จะคงไว้ซึ่งระดับความสัมพันธ์ระหว่างบุคคล

- เป้าหมายภายในบุคคล (internalized goals) หมายถึง ลักษณะความคาดหวังเกี่ยวกับการประเมินอันเป็นผลมาจากความเชื่อและการให้คุณค่าภายในของผู้ประเมิน เช่น ผู้ประเมินที่มีความเชื่อมั่นต่อกระบวนการในการประเมิน หรือผู้ประเมินที่เชื่อว่าผลการประเมินที่แม่นยำจะมีประโยชน์ต่อการนำไปใช้

แนวทางการพัฒนาเป้าหมายของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
<p>1. ทำความเข้าใจเกี่ยวกับผลลัพธ์ของการประเมินที่มีคุณภาพ เช่น</p> <ul style="list-style-type: none"> - ทราบว่า ผลการประเมินจะช่วยอธิบายเกี่ยวกับจุดแข็งและโอกาสสำหรับการพัฒนาทั้งตัวผู้เรียนเอง และระบบการเรียนการสอน - ทราบว่า ผลการประเมินจะช่วยจุดประกายความคิดของผู้ที่เกี่ยวข้อง และจะนำไปสู่การพัฒนาอย่างต่อเนื่อง 	<p>นายแก้วทราบว่า เมื่อตนเองทำการประเมินเพื่อนในกลุ่ม คือ นางสาวปาน ได้อย่างถูกต้องแม่นยำ จะทำให้ผู้ที่เกี่ยวข้อง เช่น</p> <ul style="list-style-type: none"> - นางสาวปาน ระบุว่าตนมีจุดแข็ง คือ มีความสามารถด้านการพูด แต่มีจุดอ่อนด้านขาดทักษะการคิดวิเคราะห์แก้ปัญหา ดังนั้นนางสาวปานจึงควรเน้นพัฒนาตนเองด้านการคิดวิเคราะห์แก้ปัญหาให้มากขึ้น - อาจารย์ทราบว่า นางสาวปานมีจุดอ่อนด้านการคิดวิเคราะห์แก้ปัญหา จึงช่วยตั้งคำถามฝึกให้นางสาวปานได้ใช้ความคิดวิเคราะห์แก้ปัญหา

แนวทาง	ตัวอย่าง
<p>2. เข้าใจถึงบทบาทของการเป็นผู้ประเมินที่ดี คือ</p> <ul style="list-style-type: none"> - ทำความเข้าใจในวัตถุประสงค์และวิธีการประเมิน - จัดหาข้อมูลที่ทันสมัย น่าเชื่อถือ และมีประโยชน์ เพื่อใช้ประกอบการประเมิน - ยึดมั่นในหลักความยุติธรรมโดยแสดงตนเป็นกลางหรือเป็นเอกเทศต่อการประเมิน - หมั่นสังเกต ศึกษาเพิ่มเติม ตลอดจนแลกเปลี่ยนความรู้และมุมมองประสบการณ์ระหว่างผู้ประเมิน 	<p>นายแก้วทราบว่า ก่อนทำหน้าที่เป็นผู้ประเมิน ตนเองจะต้องศึกษาวัตถุประสงค์และวิธีการประเมินจากเอกสารประกอบการประเมินนั้นๆ จนเข้าใจถ่องถ้วนเป็นอย่างดี และจะต้องจัดหาข้อมูลที่ถูกต้อง และสอดคล้องกับวัตถุประสงค์การประเมิน เพื่อใช้ประกอบการตัดสินใจ และขณะที่นายแก้วทำหน้าที่เป็นผู้ประเมิน ก็ต้องแสดงความเป็นกลาง และมีความยุติธรรมในการประเมิน</p>

2) ความสามารถ (ability for rating) หมายถึง ศักยภาพของผู้ประเมินในการทำความเข้าใจคุณลักษณะที่ต้องการประเมิน โดยความแตกต่างของศักยภาพนี้จะขึ้นกับการได้รับความรู้หรือประสบการณ์จากการฝึกอบรมเกี่ยวกับการประเมินผลการปฏิบัติ หรือการประเมินผลบุคคล ตลอดจนโอกาสของผู้ประเมินในการสังเกตพฤติกรรมหรือคุณลักษณะที่ต้องการประเมินจากผู้ถูกประเมิน

แนวทางการพัฒนาความสามารถของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
<p>1. ศึกษาเพิ่มเติมเกี่ยวกับหลักการ และเทคนิคการประเมิน ดังนี้</p> <ul style="list-style-type: none"> - การประเมิน เป็นวิธีการให้ข้อเท็จจริงหรือเสนอสารสนเทศที่เป็นประโยชน์ เพื่อตอบสนองต่อความต้องการใช้สารสนเทศของผู้เกี่ยวข้อง เช่น ต้องการทราบปัญหาที่ควรแก้ไข ความก้าวหน้าหรือพัฒนาการ ตลอดจนผลสำเร็จที่เกิดขึ้น เป็นต้น - การประเมิน เป็นกระบวนการที่เกี่ยวข้องกับการกำหนดคุณค่าของสิ่งที่สนใจ ด้วยการสรุปอ้างอิงจากสิ่งที่สังเกตได้โดยตรง ซึ่งเป็นตัวบ่งชี้คุณค่านั้น เพื่อนำมาเปรียบเทียบกับเกณฑ์มาตรฐานที่ใช้ในการตัดสินคุณค่า หรือ 	<p>นายแก้วได้ไปศึกษาความรู้เพิ่มเติมเกี่ยวกับหลักการ และเทคนิคการประเมินจากห้องสมุด</p>

แนวทาง	ตัวอย่าง
<p>การประเมิน (evaluation) = การวัด (measurement) + การตัดสิน (judgment)</p> <p>- นักประเมิน ควรมีความสามารถในการตัดสินคุณค่าให้กับสิ่งที่ประเมินได้อย่างยุติธรรม มีความอิสระจากสิ่งที่ประเมิน เข้าใจและรอบรู้ในเรื่องที่ประเมิน สามารถวินิจฉัยคุณค่าของความสำเร็จ เช่น ประสิทธิภาพ ประสิทธิภาพ และสามารถให้ข้อเสนอแนะเพื่อการปรับปรุงพัฒนาคุณค่าของสิ่งนั้นได้อย่างเหมาะสม</p> <p>- ผลการประเมิน ควรมีความยุติธรรม น่าเชื่อถือ สมเหตุสมผล และเป็นประโยชน์</p>	
<p>2. ฝึกทักษะในการคิด วิเคราะห์ สังเคราะห์ และประเมินค่าซึ่งเป็นทักษะที่จำเป็นสำหรับผู้ประเมิน</p>	<p>- นายแก้วมักฝึกการตั้งคำถามและหาคำตอบด้วยตนเอง รวมถึงฝึกการคิดสรุปความสัมพันธ์ของสิ่งต่างๆ หรือข้อมูลที่มีความเกี่ยวข้องกัน</p>
<p>3. วางแผนและบริหารเวลาที่ใช้ในการประเมินอย่างมีประสิทธิภาพ</p>	<p>- เมื่อได้รับมอบหมายให้ทำการประเมิน นายแก้วจึงกำหนดเวลาที่จะใช้ในการประเมินให้ชัดเจน และปฏิบัติตามกำหนดเวลานั้นอย่างเคร่งครัด</p>
<p>4. หมั่นทบทวนข้อผิดพลาดตลอดจนความคงเส้นคงวาจากผลการประเมินของตน</p>	<p>- เมื่อนายแก้วทำการประเมินเสร็จได้ทำการตรวจทานผลการประเมินว่ามีข้อผิดพลาดตรงจุดไหนบ้าง และในประเด็นการประเมินที่ใกล้เคียงกันนั้นมีผลการประเมินที่เหมือนหรือแตกต่างกันหรือไม่</p>

3) **บุคลิกภาพแบบมีจิตสำนึก (conscientiousness)** หมายถึง ลักษณะของบุคคลที่แสดงถึงการตระหนักในการควบคุมหรือกระตุ้นตนเองเพื่อให้เกิดการตอบสนองอย่างมีประสิทธิภาพ เช่น มีพลังในการทำงาน ต้องการที่จะประสบความสำเร็จ ทำงานเป็นระบบ ยึดมั่นในหน้าที่ความรับผิดชอบ

แนวทางการพัฒนาบุคลิกภาพแบบมีจิตสำนึกของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. มีความซื่อสัตย์ต่อตนเอง และผู้อื่น	- นายแก้วประเมินให้คะแนนตนเองและเพื่อนในกลุ่มอย่างตรงตามความสามารถหรือการปฏิบัติที่เกิดขึ้นจริง
2. มีวินัยในตนเอง และมีความรับผิดชอบต่อนหน้าที่	- นายแก้วแสดงความรับผิดชอบต่อนหน้าที่ในการเป็นผู้ประเมินโดยให้ผลการประเมินที่ถูกต้องแม่นยำ
3. ได้รับความสามารถของตน และมีความพยายามพัฒนาตนให้ประสบความสำเร็จ	- นายแก้วทราบว่าตนยังไม่มีความรู้เกี่ยวกับหลักการประเมิน จึงได้ไปพัฒนาความรู้โดยการศึกษเพิ่มเติมในเรื่องนั้น
4. ฝึกความสามารถในการเริ่มงานและปฏิบัติงานให้สำเร็จ แม้ว่าจะมีอุปสรรค	- เมื่อนายแก้วยังไม่ทราบวิธีการในการประเมิน ก็ตั้งใจศึกษาเรียนรู้ และถึงแม้จะเจออุปสรรค ก็สามารถผ่านพ้นไปจนกระทั่งเกิดความเข้าใจแจ่มแจ้ง
5. ฝึกความเป็นระเบียบ รอบคอบ ตลอดจนมีแบบแผนในการทำงาน เพื่อให้งานออกมาสัมบูรณ์แบบ	- นายแก้วฝึกคิดทบทวน ไตร่ตรองในการทำงานต่างๆ เพื่อป้องกันความผิดพลาด เช่น การตรวจทานผลการประเมินทุกๆ ครั้ง

4) แรงจูงใจ (motivation) หมายถึง สิ่งที่มีผลผลักดันให้ผู้ประเมินเกิดความพยายามในการประเมินเพื่อให้บรรลุเป้าหมายแห่งความถูกต้องของผลการประเมิน โดยคนที่มีแรงจูงใจสูงจะมีความพยายามอย่างไม่ลดละ ต่างกับคนที่มีแรงจูงใจต่ำที่อาจล้มเลิกการกระทำก่อนที่จะบรรลุเป้าหมาย

แนวทางการพัฒนาแรงจูงใจของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. รู้จักเป้าหมายของการประเมิน คือ สารสนเทศหรือข้อมูลที่มีน้ำหนักเชื่อถือ สมเหตุสมผล และสามารถนำไปใช้ประโยชน์ได้ต่อทั้งในแง่การพัฒนาผู้เรียน ตลอดจนการเรียนการสอน	- นายแก้วตั้งเป้าหมายว่า ผลการประเมินต้องเป็นข้อมูลที่ถูกต้อง น่าเชื่อถือ และสามารถนำไปใช้ประโยชน์ เช่น ใช้เป็นข้อมูลย้อนกลับ เพื่อแสดงถึงความสามารถของผู้ถูกประเมิน

แนวทาง	ตัวอย่าง
2. ต้องการเป็นผู้ประเมินที่น่าเชื่อถือ มีคุณธรรม และจริยธรรม	- นายแก้วเห็นว่าผู้ประเมินที่ทำหน้าที่ได้อย่างมีคุณธรรมและจริยธรรมจะช่วยพัฒนาสังคมให้น่าอยู่ จึงต้องการเป็นผู้ประเมินเช่นนั้น
3. รับรู้ว่าผลการประเมินที่แม่นยำ เป็นเรื่องที่ทำทหายความสามารถของผู้ประเมิน	- นายแก้วทราบว่า การประเมินที่ให้ผลประเมินได้อย่างถูกต้อง แม่นยำ เป็นเรื่องที่ไม่ง่าย และทำทหายความสามารถของผู้ประเมิน
4. เชื่อมั่นในความสามารถของตนเองด้านการประเมิน	- นายแก้วมั่นใจในความสามารถของตนเองในการทำหน้าที่ผู้ประเมินว่าจะทำได้ถูกต้อง แม่นยำ

5) ความรับผิดชอบที่ตรวจสอบได้ (accountability) หมายถึง ความพร้อมที่จะรายงานหรือรับการตรวจสอบจากผู้ที่เกี่ยวข้องในผลจากการปฏิบัติหน้าที่เป็นผู้ประเมิน
แนวทางการพัฒนาความรับผิดชอบที่ตรวจสอบได้ของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. ทำงานอย่างใส่ใจ เป็นระบบ มีระเบียบ	- นายแก้วปฏิบัติหน้าที่เป็นผู้ประเมินอย่างใส่ใจ โดยปฏิบัติตามวิธีการประเมินอย่างถูกต้อง
2. ย้ำเตือนตนเองว่าการประเมินแต่ละรายการจะต้องมีความถูกต้อง และมีข้อมูลยืนยัน/ สนับสนุนอย่างเพียงพอ	- นายแก้วหมั่นคิดทบทวนก่อนตัดสินผลการประเมินในแต่ละรายการ ว่ามีความถูกต้องหรือยัง และมีข้อมูลเพียงพอสำหรับการตัดสินหรือไม่
3. สามารถแสดงหลักฐาน/ ข้อมูลสนับสนุนในผลการประเมินของตนได้ เช่น กรณีที่พบปัญหาความไม่สอดคล้องกันของผลประเมินจากผู้ประเมินอื่นๆ	- เมื่อผลการประเมินของตนไม่ตรงกับผู้ประเมินคนอื่นๆ นายแก้วก็สามารถให้เหตุผลหรือหลักฐานข้อมูลยืนยันในผลการประเมินของตนเองได้
4. ตรวจสอบซ้ำในผลการประเมิน เพื่อป้องกันความผิดพลาดหรือความประมาทเลินเล่อโดยไม่ได้ตั้งใจ	- เมื่อนายแก้วทำการประเมินเสร็จทุกครั้งจะทำการตรวจสอบซ้ำในขั้นตอนและผลของการประเมิน

แนวทาง	ตัวอย่าง
5. รับผิดชอบส่งผลประเมินให้ทันตามกำหนดเวลา	- เมื่อนายแก้วได้รับมอบหมายให้ทำการประเมินก็สามารถส่งผลการประเมินได้ตามกำหนดเวลา

6) การรับรู้มาตรฐาน (standards) หมายถึง การตีความของผู้ประเมินเกี่ยวกับหลักเกณฑ์หรือแนวทางการปฏิบัติที่เหมาะสมสำหรับผู้เรียนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน

แนวทางการพัฒนาการรับรู้มาตรฐานของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
<p>- ศึกษาหลักการและแนวทางการปฏิบัติที่เหมาะสมสำหรับผู้เรียนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน ดังนี้</p> <p><u>หลักการ:</u> การเรียนรู้โดยใช้ปัญหาเป็นฐาน คือ รูปแบบของการเรียนการสอนที่ใช้การขับเคลื่อนของปัญหาที่มีความใกล้เคียงกับสถานการณ์จริงเป็นจุดเริ่มต้นสำหรับการเรียนรู้ และประโยชน์ที่ได้คือ เสริมสร้างความเข้าใจและพัฒนาเป็นความรู้ที่ติดตัวไป ฝึกการทำงานร่วมกันอย่างมีประสิทธิภาพ พัฒนาทักษะการแก้ปัญหาและเกิดความรู้ที่ได้จากการลงมือกระทำ มีฐานความรู้ที่กว้าง ทราบวิธีการประยุกต์ใช้ความรู้ในสถานการณ์ต่างๆ เกิดความรู้สึกรู้ตัว มีแรงจูงใจ มีพัฒนาการของการกำหนด/ตัดสินใจได้โดยตนเอง มีความมั่นใจในทักษะการเรียนรู้และการให้เหตุผลของตน เกิดความประทับใจต่อองค์ความรู้ และในการวัดและประเมินผลการเรียนรู้จะใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อยจากกลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง เนื่องจากเป็นผู้ที่ทราบข้อมูลของสมาชิกในกลุ่มเป็นอย่างดี</p>	<p>นายแก้วได้ไปศึกษาความรู้เพิ่มเติมเกี่ยวกับหลักหรือแนวทางการปฏิบัติที่เหมาะสมสำหรับผู้เรียนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน</p>

แนวทาง	ตัวอย่าง
<p>จึงสามารถที่จะสะท้อนข้อมูลได้อย่างถูกต้องและเป็นประโยชน์</p> <p><u>กระบวนการ:</u> เน้นการอภิปรายโต้ตอบร่วมกันของสมาชิกในกลุ่มขนาดเล็ก ซึ่งประกอบด้วย 9 ขั้นตอนสำคัญ ดังนี้ 1) การรวบรวม วิเคราะห์ และทำความเข้าใจในข้อมูลจากสถานการณ์ 2) การกำหนดประเด็นปัญหา 3) การสร้างสมมติฐานเกี่ยวกับแนวทางการแก้ไขปัญหาที่เป็นไปได้ จากความรู้และประสบการณ์ที่มีอยู่ 4) การสรุปเป็นแผนภาพแสดงกลไกเหตุและผลสำหรับใช้อธิบายข้อมูล 5) การระบุประเด็นสำหรับการเรียนรู้เพื่อใช้ในการแก้ไขปัญหา 6) การศึกษาค้นคว้าด้วยตนเอง 7) การรายงานและอภิปรายแลกเปลี่ยนสิ่งที่ได้ไปเรียนรู้มาในกลุ่ม 8) การทดสอบสมมติฐานที่กำหนดไว้ หรือทำการปรับเปลี่ยนสมมติฐานตามความรู้ใหม่ที่ได้รับ และ 9) การสะท้อนความรู้ในบทเรียน โดยทำการสรุปทั้งในด้านประเด็นปัญหา ผลจากการไปศึกษาเพิ่มเติมของสมาชิกแต่ละคน และกระบวนการในการแก้ปัญหา</p>	

7) **กระบวนการเปรียบเทียบ (comparison process)** หมายถึง การใช้ความคิดอย่างเป็นลำดับของผู้ประเมินเพื่อให้ได้มาซึ่งสารสนเทศสำหรับการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ซึ่งประกอบด้วย 2 ขั้นตอน คือ 1) การรวบรวมข้อมูลเพื่อใช้ในการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย โดยข้อมูลจะได้รับการสังเกตพฤติกรรม หรือผลของการปฏิบัติ และ 2) การคัดเลือกและบูรณาการข้อมูล โดยผู้ประเมินจะพิจารณาคัดเลือก จำแนก และสรุปข้อมูลที่ตรงกับวัตถุประสงค์การประเมิน

แนวทางการพัฒนากระบวนการเปรียบเทียบของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. มีการรวบรวมหาข้อมูลทั้งเชิงปริมาณและเชิงคุณภาพ ที่ครอบคลุมครบถ้วน และเหมาะสมตรงกับรายการที่ต้องการประเมิน โดยข้อมูลอาจมาจากการสังเกตพฤติกรรม หรือผลที่เกิดขึ้นจากการปฏิบัติ	- นายแก้วรวบรวมข้อมูลทั้งความถี่ และลักษณะของคำถามที่นางสาวปาน (เพื่อนในกลุ่ม) ได้แสดงออกระหว่างการทำกิจกรรมกลุ่ม เพื่อใช้เป็นข้อมูลสำหรับการประเมินทักษะ การตั้งคำถามของนางสาวปาน
2. พิจารณาคัดเลือก จำแนก และตรวจสอบความถูกต้องของข้อมูล	- นายแก้วพิจารณาเลือกใช้ข้อมูลการมีส่วนร่วม ทำกิจกรรมกลุ่มของนางสาวปาน (เพื่อนในกลุ่ม) เพื่อให้ตรงกับรายการประเมิน คือ ความสนใจ ทำกิจกรรมกลุ่มของนางสาวปาน
3. วิเคราะห์ สรุป และแปลผลการประเมิน ให้ตรงกับวัตถุประสงค์การประเมิน	- นายแก้ววิเคราะห์ข้อมูลที่ได้จากการสังเกต เพื่อน คือ นางสาวปาน ว่าสามารถถ่ายทอด เนื้อหาความรู้ให้กับสมาชิกในกลุ่มได้อย่าง ชัดเจน เข้าใจง่าย จึงทำการสรุป และแปลผลการประเมินให้ นางสาวปานมีความสามารถ อยู่ในระดับดีมาก

3) ความคลาดเคลื่อน (error) หมายถึง ผลการให้คะแนนตามแบบประเมินผลกระบวนการเรียนรู้ในกลุ่มย่อยของผู้ประเมินมีความแตกต่างจากค่าความสามารถที่แท้จริงหรือค่าที่ยอมรับของผู้ถูกประเมิน

แนวทางการพัฒนาการป้องกันความคลาดเคลื่อนของผู้ประเมิน มีดังนี้

แนวทาง	ตัวอย่าง
1. เข้าใจสาเหตุของการประเมินที่คลาดเคลื่อน ไม่ตรงกับความจริง	นายแก้วเข้าใจว่าผลการประเมินที่คลาดเคลื่อนเกิดจากสาเหตุ เช่น - การเข้าใจผิดในวัตถุประสงค์การประเมิน เช่น แทนที่จะเข้าใจให้ถูกต้องว่าวัตถุประสงค์การประเมินก็เพื่อให้ข้อมูลย้อนกลับในการพัฒนาความสามารถ แต่กลับคิดว่าควรให้คะแนนประเมินเพื่อนมากๆ จะได้เป็นการช่วยเหลือกัน

แนวทาง	ตัวอย่าง
	<ul style="list-style-type: none"> - ไม่เข้าใจเกณฑ์ในการประเมิน ว่าแต่ละระดับคะแนนมีความแตกต่างในการตัดสินอย่างไร - มีผลประโยชน์เข้ามาเกี่ยวข้อง หรือเป็นผู้ประเมินที่มีความลำเอียง เช่น ให้คะแนนตามความประทับใจโดยรวม ชอบกตหรือปล่อยคะแนนนิยมให้ค่าคะแนนกลางๆ เป็นต้น
2. เข้าใจผลที่เกิดจากการประเมินที่คลาดเคลื่อน	นายแก้วเข้าใจผลที่เกิดจากการประเมินที่คลาดเคลื่อน คือ ทำให้เกิดการบิดเบือนในการวัดผลและส่งผลกระทบต่อความตรงหรือความยุติธรรมในการประเมิน ตลอดจนไม่สามารถนำผลคะแนนเหล่านั้นไปใช้ประโยชน์ได้
3. ทำความเข้าใจอย่างถ่องแท้ในวิธีการรายการ/ตัวบ่งชี้ของแบบประเมิน และเกณฑ์การประเมิน	นายแก้วอ่านเอกสารหรือคำอธิบายในแบบประเมินเพื่อทำความเข้าใจเกี่ยวกับวิธีการรายการที่ต้องประเมิน และเกณฑ์ในการประเมินว่าประกอบด้วยอะไรบ้าง และทำอย่างไรจึงจะถูกต้อง
4. กำจัด หรือควบคุมปัจจัยที่ส่งผลกระทบต่อความคลาดเคลื่อนในการประเมิน	นายแก้ว เตือนตนเองอยู่เสมอว่าต้องไม่ลำเอียงในการประเมิน และต้องพยายามทำหน้าที่ผู้ประเมินอย่างเป็นกลางและมีความยุติธรรมในการประเมิน
5. หมั่นตรวจสอบความถูกต้องของผล การประเมิน	นายแก้ว พยายามคิดทบทวนก่อนตัดสินผลการประเมิน และเมื่อประเมินเสร็จทุกครั้ง ก็ทำการตรวจสอบผลของการประเมินซ้ำ

ภาคผนวก: แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (กลุ่มที่ __)
คำชี้แจง โปรดกรอคะแนนที่ตรงมากที่สุด กับพฤติกรรมของนิสิต/ นักศึกษาภายในกลุ่ม โดยเกณฑ์การประเมินมีดังนี้

- 4 = ดีมาก หมายถึง มีทุกครั้ง หรือ มีมากกว่าครึ่ง มีความถี่ และเกิดประโยชน์ต่อกลุ่ม ตามรายการประเมินพฤติกรรมในข้อนี้ๆ ตลอดจนกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 3 = ดี หมายถึง มีมากกว่าครึ่ง และ มีครบถ้วน ตามรายการประเมินพฤติกรรมในข้อนี้ๆ ตลอดจนกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 2 = ปานกลาง หมายถึง มีน้อยครั้ง หรือ มีไม่ครบถ้วน ตามรายการประเมินพฤติกรรมในข้อนี้ๆ ตลอดจนกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 1 = ควรปรับปรุง หมายถึง ไม่มี พฤติกรรมตามรายการประเมินในข้อนี้ๆ ตลอดจนกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

สถานะของผู้ประเมิน 1) อาจารย์ 2) นักศึกษาชื่อ.....

	ชื่อนักศึกษา/ ระดับคะแนน								
	1	2	3	4	5	6	7	8	9
รายการประเมิน									
1	ตั้งคำถามที่นำไปสู่การทำความเข้าใจในสถานการณ์ปัญหา								
2	ชี้/ระบุข้อมูลที่สำคัญจากสถานการณ์ปัญหา								
3	คิดวิเคราะห์/ เปรียบเทียบอย่างเป็นระบบ								
4	ประยุกต์เอาความรู้หรือประสบการณ์มาใช้กับสถานการณ์ใน PBL								
5	ตั้งสมมติฐานที่เป็นไปได้สำหรับสาเหตุหรือแนวทางการแก้ปัญหา								
6	อธิบายเชื่อมโยงประเด็นต่างๆ อย่างเป็นเหตุ เป็นผล								

	ชื่อนักศึกษา/ ระดับคะแนน								
	1	2	3	4	5	6	7	8	9
	รายการประเมิน								
7									
	ระบุประเด็นที่ต้องการเรียนรู้จากสถานการณ์ใน PBL ได้								
8									
	ค้นหาข้อมูลความรู้ที่ทันสมัยและเป็นประโยชน์								
9									
	ค้นคว้าหาข้อมูลได้อย่างครอบคลุมและครบถ้วน								
10									
	ถ่ายทอดเนื้อหาความรู้ได้อย่างชัดเจน เข้าใจง่าย								
11									
	นำความรู้ที่ได้จากการค้นคว้ามาใช้อธิบายตอบคำถามหรือ สมมติฐานจากสถานการณ์ของปัญหา								
12									
	ชี้แนะหรือนำพากลุ่มไปสู่แนวทางในการเรียนรู้/เป้าหมายความสำเร็จ								
13									
	ใช้เกียรติเพื่อนสมาชิกในกลุ่มอย่างเท่าเทียมกัน								
14									
	ให้ความร่วมมือ/ ทำงานร่วมกับผู้อื่นได้								
15									
	สนใจมีส่วนร่วม/ตั้งใจทำกิจกรรมกลุ่มอย่างสม่ำเสมอ								
16									
	แสดงออกถึงการมีความคิดริเริ่ม/ สร้างสรรค์								
17									
	สามารถให้ข้อมูลย้อนกลับ/ สะท้อนความคิดแก่กลุ่มได้								
18									
	สามารถสรุปเนื้อหาที่สำคัญของการเรียนรู้ในครั้งนี้อย่าง								
19									
	แสดงความพึงพอใจต่อการเรียนแบบ PBL ในครั้งนี้								

คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย
และแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มี
คุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน
ระดับอุดมศึกษา
(ฉบับสั้น สำหรับอาจารย์ประจำกลุ่ม PBL)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

คำนำ

การประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานที่ใช้การประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (tutotal-based assessment) จากผู้ประเมินที่มีสถานะต่างกัน ได้แก่ อาจารย์ เพื่อน และตนเอง (ผู้เรียน) เพื่อที่จะได้ข้อมูลสำคัญในการใช้แนะแนวทางการพัฒนาความสามารถ หรือใช้ตัดสินความสามารถของผู้เรียน แต่การประเมินด้วยวิธีนี้มีจุดอ่อน คือ ปัญหาการขาดความแม่นยำในการให้คะแนนระหว่างกลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง ซึ่งผลการศึกษาเปรียบเทียบความสัมพันธ์ระหว่างการให้คะแนนจากอาจารย์และจากเพื่อน พบว่ามีความสัมพันธ์กันในระดับต่ำจนถึงระดับปานกลาง นอกจากนี้การให้คะแนนตนเองมีแนวโน้มที่จะมีความสัมพันธ์ระดับต่ำกับการให้คะแนนจากอาจารย์หรือจากเพื่อน และในการแก้ไขปัญหาการขาดความแม่นยำในการประเมินผลการเรียนรู้โดยใช้ปัญหาเป็นฐานจากผู้ประเมินอาจารย์ เพื่อน และตนเอง ได้มีหลายงานวิจัยที่แนะนำให้ใช้เทคนิคต่างๆ เช่น ฝึกอบรมการใช้เกณฑ์การประเมิน การสร้างแบบประเมินให้ตรงตามคุณลักษณะที่จะประเมิน เป็นต้น แต่พบว่า วิธีการเหล่านี้ยังไม่สามารถแก้ไขปัญหาคือให้หมดไปได้

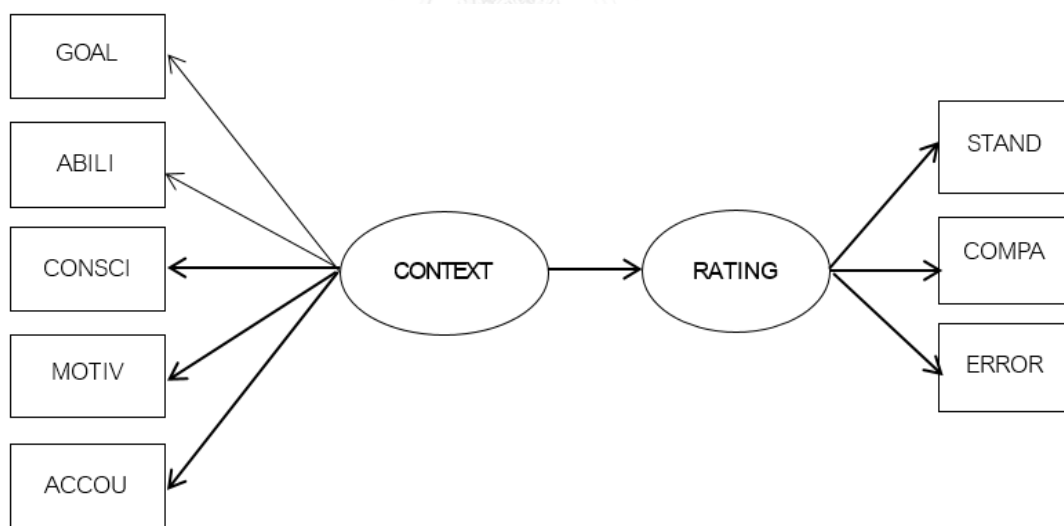
จากแนวคิดและสภาพปัญหาที่กล่าวมาในข้างต้น ผู้จัดทำจึงพัฒนาคู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา เพื่อให้เกิดการประเมินผลการเรียนรู้ได้อย่างมีคุณภาพทั้งในด้านความแม่นยำ ความเที่ยงตรง และยุติธรรม โดยนำแนวคิดมาจากตัวแปรในงานวิทยานิพนธ์ของผู้จัดทำ ซึ่งมีชื่อเรื่อง “การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสส์” มาประยุกต์ใช้ในการพัฒนาคู่มือทั้งนี้ผู้ประเมินควรศึกษาหรือทำความเข้าใจในคู่มือนี้ก่อนทำการประเมินจริง

ในโอกาสนี้ผู้จัดทำ ขอขอบพระคุณท่านอาจารย์ที่ปรึกษา ท่านอาจารย์ผู้ทรงคุณวุฒิ ตลอดจนท่านคณาจารย์ และนิสิต/นักศึกษาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน ซึ่งได้ให้ความอนุเคราะห์ช่วยเหลือเป็นอย่างดีทั้งในขั้นตอนของการพัฒนาเครื่องมือ และการเก็บรวบรวมข้อมูล

ศุภรัตน์ อิงชาติเจริญ
ผู้จัดทำคู่มือ

คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็นผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา การปฏิบัติสำหรับการพัฒนาคุณลักษณะหรือพฤติกรรมที่ส่งผลต่อการเป็นผู้ประเมินที่มีคุณภาพ

จากงานวิทยานิพนธ์เรื่อง “การพัฒนาโมเดลคุณภาพการให้คะแนนระหว่างกลุ่มผู้ประเมินในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานการประยุกต์ใช้โมเดลหลายองค์ประกอบของราสส์” ของผู้จัดทำ พบว่า โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา (ดังภาพที่ 1) ประกอบด้วย องค์ประกอบของตัวแปรต้น คือ บริบทของผู้ประเมิน (CONTEXT) ซึ่งวัดจากตัวแปรสังเกตได้ 5 ตัวแปร ได้แก่ 1)เป้าหมาย (GOAL) 2)ความสามารถ (ABILI) 3)บุคลิกภาพแบบมีจิตสำนึก (CONSCI) 4)แรงจูงใจ (MOTIV) และ 5)ความรับผิดชอบที่ตรวจสอบได้ (ACCOU) ซึ่งตัวแปรที่กล่าวมาทั้งหมดนี้จะมีอิทธิพลทางบวกต่อองค์ประกอบของตัวแปรตาม คือ คุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐาน (RATING) ซึ่งวัดจากตัวแปรสังเกตได้ 3 ตัวแปร ได้แก่ 6)การรับรู้มาตรฐาน (STAND) 7)กระบวนการเปรียบเทียบ (COMPAR) และ 8)ความคลาดเคลื่อน (ERROR)



ภาพที่ 1.โมเดลเชิงสาเหตุของคุณภาพการให้คะแนนในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานในระดับอุดมศึกษา

โดยรายละเอียดของแต่ละตัวแปรมีดังนี้

1) เป้าหมาย หมายถึง ลักษณะความคาดหวังของผู้ประเมินที่มีต่อกระบวนการและผลลัพธ์ของการประเมิน โดยลักษณะของเป้าหมายอาจมีความซับซ้อน หรือเป็นเป้าหมายในเชิงตรงกันข้าม (conflicting goals) เป้าหมายที่สำคัญของผู้ประเมินภายในบริบทของการศึกษามี 2 ประเภท คือ

- เป้าหมายระหว่างบุคคล (interpersonal goals) หมายถึง ลักษณะความคาดหวังของผู้ประเมินที่เกี่ยวข้องกับการใช้ผลการประเมินเพื่อพัฒนาความสัมพันธ์ระหว่างบุคคล เช่น การให้คะแนนประเมินผลการปฏิบัติที่สูงเพื่อให้เกิดความเสมอภาคในกลุ่ม หรือความพยายามที่จะคงไว้ซึ่งระดับความสัมพันธ์ระหว่างบุคคล

- เป้าหมายภายในบุคคล (internalized goals) หมายถึง ลักษณะความคาดหวังเกี่ยวกับการประเมินอันเป็นผลมาจากความเชื่อและการให้คุณค่าภายในของผู้ประเมิน เช่น ผู้ประเมินที่มีความเชื่อมั่นต่อกระบวนการในการประเมิน หรือผู้ประเมินที่เชื่อว่าผลการประเมินที่แม่นยำจะมีประโยชน์ต่อการนำไปใช้

2) ความสามารถ หมายถึง ศักยภาพของผู้ประเมินในการทำความเข้าใจคุณลักษณะที่ต้องการประเมิน โดยความแตกต่างของศักยภาพนี้จะขึ้นกับการได้รับความรู้หรือประสบการณ์จากการฝึกอบรมเกี่ยวกับการประเมินผลการปฏิบัติ หรือการประเมินผลบุคคล ตลอดจนโอกาสของผู้ประเมินในการสังเกตพฤติกรรมหรือคุณลักษณะที่ต้องการประเมินจากผู้ถูกประเมิน

3) บุคลิกภาพแบบมีจิตสำนึก หมายถึง ลักษณะของบุคคลที่แสดงถึงการตระหนักในการควบคุมหรือกระตุ้นตนเองเพื่อให้เกิดการตอบสนองอย่างมีประสิทธิภาพ เช่น มีพลังในการทำงาน ต้องการที่จะประสบความสำเร็จ ทำงานเป็นระบบ ยึดมั่นในหน้าที่ความรับผิดชอบ

4) แรงจูงใจ หมายถึง สิ่งที่มีผลผลักดันให้ผู้ประเมินเกิดความพยายามในการประเมินเพื่อให้บรรลุเป้าหมายแห่งความถูกต้องของผลการประเมิน โดยคนที่มีแรงจูงใจสูงจะมีความพยายามอย่างไม่ลดละ ต่างกับคนที่มีแรงจูงใจต่ำที่อาจล้มเลิกการกระทำก่อนที่จะบรรลุเป้าหมาย

5) ความรับผิดชอบที่ตรวจสอบได้ หมายถึง ความพร้อมที่จะรายงานหรือรับการตรวจสอบจากผู้ที่เกี่ยวข้องในผลจากการปฏิบัติหน้าที่เป็นผู้ประเมิน

6) การรับรู้มาตรฐาน หมายถึง การตีความของผู้ประเมินเกี่ยวกับหลักเกณฑ์หรือแนวทางการปฏิบัติที่เหมาะสมสำหรับผู้เรียนในการเรียนรู้โดยใช้ปัญหาเป็นฐาน

7) กระบวนการเปรียบเทียบ หมายถึง การใช้ความคิดอย่างเป็นลำดับของผู้ประเมินเพื่อให้ได้มาซึ่งสารสนเทศสำหรับใช้ในการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย ซึ่งประกอบด้วย 2 ขั้นตอน คือ 1) การรวบรวมข้อมูลเพื่อใช้ในการประเมินกระบวนการเรียนรู้ในกลุ่มย่อย โดยข้อมูลจะได้จาก

การสังเกตพฤติกรรม หรือผลของการปฏิบัติ และ 2) การคัดเลือกและบูรณาการข้อมูล โดยผู้ประเมิน จะพิจารณาคัดเลือก จำแนก และสรุปข้อมูลที่ตรงกับวัตถุประสงค์การประเมิน

8) ความคลาดเคลื่อน หมายถึง ผลการให้คะแนนตามแบบประเมินผลกระบวนการเรียนรู้ในกลุ่มย่อยของผู้ประเมินมีความแตกต่างจากค่าความสามารถที่แท้จริงหรือค่าที่ยอมรับของผู้ถูกประเมิน

ในการแก้ไขปัญหาการขาดความแม่นยำในการให้คะแนนระหว่างกลุ่มผู้ประเมิน ได้แก่ อาจารย์ เพื่อน และตนเอง (ผู้เรียน) จึงควรมุ่งเน้นการพัฒนาคุณลักษณะหรือพฤติกรรมที่เป็นตัวแปรต่างๆ ภายในโมเดลที่กล่าวมาข้างต้นนี้ ซึ่งสามารถสรุปได้เป็นวิธีการ/ แนวทางการปฏิบัติจำนวน 12 ข้อ ได้แก่

1. ศึกษาค้นคว้าเพิ่มเติมด้วยตนเอง (self-directed learning) ให้รอบรู้ในประเด็นสำหรับการเรียนรู้จากโจทย์สถานการณ์
2. ให้ความสนใจและมีส่วนร่วมในระหว่างการทำกิจกรรมกลุ่ม PBL
3. ใช้ทักษะการคิด วิเคราะห์ เชื่อมโยง เกี่ยวกับความสัมพันธ์ หรือความเป็นเหตุเป็นผล
4. ฝึกการทำงานเป็นระบบ และมีความรับผิดชอบต่อกันที่
5. เข้าใจและให้ความสำคัญกับวัตถุประสงค์ของการประเมิน นั่นคือ การให้ข้อมูลที่เป็นประโยชน์เพื่อเป็นแนวทางสำหรับพัฒนาความสามารถของผู้เรียน
6. รวบรวมข้อมูลการทำกิจกรรมกลุ่มของสมาชิกทุกคน โดยใช้การสังเกตพฤติกรรม ได้แก่ ความถี่ของการมีส่วนร่วมในกลุ่ม การสนใจฟัง การตั้งคำถาม การตอบคำถาม และการแสดงความคิดเห็นในเรื่องต่างๆ
7. ประเมินเนื้อหาการนำเสนอของสมาชิกกลุ่มทุกคน เช่น ความถูกต้อง ตรงประเด็น มีความทันสมัย เป็นประโยชน์ต่อกลุ่ม
8. ประเมินรูปแบบการนำเสนอของสมาชิกกลุ่ม เช่น พูดชัดเจน เข้าใจง่าย มีการสรุปใจความสำคัญ ใช้เวลาเหมาะสม
9. หลีกเลี่ยงการให้คะแนนตามความพอใจของตน การกด-ปล่อยคะแนน หรือให้คะแนนช่วงกลางๆ เพราะผลคะแนนที่คลาดเคลื่อนเหล่านี้ไม่สามารถนำไปใช้ประโยชน์ได้
10. สามารถให้เหตุผลหรือข้อมูลสนับสนุนผลประเมินได้ หากคะแนนที่ตนประเมินไม่สอดคล้องกับผู้ประเมินคนอื่นๆ
11. ตรวจสอบผลการประเมินทุกครั้ง เพื่อป้องกันข้อผิดพลาด
12. แสดงตนเป็นผู้ประเมินที่เป็นกลาง มีความยุติธรรม และมีความรับผิดชอบในผลการประเมิน

	ชื่อนักศึกษา/ ระดับคะแนน								
	1	2	3	4	5	6	7	8	9
	รายการประเมิน								
7									
	ระบุประเด็นที่ต้องการเรียนรู้จากสถานการณ์ใน PBL ได้								
8									
	ค้นหาข้อมูลความรู้ที่ทันสมัยและเป็นประโยชน์								
9									
	ค้นคว้าหาข้อมูลได้อย่างครอบคลุมและครบถ้วน								
10									
	ถ่ายทอดเนื้อหาความรู้ได้อย่างชัดเจน เข้าใจง่าย								
11									
	นำความรู้ที่ได้จากการค้นคว้ามาใช้อธิบายตอบคำถามหรือ สมมติฐานจากสถานการณ์ของปัญหา								
12									
	ชี้แนะหรือนำพากลุ่มไปสู่แนวทางในการเรียนรู้/เป้าหมายความสำเร็จ								
13									
	ให้เกิดยรติเพื่อนสมาชิกในกลุ่มอย่างเท่าเทียมกัน								
14									
	ให้ความร่วมมือ/ ทำงานร่วมกับผู้อื่นได้								
15									
	สนใจมีส่วนร่วม/ตั้งใจทำกิจกรรมกลุ่มอย่างสม่ำเสมอ								
16									
	แสดงออกถึงการมีความคิดริเริ่ม/ สร้างสรรค์								
17									
	สามารถให้ข้อมูลย้อนกลับ/ สะท้อนความคิดแก่กลุ่มได้								
18									
	สามารถสรุปเนื้อหาที่สำคัญของการเรียนรู้ในครั้งนั้น								
19									
	แสดงความพึงพอใจต่อการเรียนแบบ PBL ในครั้งนี้								

**คู่มือการประเมินกระบวนการเรียนรู้ในกลุ่มย่อยและแนวทางการปฏิบัติสำหรับการเป็น
ผู้ประเมินที่มีคุณภาพในวิชาที่มีการเรียนรู้โดยใช้ปัญหาเป็นฐานระดับอุดมศึกษา
(ฉบับสั้น สำหรับนักศึกษาในกลุ่ม PBL)**

กลุ่มเป้าหมาย: อาจารย์ และนักศึกษา ที่ใช้การเรียนการสอนแบบ PBL

วัตถุประสงค์: เพื่อพัฒนาวิธีการ และแบบประเมิน สำหรับใช้ในการประเมินกระบวนการกลุ่มของ PBL

ประโยชน์ที่จะได้รับ: การประเมินผลผู้เรียนได้อย่างถูกต้องใกล้เคียงกับความสามารถที่แท้จริง

คำชี้แจง: แนวทางการปฏิบัติฯ นี้เป็นส่วนหนึ่งของงานวิทยานิพนธ์หลักสูตรครุศาสตรดุษฎีบัณฑิตกิตติมศักดิ์ของ
ผู้จัดทำ สำหรับเป็นแนวทางในการประเมินกระบวนการกลุ่มใน PBL นี้ ทั้งนี้ผู้ประเมินควรศึกษาหรือ
ทำความเข้าใจในคู่มือนี้ก่อนทำการประเมินจริง

วิธีการประเมินกระบวนการกลุ่มใน PBL

การเรียนรู้โดยใช้ปัญหาเป็นฐาน (PBL) คือ รูปแบบของการเรียนที่ใช้การ
ขับเคลื่อนของปัญหาที่ใกล้เคียงสถานการณ์จริง เป็นจุดเริ่มต้นสำหรับการเรียนรู้ เน้นการ
อภิปรายโต้ตอบร่วมกันของสมาชิกในกลุ่มขนาดเล็ก โดยประโยชน์จากการเรียนวิธีนี้ เช่น
ความรู้ที่ได้จะพัฒนาเป็นความจำระยะยาว มีฐานความรู้กว้าง ฝึกทักษะการแก้ปัญหา
ฝึกการทำงานร่วมกัน ฝึกการตัดสินใจ มีความตื่นตัวและมีแรงจูงใจในการเรียน ในการ
ประเมินผลการเรียนรู้ของ PBL จะเน้นประเมินกระบวนการเรียนรู้ในกลุ่มย่อย
(tutorial-based assessment) จากกลุ่มผู้ประเมิน คือ อาจารย์ เพื่อน และตนเอง
(ผู้เรียน) เนื่องจากเป็นผู้ที่ทราบถึงข้อมูลสมาชิกในกลุ่มเป็นอย่างดี และการประเมินจะ
ใช้การพิจารณาตัดสินคุณค่าโดยเปรียบเทียบกับเกณฑ์ที่กำหนดไว้ ซึ่งผลของการ
ประเมินนี้จะช่วยให้ข้อเท็จจริงอันเป็นประโยชน์ เช่น ระดับความสามารถ ผลสำเร็จ
ความก้าวหน้า/พัฒนาการ รวมถึงปัญหา/จุดอ่อนที่ต้องปรับปรุง นอกจากนี้ผู้ประเมิน
ควรมีคุณสมบัติ คือ มีความรอบรู้ในเรื่องที่จะประเมิน สามารถให้ข้อเสนอแนะเพื่อการ
พัฒนา ตลอดจนมีความซื่อสัตย์ ยุติธรรม

วิธีการประเมินกระบวนการกลุ่มให้มีความแม่นยำและลดความคลาดเคลื่อน โดยสรุปมีจำนวน 12 ข้อ ได้แก่

1. ศึกษาค้นคว้าเพิ่มเติมด้วยตนเอง (self-directed learning) ให้อบรูู้
ในประเด็นสำหรับการเรียนรู้จากโจทย์สถานการณ์
2. ให้ความสนใจและมีส่วนร่วมในระหว่างการทำกิจกรรมกลุ่ม PBL
3. ใช้ทักษะการคิด วิเคราะห์ เชื่อมโยง เกี่ยวกับความสัมพันธ์ หรือความเป็นเหตุเป็นผล
4. ฝึกการทำงานเป็นระบบ และมีความรับผิดชอบต่อกันที่
5. เข้าใจและให้ความสำคัญกับวัตถุประสงค์ของการประเมิน นั่นคือ การให้ข้อมูล
ที่เป็นประโยชน์เพื่อเป็นแนวทางสำหรับพัฒนาความสามารถของผู้เรียน
6. รวบรวมข้อมูลการทำกิจกรรมกลุ่มของสมาชิกทุกคน โดยใช้การสังเกต
พฤติกรรม ได้แก่ ความถี่ของการมีส่วนร่วมในกลุ่ม การสนใจฟัง การตั้ง
คำถาม การตอบคำถาม และการแสดงความคิดเห็นในเรื่องต่างๆ
7. ประเมินเนื้อหาการนำเสนอของสมาชิกกลุ่มทุกคน เช่น ความถูกต้อง
ตรงประเด็น มีความทันสมัย เป็นประโยชน์ต่อกลุ่ม
8. ประเมินรูปแบบการนำเสนอของสมาชิกกลุ่ม เช่น พูดชัดเจน เข้าใจง่าย
มีการสรุปใจความสำคัญ ใช้เวลาเหมาะสม
9. หลีกเลี่ยงการให้คะแนนตามความพอใจของตน การกด-ปล่อยคะแนน
หรือให้คะแนนช่วงกลางๆ เพราะผลคะแนนที่คลาดเคลื่อนเหล่านี้ไม่สามารถ
นำไปใช้ประโยชน์ต่อได้
10. สามารถให้เหตุผลหรือข้อมูลสนับสนุนผลประเมินได้ หากคะแนนที่ตน
ประเมินไม่สอดคล้องกับผู้ประเมินคนอื่นๆ
11. ตรวจสอบผลการประเมินทุกครั้ง เพื่อป้องกันข้อผิดพลาด
12. แสดงตนเป็นผู้ประเมินที่เป็นกลาง มีความยุติธรรม และมีความรับผิดชอบ
ในผลการประเมิน

ภาพผนวก: แบบประเมินกระบวนการเรียนรู้ในกลุ่มย่อย (กลุ่มที่ __)

คำชี้แจง โปรดกรอคะแนนที่ตรงมากที่สุด กับพฤติกรรมของนิสิต/ นักศึกษาภายในกลุ่ม โดยเกณฑ์การประเมินมีดังนี้

- 4 = ดีมาก หมายถึง มีทุกครั้ง หรือ มีมากครั้ง มีความรู้ มีควมดี และเกิดประโยชน์ต่อกลุ่ม ตามรายการประเมินพฤติกรรมในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 3 = ดี หมายถึง มีมากครั้ง และ มีครั้ง และ มีครั้ง ในพฤติกรรมในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 2 = ปานกลาง หมายถึง มีน้อยครั้ง หรือ มีไม่ครบถ้วน ตามรายการประเมินพฤติกรรมในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป
- 1 = ควรปรับปรุง หมายถึง ไม่มี พฤติกรรมตามรายการประเมินในข้อนั้นๆ ตลอดกระบวนการกลุ่มที่พึงเสร็จสิ้นไป

สถานะของผู้ประเมิน 1) อาจารย์ 2) นักศึกษาชื่อ.....

	ชื่อนักศึกษา/ ระดับคะแนน								
	1	2	3	4	5	6	7	8	9
1	ตั้งคำถามที่นำไปสู่การทำความเข้าใจในสถานการณ์ปัญหา								
2	ชี้/ระบุข้อมูลที่สำคัญจากสถานการณ์ปัญหา								
3	คิดวิเคราะห์ไตร่ตรองอย่างเป็นระบบ								
4	ประยุกต์ใช้ความรู้หรือประสบการณ์มาใช้กับสถานการณ์ใน PBL								
5	ตั้งสมมติฐานที่เป็นไปได้สำหรับสาเหตุหรือแนวทางการแก้ปัญหา								
6	อธิบายเชื่อมโยงประเด็นต่างๆ อย่างเป็นเหตุ เป็นผล								

รายการประเมิน

	ชื่อนักศึกษา/ ระดับคะแนน								
	1	2	3	4	5	6	7	8	9
	รายการประเมิน								
7									
	ระบุประเด็นที่ต้องการเรียนรู้จากสถานการณ์ใน PBL ได้								
8									
	ค้นหาข้อมูลความรู้ที่ทันสมัยและเป็นประโยชน์								
9									
	ค้นคว้าหาข้อมูลได้อย่างครอบคลุมและครบถ้วน								
10									
	ถ่ายทอดเนื้อหาความรู้ได้อย่างชัดเจน เข้าใจง่าย								
11									
	นำความรู้ที่ได้จากการค้นคว้ามาใช้อธิบายตอบคำถามหรือ สมมติฐานจากสถานการณ์ของปัญหา								
12									
	ชี้แนะหรือนำพากลุ่มไปสู่แนวทางในการเรียนรู้/เป้าหมายความสำเร็จ								
13									
	ให้เกิดการเพื่อนสมาชิกในกลุ่มอย่างเท่าเทียมกัน								
14									
	ให้ความร่วมมือ/ ทำงานร่วมกับผู้อื่นได้								
15									
	สนใจมีส่วนร่วม/ตั้งใจทำกิจกรรมกลุ่มอย่างสม่ำเสมอ								
16									
	แสดงออกถึงการมีความคิดริเริ่ม/ สร้างสรรค์								
17									
	สามารถให้ข้อมูลย้อนกลับ/ สะท้อนความคิดแก่กลุ่มได้								
18									
	สามารถสรุปเนื้อหาที่สำคัญของการเรียนรู้ในครั้งนี้								
19									
	แสดงความพึงพอใจต่อการเรียนแบบ PBL ในครั้งนี้								

ภาคผนวก จ
ผลการวิเคราะห์โปรแกรม Facets

Facets (Many-Facet Rasch Measurement) Version No. 3.71.4 Copyright ?(c) 1987-2014, John M. Linacre.
All rights reserved.

PBL group

Table 1. Specifications from file "F:\facet2\thesis draft\file data\MFRM\120 group_analyze.txt".

Title = PBL group

Data file = (F:\facet2\thesis draft\file data\MFRM\120 group_analyze.txt)

Output file = F:\facet2\thesis draft\file data\MFRM\120 group_analyze.out.txt

; Data specification

Facets = 3

Delements = N

Dvalues = 3,1-19 ; fixed values for data facets

Non-centered = 1

Positive = 1, 3

Labels =

1,Ratee ; (elements = 120)

2,Rater-type ; (elements = 3)

3,Item ; (elements = 19)

Model = ?,?,?,R4,1

; Output description

Arrange tables in order = MN

Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit

Fair score = Mean

Pt-biserial = Measure

Heading lines in output data files = Y

Inter-rater coefficients reported for facet = 2

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 ≥ 2

Usort unexpected observations sort order = 1,2,3

Vertical ruler definitions = 1N,2A,3N,5

WHexact - Wilson-Hilferty standardization = Y

; Convergence control

Convergence = .1, .01

Iterations (maximum) = 0 ; unlimited

Xtreme scores adjusted by = .3, .5 ;(estimation, bias)

Table 2. Data Summary Report.

Assigning models to Data= "F:\facet2\thesis draft\file data\MFRM\120 group_analyze.txt"

Total lines in data file = 361

Total data lines = 360

Responses matched to model: ?,?,?,R4,1 = 6840

Total non-blank responses found = 6840

Valid responses used for estimation = 6840

Table 3. Iteration Report.

Iteration	Max. Score Elements	Residual %	Categories	Max. Logit Elements	Change Steps
PROX 1				2.4976	
PROX 2				.0904	
JMLE 3	-535.7856	-8.6	1978.5186	-.4208	-2.0792
JMLE 4	666.6211	12.1	-590.1047	.5095	-.0302
JMLE 5	421.8496	7.2	-389.0645	.3017	.1110
JMLE 6	324.5747	5.4	-299.8229	.2215	.0795
JMLE 7	252.7729	4.1	-234.5220	.1639	.0645
JMLE 8	200.9790	3.2	-187.0109	.1244	.0522
JMLE 9	161.9883	2.5	-151.1354	.0964	.0426
JMLE 10	131.8965	2.0	-123.3850	.0759	.0350
JMLE 11	108.2207	1.7	-101.5105	.0607	.0289
JMLE 12	89.3179	1.4	-84.0117	.0490	.0240
JMLE 13	74.0537	1.1	-69.8503	.0399	.0199
JMLE 14	61.6167	.9	-58.2845	.0328	.0166
JMLE 15	51.4121	.8	-48.7698	.0270	.0139
JMLE 16	42.9922	.6	-40.8962	.0224	.0116
JMLE 17	36.0132	.5	-34.3519	.0186	.0097
JMLE 18	30.2075	.5	-28.8914	.0155	.0081
JMLE 19	25.3652	.4	-24.3226	.0130	.0068
JMLE 20	21.3169	.3	-20.4912	.0109	-.0057
JMLE 21	17.9258	.3	-17.2722	.0091	-.0048
JMLE 22	15.0811	.2	-14.5643	.0076	-.0041
JMLE 23	12.6924	.2	-12.2839	.0064	-.0034
JMLE 24	10.6846	.2	-10.3618	.0054	-.0029
JMLE 25	8.9956	.1	-8.7410	.0045	-.0024
JMLE 26	7.5747	.1	-7.3737	.0038	-.0020
JMLE 27	6.3784	.1	-6.2196	.0032	-.0017
JMLE 28	5.3716	.1	-5.2461	.0027	-.0014
JMLE 29	4.5229	.1	-4.4241	.0023	-.0012
JMLE 30	3.8081	.1	-3.7305	.0019	-.0010
JMLE 31	3.2065	.0	-3.1451	.0016	-.0009
JMLE 32	2.6997	.0	-2.6512	.0013	-.0007
JMLE 33	2.2725	.0	-2.2341	.0011	-.0006
JMLE 34	1.9126	.0	-1.8827	.0009	-.0005
JMLE 35	1.6099	.0	-1.5861	.0008	-.0004
JMLE 36	1.3550	.0	-1.3361	.0007	-.0004
JMLE 37	1.1396	.0	-1.1251	.0006	-.0003
JMLE 38	.9590	.0	-.9474	.0005	-.0003
JMLE 39	.8066	.0	-.7977	.0004	-.0002
JMLE 40	.6787	.0	-.6715	.0003	-.0002
JMLE 41	.5708	.0	-.5651	.0003	-.0002
JMLE 42	.4800	.0	-.4756	.0002	-.0001
JMLE 43	.4038	.0	-.4000	.0002	-.0001
JMLE 44	.3394	.0	-.3367	.0002	-.0001
JMLE 45	.2852	.0	-.2832	.0001	-.0001
JMLE 46	.2397	.0	-.2382	.0001	-.0001
JMLE 47	.2017	.0	-.2003	.0001	-.0001
JMLE 48	.1694	.0	-.1685	.0001	.0000
JMLE 49	.1426	.0	-.1416	.0001	.0000
JMLE 50	.1196	.0	-.1191	.0001	.0000
JMLE 51	.1006	.0	-.1001	.0000	.0000
JMLE 52	.0845	.0	-.0841	.0000	.0000

Subset connection O.K.

Table 4. Unexpected Responses - appears after Table 8.

Table 5. Measurable Data Summary.

Cat	Score	Exp.	Resd	StRes	
2.94	2.94	2.94	.00	.00	Mean (Count: 6840)
.72	.72	.40	.60	1.00	S.D. (Population)
.72	.72	.40	.60	1.00	S.D. (Sample)

Data log-likelihood chi-square = 12305.3574

Approximate degrees of freedom = 6698

Chi-square significance prob. = .0000

	Count	Mean	S.D.	Params
Responses used for estimation	= 6840	2.94	0.72	142
Count of measurable responses	= 6840			
Raw-score variance of observations	= 0.53	100.00%		
Variance explained by Rasch measures	= 0.16	30.93%		
Variance of residuals	= 0.36	69.07%		



Table 6.0 All Facet Vertical "Rulers".

Vertical = (1N,2A,3N,S) Yardstick (columns lines low high extreme)= 0,8,-1,5,End

Measr	+Ratee	-Rater-type	+Item	Scale
5	+	+	+	(4)
4	+ 99	+	+	+
3	+	+	+	+
	53 101 42 65 62 30 24 63 83			---
2	+ 7 12 69 28 50 57 68 74 75 84 85 95 98 100 105 34 46 56 67 4 16 27 35 45 60 61 64 102 103 120 18 51 55 66 71 87 89 94 111 88 108 113 115 5 20 21 32 37 41 49 86 104 112 116 73 79 96 117	+	+	+
			14 13	3
			15	
1	+ 1 2 23 31 33 39 77 80 90 92 97 106 110 9 48 78 109 119 6 8 10 14 40 43 47 29 44 54 59 76 93 107 118 3 15 52 72 13 81 58 19 26 36 114	+	+	+
		Self-rater	19	
		Tutor-rater		
*	0 *	*	* 8 18 6 10 11 3 7 16 2 4 5 9 12 17 1	* --- *
	17 22 91 11 82 25 38 70			
		Peer-rater		
-1	+	+	+	(1)
Measr	+Ratee	-Rater-type	+Item	Scale

Table 7.1.1 Ratee Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Averag	Fair(M) Averag	Model		Infit		Outfit		Estim. Discrm	Correlation		Num Ratee
				Measure	S.E.	MnSq	ZStd	MnSq	ZStd		PtMea	PtExp	
125	57	2.19	2.19	-.79	.22	.52	-3.1	.52	-3.1	1.51	.24	.46	38 Ratee
126	57	2.21	2.20	-.74	.22	1.05	.3	1.06	.3	.95	.76	.46	25 Ratee
126	57	2.21	2.20	-.74	.22	3.24	7.8	3.24	7.8	-1.42	-.42	.46	70 Ratee
129	57	2.26	2.26	-.60	.22	.94	-.2	.93	-.3	1.08	.69	.46	82 Ratee
136	57	2.39	2.38	-.27	.22	1.07	.4	1.07	.4	.93	.69	.46	11 Ratee
138	57	2.42	2.42	-.18	.21	.87	-.6	.86	-.7	1.15	.64	.46	22 Ratee
140	57	2.46	2.45	-.09	.21	.74	-1.5	.74	-1.5	1.29	.75	.46	17 Ratee
140	57	2.46	2.45	-.09	.21	1.84	3.7	1.84	3.7	.08	.31	.46	91 Ratee
144	57	2.53	2.53	.09	.21	1.49	2.4	1.47	2.3	.49	.69	.46	114 Ratee
145	57	2.54	2.54	.14	.21	.67	-1.9	.67	-1.9	1.36	.65	.46	19 Ratee
145	57	2.54	2.54	.14	.21	.69	-1.8	.70	-1.7	1.34	.50	.46	36 Ratee
146	57	2.56	2.56	.19	.21	.61	-2.4	.61	-2.4	1.44	.81	.46	26 Ratee
148	57	2.60	2.60	.28	.21	.84	-.8	.84	-.9	1.17	.33	.46	58 Ratee
149	57	2.61	2.62	.32	.21	1.48	2.4	1.48	2.4	.49	.54	.46	81 Ratee
150	57	2.63	2.63	.37	.21	1.03	.2	1.03	.2	.97	.60	.46	13 Ratee
152	57	2.67	2.67	.46	.21	.84	-.8	.84	-.8	1.18	.50	.46	3 Ratee
153	57	2.68	2.69	.51	.21	2.51	6.1	2.51	6.1	-.71	.44	.46	72 Ratee
154	57	2.70	2.71	.55	.21	.95	-.2	.96	-.1	1.08	.78	.46	15 Ratee
154	57	2.70	2.71	.55	.21	.51	-3.3	.51	-3.3	1.54	.47	.46	52 Ratee
155	57	2.72	2.72	.60	.21	.74	-1.5	.74	-1.5	1.28	.34	.46	29 Ratee
155	57	2.72	2.72	.60	.21	.90	-.5	.89	-.5	1.10	.41	.46	59 Ratee
155	57	2.72	2.72	.60	.21	2.29	5.4	2.26	5.3	-.47	.42	.46	93 Ratee
155	57	2.72	2.72	.60	.21	.87	-.6	.87	-.6	1.16	.63	.46	118 Ratee
156	57	2.74	2.74	.64	.21	.88	-.6	.88	-.6	1.13	.19	.46	44 Ratee
156	57	2.74	2.74	.64	.21	.61	-2.5	.61	-2.5	1.46	.62	.46	54 Ratee
156	57	2.74	2.74	.64	.21	.80	-1.1	.80	-1.1	1.24	.56	.46	76 Ratee
156	57	2.74	2.74	.64	.21	.68	-1.9	.68	-1.9	1.38	.62	.46	107 Ratee
157	57	2.75	2.76	.69	.21	1.15	.8	1.15	.8	.83	.48	.46	43 Ratee
158	57	2.77	2.78	.74	.21	.68	-2.0	.68	-1.9	1.34	.17	.45	8 Ratee
158	57	2.77	2.78	.74	.21	1.41	2.1	1.42	2.1	.56	.60	.45	40 Ratee
158	57	2.77	2.78	.74	.21	.61	-2.5	.61	-2.5	1.45	.65	.45	47 Ratee
159	57	2.79	2.80	.78	.21	.81	-1.1	.81	-1.1	1.24	.68	.45	6 Ratee
159	57	2.79	2.80	.78	.21	1.13	.7	1.14	.8	.83	.11	.45	10 Ratee
159	57	2.79	2.80	.78	.21	.74	-1.5	.74	-1.5	1.31	.58	.45	14 Ratee
161	57	2.82	2.83	.87	.22	.89	-.5	.93	-.3	1.12	.45	.45	78 Ratee
162	57	2.84	2.85	.92	.22	.72	-1.6	.72	-1.7	1.33	.53	.45	9 Ratee
162	57	2.84	2.85	.92	.22	.63	-2.3	.63	-2.3	1.44	.62	.45	48 Ratee
162	57	2.84	2.85	.92	.22	1.36	1.9	1.36	1.8	.59	.42	.45	109 Ratee
162	57	2.84	2.85	.92	.22	1.20	1.1	1.22	1.2	.76	.22	.45	119 Ratee
163	57	2.86	2.87	.97	.22	.92	-.4	.92	-.3	1.10	.46	.45	1 Ratee
163	57	2.86	2.87	.97	.22	1.02	.1	1.02	.1	.99	.49	.45	77 Ratee
163	57	2.86	2.87	.97	.22	.81	-1.0	.81	-1.0	1.24	.67	.45	92 Ratee
163	57	2.86	2.87	.97	.22	1.24	1.3	1.29	1.5	.69	.14	.45	106 Ratee
164	57	2.88	2.89	1.01	.22	1.52	2.6	1.51	2.5	.40	.19	.45	2 Ratee
164	57	2.88	2.89	1.01	.22	.84	-.9	.84	-.8	1.21	.63	.45	31 Ratee
164	57	2.88	2.89	1.01	.22	.80	-1.1	.81	-1.0	1.26	.72	.45	80 Ratee
164	57	2.88	2.89	1.01	.22	.70	-1.8	.70	-1.8	1.31	.15	.45	90 Ratee
165	57	2.89	2.90	1.06	.22	.63	-2.3	.63	-2.3	1.43	.42	.45	23 Ratee
165	57	2.89	2.90	1.06	.22	.67	-2.0	.66	-2.1	1.41	.65	.45	33 Ratee
165	57	2.89	2.90	1.06	.22	.51	-3.3	.52	-3.2	1.54	.45	.45	39 Ratee
165	57	2.89	2.90	1.06	.22	1.05	.3	1.04	.2	.94	.36	.45	97 Ratee
165	57	2.89	2.90	1.06	.22	1.33	1.7	1.32	1.6	.64	.66	.45	110 Ratee
166	57	2.91	2.92	1.11	.22	.85	-.8	.83	-.9	1.19	.46	.45	79 Ratee
166	57	2.91	2.92	1.11	.22	.73	-1.6	.74	-1.6	1.32	.62	.45	117 Ratee

Table 7.1.1 (Continue)

Total Score	Total Count	Obsvd Averag	Fair(M) Averag	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Num Ratee
167	57	2.93	2.94	1.15	.22	.43	-4.0	.44	-4.0	1.63	.40	.45	73 Ratee
167	57	2.93	2.94	1.15	.22	1.22	1.2	1.27	1.4	.70	.04	.45	96 Ratee
168	57	2.95	2.96	1.20	.22	.63	-2.3	.62	-2.4	1.45	.68	.45	20 Ratee
168	57	2.95	2.96	1.20	.22	.55	-2.9	.56	-2.9	1.52	.40	.45	41 Ratee
168	57	2.95	2.96	1.20	.22	.77	-1.3	.86	-.7	1.21	.19	.45	86 Ratee
169	57	2.96	2.97	1.25	.22	.78	-1.2	.78	-1.2	1.28	.69	.44	5 Ratee
169	57	2.96	2.97	1.25	.22	.67	-2.0	.67	-2.0	1.40	.61	.44	32 Ratee
169	57	2.96	2.97	1.25	.22	.88	-.6	.96	-.1	1.10	.37	.44	49 Ratee
169	57	2.96	2.97	1.25	.22	.97	-.1	1.01	.0	1.03	.39	.44	104 Ratee
169	57	2.96	2.97	1.25	.22	1.07	.4	1.09	.5	.89	.12	.44	116 Ratee
170	57	2.98	2.99	1.29	.22	.73	-1.6	.74	-1.5	1.31	.44	.44	21 Ratee
170	57	2.98	2.99	1.29	.22	.76	-1.4	.76	-1.4	1.30	.62	.44	37 Ratee
170	57	2.98	2.99	1.29	.22	.75	-1.4	.78	-1.2	1.26	.40	.44	112 Ratee
171	57	3.00	3.01	1.34	.22	1.37	1.9	1.59	2.9	.45	-.45	.44	88 Ratee
171	57	3.00	3.01	1.34	.22	1.17	.9	1.17	.9	.80	.29	.44	108 Ratee
171	57	3.00	3.01	1.34	.22	.50	-3.4	.52	-3.2	1.57	.51	.44	113 Ratee
172	57	3.02	3.03	1.39	.22	1.35	1.8	1.36	1.9	.59	.32	.44	115 Ratee
173	57	3.04	3.05	1.44	.22	1.03	.2	1.06	.3	.92	.17	.44	51 Ratee
173	57	3.04	3.05	1.44	.22	.87	-.7	.91	-.4	1.09	.04	.44	94 Ratee
173	57	3.04	3.05	1.44	.22	1.16	.9	1.16	.9	.79	.22	.44	111 Ratee
174	57	3.05	3.06	1.49	.22	1.17	.9	1.19	1.0	.79	.26	.44	18 Ratee
174	57	3.05	3.06	1.49	.22	.56	-2.9	.57	-2.8	1.51	.48	.44	66 Ratee
174	57	3.05	3.06	1.49	.22	.85	-.8	.83	-.9	1.20	.49	.44	87 Ratee
175	57	3.07	3.08	1.53	.22	1.09	.5	1.07	.4	.92	.62	.44	55 Ratee
175	57	3.07	3.08	1.53	.22	1.08	.4	1.10	.5	.85	.02	.44	71 Ratee
175	57	3.07	3.08	1.53	.22	.77	-1.3	.79	-1.2	1.23	.26	.44	89 Ratee
176	57	3.09	3.10	1.58	.22	1.05	.3	1.02	.1	.96	.50	.44	4 Ratee
176	57	3.09	3.10	1.58	.22	1.09	.5	1.06	.3	.93	.54	.44	16 Ratee
176	57	3.09	3.10	1.58	.22	1.59	2.8	1.55	2.7	.34	.51	.44	35 Ratee
177	57	3.11	3.12	1.63	.22	.91	-.4	.96	-.1	1.09	.44	.43	45 Ratee
177	57	3.11	3.12	1.63	.22	.95	-.2	1.01	.1	1.00	.14	.43	61 Ratee
177	57	3.11	3.12	1.63	.22	1.74	3.4	1.70	3.3	.11	.25	.43	102 Ratee
177	57	3.11	3.12	1.63	.22	.56	-2.9	.56	-2.9	1.51	.43	.43	103 Ratee
177	57	3.11	3.12	1.63	.22	1.22	1.1	1.18	1.0	.79	.59	.43	120 Ratee
178	57	3.12	3.14	1.68	.22	.88	-.6	.86	-.7	1.16	.52	.43	27 Ratee
178	57	3.12	3.14	1.68	.22	.78	-1.2	.78	-1.2	1.26	.46	.43	60 Ratee
178	57	3.12	3.14	1.68	.22	.72	-1.7	.71	-1.7	1.39	.73	.43	64 Ratee
179	57	3.14	3.16	1.73	.22	1.00	.0	.98	.0	1.03	.57	.43	34 Ratee
179	57	3.14	3.16	1.73	.22	1.14	.7	1.16	.9	.82	.37	.43	46 Ratee
179	57	3.14	3.16	1.73	.22	1.14	.8	1.16	.9	.79	.13	.43	67 Ratee
180	57	3.16	3.17	1.78	.22	.81	-1.1	.79	-1.2	1.28	.65	.43	56 Ratee
181	57	3.18	3.19	1.83	.22	.62	-2.4	.61	-2.5	1.43	.33	.43	57 Ratee
181	57	3.18	3.19	1.83	.22	1.11	.6	1.07	.4	.90	.54	.43	75 Ratee
181	57	3.18	3.19	1.83	.22	2.20	5.1	2.21	5.1	-.43	.18	.43	105 Ratee
182	57	3.19	3.21	1.88	.22	.48	-3.6	.49	-3.5	1.64	.68	.43	28 Ratee
182	57	3.19	3.21	1.88	.22	.97	-.1	1.02	.1	.99	.30	.43	50 Ratee
182	57	3.19	3.21	1.88	.22	1.28	1.5	1.29	1.5	.68	.36	.43	74 Ratee
182	57	3.19	3.21	1.88	.22	.55	-2.9	.56	-2.9	1.57	.68	.43	84 Ratee
182	57	3.19	3.21	1.88	.22	1.19	1.0	1.18	1.0	.81	.50	.43	85 Ratee
182	57	3.19	3.21	1.88	.22	.57	-2.8	.59	-2.7	1.52	.57	.43	98 Ratee
183	57	3.21	3.23	1.93	.23	.85	-.8	.82	-1.0	1.21	.55	.42	68 Ratee
183	57	3.21	3.23	1.93	.23	.95	-.2	.92	-.4	1.06	.38	.42	95 Ratee
183	57	3.21	3.23	1.93	.23	1.24	1.3	1.27	1.4	.66	.24	.42	100 Ratee
184	57	3.23	3.25	1.98	.23	1.88	4.0	1.84	3.8	-.09	.15	.42	12 Ratee

Table 7.1.1 (Continue)

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim Discrm	Correlation PtMea	Correlation PtExp	Num Ratee
185	57	3.25	3.27	2.03	.23	1.77	3.6	1.70	3.3	.12	.34	.42	7 Ratee
185	57	3.25	3.27	2.03	.23	.83	-.9	.81	-1.0	1.23	.54	.42	69 Ratee
187	57	3.28	3.30	2.14	.23	.79	-1.2	.88	-.6	1.17	.17	.41	63 Ratee
187	57	3.28	3.30	2.14	.23	.98	.0	1.07	.4	1.00	.36	.41	83 Ratee
190	57	3.33	3.36	2.30	.23	.85	-.8	.90	-.4	1.18	.47	.41	24 Ratee
192	57	3.37	3.39	2.41	.23	.80	-1.1	.79	-1.1	1.23	.41	.40	30 Ratee
194	57	3.40	3.43	2.52	.24	1.30	1.6	1.20	1.1	.70	.48	.39	62 Ratee
195	57	3.42	3.45	2.57	.24	1.26	1.4	1.20	1.0	.72	.45	.39	65 Ratee
196	57	3.44	3.47	2.63	.24	1.28	1.5	1.15	.8	.76	.48	.39	42 Ratee
198	57	3.47	3.51	2.75	.24	.93	-.3	.99	.0	1.05	.36	.38	53 Ratee
198	57	3.47	3.51	2.75	.24	1.04	.2	1.04	.2	.93	.27	.38	101 Ratee
215	57	3.77	3.81	4.01	.32	.94	-.1	.98	.0	1.02	.24	.27	99 Ratee
167.7	57	2.94	2.95	1.21	.22	1.00	-.2	1.00	-.2		.43		Mean (Count: 120)
15.7	.0	.28	.28	.77	.01	.42	2.1	.42	2.0		.22		S.D. (Population)
15.8	.0	.28	.29	.77	.01	.42	2.1	.42	2.1		.22		S.D. (Sample)

Model, Populn: RMSE .22 Adj (True) S.D. .73 Separation 3.32 Strata 4.76 Reliability .92
 Model, Sample: RMSE .22 Adj (True) S.D. .74 Separation 3.34 Strata 4.78 Reliability .92
 Model, Fixed (all same) chi-square: 1343.7 d.f.: 119 significance (probability): .00
 Model, Random (normal) chi-square: 109.3 d.f.: 118 significance (probability): .70

PBL group 10/5/2558 20:09:06

Table 7.2.1 Rater-type Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim Discrm	Correlation PtMea	Correlation PtExp	Exact Obs %	Agree Exp %	Inter Rater
7323	2280	3.21	3.24	-.75	.04	1.05	1.8	1.06	2.0	.93	.46	.48	40.6	41.7	2 Peer
6483	2280	2.84	2.85	.28	.03	.98	-.5	.99	-.2	1.02	.53	.51	43.1	44.1	3 Tutor
6318	2280	2.77	2.78	.47	.03	.97	-1.2	.96	-1.2	1.04	.51	.51	42.5	43.1	1 Self
6708.0	2280	2.94	2.96	.00	.03	1.00	.0	1.00	.2		.50				Mean(Count:3)
440.1	.0	.19	.20	.53	.00	.04	1.3	.04	1.4		.03				S.D. (Population)
539.0	.0	.24	.25	.66	.00	.05	1.6	.05	1.7		.04				S.D. (Sample)

Model, Populn: RMSE .03 Adj (True) S.D. .53 Separation 15.35 Strata 20.80 Reliability(not interrater)1.00
 Model, Sample: RMSE .03 Adj (True) S.D. .65 Separation 18.81 Strata 25.41 Reliability(not interrater)1.00
 Model, Fixed (all same) chi-square: 687.7 d.f.: 2 significance (probability): .00
 Model, Random (normal) chi-square: 2.0 d.f.: 1 significance (probability): .16
 Inter-Rater agreement opportunities: 6840 Exact agreements: 2878 = 42.1% Expected: 2939.7 = 43.0%

Table 7.3.1 Item Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu Item
980	360	2.72	2.73	-.60	.09	.91	-1.3	.91	-1.2	1.10	.47	.50	1 1
994	360	2.76	2.77	-.50	.09	1.05	.6	1.04	.6	.95	.54	.50	12 12
997	360	2.77	2.78	-.48	.09	.92	-1.0	.92	-1.1	1.09	.53	.50	17 17
1006	360	2.79	2.80	-.41	.09	.73	-4.2	.73	-4.2	1.31	.54	.50	2 2
1006	360	2.79	2.80	-.41	.09	.90	-1.5	.90	-1.4	1.11	.44	.50	5 5
1008	360	2.80	2.81	-.40	.09	.99	.0	1.00	.0	1.01	.50	.50	4 4
1015	360	2.82	2.83	-.35	.09	1.18	2.3	1.18	2.3	.80	.49	.50	9 9
1022	360	2.84	2.85	-.30	.09	.86	-1.9	.87	-1.9	1.16	.55	.50	3 3
1023	360	2.84	2.85	-.29	.09	.89	-1.5	.89	-1.5	1.12	.50	.49	7 7
1036	360	2.88	2.89	-.19	.09	1.02	.3	1.02	.3	.97	.52	.49	16 16
1038	360	2.88	2.89	-.18	.09	.94	-.8	.96	-.5	1.06	.51	.49	6 6
1040	360	2.89	2.90	-.16	.09	1.13	1.7	1.14	1.9	.85	.49	.49	10 10
1041	360	2.89	2.90	-.15	.09	.91	-1.3	.91	-1.3	1.10	.48	.49	11 11
1057	360	2.94	2.95	-.03	.09	1.20	2.7	1.20	2.7	.77	.47	.49	8 8
1069	360	2.97	2.98	.06	.09	.89	-1.5	.89	-1.6	1.13	.57	.49	18 18
1112	360	3.09	3.10	.38	.09	.93	-1.0	.96	-.4	1.06	.45	.48	19 19
1208	360	3.36	3.39	1.17	.09	1.21	2.8	1.19	2.5	.75	.42	.46	15 15
1232	360	3.42	3.46	1.39	.10	1.26	3.5	1.25	3.1	.67	.35	.45	13 13
1240	360	3.44	3.48	1.46	.10	1.17	2.3	1.13	1.7	.81	.42	.45	14 14
1059.2	360	2.94	2.95	.00	.09	1.00	.0	1.00	.0		.49		Mean(Count:19)
78.3	.0	.22	.23	.62	.00	.14	2.0	.14	1.9		.05		S.D.(Population)
80.5	.0	.22	.23	.64	.00	.15	2.1	.14	2.0		.05		S.D.(Sample)

Model, Populn: RMSE .09 Adj (True) S.D. .62 Separation 7.01 Strata 9.69 Reliability .98
 Model, Sample: RMSE .09 Adj (True) S.D. .63 Separation 7.21 Strata 9.95 Reliability .98
 Model, Fixed (all same) chi-square: 846.1 d.f.: 18 significance (probability): .00
 Model, Random (normal) chi-square: 17.6 d.f.: 17 significance (probability): .41

Table 8.1 Category Statistics.

Model = ?,?,?,R4

DATA				QUALITY CONTROL			RASCH-ANDRICH		EXPECTATION		MOST	RASCH-	Cat
Score	Category Total	Counts Used	Cum. %	Avg Exp. Meas	OUTFIT Meas	MnSq	Thresholds Measure	at S.E.	Category -0.5	PROBABLE rom	THURSTONE Threshold	PEAK Prob	
1	116	116	2%	-.20	-.28	1.1			(-3.71)	low	low	100%	
2	1658	1658	4%	.43	.43	1.0	-2.59	.10	-1.28 -2.74	-2.59	-2.65	65%	
3	3572	3572	52%	1.21	1.21	1.0	.04	.03	1.31 .03	.04	.03	63%	
4	1494	1494	22%	2.18	2.17	1.0	2.54	.03	(3.68) 2.72	2.54	2.60	100%	

(Mean)------(Modal)---(Median)

Table 8.2 Category Statistics.

Model = ?,1,?,R4 ; Rater-type: Self-rater

DATA				QUALITY CONTROL			RASCH-ANDRICH		EXPECTATION		MOST	RASCH-	Cat
Score	Category Total	Counts Used	Cum. %	Avg Exp. Meas	OUTFIT Meas	MnSq	Thresholds Measure	at S.E.	Category -0.5	PROBABLE from	THURSTONE Threshold	PEAK Prob	
1	50	50	2%	-.17	-.37	1.1			(-3.85)	low	low	100%	
2	732	732	32%	.24	.26	1.0	-2.74	.15	-1.33 -2.86	-2.74	-2.79	67%	
3	1188	1188	52%	.90	.91	1.0	.09	.05	1.38 .05	.09	.06	64%	
4	310	310	14%	1.80	1.75	1.0	2.65	.07	(3.78) 2.82	2.65	2.71	100%	

(Mean)------(Modal)---(Median)

Table 8.3 Category Statistics.

Model = 2,2,2,R4 ; Rater-type: Peer-rater

DATA				QUALITY CONTROL			RASCH-ANDRICH		EXPECTATION		MOST	RASCH-	Cat
Score	Category	Counts	Cum.	Avg	Exp.	OUTFIT	Thresholds	Measure	at	PROBABLE	THURSTONE	PEAK	
Total	Used	%	%	Meas	Meas	MnSq	Measure	S.E.	Category	-0.5	from	Threshold	Prob
1	15	15	1%	.08	.38	.9			(-3.41)		low	low	100%
2	285	285	13%	1.08	1.00	1.1	-2.26	.26	-1.21	-2.49	-2.26	-2.36	59%
3	1182	1182	52%	1.62	1.62	1.0	-.11	.07	1.16	-.05	-.11	-.09	63%
4	798	798	35%	2.36	2.39	1.0	2.37	.05	(3.51)	2.56	2.37	2.44	100%
									(Mean)	(Modal)		(Median)	

Table 8.4 Category Statistics.

Model = 2,3,2,R4 ; Rater-type: Tutor-rater

DATA				QUALITY CONTROL			RASCH-ANDRICH		EXPECTATION		MOST	RASCH-	Cat
Score	Category	Counts	Cum.	Avg	Exp.	OUTFIT	Thresholds	Measure	at	PROBABLE	THURSTONE	PEAK	
Total	Used	%	%	Meas	Meas	MnSq	Measure	S.E.	Category	-0.5	from	Threshold	Prob
1	51	51	2%	-.34	-.29	1.0			(-3.62)		low	low	100%
2	641	641	28%	.31	.33	1.0	-2.51	.15	-1.26	-2.68	-2.51	-2.58	64%
3	1202	1202	53%	.98	.97	.9	.02	.05	1.27	.01	.02	.00	63%
4	386	386	17%	1.82	1.79	1.0	2.49	.06	(3.63)	2.67	2.49	2.55	100%
									(Mean)	(Modal)		(Median)	

PBL group 10/5/2558 20:09:06

Table 4.1 Unexpected Responses (100 residuals sorted by 1,2,3).

Cat	Score	Exp.	Resd	StRes	Num	Ratee	N	Rater-type	Nu	It
4	4	2.5	1.5	2.4	2	Ratee 1	Self-rater	17	17	
4	4	2.5	1.5	2.4	3	Ratee 3	Tutor-rater	10	10	
2	2	3.4	-1.4	-2.5	7	Ratee 2	Peer-rater	3	3	
2	2	3.4	-1.4	-2.4	7	Ratee 2	Peer-rater	4	4	
2	2	3.4	-1.4	-2.4	7	Ratee 2	Peer-rater	5	5	
2	2	3.4	-1.4	-2.4	7	Ratee 2	Peer-rater	9	9	
2	2	3.5	-1.5	-2.6	7	Ratee 2	Peer-rater	16	16	
4	4	2.5	1.5	2.4	8	Ratee 3	Tutor-rater	17	17	
4	4	2.5	1.5	2.4	10	Ratee 1	Self-rater	3	3	
1	1	2.9	-1.9	-3.1	12	Ratee 1	Self-rater	9	9	
2	2	3.4	-1.4	-2.4	12	Ratee 2	Peer-rater	3	3	
2	2	3.4	-1.4	-2.5	12	Ratee 2	Peer-rater	6	6	
1	1	3.3	-2.3	-3.8	18	Ratee 2	Peer-rater	10	10	
1	1	2.7	-1.7	-2.7	20	Ratee 3	Tutor-rater	4	4	
3	3	3.9	-.9	-2.6	24	Ratee 2	Peer-rater	14	14	
4	4	2.5	1.5	2.4	29	Ratee 3	Tutor-rater	9	9	
1	1	2.7	-1.7	-2.8	35	Ratee 1	Self-rater	12	12	
1	1	2.9	-1.9	-3.0	35	Ratee 1	Self-rater	16	16	
2	2	3.5	-1.5	-2.7	35	Ratee 2	Peer-rater	19	19	
2	2	3.4	-1.4	-2.4	41	Ratee 3	Tutor-rater	14	14	
1	1	2.5	-1.5	-2.4	43	Ratee 1	Self-rater	16	16	
2	2	3.5	-1.5	-2.7	45	Ratee 3	Tutor-rater	13	13	
2	2	3.5	-1.5	-2.8	45	Ratee 3	Tutor-rater	14	14	
2	2	3.7	-1.7	-3.4	49	Ratee 2	Peer-rater	13	13	

Table 4.1 (Continue)

Cat	Score	Exp.	Resd	StRes	Num	Ratee	N	Rater-type	Nu	It
3	3	3.9	-.9	-2.8	53	Ratee	2	Peer-rater	15	15
1	1	2.9	-1.9	-3.1	55	Ratee	1	Self-rater	8	8
1	1	2.8	-1.8	-2.8	55	Ratee	1	Self-rater	9	9
1	1	2.5	-1.5	-2.4	59	Ratee	3	Tutor-rater	7	7
2	2	3.4	-1.4	-2.4	60	Ratee	2	Peer-rater	8	8
2	2	3.5	-1.5	-2.7	61	Ratee	2	Peer-rater	19	19
2	2	3.5	-1.5	-2.7	63	Ratee	2	Peer-rater	11	11
1	1	3.2	-2.2	-3.7	65	Ratee	1	Self-rater	10	10
2	2	3.4	-1.4	-2.4	67	Ratee	1	Self-rater	15	15
2	2	3.5	-1.5	-2.8	67	Ratee	2	Peer-rater	19	19
4	4	1.9	2.1	3.5	70	Ratee	1	Self-rater	4	4
4	4	1.9	2.1	3.6	70	Ratee	1	Self-rater	5	5
4	4	2.0	2.0	3.3	70	Ratee	1	Self-rater	8	8
4	4	2.0	2.0	3.4	70	Ratee	1	Self-rater	11	11
4	4	2.2	1.8	3.0	70	Ratee	1	Self-rater	19	19
1	1	2.5	-1.5	-2.4	70	Ratee	2	Peer-rater	8	8
1	1	3.0	-2.0	-3.3	70	Ratee	2	Peer-rater	13	13
1	1	3.1	-2.1	-3.4	70	Ratee	2	Peer-rater	14	14
1	1	2.9	-1.9	-3.2	70	Ratee	2	Peer-rater	15	15
1	1	2.6	-1.6	-2.6	70	Ratee	2	Peer-rater	19	19
1	1	2.6	-1.6	-2.5	72	Ratee	3	Tutor-rater	8	8
1	1	3.1	-2.1	-3.5	72	Ratee	3	Tutor-rater	13	13
1	1	3.1	-2.1	-3.5	72	Ratee	3	Tutor-rater	14	14
1	1	3.0	-2.0	-3.3	72	Ratee	3	Tutor-rater	15	15
1	1	2.5	-1.5	-2.4	72	Ratee	3	Tutor-rater	16	16
1	1	2.7	-1.7	-2.8	72	Ratee	3	Tutor-rater	19	19
2	2	3.5	-1.5	-2.6	74	Ratee	1	Self-rater	15	15
2	2	3.5	-1.5	-2.6	74	Ratee	2	Peer-rater	8	8
2	2	3.5	-1.5	-2.7	74	Ratee	3	Tutor-rater	15	15
2	2	3.6	-1.6	-3.0	78	Ratee	2	Peer-rater	13	13
2	2	3.6	-1.6	-3.3	83	Ratee	3	Tutor-rater	13	13
2	2	3.7	-1.7	-3.4	83	Ratee	3	Tutor-rater	14	14
2	2	3.6	-1.6	-3.0	85	Ratee	3	Tutor-rater	13	13
2	2	3.7	-1.7	-3.5	86	Ratee	2	Peer-rater	14	14
2	2	3.7	-1.7	-3.6	88	Ratee	2	Peer-rater	13	13
2	2	3.7	-1.7	-3.7	88	Ratee	2	Peer-rater	14	14
2	2	3.7	-1.7	-3.3	88	Ratee	2	Peer-rater	15	15
1	1	2.8	-1.8	-2.9	91	Ratee	1	Self-rater	13	13
1	1	2.8	-1.8	-3.0	91	Ratee	1	Self-rater	14	14
1	1	2.7	-1.7	-2.8	91	Ratee	1	Self-rater	15	15
4	4	2.2	1.8	2.9	91	Ratee	3	Tutor-rater	9	9
1	1	2.5	-1.5	-2.4	93	Ratee	1	Self-rater	8	8
1	1	2.5	-1.5	-2.4	93	Ratee	1	Self-rater	11	11
1	1	3.0	-2.0	-3.2	93	Ratee	1	Self-rater	15	15
1	1	2.6	-1.6	-2.5	93	Ratee	1	Self-rater	18	18
1	1	2.7	-1.7	-2.7	93	Ratee	1	Self-rater	19	19
4	4	2.4	1.6	2.6	93	Ratee	3	Tutor-rater	1	1
4	4	2.5	1.5	2.5	93	Ratee	3	Tutor-rater	4	4
4	4	2.5	1.5	2.5	93	Ratee	3	Tutor-rater	5	5
4	4	2.5	1.5	2.4	93	Ratee	3	Tutor-rater	9	9
2	2	3.6	-1.6	-3.0	96	Ratee	2	Peer-rater	15	15
1	1	2.6	-1.6	-2.6	97	Ratee	1	Self-rater	16	16
3	3	3.9	-.9	-2.7	99	Ratee	2	Peer-rater	10	10
3	3	3.9	-.9	-3.6	99	Ratee	2	Peer-rater	19	19
2	2	3.7	-1.7	-3.4	101	Ratee	2	Peer-rater	6	6
1	1	2.9	-1.9	-3.1	102	Ratee	3	Tutor-rater	7	7
2	2	3.4	-1.4	-2.5	102	Ratee	3	Tutor-rater	15	15

Table 4.1 (Continue)

Cat	Score	Exp.	Resd	StRes	Num	Ratee	N	Rater-type	Nu	It
1	1	2.8	-1.8	-2.9	102	Ratee	3	Tutor-rater	17	17
2	2	3.6	-1.6	-3.2	104	Ratee	2	Peer-rater	15	15
2	2	3.6	-1.6	-2.9	105	Ratee	3	Tutor-rater	13	13
2	2	3.6	-1.6	-3.0	105	Ratee	3	Tutor-rater	14	14
1	1	3.5	-2.5	-4.5	105	Ratee	3	Tutor-rater	15	15
4	4	2.5	1.5	2.4	106	Ratee	1	Self-rater	5	5
2	2	3.6	-1.6	-3.1	106	Ratee	2	Peer-rater	13	13
2	2	3.4	-1.4	-2.4	108	Ratee	3	Tutor-rater	13	13
1	1	2.5	-1.5	-2.4	109	Ratee	3	Tutor-rater	1	1
1	1	2.6	-1.6	-2.6	110	Ratee	3	Tutor-rater	12	12
1	1	2.6	-1.6	-2.6	110	Ratee	3	Tutor-rater	17	17
1	1	2.8	-1.8	-2.9	110	Ratee	3	Tutor-rater	18	18
1	1	2.5	-1.5	-2.4	114	Ratee	1	Self-rater	19	19
2	2	3.4	-1.4	-2.5	115	Ratee	3	Tutor-rater	13	13
2	2	3.5	-1.5	-2.5	115	Ratee	3	Tutor-rater	14	14
2	2	3.4	-1.4	-2.4	116	Ratee	3	Tutor-rater	13	13
4	4	2.5	1.5	2.4	119	Ratee	1	Self-rater	12	12
2	2	3.5	-1.5	-2.8	119	Ratee	2	Peer-rater	15	15
1	1	2.7	-1.7	-2.8	120	Ratee	1	Self-rater	12	12
Cat	Score	Exp.	Resd	StRes	Num	Ratee	N	Rater-type	Nu	It



ภาคผนวก ฉ

ตัวอย่างผลการวิเคราะห์โมเดลสมการโครงสร้างกลุ่มพหุด้วยโปรแกรม Lisrel

LISREL 8.52

BY

Karl G. Jöreskog & Dag Sörbom

This program is published exclusively by
 Scientific Software International, Inc.
 7383 N. Lincoln Avenue, Suite 100
 Lincolnwood, IL 60712, U.S.A.

Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140

Copyright by Scientific Software International, Inc., 1981-2002

Use of this program is subject to the terms specified in the
 Universal Copyright Convention.

Website: www.ssicentral.com

The following lines were read from file G:\CU ETHESIS\DATA_PH1_2\1 MG_MO.LS8:

TUTOR

TI MULTIPLE GROUP PBL

DA NI=8 NO=120 NG=3 MA=KM

LA

GOAL ABILI CONSCI MOTIV ACCOU STAND COMPAR ERROR

KM

1.000

0.219 1.000

0.003 0.360 1.000

0.256 0.422 0.375 1.000

0.094 0.258 0.363 0.462 1.000

0.025 0.168 0.410 0.474 0.448 1.000

0.023 0.371 0.422 0.493 0.529 0.545 1.000

0.091 -0.076 -0.255 -0.008 0.018 -0.256 -0.042 1.000

ME

3.983 4.275 38.700 19.200 21.192 22.233 16.508 0.382

SD

0.917 0.860 4.283 2.630 2.677 2.340 1.983 1.583

SE

6 7 8 1 2 3 4 5 /

MO NX=5 NY=3 NK=1 NE=1 LY=FU,FI LX=FU,FR GA=FR PH=DI PS=DI TE=SY TD=SY

LE

RATING

LK

CONTEXT

FR LX(1,1) LX(2,1) LX(3,1) LX(4,1) LX(5,1)

FR LY(1,1) LY(2,1) LY(3,1) GA(1,1)
 FI TE(3,1)
 ST 0.12 TE(3,2)
 PD
 OU ME=UL PC RS EF FS SS SC PT MR ND=3 AD=OFF IT=250 MI

TUTOR

Number of Input Variables 8
 Number of Y - Variables 3
 Number of X - Variables 5
 Number of ETA - Variables 1
 Number of KSI - Variables 1
 Number of Observations 120
 Number of Groups 3

PEER

DA NI=8 NO=120 NG=3 MA=KM

LA

GOAL ABILI CONSCI MOTIV ACCOU STAND COMPAR ERROR

KM

1.000

0.124 1.000

-0.002 0.181 1.000

0.263 0.270 0.350 1.000

0.144 0.127 0.274 0.417 1.000

0.122 0.171 0.138 0.408 0.451 1.000

0.124 0.157 0.234 0.456 0.475 0.523 1.000

0.025 -0.227 -0.075 -0.085 -0.078 -0.234 -0.272 1.000

ME

2.908 3.392 35.908 18.008 20.567 21.917 15.558 1.048

SD

1.037 0.802 3.658 2.310 2.595 2.266 1.846 1.408

SE

6 7 8 1 2 3 4 5 /

MO NX=5 NY=3 NK=1 NE=1 LY=PS LX=PS GA=PS PH=PS PS=PS TE=PS TD=PS

LE

RATING

LK

CONTEXT

PD

OU

PEER

Number of Input Variables 8
 Number of Y - Variables 3
 Number of X - Variables 5
 Number of ETA - Variables 1
 Number of KSI - Variables 1
 Number of Observations 120

Number of Groups 3

SELF
 DA NI=8 NO=120 NG=3 MA=KM
 LA
 GOAL ABILI CONSCI MOTIV ACCOU STAND COMPAR ERROR
 KM
 1.000
 0.234 1.000
 0.173 0.012 1.000
 0.263 0.160 0.327 1.000
 0.163 -0.006 0.177 0.360 1.000
 0.171 0.039 0.317 0.322 0.334 1.000
 0.115 0.263 0.220 0.467 0.386 0.332 1.000
 -0.154 -0.103 -0.315 -0.220 -0.231 -0.387 -0.160 1.000
 ME
 2.992 3.392 35.883 17.492 20.317 21.442 15.283 0.665
 SD
 1.041 0.910 4.197 2.473 2.738 2.435 1.567 1.312
 SE
 6 7 8 1 2 3 4 5 /
 MO NX=5 NY=3 NK=1 NE=1 LY=PS LX=PS GA=PS PH=PS PS=PS TE=PS TD=PS
 LE
 RATING
 LK
 CONTEXT
 PD
 OU

SELF
 Number of Input Variables 8
 Number of Y - Variables 3
 Number of X - Variables 5
 Number of ETA - Variables 1
 Number of KSI - Variables 1
 Number of Observations 120
 Number of Groups 3

TUTOR
 Parameter Specifications

LAMBDA-Y

RATING

STAND 1
 COMPAR 2
 ERROR 3

LAMBDA-X

CONTEXT

GOAL	4
ABILI	5
CONSCI	6
MOTIV	7
ACCOU	8

GAMMA

CONTEXT

RATING	9
--------	---

PHI EQUALS PHI IN THE FOLLOWING GROUP

PSI

RATING

10

THETA-EPS

STAND	COMPAR	ERROR
-------	--------	-------

11	12	13
----	----	----

THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU
------	-------	--------	-------	-------

14	15	16	17	18
----	----	----	----	----

PEER

Parameter Specifications

LAMBDA-Y

RATING

STAND	19
COMPAR	20
ERROR	21

LAMBDA-X

CONTEXT

GOAL 22
 ABILI 23
 CONSCI 24
 MOTIV 25
 ACCOU 26

GAMMA

CONTEXT

RATING 27

PHI EQUALS PHI IN THE FOLLOWING GROUP

PSI

RATING

28

THETA-EPS

STAND	COMPAR	ERROR
-------	--------	-------

29	30	31
----	----	----

THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU
------	-------	--------	-------	-------

32	33	34	35	36
----	----	----	----	----

SELF

Parameter Specifications

LAMBDA-Y

RATING

STAND 37
 COMPAR 38
 ERROR 39

LAMBDA-X

CONTEXT

GOAL 40
 ABILI 41
 CONSCI 42
 MOTIV 43
 ACCOU 44

GAMMA

CONTEXT

 RATING 45

PSI

RATING

46

THETA-EPS

STAND	COMPAR	ERROR
47	48	49

47 48 49

THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU
50	51	52	53	54

50 51 52 53 54

TUTOR

Number of Iterations = 69

LISREL Estimates (Unweighted Least Squares)

LAMBDA-Y

RATING

STAND 0.700
 (0.050)
 13.898

COMPAR 0.810
 (0.058)
 13.860

ERROR -0.195
 (0.054)
 -3.613

LAMBDA-X

CONTEXT

GOAL 0.151
(0.053)
2.820

ABILI 0.492
(0.054)
9.146

CONSCI 0.604
(0.056)
10.814

MOTIV 0.721
(0.055)
13.152

ACCOU 0.651
(0.052)
12.469

GAMMA

CONTEXT

RATING 0.878
(0.082)
10.652



Covariance Matrix of ETA and KSI

RATING CONTEXT

RATING	1.012	
CONTEXT	0.878	1.000

PHI EQUALS PHI IN THE FOLLOWING GROUP

PSI

RATING

0.242
(0.185)
1.306

Squared Multiple Correlations for Structural Equations

RATING

0.761

THETA-EPS

STAND COMPAR ERROR

----- ----- -----

STAND 0.503
 (0.093)
 5.412

COMPAR -- 0.336
 (0.149)
 2.263

ERROR -- 0.120 0.962
 (0.096)
 10.033

Squared Multiple Correlations for Y - Variables

STAND COMPAR ERROR

----- ----- -----

0.497 0.664 0.038

THETA-DELTA

GOAL ABILI CONSCI MOTIV ACCOU

----- ----- ----- ----- -----

0.977 0.758 0.636 0.480 0.576
 (0.076) (0.077) (0.095) (0.084) (0.068)
 12.942 9.903 6.683 5.701 8.533

Squared Multiple Correlations for X - Variables

GOAL ABILI CONSCI MOTIV ACCOU

----- ----- ----- ----- -----

0.023 0.242 0.364 0.520 0.424

Group Goodness of Fit Statistics

Root Mean Square Residual (RMR) = 0.0724

Standardized RMR = 0.0724

Goodness of Fit Index (GFI) = 0.973

Summary Statistics for Fitted Residuals

Smallest Fitted Residual = -0.152
 Median Fitted Residual = 0.000
 Largest Fitted Residual = 0.147

Stemleaf Plot

```
- 1|5
- 1|32
- 0|98766
- 0|332110000000000
 0|1234
 0|5677
 1|2234
 1|5
```

Standardized Residuals

	STAND	COMPAR	ERROR	GOAL	ABILI	CONSCI
STAND	0.000					
COMPAR	-0.701	0.000				
ERROR	-1.438	-0.030	0.000			
GOAL	-0.794	-1.011	1.284	0.000		
ABILI	-1.656	0.271	0.092	1.657	0.000	
CONSCI	0.503	-0.090	-1.755	-1.031	0.781	0.000
MOTIV	0.410	-0.259	1.364	1.782	0.862	-0.786
ACCOU	0.625	0.864	1.508	-0.050	-0.766	-0.381

Standardized Residuals

	MOTIV	ACCOU
MOTIV	0.000	
ACCOU	-0.094	0.000

Summary Statistics for Standardized Residuals

Smallest Standardized Residual = -1.755
 Median Standardized Residual = 0.000
 Largest Standardized Residual = 1.782

Stemleaf Plot

```
- 1|87
- 1|400
- 0|8887
- 0|431100000000000
 0|134
 0|56899
 1|34
 1|578
```

TUTOR

Qplot of Standardized Residuals



TUTOR

Modification Indices and Expected Change

No Non-Zero Modification Indices for LAMBDA-Y

No Non-Zero Modification Indices for LAMBDA-X

No Non-Zero Modification Indices for GAMMA

No Non-Zero Modification Indices for PHI

No Non-Zero Modification Indices for PSI

Modification Indices for THETA-EPS

STAND	COMPAR	ERROR
STAND	--	
COMPAR	0.491	--
ERROR	2.067	0.001

Expected Change for THETA-EPS

STAND	COMPAR	ERROR
STAND	--	
COMPAR	-0.142	--
ERROR	-0.147	-0.003

Completely Standardized Expected Change for THETA-EPS

STAND	COMPAR	ERROR
STAND	--	
COMPAR	-0.142	--
ERROR	-0.147	-0.003

Modification Indices for THETA-DELTA-EPS

STAND	COMPAR	ERROR
GOAL	0.630	1.022
ABILI	2.742	0.073
CONSCI	0.253	0.008
MOTIV	0.169	0.067
ACCOU	0.390	0.746

Expected Change for THETA-DELTA-EPS

STAND	COMPAR	ERROR
GOAL	-0.078	-0.102
ABILI	-0.172	0.028
CONSCI	0.055	-0.010

MOTIV	0.046	-0.029	0.136
ACCOU	0.069	0.095	0.148

Completely Standardized Expected Change for THETA-DELTA-EPS

	STAND	COMPAR	ERROR
GOAL	-0.078	-0.102	0.119
ABILI	-0.172	0.028	0.009
CONSCI	0.055	-0.010	-0.171
MOTIV	0.046	-0.029	0.136
ACCOU	0.069	0.095	0.148

Modification Indices for THETA-DELTA

	GOAL	ABILI	CONSCI	MOTIV	ACCOU
GOAL	--				
ABILI	2.745	--			
CONSCI	1.064	0.610	--		
MOTIV	3.174	0.743	0.617	--	
ACCOU	0.002	0.587	0.145	0.009	--

Expected Change for THETA-DELTA

	GOAL	ABILI	CONSCI	MOTIV	ACCOU
GOAL	--				
ABILI	0.159	--			
CONSCI	-0.102	0.081	--		
MOTIV	0.181	0.092	-0.086	--	
ACCOU	-0.005	-0.080	-0.041	-0.010	--

Completely Standardized Expected Change for THETA-DELTA

	GOAL	ABILI	CONSCI	MOTIV	ACCOU
GOAL	--				
ABILI	0.159	--			
CONSCI	-0.102	0.081	--		
MOTIV	0.181	0.092	-0.086	--	
ACCOU	-0.005	-0.080	-0.041	-0.010	--

TUTOR

Factor Scores Regressions

ETA

	STAND	COMPAR	ERROR	GOAL	ABILI	CONSCI
RATING	0.257	0.480	-0.097	0.015	0.063	0.092

ETA

	MOTIV	ACCOU
RATING	0.146	0.110

KSI

	STAND	COMPAR	ERROR	GOAL	ABILI	CONSCI
CONTEXT	0.135	0.252	-0.051	0.030	0.127	0.186

KSI

	MOTIV	ACCOU
CONTEXT	0.293	0.221

TUTOR

Within Group Completely Standardized Solution

LAMBDA-Y	RATING
STAND	0.705
COMPAR	0.815
ERROR	-0.196

LAMBDA-X	CONTEXT
GOAL	0.151
ABILI	0.492
CONSCI	0.604
MOTIV	0.721
ACCOU	0.651

GAMMA	CONTEXT
RATING	0.873

Correlation Matrix of ETA and KSI

	RATING	CONTEXT
RATING	1.000	
CONTEXT	0.873	1.000



PSI

RATING

0.239

THETA-EPS

STAND COMPAR ERROR

STAND	0.503		
COMPAR	--	0.336	
ERROR	--	0.120	0.962

THETA-DELTA

GOAL ABILI CONSCI MOTIV ACCOU

0.977 0.758 0.636 0.480 0.576

Regression Matrix ETA on KSI (Standardized)

CONTEXT

RATING 0.873

TUTOR

Total and Indirect Effects

Total Effects of KSI on Y

CONTEXT

STAND	0.615
	(0.067)
	9.116
COMPAR	0.711
	(0.059)
	11.954
ERROR	-0.171
	(0.049)
	-3.525

TUTOR

Standardized Total and Indirect Effects

Standardized Total Effects of KSI on Y

CONTEXT

STAND 0.615
 COMPAR 0.711
 ERROR -0.171

Completely Standardized Total Effects of KSI on Y

CONTEXT

STAND 0.615
 COMPAR 0.711
 ERROR -0.171

PEER

Number of Iterations = 69

LISREL Estimates (Unweighted Least Squares)

LAMBDA-Y
RATING

STAND 0.904
 (0.085)
 10.607

COMPAR 1.095
 (0.092)
 11.925

ERROR -0.444
 (0.076)
 -5.799

LAMBDA-X
CONTEXT

GOAL 0.234
 (0.056)
 4.153

ABILI 0.327
 (0.056)
 5.847

CONSCI 0.392
 (0.062)
 6.360



MOTIV 0.726
(0.073)
9.974

ACCOU 0.655
(0.060)
11.004

GAMMA

CONTEXT

RATING 0.585
(0.062)
9.409

Covariance Matrix of ETA and KSI

	RATING	CONTEXT
RATING	0.583	
CONTEXT	0.585	1.000

PHI EQUALS PHI IN THE FOLLOWING GROUP

PSI

RATING

0.241
(0.089)
2.695

Squared Multiple Correlations for Structural Equations

RATING

0.586

THETA-EPS

	STAND	COMPAR	ERROR
STAND	0.524 (0.123) 4.255		



COMPAR -- 0.301
 (0.149)
 2.025

ERROR -- 0.120 0.885
 (0.074)
 11.907

Squared Multiple Correlations for Y - Variables

STAND	COMPAR	ERROR
0.476	0.699	0.115

THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU
0.945	0.893	0.846	0.473	0.570
(0.064)	(0.072)	(0.097)	(0.135)	(0.082)
14.875	12.399	8.726	3.512	6.925

Squared Multiple Correlations for X - Variables

GOAL	ABILI	CONSCI	MOTIV	ACCOU
0.055	0.107	0.154	0.527	0.430

Group Goodness of Fit Statistics

Root Mean Square Residual (RMR) = 0.0588

Standardized RMR = 0.0588

Goodness of Fit Index (GFI) = 0.979

Summary Statistics for Fitted Residuals

Smallest Fitted Residual = -0.142

Median Fitted Residual = 0.000

Largest Fitted Residual = 0.105

Stemleaf Plot

-14|2
 -12|
 -10|9
 -8|48

- 6|9
 - 4|943
 - 2|6
 - 0|70922000000000
 0|7
 2|472
 4|735
 6|6
 8|623
 10|35

Standardized Residuals

	STAND	COMPAR	ERROR	GOAL	ABILI	CONSCI
STAND	0.000					
COMPAR	-0.866	0.000				
ERROR	-0.005	-1.538	0.000			
GOAL	-0.021	-0.313	0.954	0.000		
ABILI	-0.023	-0.648	-1.595	0.533	0.000	
CONSCI	-0.825	-0.215	0.301	-1.072	0.608	0.000
MOTIV	0.372	-0.122	1.233	1.160	0.411	0.878
ACCOU	1.387	0.878	1.085	-0.116	-1.069	0.216

Standardized Residuals

	MOTIV	ACCOU
MOTIV	0.000	
ACCOU	-0.819	0.000

Summary Statistics for Standardized Residuals

Smallest Standardized Residual = -1.595
 Median Standardized Residual = 0.000
 Largest Standardized Residual = 1.387

Stemleaf Plot

- 1|65
 - 1|11
 - 0|9886
 - 0|321100000000000
 0|2344
 0|5699
 1|01224

PEER

Qplot of Standardized Residuals



PEER

Modification Indices and Expected Change

No Non-Zero Modification Indices for LAMBDA-Y

No Non-Zero Modification Indices for LAMBDA-X

No Non-Zero Modification Indices for GAMMA

No Non-Zero Modification Indices for PHI

No Non-Zero Modification Indices for PSI

Modification Indices for THETA-EPS

	STAND	COMPAR	ERROR
	-----	-----	-----
STAND	--		
COMPAR	0.750	--	
ERROR	0.000	2.366	--

Expected Change for THETA-EPS

	STAND	COMPAR	ERROR
	-----	-----	-----
STAND	--		
COMPAR	-0.117	--	
ERROR	-0.001	-0.183	--

Completely Standardized Expected Change for THETA-EPS

	STAND	COMPAR	ERROR
	-----	-----	-----
STAND	--		
COMPAR	-0.117	--	
ERROR	-0.001	-0.183	--

Modification Indices for THETA-DELTA-EPS

	STAND	COMPAR	ERROR
	-----	-----	-----
GOAL	0.000	0.098	0.910
ABILI	0.001	0.420	2.545
CONSCI	0.681	0.046	0.091
MOTIV	0.139	0.015	1.521
ACCOU	1.925	0.771	1.177

Expected Change for THETA-DELTA-EPS

STAND	COMPAR	ERROR	
GOAL	-0.002	-0.032	0.089
ABILI	-0.002	-0.067	-0.151
CONSCI	-0.083	-0.023	0.029
MOTIV	0.048	-0.014	0.124
ACCOU	0.154	0.117	0.108

Completely Standardized Expected Change for THETA-DELTA-EPS

STAND	COMPAR	ERROR	
GOAL	-0.002	-0.032	0.089
ABILI	-0.002	-0.067	-0.151
CONSCI	-0.083	-0.023	0.029
MOTIV	0.048	-0.014	0.124
ACCOU	0.154	0.117	0.108

Modification Indices for THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU	
GOAL	--				
ABILI	0.284	--			
CONSCI	1.150	0.370	--		
MOTIV	1.346	0.169	0.770	--	
ACCOU	0.013	1.143	0.047	0.671	--

Expected Change for THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU	
GOAL	--				
ABILI	0.050	--			
CONSCI	-0.103	0.059	--		
MOTIV	0.122	0.044	0.099	--	
ACCOU	-0.012	-0.110	0.023	-0.096	--

Completely Standardized Expected Change for THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU	
GOAL	--				
ABILI	0.050	--			
CONSCI	-0.103	0.059	--		
MOTIV	0.122	0.044	0.099	--	
ACCOU	-0.012	-0.110	0.023	-0.096	--

PEER

Factor Scores Regressions

ETA

	STAND	COMPAR	ERROR	GOAL	ABILI	CONSCI
RATING	0.185	0.434	-0.112	0.014	0.020	0.026

ETA

	MOTIV	ACCOU
RATING	0.086	0.064

KSI

	STAND	COMPAR	ERROR	GOAL	ABILI	CONSCI
CONTEXT	0.096	0.226	-0.059	0.061	0.090	0.113

KSI

	MOTIV	ACCOU
CONTEXT	0.375	0.281

PEER

Within Group Standardized Solution

LAMBDA-Y

	RATING
STAND	0.690
COMPAR	0.836
ERROR	-0.339

LAMBDA-X

	CONTEXT
GOAL	0.234
ABILI	0.327
CONSCI	0.392
MOTIV	0.726
ACCOU	0.655

GAMMA

CONTEXT

RATING 0.766

Correlation Matrix of ETA and KSI

RATING CONTEXT

RATING	1.000	
CONTEXT	0.766	1.000

PSI

RATING

0.414

Regression Matrix ETA on KSI (Standardized)

CONTEXT

RATING 0.766

PEER

Within Group Completely Standardized Solution

LAMBDA-Y

RATING

STAND	0.690
COMPAR	0.836
ERROR	-0.339

LAMBDA-X

CONTEXT

GOAL	0.234
ABILI	0.327
CONSCI	0.392
MOTIV	0.726
ACCOU	0.655

GAMMA

CONTEXT

RATING 0.766

Correlation Matrix of ETA and KSI

RATING CONTEXT

RATING	1.000	
CONTEXT	0.766	1.000

PSI

RATING

0.414

THETA-EPS

STAND COMPAR ERROR

STAND	0.524		
COMPAR	--	0.301	
ERROR	--	0.120	0.885

THETA-DELTA

GOAL ABILI CONSCI MOTIV ACCOU

0.945	0.893	0.846	0.473	0.570
-------	-------	-------	-------	-------

Regression Matrix ETA on KSI (Standardized)

CONTEXT

RATING 0.766

PEER

Total and Indirect Effects

Total Effects of KSI on Y

CONTEXT

STAND	0.528
	(0.076)
	6.949

COMPAR 0.640
 (0.078)
 8.186

ERROR -0.259
 (0.046)
 -5.645

PEER

Standardized Total and Indirect Effects

Standardized Total Effects of KSI on Y

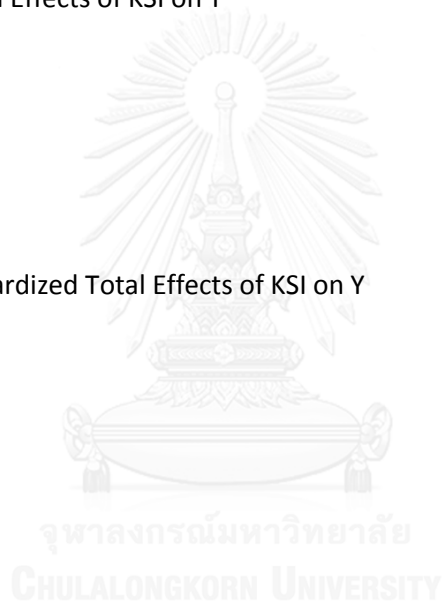
CONTEXT

STAND 0.528
 COMPAR 0.640
 ERROR -0.259

Completely Standardized Total Effects of KSI on Y

CONTEXT

STAND 0.528
 COMPAR 0.640
 ERROR -0.259



SELF

Number of Iterations = 69

LISREL Estimates (Unweighted Least Squares)

LAMBDA-Y

RATING

STAND 0.576
 (0.066)
 8.717

COMPAR 0.616
 (0.065)
 9.480

ERROR -0.483
 (0.062)
 -7.781

LAMBDA-X

CONTEXT

GOAL 0.331
 (0.063)
 5.219

ABILI 0.212
 (0.062)
 3.401

CONSCI 0.465
 (0.061)
 7.597

MOTIV 0.656
 (0.073)
 9.006

ACCOU 0.524
 (0.062)
 8.508

GAMMA

CONTEXT

RATING 0.997
 (0.067)
 14.934

Covariance Matrix of ETA and KSI

RATING CONTEXT

RATING	1.058	
CONTEXT	0.997	1.000

PHI

CONTEXT

1.000



PSI

RATING

0.064
(0.292)
0.220

Squared Multiple Correlations for Structural Equations

RATING

0.939

THETA-EPS

	STAND	COMPAR	ERROR
	-----	-----	-----
STAND	0.649 (0.123) 5.298		
COMPAR	--	0.599 (0.076) 7.834	
ERROR	--	0.120 (0.087) 8.630	0.753



Squared Multiple Correlations for Y - Variables

STAND COMPAR ERROR

0.351 0.401 0.247

THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU
-----	-----	-----	-----	-----

0.890	0.955	0.784	0.570	0.726
(0.078)	(0.056)	(0.078)	(0.124)	(0.080)
11.437	17.183	9.991	4.603	9.122

Squared Multiple Correlations for X - Variables

GOAL	ABILI	CONSCI	MOTIV	ACCOU
-----	-----	-----	-----	-----

0.110	0.045	0.216	0.430	0.274
-------	-------	-------	-------	-------

Global Goodness of Fit Statistics

Degrees of Freedom = 54

Normal Theory Weighted Least Squares Chi-Square = 34.804 (P = 0.980)

Estimated Non-centrality Parameter (NCP) = 0.0

90 Percent Confidence Interval for NCP = (0.0 ; 0.0)

Minimum Fit Function Value = 0.150

Population Discrepancy Function Value (F0) = 0.0

90 Percent Confidence Interval for F0 = (0.0 ; 0.0)

Root Mean Square Error of Approximation (RMSEA) = 0.0

90 Percent Confidence Interval for RMSEA = (0.0 ; 0.0)

P-Value for Test of Close Fit (RMSEA < 0.05) = 0.998

Expected Cross-Validation Index (ECVI) = 0.454

90 Percent Confidence Interval for ECVI = (0.454 ; 0.454)

ECVI for Saturated Model = 0.202

ECVI for Independence Model = 2.933

Chi-Square for Independence Model with 84 Degrees of Freedom = 1030.907

Independence AIC = 1078.907

Model AIC = 142.804

Saturated AIC = 216.000

Independence CAIC = 1196.174

Model CAIC = 406.654

Saturated CAIC = 743.699

Normed Fit Index (NFI) = 0.948

Non-Normed Fit Index (NNFI) = 1.001

Parsimony Normed Fit Index (PNFI) = 0.609

Comparative Fit Index (CFI) = 1.000

Incremental Fit Index (IFI) = 1.000

Relative Fit Index (RFI) = 0.919

Critical N (CN) = 540.319

Group Goodness of Fit Statistics

Root Mean Square Residual (RMR) = 0.0618

Standardized RMR = 0.0618

Goodness of Fit Index (GFI) = 0.976

Summary Statistics for Fitted Residuals

Smallest Fitted Residual = -0.117

Median Fitted Residual = 0.000

Largest Fitted Residual = 0.164

Stemleaf Plot

```

- 1|2
- 0|99998775
- 0|421000000000
  0|12222233
  0|5566
  1|03
  1|6

```

Standardized Residuals

	STAND	COMPAR	ERROR	GOAL	ABILI	CONSCI
STAND	0.000					
COMPAR	-0.646	0.000				
ERROR	-1.274	0.469	0.000			
GOAL	-0.235	-1.086	0.067	0.000		
ABILI	-0.997	1.626	-0.007	1.862	0.000	
CONSCI	0.637	-0.829	-1.133	0.225	-1.005	0.000
MOTIV	-0.736	0.906	1.207	0.588	0.258	0.295
ACCOU	0.430	0.848	0.267	-0.127	-1.376	-0.823

Standardized Residuals

	MOTIV	ACCOU
MOTIV	0.000	
ACCOU	0.229	0.000

Summary Statistics for Standardized Residuals

Smallest Standardized Residual = -1.376
Median Standardized Residual = 0.000
Largest Standardized Residual = 1.862

Stemleaf Plot

```

- 1|431100
- 0|8876
- 0|210000000000
  0|1223334
  0|56689
  1|2
  1|69

```

SELF

Qplot of Standardized Residuals



SELF

Modification Indices and Expected Change

No Non-Zero Modification Indices for LAMBDA-Y

No Non-Zero Modification Indices for LAMBDA-X

No Non-Zero Modification Indices for GAMMA

No Non-Zero Modification Indices for PHI

No Non-Zero Modification Indices for PSI

Modification Indices for THETA-EPS

	STAND	COMPAR	ERROR
	-----	-----	-----
STAND	--		
COMPAR	0.417	--	
ERROR	1.624	0.220	--

Expected Change for THETA-EPS

	STAND	COMPAR	ERROR
	-----	-----	-----
STAND	--		
COMPAR	-0.081	--	
ERROR	-0.148	0.053	--

Completely Standardized Expected Change for THETA-EPS

	STAND	COMPAR	ERROR
	-----	-----	-----
STAND	--		
COMPAR	-0.081	--	
ERROR	-0.148	0.053	--

Modification Indices for THETA-DELTA-EPS

	STAND	COMPAR	ERROR
	-----	-----	-----
GOAL	0.055	1.179	0.004
ABILI	0.994	2.644	0.000
CONSCI	0.406	0.687	1.283
MOTIV	0.542	0.821	1.458
ACCOU	0.185	0.718	0.071

Expected Change for THETA-DELTA-EPS

STAND	COMPAR	ERROR	
GOAL	-0.024	-0.112	0.007
ABILI	-0.101	0.168	-0.001
CONSCI	0.068	-0.088	-0.118
MOTIV	-0.083	0.108	0.127
ACCOU	0.047	0.094	0.028

Completely Standardized Expected Change for THETA-DELTA-EPS

STAND	COMPAR	ERROR	
GOAL	-0.024	-0.112	0.007
ABILI	-0.101	0.168	-0.001
CONSCI	0.068	-0.088	-0.118
MOTIV	-0.083	0.108	0.127
ACCOU	0.047	0.094	0.028

Modification Indices for THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU	
GOAL	--				
ABILI	3.467	--			
CONSCI	0.050	1.011	--		
MOTIV	0.345	0.066	0.087	--	
ACCOU	0.016	1.893	0.677	0.053	--

Expected Change for THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU	
GOAL	--				
ABILI	0.178	--			
CONSCI	0.022	-0.098	--		
MOTIV	0.064	0.027	0.033	--	
ACCOU	-0.013	-0.136	-0.086	0.027	--

Completely Standardized Expected Change for THETA-DELTA

GOAL	ABILI	CONSCI	MOTIV	ACCOU	
GOAL	--				
ABILI	0.178	--			
CONSCI	0.022	-0.098	--		
MOTIV	0.064	0.027	0.033	--	
ACCOU	-0.013	-0.136	-0.086	0.027	--

Max. Mod. Index is 3.47 for Element (2, 1) of THETA-DELTA in Group 3

SELF

Factor Scores Regressions

ETA

	STAND	COMPAR	ERROR	GOAL	ABILI	CONSCI
RATING	0.229	0.308	-0.215	0.083	0.049	0.132

ETA

	MOTIV	ACCOU
RATING	0.256	0.160

KSI

	STAND	COMPAR	ERROR	GOAL	ABILI	CONSCI
CONTEXT	0.197	0.265	-0.185	0.092	0.055	0.146

KSI

	MOTIV	ACCOU
CONTEXT	0.284	0.178

SELF

Within Group Completely Standardized Solution

LAMBDA-Y

RATING

STAND	0.592
COMPAR	0.633
ERROR	-0.497

LAMBDA-X

CONTEXT

GOAL	0.331
ABILI	0.212

CONSCI 0.465
 MOTIV 0.656
 ACCOU 0.524

GAMMA

CONTEXT

RATING 0.969

Correlation Matrix of ETA and KSI

RATING CONTEXT

RATING 1.000
 CONTEXT 0.969 1.000

PSI

RATING

0.061

THETA-EPS

STAND COMPAR ERROR

STAND 0.649
 COMPAR -- 0.599
 ERROR -- 0.120 0.753

THETA-DELTA

GOAL ABILI CONSCI MOTIV ACCOU

0.890 0.955 0.784 0.570 0.726

Regression Matrix ETA on KSI (Standardized)

CONTEXT

RATING 0.969

SELF

Total and Indirect Effects

Total Effects of KSI on Y

CONTEXT

STAND 0.574
 (0.064)
 8.983

COMPAR 0.614
 (0.054)
 11.393

ERROR -0.482
 (0.064)
 -7.554

SELF

Standardized Total and Indirect Effects

Standardized Total Effects of KSI on Y

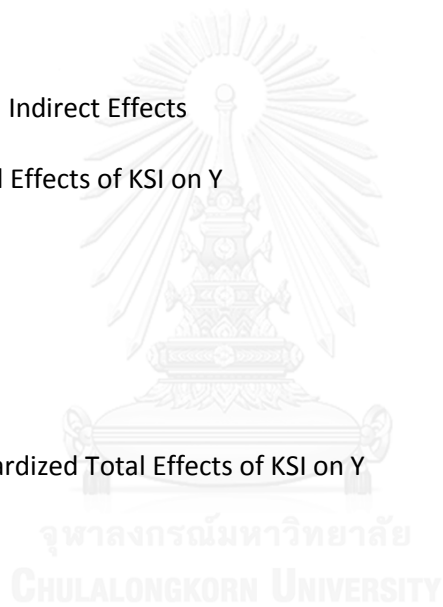
CONTEXT

STAND 0.574
 COMPAR 0.614
 ERROR -0.482

Completely Standardized Total Effects of KSI on Y

CONTEXT

STAND 0.574
 COMPAR 0.614
 ERROR -0.482



Time used: 0.109 Seconds

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวศุภกรรัตน์ อิงชาติเจริญ เกิดเมื่อวันที่ 22 กรกฎาคม พ.ศ.2524 สำเร็จการศึกษาพยาบาลศาสตรบัณฑิต มหาวิทยาลัยธรรมศาสตร์ ในปีการศึกษา 2544 ประกาศนียบัตรวิชาชีพพยาบาล คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล ในปี พ.ศ.2546 วิทยาศาสตร์มหาบัณฑิต สาขาการวิจัยพฤติกรรมศาสตร์ประยุกต์ มหาวิทยาลัยศรีนครินทรวิโรฒ ในปีการศึกษา 2551 ประกาศนียบัตรหลักสูตรฝึกอบรมระยะสั้น สาขาศาสตร์และศิลปการสอนทางการพยาบาล คณะพยาบาลศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ในปี พ.ศ. 2552 และศึกษาต่อในระดับครุศาสตรดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554 ปัจจุบันเป็นอาจารย์ประจำคณะพยาบาลศาสตร์ มหาวิทยาลัยมหิดล

