

การพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6
โดยใช้แฟ้มสะสมงาน: การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร

นางสาวดารุวรรณ ศรีแก้ว



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

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.

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2557

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

DEVELOPMENT OF AN ENGLISH SPEAKING SKILL ASSESSMENT MODEL FOR
GRADE 6 STUDENTS BY USING PORTFOLIO: AN APPLICATION OF
MULTIVARIATE GENERALIZABILITY THEORY

Miss Daruwan Srikaew



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

การพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษ
ของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน:
การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบ
พหุตัวแปร

โดย

นางสาวดารุวรรณ ศรีแก้ว

สาขาวิชา

การวัดและประเมินผลการศึกษา

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

ผู้ช่วยศาสตราจารย์ ดร. กมลวรรณ ตังธนากานนท์

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี

คณะกรรมการ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยเป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญาศึกษาศาสตรบัณฑิต

..... คณบดีคณะครุศาสตร์

(รองศาสตราจารย์ ดร. บัญชา ชลาภิรมย์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร. กมลวรรณ ตังธนากานนท์)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

..... กรรมการ

(รองศาสตราจารย์ ดร. โชติกา ภาชีผล)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ณิชฐภรณ์ หลาวทอง)

..... กรรมการภายนอกมหาวิทยาลัย

(ดร. ชนะศึก นิชานนท์)

दारूरرن ศรีแก้ว : การพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน: การประยุกต์ใช้ทฤษฎีการสรุปร่างอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร (DEVELOPMENT OF AN ENGLISH SPEAKING SKILL ASSESSMENT MODEL FOR GRADE 6 STUDENTS BY USING PORTFOLIO: AN APPLICATION OF MULTIVARIATE GENERALIZABILITY THEORY) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ. ดร. กมลวรรณ ตั้งธนานนท์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ศ. ดร. ศิริชัย กาญจนวาสี, 319 หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน และเพื่อเปรียบเทียบรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานด้วยทฤษฎีการสรุปร่างอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร กลุ่มตัวอย่างที่ใช้ในการเก็บรวบรวมข้อมูล คือ นักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 34 คน ในกรุงเทพมหานคร ครูผู้สอนวิชาภาษาอังกฤษ จำนวน 1 คน และผู้ประเมิน จำนวน 4 คน วิเคราะห์ข้อมูลเพื่อหาความสัมพันธ์การสรุปร่างอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร โดยใช้โปรแกรม mGENOVA version 2.1 และทำวิเคราะห์ความก้าวหน้าด้านทักษะการพูดภาษาอังกฤษของนักเรียนโดยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำชนิดทางเดียว

ผลการวิจัย พบว่า

ขั้นตอนของแฟ้มสะสมงานสำหรับการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วย 7 ขั้นตอน ดังนี้ 1) ขั้นตอนการวางแผน 2) ขั้นตอนเตรียมความพร้อมนักเรียน 3) ขั้นตอนเก็บรวบรวมหลักฐาน 4) ขั้นตอนติดตามความก้าวหน้า 5) ขั้นตอนปรับปรุงผลงาน 6) ขั้นตอนสะท้อนคิด และ 7) ขั้นตอนจัดแสดงผลงาน ส่วนประกอบของแฟ้มสะสมงาน ประกอบด้วย 3 ส่วน ได้แก่ 1) ส่วนเนื้อหา และ 3) ส่วนภาคผนวก ภาระงานการพูดในการประเมินจำนวนทั้งสิ้น 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ (interview) 2) การนำเสนอปากเปล่า (oral presentation) 3) การเล่านิทาน (storytelling) และ 4) การบรรยายภาพ (picture description) เกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) มีขอบเขตเนื้อหา 6 ด้าน ได้แก่ คำศัพท์ (vocabulary) การสร้างประโยค (syntax) การเชื่อมโยงเนื้อหา (cohesion) การออกเสียง (pronunciation) บรรลุวัตถุประสงค์ (ideational function) และความคล่องแคล่ว (fluency) ให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี ตามลำดับ ผลของความก้าวหน้าของความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนหลังจากใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษจากการวัดซ้ำ จำนวน 3 ครั้ง พบว่านักเรียนมีพัฒนาการในการพูดภาษาอังกฤษสูงขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้ง 4 ภาระงาน

การศึกษาสัมประสิทธิ์การสรุปร่างอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ด้วยทฤษฎีการสรุปร่างอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร พบว่า สำหรับเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษนั้น ความแปรปรวนของนักเรียนมีขนาดใหญ่ที่สุด เกณฑ์การให้คะแนนที่มีความแปรปรวนมากที่สุดคือ การออกเสียง รองลงมาคือ ความคล่องแคล่ว การบรรลุวัตถุประสงค์ การสร้างประโยค คำศัพท์ และการเชื่อมโยงเนื้อหา ความเที่ยงของคะแนนรวม เมื่อใช้ภาระงาน จำนวน 4 ภาระงาน และผู้ประเมิน จำนวน 4 คน พบว่า ค่าสัมประสิทธิ์การสรุปร่างอ้างอิงเชิงสัมพัทธ์ เท่ากับ .846 ส่วนค่าสัมประสิทธิ์การสรุปร่างอ้างอิงเชิงสัมบูรณ์ เท่ากับ .745 รูปแบบการประเมินทักษะการพูดภาษาอังกฤษ พบว่า เมื่อใช้ภาระงานจำนวน 4 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ สามารถใช้ผู้ประเมินจำนวน 2 คน และมีความเที่ยงในการประเมินของคะแนนรวมในระดับที่สูง คือ ค่าสัมประสิทธิ์การสรุปร่างอ้างอิงเชิงสัมพัทธ์ เท่ากับ .939 ส่วนค่าสัมประสิทธิ์การสรุปร่างอ้างอิงเชิงสัมบูรณ์ เท่ากับ .925 สำหรับรูปแบบการตรวจให้คะแนนพบว่า การประเมินที่ผู้ประเมินตรวจให้คะแนนเฉพาะภาระงานมีความน่าเชื่อถือสูงกว่าการให้ผู้ประเมินทุกคนตรวจให้คะแนนซ้ำกันทุกภาระงาน

ภาควิชา วิชาวิจัยและจิตวิทยาการศึกษา

ลายมือชื่อ นิสิต

สาขาวิชา การวัดและประเมินผลการศึกษา

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ปีการศึกษา 2557

ลายมือชื่อ อ.ที่ปรึกษารวม

5384220227 : MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION

KEYWORDS: ENGLISH SPEAKING SKILL / ANALYTIC RATING SCALE

DARUWAN SRIKAEW: DEVELOPMENT OF AN ENGLISH SPEAKING SKILL ASSESSMENT MODEL FOR GRADE 6 STUDENTS BY USING PORTFOLIO: AN APPLICATION OF MULTIVARIATE GENERALIZABILITY THEORY. ADVISOR: ASST. PROF. KAMONWAN TANGDHANAKANOND, Ph.D., CO-ADVISOR: PROF. SIRICHAJ KANJANAWASEE, Ph.D., 319 pp.

The objectives of this research were to develop an English speaking assessment model using portfolios for grade 6 students as well as to use the multivariate generalizability theory to compare among English speaking assessment models using portfolios for grade 6 students. The research samples for data collection were 34 grade 6 students studying in Bangkok, one English teacher, and four raters. The multivariate generalizability coefficients were analyzed using mGENOVA version 2.1 whereas one-way repeated measure ANOVA was used to analyze students' progress in English speaking skills.

It was found that the working procedures on developing the portfolio on English speaking skill of grade 6 students were composed of 7 steps: 1) Planning 2) Preparing students 3) Evidence collection 4) Progress monitoring 5) Improving students' works 6) Reflection and 7) Displaying students' works. The portfolio was divided into three parts: 1) Introduction 2) Contents and 3) Appendix. The four speaking tasks assessed were 1) interview 2) oral presentation 3) storytelling and 4) picture description. The analytic rating scale covered six aspects which were vocabulary, syntax, cohesion, pronunciation, ideational function and fluency. Scoring criteria of 1-4 was applied ranging from very poor, poor, moderate to good, respectively. The results on the progress in English speaking skill of the students shown from three time repeated measurement revealed that English speaking skill of the students was significantly higher at the .01 level for all of the four tasks.

From the study on generalizability coefficients of the English speaking assessment model using portfolios for grade 6 students that applied the multivariate generalizability theory, it was found that for the English speaking assessment criteria, student variance was the greatest. The assessment criteria that showed the greatest variance was pronunciation, followed by fluency, ideational function, syntax, vocabulary and cohesion, respectively. Concerning the reliability of the composite score for four tasks with four raters, it was found that multivariate generalizability coefficient was at .846 while the multivariate phi coefficient was at .745. Regarding the four speaking tasks (interview, oral presentation, storytelling, picture description), it could involve two raters and the reliability of composite score was at .939 while the multivariate phi coefficient was at .925. As for the scores given by the raters, it was found that the reliability of the assessment with different raters for each task was higher than that with the same raters for every task

Department: Educational Research and Psychology Student's Signature

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

Evaluation Co-Advisor's Signature

Academic Year: 2014

กิตติกรรมประกาศ

ปริญญาบัตรฉบับนี้ สำเร็จเป็นอย่างดี ด้วยความกรุณาอย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ ดร.กมลวรรณ ตั้งธนกานนท์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ และ ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสิ อาจารย์ที่ปรึกษาร่วม ซึ่งกรุณาให้คำปรึกษาและแนะนำ แก้ไขข้อบกพร่องต่างๆ และให้กำลังใจแก่ผู้วิจัยด้วยดีมาโดยตลอดระยะเวลาที่วิทยานิพนธ์ รวมทั้ง รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ ประธานคณะกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.โชติกา ภาษีผล ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง และอาจารย์ ดร.ชนะศึก นิซานนท์ คณะกรรมการสอบวิทยานิพนธ์ ที่ให้คำแนะนำ และชี้แนะแนวทางในการแก้ไขข้อบกพร่องต่างๆ ของงานวิจัย ทำให้งานวิจัยสมบูรณ์มากขึ้น ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร.โชติกา ภาษีผล ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง และดร.ชนะศึก นิซานนท์ กรรมการสอบวิทยานิพนธ์ ที่ได้เสียสละเวลาให้ข้อคิด ข้อเสนอแนะ อันมีค่าและมีประโยชน์ในการแก้ไขปรับปรุงเพื่อให้วิทยานิพนธ์เล่มนี้มีความสมบูรณ์ ทั้งด้านเนื้อหาและคุณค่าทางวิชาการมากยิ่งขึ้น

ขอกราบขอบพระคุณผู้ทรงคุณวุฒิทุกท่านดังรายนามในภาคผนวก ที่กรุณาเสียสละเวลาในการให้สัมภาษณ์ ตอบแบบสอบถาม แบบประเมิน และตรวจสอบคุณภาพเครื่องมือที่ใช้ในการวิจัย ขอขอบพระคุณท่านผู้อำนวยการโรงเรียน คุณครูและนักเรียนโรงเรียนสายน้ำทิพย์ และโรงเรียนอนุบาลพิบูลเวศม์ ที่ให้ความกรุณาเสียสละเวลาในการให้ผู้วิจัยได้เก็บรวบรวมข้อมูลและทำการทดลองในการวิจัยครั้งนี้

ขอกราบขอบพระคุณ ผศ.ดร.จิรดา วุฒินิยามกร ที่คอยให้คำชี้แนะและแนะนำด้านการทดสอบทางภาษา ขอขอบพระคุณพี่ๆ นิสิตระดับปริญญาเอกสาขาภาษาอังกฤษเป็นภาษานานาชาติ (English as an International Language) สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย ที่เสียสละเวลามาเป็นผู้ประเมินการพูดภาษาอังกฤษและให้ข้อเสนอเพื่อปรับแก้เกณฑ์การประเมินให้ดียิ่งขึ้น รวมทั้ง ดร.ยุวดี อยู่สบาย อาจารย์กฤษกรรัตน์ รัตนสิงห์ Miss Lottie Baker และพี่ๆ ที่สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ที่ช่วยตรวจสอบความถูกต้องทางภาษาอังกฤษและให้ความช่วยเหลือในทุกๆ เรื่อง

ขอขอบคุณเพื่อน รุ่นพี่ รุ่นน้อง และเจ้าหน้าที่ในภาควิชาวิจัยและจิตวิทยาการศึกษาทุกคนที่ให้ความช่วยเหลือจนวิทยานิพนธ์สำเร็จลุล่วงได้ด้วยดี โดยเฉพาะอย่างยิ่ง นางชจี พงศธรพิบูลย์ นายปิยพงษ์ คล้ายคลึง และนางสาวสิริกร โดสดี ที่ช่วยเก็บรวบรวมข้อมูล และเป็นกำลังใจในการทำวิทยานิพนธ์ในทุกขั้นตอน และขอขอบคุณคนอื่นๆ อีกหลายคนที่ไม่ได้เอียนาม มา ณ โอกาสนี้

ขอกราบขอบพระคุณคุณแม่ และทุกคนในครอบครัวที่ให้กำลังใจ และโอกาสที่ดีเสมอมา ขอขอบคุณ นายชานนท์ ภูริพันธ์ภิญโญ ที่ให้กำลังใจคอยช่วยเหลือ และสนับสนุนทั้งในระหว่างเรียนและในการทำวิทยานิพนธ์เป็นอย่างดี จนวิทยานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยดี

ท้ายที่สุดนี้ ขอขอบพระคุณ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย ในการสนับสนุนทุนวิจัย “ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย” จากกองทุนรัชดาภิเษกสมโภช ประจำปี 2556

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ	ฐ
บทที่ 1 บทนำ	1
ความเป็นมาและความสำคัญของปัญหา	1
คำถามวิจัย	6
วัตถุประสงค์ของการวิจัย.....	6
ขอบเขตของการวิจัย.....	7
คำจำกัดความของการวิจัย.....	8
ประโยชน์ที่คาดว่าจะได้รับ.....	9
บทที่ 2 แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	11
ตอนที่ 1 การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษา	11
ตอนที่ 2 การประเมินโดยแฟ้มสะสมงาน	43
ตอนที่ 3 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร	77
ตอนที่ 4 กรอบแนวคิดการวิจัย	112
บทที่ 3 วิธีดำเนินการวิจัย.....	114
ระยะที่ 1 การพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน	114
ระยะที่ 2 การทดลองใช้รูปแบบการประเมินแฟ้มสะสมงานทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน	124

ระยะที่ 3 การวิเคราะห์ผลการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน.....	130
ระยะที่ 4 การพัฒนาคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของ นักเรียนชั้นประถมศึกษาปีที่ 6	133
บทที่ 4 ผลการวิเคราะห์ข้อมูล.....	135
ตอนที่ 1 ผลการพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน	136
ตอนที่ 2 ผลการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานสำหรับประเมินทักษะการพูดภาษาอังกฤษ	165
ตอนที่ 3 ผลการเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการ ประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสม งาน	181
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	211
สรุปผลวิจัย	212
อภิปรายผลวิจัย	218
ข้อเสนอแนะ	224
ข้อเสนอแนะในการนำไปใช้.....	224
ข้อเสนอแนะในการวิจัยครั้งต่อไป	225
รายการอ้างอิง	227
ภาคผนวก.....	228
ประวัติผู้เขียนวิทยานิพนธ์	319

สารบัญตาราง

ตารางที่ 2.1	ระดับของความสามารถตามกรอบ CEFR.....	19
ตารางที่ 2.2	คำอธิบายความสามารถทางภาษาจากกรอบ CEFR.....	19
ตารางที่ 2.3	ระดับความสามารถทางภาษามาตรแบบรวม (global scale) ของ 3 อันดับแรกของ CEFR	20
ตารางที่ 2.4	ตัวอย่างโครงสร้างไวยากรณ์ของแบบวัดความสามารถทางภาษาของเด็กของ Cambridge ESOL (2007).....	24
ตารางที่ 2.5	รายละเอียดกิจกรรมการประเมินการพูด จุดเด่น และจุดด้อย	26
ตารางที่ 2.6	เกณฑ์การให้คะแนนทักษะการพูดของ CEFR (2001).....	37
ตารางที่ 2.7	สรุปผลจากการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องกับการประเมินทักษะการพูด ภาษาอังกฤษ.....	42
ตารางที่ 2.8	ตัวอย่างการจับคู่กิจกรรมกับวัตถุประสงค์.....	59
ตารางที่ 2.9	ตัวอย่างเกณฑ์การประเมินเพิ่มสะสมงานเพื่อให้เกรด (grading).....	64
ตารางที่ 2.10	ตัวอย่างรูบริกแบบองค์รวมเพื่อประเมินการนำเสนอปากเปล่า.....	66
ตารางที่ 2.11	การให้คะแนนแบบรูบริกเพื่อประเมินเพิ่มสะสมงานแบบแยกส่วน	67
ตารางที่ 2.12	ตัวอย่างรูบริกแบบกำหนดน้ำหนัก.....	68
ตารางที่ 2.13	ประเภทของการให้ข้อมูลย้อนกลับ	71
ตารางที่ 2.14	ตัวอย่างการออกแบบแบบพหุตัวแปร (multivariate designs) และการออกแบบ univariate	85
ตารางที่ 2.15	สรุปผลจากการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องกับทฤษฎีการสรุปอ้างอิงความ น่าเชื่อถือแบบพหุตัวแปร	109
ตารางที่ 3.1	ข้อเสนอแนะและข้อคิดเห็นจากผู้ประเมินจากการทดลองใช้เกณฑ์การให้คะแนน การพูดภาษาอังกฤษ	118
ตารางที่ 3.2	ผลการวิเคราะห์คุณภาพเกณฑ์การให้คะแนนแบบแยกส่วน	119
ตารางที่ 3.3	ค่าสหสัมพันธ์ระหว่างคะแนนของผู้ประเมินที่ใช้เกณฑ์การให้คะแนนแบบแยกส่วน ..	126

ตารางที่ 3.4	ช่วงเวลาและการดำเนินกิจกรรม.....	128
ตารางที่ 4.1	การสังเคราะห์ส่วนประกอบของแฟ้มสะสมงานทางภาษาอังกฤษ.....	137
ตารางที่ 4.2	การสังเคราะห์ขั้นตอนของแฟ้มสะสมงานทางภาษาอังกฤษ	138
ตารางที่ 4.3	การสังเคราะห์เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ	142
ตารางที่ 4.4	ผลการตรวจสอบความสอดคล้องของนิยามภาระงานการพูดภาษาอังกฤษ และความ คิดเห็น ข้อเสนอแนะของผู้ทรงคุณวุฒิเกี่ยวกับภาระงานการพูดภาษาอังกฤษ.....	143
ตารางที่ 4.5	เกณฑ์การให้คะแนนการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6	144
ตารางที่ 4.6	ผลการตรวจสอบความสอดคล้องของนิยามของเกณฑ์การให้คะแนนการพูด ภาษาอังกฤษ และความคิดเห็น ข้อเสนอแนะของผู้ทรงคุณวุฒิเกี่ยวกับภาระงานการพูด ภาษาอังกฤษ.....	146
ตารางที่ 4.7	เครื่องมือที่ใช้ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ.....	151
ตารางที่ 4.8	ความเที่ยงแบบอิงกลุ่มในการประเมินทักษะการพูดรายองค์ประกอบ	153
ตารางที่ 4.9	ความเที่ยงแบบอิงเกณฑ์ในการประเมินทักษะการพูดรายด้าน	154
ตารางที่ 4.10	ระดับความคลาดเคลื่อนและความเที่ยงของคะแนนรวมในการประเมินการพูดฯ รายองค์ประกอบเพื่อการตัดสินใจจำนวนผู้ประเมินและภาระงาน	155
ตารางที่ 4.11	ความเที่ยงในการประเมินเมื่อประเมินรายภาระงาน	156
ตารางที่ 4.12	ระดับความคลาดเคลื่อนและความเที่ยงของคะแนนรวมในการประเมินการพูดฯ เพื่อการตัดสินใจจำนวนผู้ประเมินและภาระงาน	156
ตารางที่ 4.13	ระดับความคลาดเคลื่อนและความเที่ยงของคะแนนรวมในการประเมินการพูดฯ เพื่อการตัดสินใจเลือกรูปแบบการผู้ประเมิน.....	157
ตารางที่ 4.14	ระดับความเที่ยงของการประเมินการพูดฯ เพื่อการตัดสินใจรูปแบบการประเมิน ..	157
ตารางที่ 4.15	ผลการตรวจสอบความสอดคล้องของคู่มือแฟ้มสะสมงานการประเมินทักษะการ พูดภาษาอังกฤษ	159
ตารางที่ 4.16	การวิเคราะห์สถิติพื้นฐานและการแปลผลของคะแนน ความสามารถด้านทักษะ การพูด 4 ภาระงาน.....	169

ตารางที่ 4.17 ผลการทดสอบ sphericity ของคะแนนความสามารถในทักษะการพูด ภาษาอังกฤษจำแนกตามภาระงาน.....	170
ตารางที่ 4.18 การเปรียบเทียบคะแนนเฉลี่ยจากการสัมภาษณ์โดยรวม	171
ตารางที่ 4.19 การเปรียบเทียบคะแนนเฉลี่ยการสัมภาษณ์รายคู่ ด้วยวิธี Bonferroni	171
ตารางที่ 4.20 การเปรียบเทียบคะแนนเฉลี่ยจากการนำเสนอปากเปล่าโดยรวม	172
ตารางที่ 4.21 การเปรียบเทียบคะแนนเฉลี่ยการนำเสนอปากเปลารายคู่ วิธี Bonferroni	172
ตารางที่ 4.22 การเปรียบเทียบคะแนนเฉลี่ยจากการเล่านิทานโดยรวม	173
ตารางที่ 4.23 การเปรียบเทียบคะแนนเฉลี่ยการเล่านิทานรายคู่ ด้วยวิธี Bonferroni	174
ตารางที่ 4.24 การเปรียบเทียบคะแนนเฉลี่ยการบรรยายภาพในภาพรวม	175
ตารางที่ 4.25 การเปรียบเทียบคะแนนเฉลี่ยการบรรยายภาพรายคู่ ด้วยวิธี Bonferroni	175
ตารางที่ 4.26 คะแนนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของภาระงาน และผู้ประเมิน จำแนก ตามเกณฑ์การให้คะแนนแบบแยกส่วน	182
ตารางที่ 4.27 องค์ประกอบความแปรปรวนและความแปรปรวนร่วมสำหรับ $p \times t \times r \bullet$ Design	184
ตารางที่ 4.28 องค์ประกอบความแปรปรวนและความแปรปรวนร่วมของการประเมินสำหรับ $p \times t \times r \bullet$ Design	185
ตารางที่ 4.29 องค์ประกอบความแปรปรวนและความแปรปรวนร่วมสำหรับ $p \times t^{\circ} \times r \bullet$ Design.....	187
ตารางที่ 4.30 ผลการวิเคราะห์คะแนนรวมของการประเมินสำหรับ $p \times t^{\circ} \times r \bullet$ Design	188
ตารางที่ 4.31 องค์ประกอบความแปรปรวนและความแปรปรวนร่วมของการประเมินระหว่าง เรียนและการประเมินสรุปรวม สำหรับ $p \times t^{\circ} \times r \bullet$ Design.....	190
ตารางที่ 4.32 ผลการวิเคราะห์คะแนนรวมของการประเมินของการประเมินระหว่างเรียนและ การประเมินสรุปรวม สำหรับ $p \times t^{\circ} \times r \bullet$ Design.....	191
ตารางที่ 4.33 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ของเกณฑ์การให้คะแนนการพูดแบบแยก ส่วนเมื่อเพิ่มจำนวนผู้ประเมิน	193

ตารางที่ 4.34 สัมประสิทธิ์การสรุปร่างเชิงสมบูรณ์ของเกณฑ์การให้คะแนนการพูดแบบแยก
ส่วนเมื่อเพิ่มจำนวนผู้ประเมิน 194

ตารางที่ 4.35 คะแนนรวมของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนผู้
ประเมิน..... 195

ตารางที่ 4.36 สัมประสิทธิ์การสรุปร่างเชิงสัมพัทธ์ของเกณฑ์การให้คะแนนการพูดแบบแยก
ส่วนเมื่อเพิ่มจำนวนภาระงาน 197

ตารางที่ 4.37 ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมบูรณ์ของเกณฑ์การให้คะแนนการพูดแบบแยก
ส่วนเมื่อเพิ่มจำนวนภาระงาน..... 198

ตารางที่ 4.38 คะแนนรวมของเกณฑ์การให้คะแนนการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนภาระงาน ... 199

ตารางที่ 4.39 ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพัทธ์เมื่อจำนวนผู้ประเมินต่างกัน 202

ตารางที่ 4.40 ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมบูรณ์เมื่อจำนวนผู้ประเมินต่างกัน..... 202

ตารางที่ 4.41 คะแนนรวมของการประเมินเมื่อจำนวนผู้ประเมินต่างกัน..... 202

ตารางที่ 4.42 คะแนนรวมสำหรับรูปแบบการให้คะแนนที่ต่างกันของผู้ประเมิน..... 203

ตารางที่ 4.43 ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพัทธ์รูปแบบการประเมินทักษะการพูด
ภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น..... 207

ตารางที่ 4.44 ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมบูรณ์ของรูปแบบการประเมินทักษะการพูด
ภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น..... 208

ตารางที่ 4.45 คะแนนรวมของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษเมื่อจำนวนผู้ประเมิน
และภาระงานเพิ่มขึ้น 209

สารบัญภาพ

ภาพที่ 2.1	มาตรฐานในการเรียนการสอนภาษาต่างประเทศ 5 มาตรฐาน (The Five Cs) ของ ACTFL, 1999.....	14
ภาพที่ 2.2	องค์ประกอบความสามารถทางภาษาตามโมเดลของ Bachman and Palmer (2010).....	15
ภาพที่ 2.3	กระบวนการพูดของ Bygate (1987).....	22
ภาพที่ 2.4	กระบวนการประเมินทักษะการพูด.....	23
ภาพที่ 2.5	ลักษณะทั่วไปของแฟ้มสะสมงาน.....	47
ภาพที่ 2.6	ขั้นตอนการพัฒนาแฟ้มสะสมงาน.....	56
ภาพที่ 2.7	กระบวนการพัฒนาแฟ้มสะสมงานทางภาษาของ Lam and Lee (2010).....	58
ภาพที่ 2.8	กระบวนการพัฒนาและดำเนินการโครงการแฟ้มสะสมงานของ Lo (2010).....	60
ภาพที่ 2.9	กรอบแนวคิดของ G-Theory.....	79
ภาพที่ 2.10	การออกแบบ แบบ $p \bullet \times i^\circ$ Design.....	87
ภาพที่ 2.11	การออกแบบ แบบ $p \bullet \times i \bullet$ Design.....	88
ภาพที่ 2.12	การออกแบบ $p \bullet \times i \bullet \times h \bullet$ Design.....	90
ภาพที่ 2.13	การออกแบบ $p \bullet \times i \bullet \times h^\circ$ Design.....	91
ภาพที่ 2.14	การออกแบบ $p \bullet \times (i^\circ : h \bullet)$ Design.....	93
ภาพที่ 2.15	การออกแบบ $p \bullet \times (i^\circ : h^\circ)$ Design.....	94
ภาพที่ 2.16	กรอบแนวคิดการวิจัย.....	113
ภาพที่ 4.1	คะแนนเฉลี่ยการสัมภาษณ์ในการประเมิน 3 ครั้ง.....	171
ภาพที่ 4.2	คะแนนเฉลี่ยการนำเสนอปากเปล่าในการประเมิน 3 ครั้ง.....	173
ภาพที่ 4.3	คะแนนเฉลี่ยการเล่านิทานในการประเมิน 3 ครั้ง.....	174
ภาพที่ 4.4	คะแนนเฉลี่ยการบรรยายภาพในการประเมิน 3 ครั้ง.....	176

ภาพที่ 4.5 ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพัทธ์ของเกณฑ์การให้คะแนนการพูดแบบแยก ส่วนเมื่อเพิ่มจำนวนผู้ประเมิน.....	193
ภาพที่ 4.6 ค่าสัมประสิทธิ์การสรุปร่างเชิงสมบูรณ์ของเกณฑ์การประเมินการพูดแบบแยก ส่วนเมื่อเพิ่มจำนวนผู้ประเมิน.....	194
ภาพที่ 4.7 ความคลาดเคลื่อนของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนผู้ ประเมิน.....	195
ภาพที่ 4.8 คะแนนรวมของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนผู้ ประเมิน.....	196
ภาพที่ 4.9 ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพัทธ์ ของเกณฑ์การประเมินการพูดแบบแยกส่วน เมื่อเพิ่มจำนวนภาระงาน.....	198
ภาพที่ 4.10 ค่าสัมประสิทธิ์การสรุปร่างเชิงสมบูรณ์ของเกณฑ์การให้คะแนนการพูดแบบแยก ส่วนเมื่อเพิ่มจำนวนภาระงาน.....	199
ภาพที่ 4.11 ความคลาดเคลื่อนของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่ม จำนวนภาระงาน.....	200
ภาพที่ 4.12 คะแนนรวมของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนภาระ งาน.....	200
ภาพที่ 4.13 ความคลาดเคลื่อนเชิงสัมพัทธ์ของการประเมินที่ใช้ผู้ประเมินต่างกัน.....	204
ภาพที่ 4.14 ความคลาดเคลื่อนเชิงสมบูรณ์ของการประเมินที่ใช้ผู้ประเมินต่างกัน.....	204
ภาพที่ 4.15 ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพัทธ์ของการประเมินที่ใช้ผู้ประเมินต่างกัน.....	205
ภาพที่ 4.16 ค่าสัมประสิทธิ์การสรุปร่างเชิงสมบูรณ์ของการประเมินที่ใช้ผู้ประเมินต่างกัน.....	205
ภาพที่ 4.17 ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพัทธ์ของรูปแบบการประเมินทักษะการพูด ภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น.....	207
ภาพที่ 4.18 ค่าสัมประสิทธิ์การสรุปร่างเชิงสมบูรณ์ของรูปแบบการประเมินทักษะการพูด ภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น.....	208
ภาพที่ 4.19 ความคลาดเคลื่อนของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษเมื่อจำนวนผู้ ประเมินและภาระงานเพิ่มขึ้น.....	210

ภาพที่ 4.20 คะแนนรวมของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษเมื่อจำนวนผู้ประเมิน
และภาระงานเพิ่มขึ้น..... 210



บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ภาษาอังกฤษเป็นภาษาที่สำคัญเพราะเป็นเครื่องมือที่ใช้ในการสื่อสารกับสังคมโลก และเป็นเครื่องมือในการค้นหาความรู้ต่าง ๆ นอกจากนี้ประเทศไทยกำลังเตรียมความพร้อมสู่ประชาคมอาเซียนในปี พ.ศ.2558 ดังนั้นความสามารถทางภาษาอังกฤษจึงมีความสำคัญยิ่งขึ้นไปอีก ทักษะการพูดเป็นทักษะหนึ่ง ในทักษะทางภาษา ซึ่งแต่ระดับประถมศึกษาควรได้รับการพัฒนาควบคู่ไปกับทักษะอื่นๆ และใช้ในการเรียนและการสื่อสารระดับสูงต่อไปได้ และการที่ครูจะจัดการเรียนการสอนให้มีคุณภาพ การประเมินผลที่สามารถประเมินความสามารถที่แท้จริงของนักเรียนได้ และมีความน่าเชื่อถือย่อมมีความจำเป็นเพื่อเป็นภาพสะท้อนในการจัดการเรียนรู้ และความสามารถทางภาษาเพื่อการสื่อสาร การที่จะประเมินความสามารถของผู้เรียนได้นั้น ครูหรือผู้ประเมินควรประเมินความสามารถของนักเรียนโดยการให้นักเรียนได้แสดงความสามารถออกมาด้วยการปฏิบัติ (performance assessment) เครื่องมือที่ใช้จึงควรเป็นกิจกรรมที่ให้นักเรียนได้พูด หรือ สนทนา แสดง และกิจกรรมที่ประเมินควรเป็นสิ่งที่สมจริง (authentic assessment) และเป็นสิ่งที่มีอยู่ในชีวิตจริง (real life) (Bachman & Palmer, 2010)

แฟ้มสะสมงานเป็นแฟ้มที่ใช้ในการรวบรวมผลงานต่างๆ ที่นักเรียนได้จัดทำขึ้นระหว่างเรียน ซึ่งแฟ้มสะสมงานได้เข้ามามีบทบาทในการเรียนการสอนตั้งแต่ปี ค.ศ.1988 โดยเป็นเครื่องมือชนิดหนึ่งในการประเมินความสามารถที่แท้จริง (authentic assessment) ของนักเรียน (อุทุมพร (ทองอุไทย) จามรมาน, 2540; Payne, 2003) มีการรวบรวมหลักฐานอย่างเป็นระบบ ใช้ในการกำกับติดตามความก้าวหน้า ความรู้ ทักษะและเจตคติของนักเรียน บันทึกความก้าวหน้าของนักเรียนเป็นรายบุคคลเกี่ยวกับสิ่งที่นักเรียนได้เรียนรู้ ช่วยการเรียนในอนาคตแก่นักเรียน ช่วยให้ครูได้สะท้อนคิดเกี่ยวกับผลงานของนักเรียนด้วยและนักเรียนมีส่วนร่วมในการประเมิน (Vavrus, 1990 cited in Cole et al., 2000; Barabouti, 2012) มีการประเมินระหว่างเรียน (formative assessment) อย่างต่อเนื่อง กล่าวคือครูให้ข้อมูลย้อนกลับแก่นักเรียนอย่างต่อเนื่อง และนักเรียนได้ปรับปรุงผลงานจนได้ผลงานที่สมบูรณ์ ทำให้นักเรียนมีความก้าวหน้าในการเรียน และการประเมินสรุปรวม (summative assessment) ของแฟ้มสะสมงานของนักเรียนยังทำให้ทราบถึงระดับความสามารถของนักเรียนในสิ่งที่เรียนว่าอยู่ในระดับใด มีจุดแข็งจุดอ่อนในผลงานของตนอย่างไร และเป็นประโยชน์สำหรับการวางแผนการเรียนในอนาคตของผู้เรียนเพื่อทำให้ตนบรรลุเป้าหมายในการเรียนได้ (Lam & Lee, 2010) นอกจากนี้แฟ้มสะสมงานยังให้ข้อมูลเกี่ยวกับผู้ที่เป็นเจ้าของแฟ้มอย่างครบถ้วนมากกว่าการทดสอบด้วยแบบทดสอบ

เนื่องจากการจัดทำแฟ้มสะสมงานต้องใช้ความคิด ความรู้ ความอดทน วิจารณ์ญาณ ความพยายาม และทักษะต่างๆ ซึ่งครอบคลุมทั้งทางสมองและจิตใจ ซึ่งเป็นการพัฒนาผู้เรียนแต่ละคนในหลายๆ ด้านไปพร้อมๆ กัน (อุทุมพร (ทองอุไทย) จามรมาน, 2540ก) เป็นสื่อกลางที่นำนักเรียนไปสู่การสร้างผลสัมฤทธิ์ในการเรียนที่มีความสัมพันธ์ระหว่างนักเรียน ครู และผู้ปกครอง การเกาะติดความก้าวหน้า และพัฒนาการตลอดทั้งภาคเรียนหรือทั้งปีการศึกษา เป็นการเรียนการสอนที่เน้นนักเรียนเป็นศูนย์กลาง และภาระงานเป็นตัวแทนของหลักฐานที่มีอยู่จริง (Payne, 2003) ดังนั้นจึงกล่าวได้ว่าการประเมินโดยแฟ้มสะสมงานมีความเหมาะสมที่จะนำมาใช้ในการเรียนการสอนภาษาอังกฤษ และการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษา เพราะการประเมินโดยแฟ้มสะสมงานสามารถทำได้ทั้งระหว่างเรียนและประเมินภาพรวม ทำให้นักเรียนมีเจตคติที่ดีต่อการเรียน นักเรียนสามารถประเมินตนเอง และประเมินเพื่อนในชั้นได้ มีการสะท้อนคิดต่อผลงานของตนเองได้ ทำให้เกิดการเรียนรู้ที่เป็นอิสระมีความเข้าใจในปัญหาที่เกิดขึ้นในการเรียนได้อย่างถ่องแท้ และสามารถจัดการกับปัญหาได้ (Kouzouli, 2012) ซึ่งการให้ข้อมูลย้อนกลับของครู (teacher feedback) ในกระบวนการประเมินระหว่างเรียนนั้นเป็นสิ่งสำคัญของแฟ้มสะสมงานที่จะทำให้ นักเรียนประสบความสำเร็จ ดังนั้นในการวิจัยครั้งนี้จึงใช้การให้ข้อมูลย้อนกลับแบบเชิงบรรยายที่ใช้การให้ข้อมูลย้อนกลับแบบการระบุความสำเร็จและการระบุสิ่งที่ควรปรับปรุง (corrective feedback) เพื่อเป็นการเสริมแรงในการเรียน รวมทั้งรู้เกี่ยวกับสิ่งที่จะนำตนเองไปสู่ความสำเร็จในการเรียนได้ และรู้ถึงข้อผิดพลาดและแก้ไขให้ถูกต้อง (correction of errors) (Tunstall & Gipps, 1996)

แม้แฟ้มสะสมงานจะมีประโยชน์หลายประการตามที่กล่าวมาข้างต้น แต่ยังมีประเด็นที่ต้องพิจารณาซึ่งส่งผลต่อคุณภาพของการประเมินโดยใช้แฟ้มสะสมงาน ได้แก่ ความสอดคล้องระหว่างคะแนนจากผู้ประเมินงานแฟ้มของผู้เรียนมีมากน้อยเพียงใด คะแนนที่ได้จากแฟ้มสะสมงานในช่วงระยะเวลาต่างๆ (ต้นปี ถึง ปลายปี) ควรมีค่าคงที่หรือไม่ ความเป็นมาตรฐานของค่าที่แสดงความสามารถของนักเรียนอยู่ที่ใด ผู้ประเมินควรมีจำนวนกี่คน และจำนวนชิ้นงานของแฟ้มสะสมงานที่ต้องพิจารณา มีปริมาณมากจะเกิดความคุ้มค่าเวลา แรงงานหรือไม่ (อุทุมพร (ทองอุไทย) จามรมาน, 2540ก) สอดคล้องกับ Payne (2003) ที่ให้แง่คิดว่าเป็นการยากที่จะยอมรับความน่าเชื่อถือของผู้อ่าน ผู้ให้คะแนน และผู้ประเมินในประเด็นเรื่องความตรงและความเที่ยง ในทำนองเดียวกัน Elbow (1991, cited in Barton and Collins (1997) กล่าวว่า การประเมินผลด้วยแฟ้มสะสมงานมีความตรงสูงอย่างยิ่ง (external valid) เมื่อเทียบกับการวัดผลในปัจจุบัน โดยให้เหตุผลว่าการประเมินด้วยแฟ้มสะสมงานได้วัดตัวแปรต่างๆ ที่ซับซ้อน (complex variables) อย่างละเอียดรอบด้าน จึงส่งผลให้เห็นความสามารถที่แท้จริงของนักเรียนอย่างมีความแม่นยำ (accurately) และการวัดตัวแปรที่มีความซับซ้อนนี้เอง ทำให้ความเที่ยง (reliability) ในการลงความเห็นในการตัดสินให้คะแนนของครูหรือผู้ให้ระดับคะแนน (graders) แต่ละคนค่อนข้างต่ำ และการใช้เกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating

scale) จะเป็นวิธีที่ถูกนำมาใช้เพื่อให้การประเมินมีความน่าเชื่อถือยิ่งขึ้น เนื่องจากให้สารสนเทศในการประเมินที่ชัดเจนและเป็นประโยชน์ในการพัฒนาการเรียนการสอนสูง ส่วนการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนนั้นต้องให้นักเรียนแสดงความสามารถในการสื่อสารออกมาและต้องให้นักเรียนได้ปฏิบัติกิจกรรมหรือภาระงานเพื่อแสดงความสามารถที่มีออกมา แบบสอบที่ใช้ในการประเมินจึงเน้นการปฏิบัติ และหลีกเลี่ยงไม่ได้ในการใช้ผู้เชี่ยวชาญในการตัดสินระดับความสามารถของนักเรียน ดังนั้นการให้คะแนนจากผู้เชี่ยวชาญทางภาษาอังกฤษ (raters) ที่เชื่อถือได้เกณฑ์การให้คะแนนที่เหมาะสม และภาระงาน (tasks) ที่เหมาะสมจึงมีความสำคัญ ซึ่งในช่วงเวลาที่ผ่านมามีความเที่ยงของการประเมินผลทางภาษาอังกฤษไม่ว่าจะเป็นการวัดความสามารถด้านการพูด ด้านการอ่าน และด้านการเขียน มีความเที่ยงค่อนข้างต่ำ เนื่องจากมีความสับสนเกี่ยวกับจำนวนตัวแปรที่เกี่ยวข้อง เช่น จำนวนชิ้นงาน (tasks) จำนวนผู้ประเมิน (raters) เกณฑ์การให้คะแนน (scoring scales) และแนวการเขียน (writing mode) เป็นต้น (Cooper, 1984; Gebriel, 2009; Huot, 1990; Lee and Kantor, 2005; Schoonen, 2005; cited in Gebriel, 2010)

แนวคิดการวัดแนวใหม่ (modern test theory) เป็นแนวคิดที่สามารถแก้ปัญหาที่เกิดขึ้นในการทดสอบแบบดั้งเดิมได้ กล่าวคือ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ (generalizability theory: G-theory) ได้นำเสนอทางเลือกที่เหมาะสม คือ เสนอกรอบแนวคิด และเป็นเครื่องมือการวิเคราะห์สำหรับรูปแบบการแยกคะแนนรวมถึงความสัมพันธ์ขององค์ประกอบระหว่างผู้สอบ จำนวนข้อสอบ ผู้ตรวจให้คะแนน และแหล่งอื่นๆ ของความคลาดเคลื่อนของรูปแบบการจัดการทดสอบ (Brennan, 2001; Lisa et al., 2010) และพบว่าทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร (multivariate generalizability theory: MGT) สามารถให้สารสนเทศที่มากกว่าทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ คือสามารถตรวจสอบความสัมพันธ์ของคะแนนเอกภพระหว่างองค์ประกอบของแบบสอบหรือการทดสอบองค์ประกอบ และเงื่อนไขของการจัดกลุ่มเพื่อสร้างความเที่ยงสูงสุดสำหรับคะแนนรวม (Brennan, 2001; Lee, 2006; Webb et al., 2007) และให้สารสนเทศเกี่ยวกับความสัมพันธ์ระหว่างส่วนประกอบของแบบสอบซึ่งเป็นสิ่งสำคัญเพื่อตรวจสอบความเหมาะสมของความสัมพันธ์ขององค์ประกอบแต่ละองค์ประกอบของแบบสอบที่นำไปสู่คะแนนรวม (Robert L. Brennan, 2011; Lee, 2006) มีความเหมาะสมสำหรับการวิเคราะห์ความน่าเชื่อถือของคะแนนรวม (composite scores) ในการสอบที่มีเนื้อหาหลายส่วน (multiple content domains) หรือ หลายคุณลักษณะ (traits) ที่มีความสัมพันธ์กันโดยมีความสำคัญเกี่ยวกับความเหมาะสมในสองเรื่อง คือ 1) การประมาณค่าความน่าเชื่อถือของคะแนนรวมที่ใช้รูปแบบที่แตกต่างของการถ่วงน้ำหนัก (different weighting schemes) และ 2) คะแนนเอกภพที่แสดงให้เห็นว่ามีความสัมพันธ์ที่แท้จริงระหว่างคะแนนของการวัดที่มีหลายองค์ประกอบ (Brennan, 2001) นอกจากนี้แล้วยังช่วยพิจารณาค่าความน่าเชื่อถือและกำหนดกลยุทธ์เพื่อการปรับปรุงการประเมินต่อไปใน

อนาคต และยังทำให้การประเมินมีความน่าเชื่อถือในองค์ประกอบย่อยของการสอบและความน่าเชื่อถือในภาพรวมของการสอบ (Brurch, Norman, Schmidt, & Van der vleuten, 2008) ดังนั้นวิธีการวัดแบบใหม่จึงน่าจะเป็นวิธีที่จะช่วยเพิ่มความเที่ยงหรือความน่าเชื่อถือของการวัดและการประเมินความสามารถทางภาษาอังกฤษได้ และทำให้ทราบแหล่งความคลาดเคลื่อนของการวัดได้ (Gebriel, 2010)

จากเหตุผลที่กล่าวมาข้างต้น ผู้วิจัยจึงได้สนใจที่จะพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานขึ้น เพื่อให้ผลการประเมินที่ได้เป็นข้อมูลย้อนกลับไปยังนักเรียนในการปรับปรุงการเรียนของตน รวมทั้งครูผู้สอนและผู้เกี่ยวข้องใช้เป็นสารสนเทศเพื่อการตัดสินใจในการออกแบบการประเมินทักษะการพูดภาษาอังกฤษที่เหมาะสมและนำผลการประเมินไปใช้ในการพัฒนาทักษะการพูดภาษาอังกฤษของนักเรียนได้อย่างมีประสิทธิภาพ ซึ่งรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานที่พัฒนาขึ้นในครั้งนี้มีประเด็นที่เกี่ยวข้องกับการประเมินทักษะการพูด 3 ประเด็นที่นำมาพิจารณาถึงแหล่งความคลาดเคลื่อนต่างๆ ที่เกิดขึ้น และส่งผลต่อความเที่ยงของการประเมิน คือ ภาระงานการพูด โดยใช้กิจกรรมการพูด (tasks) ผู้ประเมิน (raters) และเกณฑ์การให้คะแนนแยกส่วน ได้แก่ เกณฑ์การให้คะแนนภาระงานการพูดระหว่างเรียนและสรุปรวม ซึ่งภาระงานเกี่ยวกับการประเมินทักษะการพูดของเด็ก อายุ 7-12 ปี นั้นพบว่ามีภาระงานต่างๆ ดังนี้ 1) กิจกรรมที่เป็นการพูดอย่างเดียว ได้แก่ การเล่าเรื่อง (new telling) การเล่านิทาน (storytelling) เล่าเรื่องจากภาพ (picture talks) การจัดหมวดหมู่ (categorization tasks) การพูดนำเสนอเรื่องต่างๆ (oral presentation) และการพูดประเภทอื่นๆ (other speaking-only genres) 2) กิจกรรมที่รวมทักษะการฟังและการพูดเข้าด้วยกัน ได้แก่ การสัมภาษณ์ (interview) และการถามและตอบคำถาม (question-and-answer tasks) (Mckay, 2006) ส่วน Efthymiou (2012) ได้ทำการประเมินทักษะการพูดของนักเรียนระดับประถมศึกษา โดยใช้กิจกรรมการประเมิน 4 กิจกรรม ดังนี้ 1) การนำเสนอเกี่ยวกับบุคคลที่เป็นที่รู้จัก 2) การเล่านิทาน และการเล่าเรื่องจากหนังสือ 3) การอธิบายขั้นตอนการปฏิบัติบางสิ่งบางอย่าง และ 4) การถ่ายโอนข้อมูลการอธิบายการจัดวางสิ่งของในตำแหน่งต่างๆ อย่างไร และสำหรับกิจกรรมการประเมินทักษะการพูดที่ O' Sullivan (2008) ได้สรุปกิจกรรมต่างๆ ที่ใช้ประเมินทักษะการพูดไว้ 8 กิจกรรม ดังนี้ การอ่านออกเสียง (reading aloud) การจำลอง (mimicry) การสนทนาแลกเปลี่ยน (conversational exchanges) การนำเสนอโดยการพูด (oral presentation) การถ่ายโอนข้อมูล (information transfer) กิจกรรมที่มีการปฏิสัมพันธ์ระหว่างบุคคล (interaction tasks) เช่น ระหว่างนักเรียนกับนักเรียน นักเรียนกับผู้ประเมิน การสัมภาษณ์ (interview) และการอภิปราย (discussion) ในประเด็นที่กำหนดให้จากงานวิจัยที่กล่าวมานี้พบว่าการประเมินทักษะการพูดนั้น

จะต้องทำการประเมินโดยให้นักเรียนได้แสดงความสามารถในการสื่อสารออกมาด้วยการพูดโดยใช้กิจกรรมต่างๆ เป็นตัวกระตุ้น

ดังนั้นการวิจัยครั้งนี้ผู้วิจัยจึงได้ประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ซึ่งใช้ภาระงานการพูดตามระดับความสามารถทางภาษาที่เหมาะสมกับอายุของนักเรียน (CEFR, 2001) ร่วมกับเกณฑ์การให้คะแนน ACTFL Oral Proficiency Interview (OPI) ของ American Council on the Teaching Foreign Language (ACTFL, 2012) ซึ่งเป็นเกณฑ์ที่มีการใช้อย่างกว้างขวาง และเป็นที่ยอมรับในการประเมินการพูดภาษาที่สองและภาษาต่างประเทศ (Johnson และ Tyler, 1998 cited in Sutthirak Sapsirin, 2006) โดยมีเกณฑ์การให้คะแนนแบบแยกองค์ประกอบ (analytic scoring) ตามกรอบแนวคิดการประเมินความสามารถทางภาษาของ Bachman and Palmer (2010) นอกจากนี้แล้วจากการศึกษาเอกสารและงานวิจัยเกี่ยวกับทักษะการพูดภาษาอังกฤษพบว่ามีการศึกษาในสองประเด็นคือ 1) การพัฒนาความสามารถในทักษะการพูดภาษาอังกฤษ และ 2) การประเมินความสามารถในทักษะการพูดภาษาอังกฤษ โดยพบว่าการวิจัยเกี่ยวกับทักษะการพูดภาษาอังกฤษนั้นส่วนใหญ่จะเป็นการประเมินในระดับอุดมศึกษาและมัธยมศึกษา ส่วนการประเมินในระดับประถมศึกษาแล้วยังพบน้อย และที่พบเป็นการประเมินในห้องเรียนเท่านั้น (Westerveld และ Moran, 2010; Efthymiou, 2012) ส่วนการวิจัยในประเทศไทยพบว่าการวิจัยเกี่ยวกับทักษะการพูดนั้นส่วนใหญ่เป็นการวิจัยเพื่อพัฒนาการเรียนการสอนทักษะการพูดภาษาอังกฤษเท่านั้น (อดิศา เทียว, 2550; จริญญา กสิวิทย์อำนวย, 2539; นริศรา ดานูงศ์, 2537; ประไพ เทียนเงิน, 2534; กัญจน์ ฤทธิประศาสน์, 2541; Janudom , 2009) ส่วนการวิจัยเกี่ยวกับการประเมินทักษะการพูดภาษาอังกฤษพบว่ายังมีการศึกษาน้อยและพบเฉพาะในระดับ อุดมศึกษาและมัธยมศึกษา (อัจฉรา วงศ์โสธร, 2537; อัจฉรา วงศ์โสธร, 2543)

จากที่กล่าวมาข้างต้นสรุปได้ว่าการวิเคราะห์ข้อมูลเกี่ยวกับรูปแบบการประเมินทักษะการพูดภาษาอังกฤษและแฟ้มสะสมงานหลักเฉียงไม่ได้ที่จะต้องใช้การประเมินด้วยภาระงานหรือการให้นักเรียนได้ปฏิบัติจริง มีเกณฑ์การให้คะแนนที่เชื่อถือได้ และต้องเกี่ยวข้องข้องกับการประเมินด้วยผู้เชี่ยวชาญ การนำทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรมาประยุกต์ใช้จึงมีความเหมาะสมในการวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ เพื่อให้สารสนเทศในการตัดสินใจการเลือกภาระงานการพูดจำนวนผู้ประเมิน วิธีการให้คะแนนของผู้ประเมิน และเกณฑ์การให้คะแนนเกี่ยวกับการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานที่มีความน่าเชื่อถือและมีความเที่ยงสูง และนำไปสู่การได้สารสนเทศที่ใช้ในการวางแผนการประเมินที่เหมาะสมและเป็นไปได้ในการปฏิบัติต่อไป

คำถามวิจัย

1. รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานควรมีขั้นตอนการดำเนินการ องค์ประกอบของแฟ้มสะสมงาน และเกณฑ์การให้คะแนนภาระงานอย่างไร

2. รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานที่พัฒนาขึ้นมีคุณภาพมากน้อยเพียงใด เมื่อศึกษาเงื่อนไขที่เกี่ยวข้อง 3 แหล่ง ได้แก่ ผู้ประเมิน ภาระงานการพูด และเกณฑ์การให้คะแนนภาระงาน ตามแนวคิดทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานสำหรับประเมินทักษะการพูดภาษาอังกฤษ

1.1 พัฒนาขั้นตอนการดำเนินการ องค์ประกอบ และคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

1.2 พัฒนาเกณฑ์การให้คะแนนและภาระงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 สำหรับการประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment)

2. เพื่อศึกษาผลการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

3. เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานสำหรับประเมินทักษะการพูดภาษาอังกฤษตามช่วงเวลาของการประเมิน (ระหว่างเรียนและสรุปรวม) และตามแหล่งความคลาดเคลื่อนของการประเมิน 2 แหล่ง ได้แก่ จำนวนผู้ประเมินและจำนวนภาระงานด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร ดังนี้

3.1 เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ

3.2 เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของภาระงานการประเมินทักษะการพูดภาษาอังกฤษ

3.3 เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment)

3.4 เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการตรวจให้คะแนนของผู้ประเมิน

ขอบเขตของการวิจัย

การวิจัยครั้งนี้ผู้วิจัยกำหนดขอบเขตของการวิจัยไว้ ดังนี้

1. การพัฒนารูปแบบการประเมินทักษะการพูดของนักเรียนโดยใช้แฟ้มสะสมงานครั้งนี้ ทำการศึกษาเกี่ยวกับนักเรียนระดับชั้นประถมศึกษาปีที่ 6 ที่เรียนอยู่ในโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ

2. รูปแบบการประเมินทักษะการพูดภาษาอังกฤษ เจ็อนไซที่มุ่งศึกษาประกอบแหล่งของความคลาดเคลื่อนจาก 2 แหล่ง ดังนี้

2.1 ผู้ประเมิน (raters)

2.2 ภาระงานการพูดภาษาอังกฤษที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ในครั้งนี้ ได้แก่ การสัมภาษณ์ (interview) การนำเสนอปากเปล่า (oral presentation) การพูดเล่านิทาน (storytelling) และการบรรยายภาพ (picture description)

3. รูปแบบการประเมินทักษะการพูดภาษาอังกฤษโดยใช้แฟ้มสะสมงาน ได้แก่ เจ็อนไซในการประเมินทักษะการพูดภาษาอังกฤษที่มีผลต่อความน่าเชื่อถือของการประเมิน ซึ่งแหล่งที่ทำให้เกิดความคลาดเคลื่อนในการประเมินสำหรับการวิจัยครั้งนี้ ประกอบด้วย 2 แห่ง (facets) ได้แก่ ผู้ประเมิน (raters) และภาระงาน (tasks) โดยการวิจัยครั้งนี้มีการประเมินแฟ้มสะสมงาน 2 ระยะ คือ การประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) โดยรูปแบบที่ทำการศึกษาค้นคว้าได้แก่

3.1 ระดับความคลาดเคลื่อนและความเที่ยงของภาระการประเมินทักษะการพูดภาษาอังกฤษ แต่ละภาระงาน โดยใช้การออกแบบ $p^{\circ} \times t^{\circ} \times r^{\circ}$ Design หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกภาระงาน โดยแต่ละภาระงานไม่ซ้ำกัน และผู้ประเมิน (r) ทุกคนประเมินนักเรียนทุกคน

3.2 ระดับความคลาดเคลื่อนและความเที่ยงของเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ โดยใช้การออกแบบ $p^{\circ} \times t^{\circ} \times r^{\circ}$ Design หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกภาระงาน โดยแต่ละภาระงานซ้ำกัน และผู้ประเมิน (r) ทุกคนประเมินนักเรียนทุกคน

3.3 ความเหมาะสมของจำนวนภาระงานและจำนวนผู้ประเมิน และการให้คะแนนของผู้ประเมินแต่ละแบบ โดยใช้การออกแบบ $(p^\circ \times t^\circ \times r^\circ)$ และ $(p^\bullet \times t^\bullet \times r^\bullet)$ Design หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกภาระงาน โดยภาระงานแต่ละภาระงานไม่ซ้ำกัน และถูกประเมินจากผู้ประเมิน (r) ทุกคน โดยการให้คะแนนภาระงานจากผู้ประเมินคนเดียวกัน (r°) หรือต่างคนกัน (r^\bullet)

4. เกณฑ์การให้คะแนน เป็นเกณฑ์การให้คะแนนแบบแยกองค์ประกอบ (analytic rating scale) มีขอบเขตเนื้อหา 6 ด้าน ได้แก่ คำศัพท์ (vocabulary) การสร้างประโยค (syntax) การเชื่อมโยงเนื้อหา (cohesion) การออกเสียง (pronunciation) บรรลุวัตถุประสงค์ (ideational function) และความคล่องแคล่ว (fluency) ให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี ตามลำดับ

5. กรอบแนวคิดการวิจัยที่ใช้เพื่อการพัฒนาเกณฑ์การให้คะแนนทักษะการพูดใช้แนวคิดการประเมินทางภาษาของ Bachman and Palmer (2010) ใช้แนวคิดระดับความสามารถในการพูดภาษาต่างประเทศของ ACFTL (2012) และแนวคิดของ CEFR (2001) ใช้เป็นกรอบเนื้อหาของแบบสอบภาคปฏิบัติให้เหมาะสมกับระดับความสามารถของนักเรียนชั้นประถมศึกษาปีที่ 6 คือ ระดับเบื้องต้น (A1 (breakthrough) A2 (way stage)) และ ระดับพึ่งพาตนเองได้ (B1 (threshold))

6. กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1 ห้อง ที่เรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ครูผู้สอนกลุ่มสาระการเรียนรู้ภาษาต่างประเทศ (ภาษาอังกฤษ) ที่ทำหน้าที่เป็นผู้ประเมิน (raters) ความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้ภาระงานการพูดภาษาอังกฤษ (tasks)

คำจำกัดความของการวิจัย

1. ภาระงานการพูด หมายถึง กิจกรรมการพูดที่ใช้ในการประเมินความสามารถด้านภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ในทักษะการพูด ประกอบด้วย 4 ภาระงาน ได้แก่ การสัมภาษณ์ (interview) การนำเสนอปากเปล่า (oral presentation) การเล่านิทาน (storytelling) และการบรรยายภาพ (picture description)

2. รูปแบบการประเมินทักษะการพูดภาษาอังกฤษโดยใช้แฟ้มสะสมงาน หมายถึง เงื่อนไขในการประเมินทักษะการพูดภาษาอังกฤษที่มีผลต่อความน่าเชื่อถือของการประเมิน ซึ่งเงื่อนไขที่มีผลต่อความน่าเชื่อถือของการประเมินหรือความเที่ยงในการวิจัยครั้งนี้คือแหล่งที่ทำให้เกิดความคลาดเคลื่อนในการประเมินสำหรับการวิจัยครั้งนี้ ประกอบด้วย 2 แหล่ง (facets) ได้แก่ ผู้ประเมิน (raters) และภาระงาน (tasks) โดยการวิจัยครั้งนี้มีการประเมินแฟ้มสะสมงาน 2 ระยะ คือ การประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) โดยรูปแบบที่ทำการศึกษาค้นคว้าครั้งนี้ได้แก่

2.1 การประเมินระหว่างเรียน (formative assessment) คือ การประเมินทักษะการพูดของนักเรียนในระหว่างการดำเนินการเรียนการสอนโดยใช้แฟ้มสะสมงานฯ โดยมีจุดประสงค์เพื่อศึกษาความก้าวหน้าของผู้เรียน และปรับปรุงและพัฒนาการพูดของนักเรียน

2.2 การประเมินสรุปรวม (summative assessment) คือ การประเมินเมื่อสิ้นสุดการดำเนินการจัดการเรียนการสอนด้วยแฟ้มสะสมงานฯ เพื่อสรุปรวมและให้ระดับทักษะการพูดแก่นักเรียน

3. สถานการณ์การประเมิน หมายถึง แหล่งความคลาดเคลื่อนที่สำคัญในวิธีดำเนินการในการประเมินทักษะการพูดภาษาอังกฤษนำไปสู่การตัดสินใจเลือกสถานการณ์การประเมินที่มีความน่าเชื่อถือในระดับที่ต้องการ ประกอบด้วย จำนวนผู้ประเมิน และจำนวนภาระงาน

4. การประเมินทักษะการพูด หมายถึง การทดสอบทักษะการพูดภาษาอังกฤษของนักเรียนโดยให้นักเรียนได้แสดงความสามารถในการพูดสื่อสารความในการสื่อสารออกมา ซึ่งสำหรับการวิจัยครั้งนี้ทำการประเมินทักษะการพูดภาษาอังกฤษโดยให้นักเรียนพูดภาษาอังกฤษในสถานการณ์หรือภาระงานการพูด (tasks) ที่กำหนดให้

5. ภาระงานการพูด (task) เป็นกิจกรรมที่ให้นักเรียนใช้สำหรับการทดสอบทักษะการพูดของนักเรียนโดยการให้นักเรียนได้แสดงการพูดสื่อสารหรือพูดสื่อความออกมาตามความรู้ความสามารถทางภาษาอังกฤษ ซึ่งในการวิจัยครั้งนี้ภาระงานการพูดที่ใช้ได้แก่ 1) การสัมภาษณ์ (interview) 2) การนำเสนอปากเปล่า (oral presentation) 3) การเล่านิทาน (storytelling) และ 4) การบรรยายภาพ (picture description) โดยมีเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale)

6. แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ หมายถึง สิ่งที่ใช้สำหรับใส่เอกสารหลักฐานผลงานของนักเรียน รวมถึง CD บันทึกเสียงการพูดภาษาอังกฤษของนักเรียนที่แสดงถึงผลการเรียนรู้และความสามารถในการพูดภาษาอังกฤษของนักเรียน ซึ่งประกอบไปด้วย 4 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ

ประโยชน์ที่คาดว่าจะได้รับ

1. ได้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานที่มีทั้งการประเมินระหว่างเรียนและสรุปรวม ซึ่งจะเป็นประโยชน์แก่ครูผู้สอนและผู้เกี่ยวข้องในการจัดการเรียนการสอนภาษาอังกฤษในโรงเรียนระดับประถมศึกษาต่อไป

2. ได้รูปแบบการให้คะแนนของผู้ประเมินในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ซึ่งจะเป็นประโยชน์ต่อครูผู้สอน ผู้บริหาร และผู้ที่ทำหน้าที่ในการประเมินทักษะการพูดที่จะใช้เพื่อวางแผนการประเมินได้อย่างเหมาะสม มีความน่าเชื่อถือ และสามารถปฏิบัติได้จริงต่อไป

3. ได้เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานทั้งระหว่างเรียนและสรุปรวม เพื่อเป็นแนวทางในการประเมินทักษะการพูดภาษาอังกฤษของครูและผู้ประเมินที่มีความน่าเชื่อถือและเป็นไปได้ในทางปฏิบัติ



บทที่ 2

แนวคิด ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

การศึกษาทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องกับการพัฒนาหัวข้อการวิจัย ผู้วิจัยได้นำเสนอประเด็นเนื้อหาสาระเป็น 4 ตอน ได้แก่ ตอนที่ 1 การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษา ตอนที่ 2 การประเมินโดยใช้แฟ้มสะสมงาน (portfolio assessment) ตอนที่ 3 ทฤษฎีการสรุปร่างอิงความน่าเชื่อถือของการวัดผลแบบพหุตัวแปร และตอนที่ 4 กรอบแนวคิดในการวิจัยรายละเอียดของเนื้อหาสาระแต่ละตอนมีดังนี้

ตอนที่ 1 การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษา

การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษาเป็นสิ่งที่มีความสำคัญ เพราะเป็นพื้นฐานในการสื่อสารเบื้องต้น และเป็นพื้นฐานในการเรียนในระดับสูงต่อไป ซึ่งแนวคิดและทฤษฎีที่เกี่ยวข้องมีรายละเอียด ดังนี้

1.1 ความสามารถทางภาษา (language proficiency)

ความสามารถทางภาษาอังกฤษ (language proficiency) เป็นสิ่งที่นักภาษาศาสตร์ให้ความสำคัญมาก ทั้งนี้เนื่องจากภาษาเป็นเครื่องมือในการสื่อสาร เป็นการเชื่อมโยงระหว่างคนหนึ่งไปยังอีกคนหนึ่ง และใช้สื่อสารระหว่างประเทศเพื่อประกอบธุรกิจต่างๆ ดังนั้นภาษาอังกฤษจึงเป็นภาษาสำคัญ ในการวัดและประเมินทางภาษามีคำที่นักภาษาศาสตร์ใช้อยู่หลายคำ ได้แก่ language proficiency, communicative competence หรือ communicative language ability (CLA) (Fulcher & Davidsoy, 2007) Bachman and Palmer (2010) ได้ให้ความหมายของความสามารถทางภาษา (language ability) ไว้ว่า ความสามารถทางภาษาเป็นความสามารถในการใช้ภาษาในการสื่อสาร โดยมีส่วนประกอบ 2 ส่วน คือ 1) สมรรถนะทางภาษา หรือ ความรู้ทางภาษา (language knowledge) และ 2) กลยุทธ์ทางภาษา (language strategic) สอดคล้องกับ Mckay (2006) ได้ให้ความหมายความสามารถในการใช้ภาษา (language use ability) ไว้ว่าเป็นความสามารถในการใช้ภาษาเพื่อบรรลุเป้าหมายหรือวัตถุประสงค์ในสถานการณ์ที่เหมาะสม ซึ่งความสามารถทางภาษาเรียกได้อีกอย่างหนึ่งว่าเป็นความสามารถในการสื่อสารภาษา (language communicatively) ซึ่งหมายความว่านักเรียนได้แลกเปลี่ยนโดยการสร้างสรรค์การใช้ภาษาของตนให้สอดคล้องกับสถานการณ์ที่มีปฏิสัมพันธ์ด้วย เช่น เพื่อการตอบคำถามเกี่ยวกับภารกิจหรือกิจกรรมที่ทำ (task) อย่างเหมาะสมกับบริบท โดยผู้สอนสามารถสังเกตได้จากหลักฐานต่างๆ ดังนี้

- 1) เข้าใจภาษาใหม่ที่ครูออกเสียง การพูดกับนักเรียนคนอื่นๆ หรือการเขียนเรื่อง (การใช้กลยุทธ์เพื่อเดาว่าจะเริ่มพูดอะไรจากบริบท)
- 2) ตอบสนองต่อคำสั่งได้อย่างเหมาะสม (อาจเป็นการตอบสนองด้วยการเคลื่อนไหวทางกาย)
- 3) สร้างสรรค์คำพูดของตนเอง การพูดด้วยคำพูดของตนเองในการสร้างประโยค หรือเปลี่ยนรูปแบบด้วยประโยคของตนเองบนพื้นฐานของคำศัพท์และโครงสร้างที่เคยเรียนมา
- 4) ใช้ภาษาได้อย่างเหมาะสมในปฏิสัมพันธ์ที่ไม่ได้ฝึกซ้อม กล่าวคือในสถานการณ์ที่นักเรียนไม่เคยฝึกมาก่อนในการเรียน

1.1.1 มาตรฐานเกี่ยวกับความสามารถทางภาษาอังกฤษ

1) สาระการเรียนรู้แกนกลางกลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

สาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ (ภาษาอังกฤษ) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ประกอบด้วย 4 สาระ ได้แก่ ภาษาเพื่อการสื่อสาร ภาษาและวัฒนธรรม ภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น และภาษากับความสัมพันธ์กับชุมชนโลก (สำนักวิชาการและมาตรฐานการศึกษา, 2553) ซึ่งแต่ละสาระมีรายละเอียด ดังนี้

สาระที่ 1 ภาษาเพื่อการสื่อสาร นักเรียนสามารถใช้ภาษาในการสื่อสารทั้งการฟัง การพูด การอ่าน และการเขียน เพื่อแลกเปลี่ยนข้อมูลข่าวสาร เข้าใจ ติความจากการอ่านและการฟังได้ และสามารถนำเสนอความคิดโดยการพูดหรือการเขียน

สาระที่ 2 ภาษาและวัฒนธรรม นักเรียนเข้าใจความสัมพันธ์ระหว่างภาษาและวัฒนธรรมของเจ้าของภาษา สามารถนำมาปรับใช้ในการสื่อสารได้อย่างเหมาะสมและเข้าใจความแตกต่างระหว่างวัฒนธรรมของเจ้าของภาษาและภาษาและวัฒนธรรมของไทย

สาระที่ 3 ภาษากับความสัมพันธ์กับกลุ่มสาระการเรียนรู้อื่น สามารถใช้ภาษาต่างประเทศในการเชื่อมโยงไปยังการเรียนวิชาอื่นๆ และเป็นพื้นฐานในการพัฒนาหาความรู้เพื่อเรียนรู้ของตน

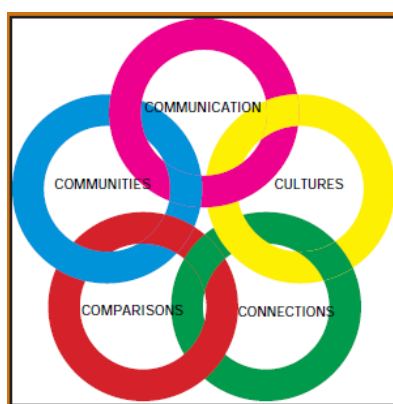
สาระที่ 4 ภาษากับความสัมพันธ์กับชุมชนโลก นักเรียนใช้ภาษาต่างประเทศในสถานการณ์ต่างๆ ทั้งในโรงเรียนและนอกโรงเรียน เช่น ชุมชน รวมทั้งใช้ภาษาต่างประเทศในการศึกษาต่อ ประกอบอาชีพ การทำงาน และเรียนรู้สิ่งต่างๆ กับสังคมโลกได้

โดยได้กำหนดคุณภาพของนักเรียนเมื่อจบในแต่ละระดับ ได้แก่ ชั้นประถมศึกษาปีที่ 3 ชั้นประถมศึกษาปีที่ 6 ชั้นมัธยมศึกษาปีที่ 3 และชั้นมัธยมศึกษาปีที่ 6 ซึ่งในการศึกษาคั้งนี้จะศึกษาเฉพาะในระดับชั้นประถมศึกษาปีที่ 6 ในทักษะการพูดเท่านั้น ซึ่งมีรายละเอียดดังนี้

1. สามารถปฏิบัติตามคำสั่ง คำขอร้องจากการฟังและการอ่านได้ สามารถอ่านข้อความและบทกลอนสั้นๆ ได้ สามารถตอบคำถามจากการฟังและการอ่านบทสนทนาิทานเรื่องเล่าง่ายๆ ได้
2. สามารถพูด/เขียนตอบโต้ในการสื่อสารระหว่างบุคคล ใช้คำสั่ง คำขอร้อง ขออนุญาต ให้คำแนะนำ แสดงความต้องการ ขอความช่วยเหลือ ตอบรับและปฏิเสธการช่วยเหลือในสถานการณ์ง่ายๆ สามารถพูด เขียน ข้อมูลที่ใกล้ตัว เช่น ส่วนตัว เพื่อน ครอบครัว และแสดงความรู้สึกกับเรื่องต่างๆ ใกล้ตัวได้โดยการบอกเหตุผลสั้นๆ ประกอบ
3. พูด เขียน ให้ข้อมูลของตนเอง เพื่อน สิ่งแวดล้อมใกล้ตัว ภาพ แผ่นผัง แผนภูมิ ตารางแสดงข้อมูลต่างๆ ที่ใกล้ตัว
4. สามารถใช้ถ้อยคำ น้ำเสียง ที่สุภาพ เหมาะสมตามมารยาททางสังคมและวัฒนธรรมของเจ้าของภาษา ให้ข้อมูลเกี่ยวกับเทศกาล วันสำคัญ งานฉลอง ชีวิตความเป็นอยู่ของเจ้าของภาษา
5. ใช้ภาษาสื่อสารในสถานการณ์ต่างๆ ที่เกิดขึ้นในห้องเรียนและโรงเรียนได้
6. สามารถสื่อสารในหัวข้อเกี่ยวกับตนเอง ครอบครัว โรงเรียน สิ่งแวดล้อม อาหาร เครื่องดื่ม เวลาว่างและนันทนาการ สุขภาพและสวัสดิการ การซื้อ-ขาย และภูมิอากาศได้ภายในวงคำศัพท์ 1,050-1,200 คำ (คำศัพท์ที่เป็นรูปธรรมและนามธรรม)
7. สามารถใช้ประโยคเดี่ยวและประโยคผสมได้ (compound sentence)

ส่วนมาตรฐานทางการจัดการเรียนการสอนภาษาต่างประเทศที่จัดทำขึ้นนี้เป็นมาตรฐานที่ใช้เป็นมาตรฐานในการจัดการเรียนรู้ในศตวรรษ ที่ 21 โดยสมาพันธ์อเมริกาในการสอนภาษาต่างประเทศ (American Council on the Teaching of Foreign languages: ACTFL, 1999) ซึ่งเป็นมาตรฐานในการเรียนการสอนภาษาต่างประเทศ ซึ่งมาตรฐานนี้เป็นมาตรฐานที่มีการนำมาใช้อย่างแพร่หลายและเป็นกรอบในการจัดการศึกษาและจัดทำหลักสูตรการเรียนรู้อาษาต่างประเทศในหลายประเทศรวมทั้งประเทศไทยด้วย ซึ่งเนื้อหาของมาตรฐานนี้ประกอบด้วยมาตรฐาน 5 มาตรฐาน (The Five Cs) ได้แก่ ภาษาเพื่อการสื่อสาร (communication) ภาษาและวัฒนธรรม (cultures) การเชื่อมโยง (connections) การเปรียบเทียบ (comparisons) และ ภาษากับชุมชน (communities) โดยมาตรฐานทั้ง 5 นี้ มีความสัมพันธ์กันและเกี่ยวข้องซึ่งกัน ซึ่งสอดคล้องกับการใช้ภาษาในชีวิตประจำวัน ซึ่งแต่ละมาตรฐานมีรายละเอียดดังนี้ **ภาษาเพื่อการสื่อสาร (communication)** ภาษาเพื่อการสื่อสารเป็นหัวใจหลักของการเรียนภาษาที่สองซึ่งต้องเป็นสื่อสารแบบเผชิญหน้า (face to face) และในการจัดการเรียนการสอนต้องให้เป็นภาษาที่ใช้ในชีวิตจริง (real life) เพื่อให้ให้นักเรียนสามารถใช้ในชีวิตได้ ซึ่งเป็นการแสดงออกโดยการฟัง-พูด การเขียน แสดงถึงการเข้าใจในวัฒนธรรมเมื่อต้องสื่อสาร **ภาษาและวัฒนธรรม (cultures)** การเข้าใจวัฒนธรรมของเจ้าของภาษาเป็นสิ่งสำคัญเพราะทำให้นักเรียนสามารถสื่อสารได้ถูกต้องตามเหมาะสมกับภาษาที่ใช้เหมือนภาษาของตนเอง ยินดีในวัฒนธรรม

ของตนและของภาษาอื่น โดยเข้าใจในมุมมองทางความคิดวิถีชีวิตของเจ้าของภาษา **การเชื่อมโยง** (connections) การสอนภาษาต่างประเทศจำเป็นต้องมีการเชื่อมโยงกับวิชาอื่นๆ การเชื่อมต่อนี้เป็นการบูรณาการการสอนภาษากับสอนในวิชาอื่นๆ **การเปรียบเทียบ** (comparisons) นักเรียนต้องได้รับการสนับสนุนให้มีการเปรียบเทียบความขัดแย้งทางภาษาและวัฒนธรรม ทำให้นักเรียนได้พบรูปแบบ การทำนาย การวิเคราะห์สถานการณ์ และความแตกต่างระหว่างภาษาและวัฒนธรรม เพื่อความเข้าใจในธรรมชาติของภาษาที่เรียนกับภาษาแม่ของตน และ **ภาษากับชุมชน** (communities) องค์ประกอบนี้ช่วยให้การใช้ภาษาของนักเรียนได้มีส่วนร่วมในชุมชนของตน เช่น ครอบครัวยุติที่ได้หลายภาษา และทั่วโลกในหลากหลายบริบทและในรูปแบบที่เหมาะสมทางวัฒนธรรม ดังภาพที่ 2.1



ภาพที่ 2.1 มาตรฐานในการเรียนการสอนภาษาต่างประเทศ 5 มาตรฐาน (The Five Cs) ของ ACTFL, 1999. ที่มา: American Council on the Teaching of Foreign Languages: ACTFL, 1999.

จากการศึกษามาตรฐานการจัดการศึกษาเกี่ยวภาษาต่างประเทศทั้ง 2 มาตรฐานพบว่า สาระการเรียนรู้ตามหลักสูตรแกนกลางพุทธศักราช 2551 มีความสอดคล้องกับมาตรฐานการจัดการเรียนการสอนภาษาต่างประเทศของ ACTFL (1999) แต่สาระการเรียนรู้ตามหลักสูตรแกนกลางพุทธศักราช 2551 มีการปรับสาระการเรียนรู้ลงเหลือเพียง 4 สาระ จากที่ ACTFL มี 5 มาตรฐาน โดยมาตรฐานที่ถูกตัดออกคือ การเปรียบเทียบ (comparisons) ส่วนสาระและมาตรฐานอื่นพบว่า มีความสอดคล้องกัน

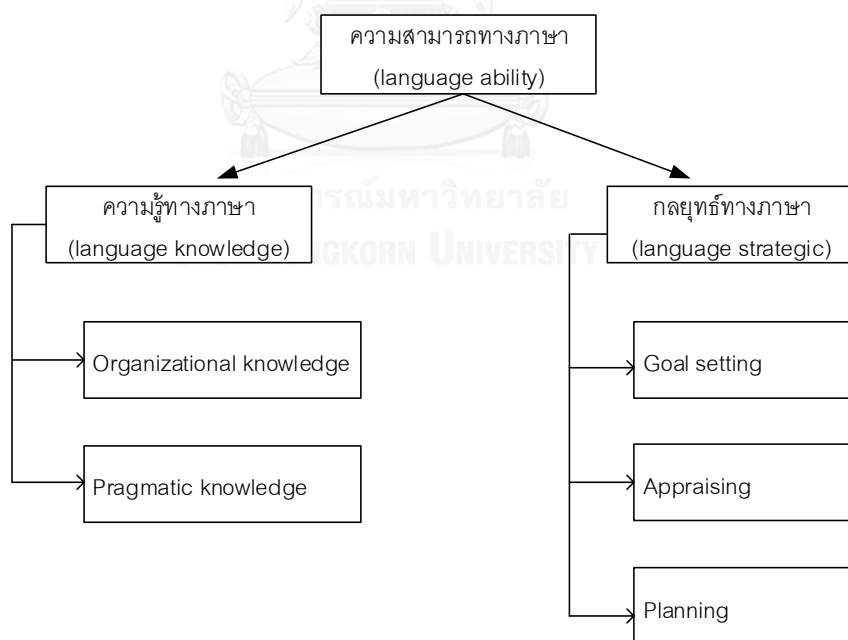
1.2 การประเมินความสามารถทางภาษาอังกฤษ

การประเมินความสามารถทางภาษา (assessing language proficiency) เป็นการประเมินความสามารถสูงสุดของบุคคลหรือผู้เรียน ซึ่งการประเมินความสามารถมีลักษณะเฉพาะและแตกต่างจากการประเมินหรือการทดสอบแบบอื่น คือเป็นการประเมินความสามารถในการใช้ภาษาในสถานการณ์ต่างๆ ที่มีอยู่จริงในสังคม เป็นการประเมินที่ไม่สัมพันธ์หรือขึ้นกับหลักสูตร แต่มุ่งวัดความสามารถของนักเรียนในปัจจุบัน เป็นการประเมินที่มุ่งไปข้างหน้าสนใจอนาคต เนื้อหาการทดสอบจึงเป็นสิ่งที่มีความจริง (authentic)

เหมาะแก่การใช้ภาษาอังกฤษเฉพาะกิจ ใช้ในการเรียนการสอน การอภิปราย การสัมมนา ซึ่งเป็นภาษาที่จำเป็นสำหรับนักเรียนในอนาคต (อัจฉรา วงศ์โสธร, 2543) ซึ่งการเข้าใจแนวคิดต่างๆ ที่เกี่ยวข้องกับการประเมินความสามารถทางภาษาจึงมีความสำคัญเพื่อเป็นความรู้ในการประเมินที่เหมาะสมต่อไป

1.2.1 โมเดล และโครงสร้างการประเมินทางภาษาอังกฤษ

การทดสอบทางภาษานั้นมักจะสับสนระหว่างคำว่า กรอบความคิด (framework) และ โมเดล (model) ซึ่งกรอบความคิดนั้นเป็นสิ่งที่ผู้ประเมินใช้ในการเลือกทักษะหรือความสามารถจากโมเดลที่มีความสัมพันธ์กับเนื้อหาการประเมิน ส่วนคำว่าโครงสร้าง (construct) เป็นสิ่งที่ใช้อธิบายส่วนประกอบในโมเดล ซึ่งสิ่งเหล่านี้เป็นเหมือนทิศทาง (เข็มทิศ) ในการพัฒนาแบบสอบ หรือเครื่องมือที่ใช้การประเมิน โดยสามารถจำแนกออกได้เป็น 3 ระดับดังนี้ โมเดลเชิงทฤษฎี (theoretical model) กรอบความคิดการประเมิน (assessment framework) และคุณลักษณะเฉพาะของแบบสอบ (test specification) ซึ่งโมเดลในการประเมินทางภาษาอังกฤษนั้น Bachman (1990) นักภาษาศาสตร์ได้นำเสนอโมเดลความสามารถทางภาษาซึ่งถือว่าเป็นโมเดลที่ได้รับการใช้อย่างแพร่หลาย โดย Bachman และ Palmer ให้ความหมายความสามารถทางภาษา (language ability) ว่าเป็นความสามารถในการใช้ภาษาเพื่อการสื่อสาร (language community) ประกอบด้วย 2 องค์ประกอบ คือ 1) ความรู้ทางภาษา (language knowledge) และ 2) กลยุทธ์ทางภาษา (language strategic) ดังภาพที่ 2.2



ภาพที่ 2.2 องค์ประกอบความสามารถทางภาษาตามโมเดลของ Bachman and Palmer (2010)
ที่มา: Bachman and Palmer (2010)

- ดั่งนี้
1. ความรู้ทางภาษา (language knowledge)
 - 1.1 ความรู้ในการเรียบเรียง (organizational knowledge) ประกอบด้วยองค์ประกอบต่างๆ
 - 1.1.1 ความรู้เกี่ยวกับหลักไวยากรณ์ (grammatical knowledge)
 - 1) ความรู้เกี่ยวกับคำศัพท์ (knowledge of vocabulary)
 - 2) ความรู้เกี่ยวกับการสร้างประโยค (knowledge of syntax)
 - 3) ความรู้เกี่ยวกับการออกเสียง/ตัวสะกด (knowledge of phonology/graphology)
 - 1.1.2 ความรู้เกี่ยวกับความสัมพันธ์ของเนื้อความ (textual knowledge)
 - 1) ความรู้เกี่ยวกับการเชื่อมโยงเนื้อความ (knowledge of cohesion)
 - 2) ความรู้เกี่ยวกับเรียบเรียงข้อความหรือการสนทนา (knowledge of rhetorical or conversational organization)
 - 1.2 ความรู้เกี่ยวกับการใช้ภาษาในทางปฏิบัติ (pragmatic knowledge)
 - 1.2.1 ความรู้เกี่ยวกับหน้าที่ทางภาษา (functional knowledge)
 - 1) ความรู้เกี่ยวกับภาษาที่เหมาะสม (knowledge of ideational functions)
 - 2) ความรู้เกี่ยวกับภาษาในเชิงจัดการ สั่งการ (knowledge of manipulative functions)
 - 3) ความรู้เกี่ยวกับหน้าที่ของภาษาที่ช่วยแก้ปัญหา (knowledge of heuristic function)
 - 4) ความรู้เกี่ยวกับหน้าที่ของภาษาในเชิงสร้างสรรค์ (knowledge of imaginative function)
 - 1.2.2 ความรู้เกี่ยวกับภาษาศาสตร์เชิงสังคม (sociolinguistic knowledge)
 - 1) ความรู้เกี่ยวกับประเภทของเนื้อความ (โวหาร) (knowledge of genres)
 - 2) ความรู้เกี่ยวกับภาษาท้องถิ่น (knowledge of dialects/varieties)
 - 3) ความรู้เกี่ยวกับทำเนียบภาษา (knowledge of registers)
 - 4) ความรู้เกี่ยวกับการใช้ภาษาอย่างเป็นธรรมชาติ (knowledge of natural or idiomatic expressions)
 - 5) ความรู้เกี่ยวกับการใช้ภาษาเชิงเปรียบเทียบโดยอ้างอิงกับวัฒนธรรม (knowledge of cultural references and figures of speech)
 2. กลยุทธ์ทางภาษา (language strategic)
 - 2.1 การตั้งเป้าหมาย (goal setting) คือ การตัดสินใจเพื่อเลือกจะทำสิ่งใดสิ่งหนึ่ง
 - 2.1.1 การระบุการใช้ภาษาหรือกิจกรรมการประเมินเพื่อหาแนวทางการใช้ภาษา
 - 2.1.2 การเลือกกิจกรรมหนึ่งกิจกรรมหรือมากกว่าหนึ่งกิจกรรมจากชุดกิจกรรมที่มีความเป็นไปได้ (บางครั้งจากสิ่งที่มีอยู่ปกติ ถ้ามีเพียงสิ่งเดียว)
 - 2.1.3 การตัดสินใจทำหรือไม่ทำเพื่อให้กิจกรรมเสร็จสิ้น

2.2 การประเมิน (appraising) คือ การให้ผู้เรียนได้ทดลองทำสิ่งที่มีความจำเป็นในการทำงานอะไรที่ต้องใช้และจะอย่างไรจึงจะสำเร็จ

2.2.1 การประเมินลักษณะของการใช้ภาษาหรือกิจกรรมการประเมินเพื่อตรวจสอบความพึงพอใจและความเป็นไปได้ของการทำให้สำเร็จ และอะไรเป็นแหล่งทรัพยากรที่จำเป็นที่จะทำให้สำเร็จ

2.2.2 ประเมินความรู้ของตนเอง องค์ประกอบของภาษาเพื่อดูความสัมพันธ์ในบริบทของความรู้ที่มีอยู่เพื่อความสะดวกความสำเร็จในการใช้ภาษา

2.2.3 ประเมินระดับความสามารถเพื่อการใช้ภาษาที่ประสบความสำเร็จ

2.3 การวางแผน (planning) คือ การตัดสินใจเพื่อจะใช้สิ่งที่มีอยู่อย่างไร

2.3.1 เริ่มต้นเลือกจากความรู้เกี่ยวกับหัวข้อและความรู้ทางภาษาสำหรับกิจกรรมการประเมินให้เสร็จสิ้น

2.3.2 กำหนดหนึ่งหรือหลายๆ แผนเพื่อการดำเนินการในการตอบสนองกิจกรรมการประเมิน

2.3.3 เลือกแผนหนึ่งแผนสำหรับการเริ่มต้นการทำกิจกรรมการประเมิน

จากที่กล่าวมาข้างต้นพบว่านักภาษาศาสตร์ได้เสนอโมเดลและองค์ประกอบของความสามารถทางภาษาไว้ต่างๆ ซึ่งเมื่อพิจารณาแล้วจะเห็นว่าโมเดลความสามารถทางภาษาของ Bachman & Palmer (2010) เป็นโมเดลที่มีความละเอียดและครอบคลุมมากที่สุดทั้งทางความรู้ทางภาษา (language knowledge) และกลยุทธ์ทางภาษา (language strategic) และถูกนำมาเป็นกรอบในการประเมินมากโมเดลหนึ่ง เช่น แบบประเมินความสามารถทางภาษาของ The Cambridge Young Learner English (YLE) (Mckay, 2006) และยังเป็นโมเดลที่ได้มาจากการทำวิจัยและมีการพัฒนาอย่างต่อเนื่อง (Bachman & Palmer, 1981 1990 1996 and 2010)

1.2.2 ระดับความสามารถทางภาษาอังกฤษ

ความสามารถทางภาษามีผู้กำหนดมาตรฐานของระดับความสามารถทางภาษาไว้หลายมาตรฐาน ซึ่งในการศึกษาครั้งนี้ผู้วิจัยจะนำเสนอระดับความสามารถทางภาษาที่เป็นที่ยอมรับระดับสากลและมีการนำมาใช้ในประเทศต่างๆ อย่างแพร่หลาย 3 มาตรฐาน ได้แก่ มาตรฐานระดับสมรรถภาพทางภาษาต่างประเทศในยุโรป (The Common European Framework of Reference for Languages: CEFR) มาตรฐานการเรียนรู้ภาษาต่างประเทศของรัฐอิลลินอยส์ (Illinois Foreign Language Learning Standards) เป็นมาตรฐานที่จัดทำขึ้นโดย Illinois state board of education ซึ่งมาตรฐานนี้จัดทำขึ้นตามความคาดหวังของเนื้อหาและความสามารถของผู้เรียนขั้นต้นตามหลักสูตรของรัฐ ซึ่งเป็นแนวทางให้แก่ครูผู้สอนและผู้ประเมินว่านักเรียนควรจะเรียนอะไรบ้าง และ The Australian

(NLIA) ESL Bandscales เป็นมาตรฐานความสามารถทางภาษาของประเทศออสเตรเลีย โดยสถาบันการรู้ทางภาษาแห่งชาติ (National languages and literacy institute of Australia) เพื่อให้เป็นมาตรฐานสำหรับครูเกี่ยวกับการเรียนการสอนภาษาอังกฤษที่เป็นภาษาที่สอง (ESL) ซึ่งสำหรับการวิจัยครั้งนี้ จะกล่าวถึงเฉพาะมาตรฐานที่จะนำมาใช้ในการวิจัยครั้งนี้ คือ มาตรฐานของ CEFR เท่านั้น ซึ่งมีรายละเอียดดังนี้

มาตรฐานระดับสมรรถภาพทางภาษาต่างประเทศในยุโรป (The Common European Framework of Reference for Languages: CEFR)

สหภาพยุโรปโดย The Council of Europe ได้กำหนดมาตรฐานของสมรรถภาพทางภาษาขั้นหรือที่เรียกว่ามาตรฐานระดับสมรรถภาพทางภาษาต่างประเทศในยุโรป โดยได้จัดพิมพ์เผยแพร่มาตรฐานนี้ขึ้นเพื่อใช้ในกลุ่มประเทศยุโรป เรียกว่า “The Common European Framework of Reference for Languages (CEFR)” เป็นมาตรฐานกลางสำหรับกลุ่มประเทศยุโรปในการจัดทำหลักสูตร ประมวลการสอน และการประเมิน เนื้อหาของ CEFR ให้รายละเอียดเกี่ยวกับผู้เรียนภาษาต้องรู้และเรียนรู้อะไรในการใช้ภาษาเพื่อการสื่อสารได้อย่างมีประสิทธิภาพในบริบทต่างๆ รวมทั้งบริบททางวัฒนธรรม โดยกำหนดมาตร (scale) สมรรถภาพทางภาษาเป็นระดับต่างๆ เพื่อให้เป็นเป้าหมายในการเรียนแก่ผู้เรียนในการพัฒนาตนเอง และเป็นเกณฑ์ในการวัดผลสัมฤทธิ์ทางภาษาของตน ซึ่ง CEFR นี้เป็นมาตรฐานความสามารถทางภาษาอังกฤษที่มีการใช้อย่างกว้างขวาง โดยอธิบายความสามารถทางภาษาในทอมของการพูด การอ่าน การฟัง และการเขียน ซึ่งมีระดับที่ใช้อ้างอิง 6 ระดับ ดังนี้ A1 A2 B1 B2 C1 และ C2 และพิมพ์เผยแพร่โดย The Council of Europe in 2001 (Cambridge ESOL, 2011) มาตรฐานของ CEFR นี้ยังถูกนำมาใช้เพื่อเทียบระดับความสามารถของการทดสอบภาษาอังกฤษของเด็กคือ The Cambridge Young Learner English Test (YLE) ซึ่งแต่ละระดับมีรายละเอียด ตารางที่ 2.1

ตารางที่ 2.1 ระดับของความสามารถตามกรอบ CEFR

ระดับ	ประเภทความเชี่ยวชาญ	ระดับความสามารถ
C2	ความเชี่ยวชาญ (mastery)	ผู้ใช้ภาษาระดับสูง (proficient user)
C1	ความสามารถในการดำเนินการได้อย่างมีประสิทธิภาพ (effective operational proficiency)	
B2	ความเป็นประโยชน์ (vantage)	ผู้ใช้ภาษาระดับอิสระ (independent user)
B1	จุดเริ่มเปลี่ยน (threshold)	
A2	ขั้นตอนและวิธีการ (way stage)	ผู้ใช้ภาษาระดับเบื้องต้น (basic users)
A1	การพัฒนา (breakthrough)	

โดยในแต่ละระดับนั้นจะมีการอธิบายความสามารถทางภาษาแยกตามระดับ ดังตารางที่ 2.2

ตารางที่ 2.2 คำอธิบายความสามารถทางภาษาจากกรอบ CEFR

ระดับความสามารถ	ระดับ CEFR	คำอธิบาย
ผู้ใช้ภาษาระดับเบื้องต้น (basic user)	A1	มีความสามารถทางภาษาที่อยู่ในขอบเขตของความต้องการส่วนตัว และภาษาที่ใช้ก็จะเป็นลักษณะของภาษาที่ท่องจำมาจากที่เรียน หรือภาษาที่เฝ้าการใช้ได้จากสถานการณ์ที่อยู่ต่อหน้าผู้พูด
	A2	เป็นระดับความสามารถทางภาษาที่สามารถสื่อสารได้ในสถานการณ์ปกติธรรมดาทั่วไปในชีวิตประจำวันหรือเป็นการสนทนาในหัวข้อที่ตนเองคุ้นเคย
ผู้ใช้ภาษาระดับอิสระ (independent user)	B1	สามารถใช้ภาษาพูดคุยในหัวข้อที่เกี่ยวกับสถานการณ์ทั่วไปในชีวิตประจำวันได้อย่างมีประสิทธิภาพโดยใช้ภาษาที่ง่ายๆ โดยเฉพาะหัวข้อที่เป็นที่รู้จักดีหรือมีสถานการณ์ประกอบในการสนทนา
	B2	สามารถใช้ภาษาในการสนทนา ให้ข้อมูล และสามารถให้ภาษาสนทนาโต้ตอบอย่างมีประสิทธิภาพรวมถึงการใช้ความรู้และความเข้าใจทางภาษาเพื่อเป็นเครื่องมือในการวางแผนและแก้ไขข้อผิดพลาดของตนเอง
ผู้ใช้ภาษาระดับสูง (proficient user)	C1	สามารถใช้ภาษาในการสนทนาโต้ตอบได้อย่างคล่องแคล่วและรวดเร็วทันต่อเหตุการณ์และภาษาที่ใช้จะนำไปสู่การทำงานที่ซับซ้อนได้สำเร็จลุล่วงได้เป็นอย่างดี
	C2	สามารถใช้ภาษาได้อย่างคล่องแคล่วถูกต้องและเหมาะสมกับสถานการณ์

ซึ่ง CEFR ยังได้อธิบายรายละเอียดความสามารถทางภาษาในแต่ละระดับแบบรวม (global scale) เพื่อเป็นแนวทางในจัดทำหลักสูตร การจัดการเรียนการสอนของครู และการประเมิน (Cambridge ESOL, 2011) ซึ่งความสามารถของนักเรียนในระดับประถมศึกษาปีที่ 6 นั้น คือ ระดับ A1-B1 (Mckay, 2006) แสดงรายละเอียดของระดับความสามารถ และเกณฑ์ ดังตารางที่ 2.3

ตารางที่ 2.3 ระดับความสามารถทางภาษามาตรแบบรวม (global scale) ของ 3 อันดับแรกของ CEFR

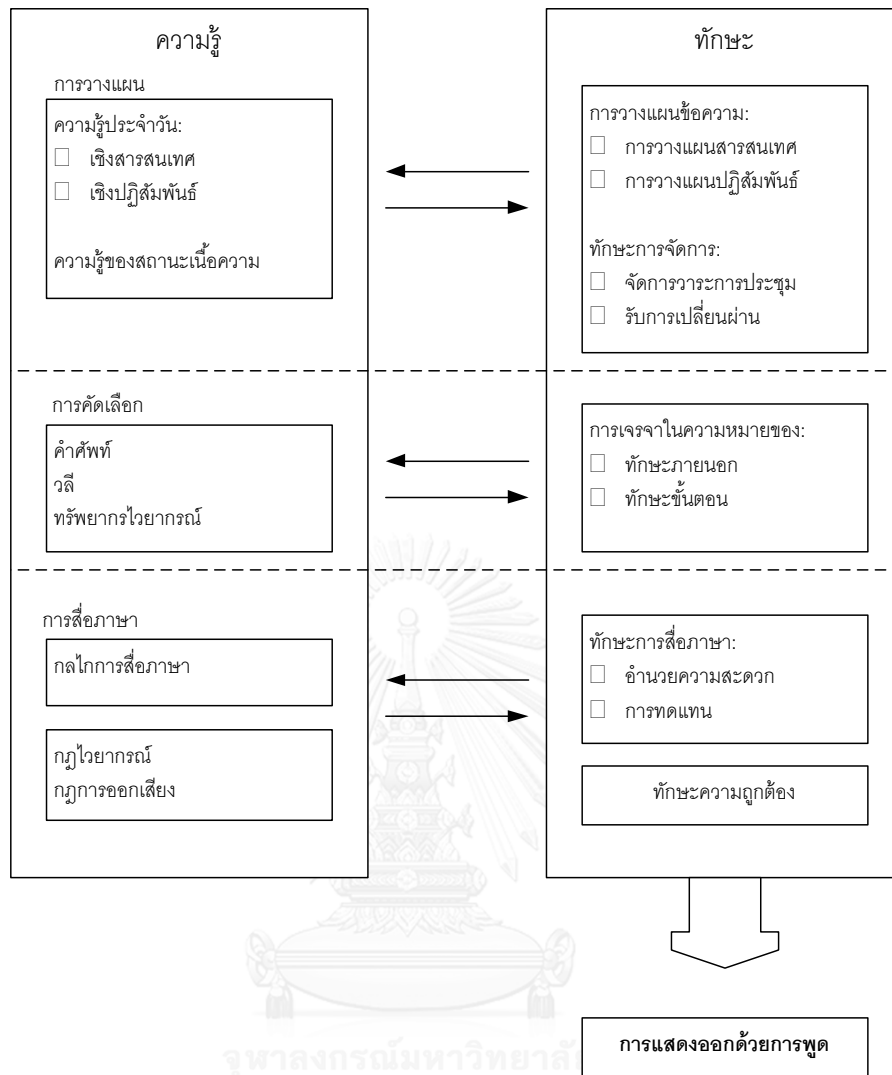
ความสามารถ	ระดับ CEFR	คำอธิบาย
ผู้ใช้ภาษาระดับอิสระ (independent user)	B1	สามารถเข้าใจจุดเน้นของมาตรฐานของสิ่งต่างๆที่คุ้นเคยทุกๆ ไปได้ เช่น งานที่ทำ โรงเรียน การใช้เวลาว่าง เป็นต้น สามารถจัดการกับสถานการณ์ที่มีแนวโน้มว่าจะเกิดขึ้นในการเดินทางในการพูด สามารถสร้างข้อความง่ายๆ โดยการเชื่อมโยงข้อความในหัวข้อที่คุ้นเคยหรือสิ่งที่สนใจ สามารถอธิบายประสบการณ์ และสถานการณ์ ความฝัน ความคาดหวัง หรือความใฝ่ฝัน และความเชื่อ โดยให้เหตุผลและอธิบายตามความคิดเห็นและสามารถอธิบายแผนได้ เช่น แผนการเดินทาง เป็นต้น
ผู้ใช้ภาษาระดับเบื้องต้น (basic user)	A2	สามารถเข้าใจประโยคและสิ่งที่กระทำบ่อยๆ ในชีวิตประจำวันทุกๆ ไป เช่น การสื่อสารกับบุคคลต่างๆ ไป และคนในครอบครัว การซื้อของ สิ่งที่มีอยู่ในท้องถิ่น การทำงาน สามารถสนทนาเรื่องง่ายๆ และภารกิจที่กระทำเป็นประจำแบบง่ายๆ ที่มีการร้องขอ และการแลกเปลี่ยนสารสนเทศโดยตรงที่คุ้นเคยและงานที่ทำเป็นประจำ รวมทั้งอธิบายในแง่มุมง่ายๆ ของภูมิหลัง สิ่งแวดล้อมแบบทันทีทันใด และเรื่องที่ต้องอธิบายในแบบทันทีทันใดได้
	A1	สามารถเข้าใจและใช้รูปแบบภาษาต่างๆ ไปที่คุ้นเคยและประโยคพื้นฐานที่ใช้เพื่อสนองความจำเป็นที่เป็นรูปธรรม สามารถแนะนำตนเอง และผู้อื่นได้ สามารถถามและตอบคำถามเกี่ยวกับรายละเอียดในเรื่องส่วนบุคคล เช่น อยู่ที่ไหน คนที่รู้จักและสิ่งที่ตนมีอยู่ สามารถโต้ตอบแบบธรรมดาได้ถ้าคุ้นเคยพูดซ้ำๆ และชัดๆ และพร้อมที่จะช่วยเหลือ

1.2 การประเมินทักษะการพูดภาษาอังกฤษ

ทักษะการพูดเป็นทักษะที่มีความสำคัญเนื่องจากเป็นช่องทางในการสื่อสารช่องทางหนึ่ง โดยเฉพาะในการเรียนการสอน และเป็นประโยชน์แก่ครูในการประเมินการเรียนรู้ของนักเรียน ซึ่งแนวคิดที่เกี่ยวข้องกับทักษะการพูดเป็นสิ่งสำคัญในการศึกษาภาษาที่สอง ซึ่งนักภาษาศาสตร์ได้กล่าวไว้ว่าการพูดเป็นกระบวนการที่เกิดขึ้นอย่างเป็นระบบ (Bygate, 1987 cited in Luoma, 2004) และการประเมินทักษะการพูดนั้นมีความสัมพันธ์ที่เป็นกระบวนการเช่นกันเนื่องจากมีความเกี่ยวข้องกับปัจจัยในการประเมินอย่างเป็นระบบ ซึ่งมีรายละเอียด ดังนี้

1.2.1 กระบวนการในการพูด (speaking process)

การพูดเป็นทักษะที่มีการแสดงออกของการสื่อสารที่เป็นกระบวนการ ซึ่งถูกนำมาใช้ในการสอนและการประเมินของครู ซึ่งการพัฒนารูปแบบของกระบวนการพูดที่เป็นขั้นตอนช่วยให้ครูสามารถวางแผนการสอนและการประเมินนักเรียนได้ (Luoma, 2004) ซึ่งโมเดลของกระบวนการพูดนั้น Bygate (1987 cited in Luoma, 2004) ได้เสนอโมเดลการพูดที่เป็นกระบวนการไว้ว่าเป็นกระบวนการภายในของผู้พูดประกอบไปด้วย 3 ขั้นตอน ดังนี้ การวางแผน (planning) การคัดเลือก (selection) และการสร้าง (production) ซึ่งทั้ง 3 ส่วนนี้เป็นส่วนประกอบของความรู้ (knowledge) และกลยุทธ์ (strategy) ที่ถูกนิยามเป็นโมเดลองค์ประกอบ ซึ่งคำว่าความรู้ (knowledge) นั้นเป็นสิ่งที่ช่วยให้นักเรียนรู้ว่าจะพูดอะไร ในขณะที่ทักษะ (skill) เป็นส่วนประกอบของกิจกรรมที่ผู้พูดถูกรวมเข้าด้วยกันในปฏิสัมพันธ์นั้นด้วย ซึ่งส่วนประกอบของความรู้ประกอบด้วย **1) การวางแผน (planning)** ในสถานการณ์การพูดเป็นกระบวนการที่นักเรียนจำเป็นต้องรู้ข้อมูลและการปฏิสัมพันธ์ที่ใช้เป็นประจำ ข้อมูลต่างๆ เช่น เรื่องราวต่างๆ การอธิบาย หรือการเปรียบเทียบ ส่วนปฏิสัมพันธ์ เช่น ความแตกต่างของสถานการณ์ต่างๆ เช่น การพูดโทรศัพท์ การสนทนา หรือการเป็นผู้เรียน เป็นต้น **2) การคัดเลือก (selection)** เป็นการคัดเลือกคำศัพท์ วลี รวมทั้งไวยากรณ์ที่จะต้องเลือกเพื่อใช้ในการพูดว่าจะพูดอย่างไร และ **3) การสร้าง (production)** หลังจากการเลือกส่วนประกอบต่างๆ ในขั้นตอนการคัดเลือกเพื่อที่จะพูดเมื่อต้องการจะพูด นักเรียนจะเลือกสิ่งที่ตนจะแสดงหรือจะพูดในสิ่งที่ตนต้องการให้ผู้ฟังได้ยินหรือสิ่งที่ตนต้องการสื่อสารออกไป ซึ่งในส่วนนี้สิ่งที่พูดออกไปนั้นนักเรียนต้องใช้ความรู้ในส่วนของกฎของไวยากรณ์ (grammar rules) และกฎของการออกเสียง (pronunciation rules) ซึ่งทั้ง 3 องค์ประกอบนี้มีความสัมพันธ์กับทักษะ ซึ่งกระบวนการพูดสามารถสรุปได้ดังภาพที่ 2.3

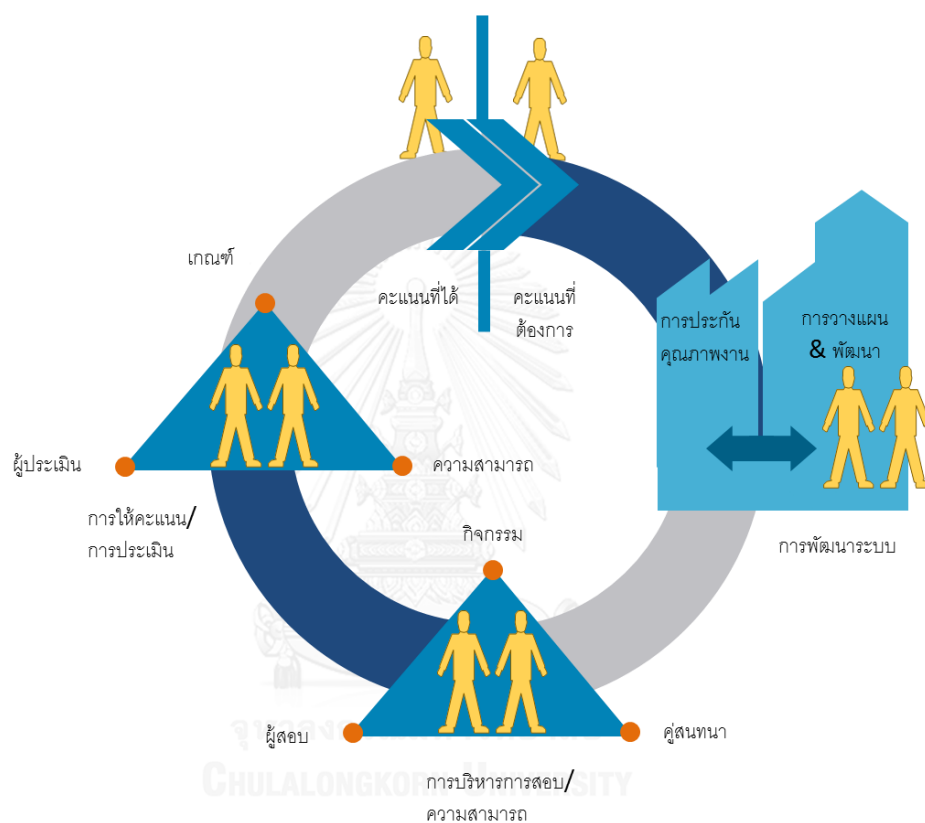


ภาพที่ 2.3 กระบวนการพูดของ Bygate (1987)

ที่มา: Luoma (2004)

1.2.2 กระบวนการประเมินทักษะการพูด

การประเมินทักษะการพูดนั้นเป็นการดำเนินการที่มีกิจกรรมต่างๆ และผู้เกี่ยวข้องหลายส่วน ซึ่งกิจกรรมต่างๆ ที่เกี่ยวข้องกับกระบวนการประเมินทักษะการพูดได้แก่ นักเรียนหรือผู้ถูกประเมิน ผู้ประเมิน แบบประเมินหรือภารกิจที่มอบหมายให้ปฏิบัติ (tasks) เกณฑ์การให้คะแนน คะแนน เกณฑ์คุณภาพในการประเมินทักษะการพูด เป็นต้น ซึ่งทุกสิ่งที่กล่าวมาข้างต้นมีความเกี่ยวข้องสัมพันธ์ซึ่งกันและกัน ดังภาพที่ 2.4



ภาพที่ 2.4 กระบวนการประเมินทักษะการพูด

ที่มา: Luoma (2004)

1.2.3 ขอบเขตการประเมินความสามารถทางภาษาของนักเรียนระดับประถมศึกษา

การทดสอบความสามารถทางภาษาของเด็กนั้นการทดสอบที่ได้รับการยอมรับและมีการใช้อย่างกว้างขวางคือแบบวัดของ The Cambridge Young Learner English Test (YLE) (Cambridge ESOL, 2007; McKay, 2006) โดย YLE แบ่งการทดสอบออกเป็น 3 ระดับ คือ STARTERS (อายุ 7-8 ปี) MOVERS (อายุ 8-11 ปี) และ FLYERS (อายุ 12 ปี) และเทียบระดับความสามารถตามกรอบความสามารถทางภาษาของ The Common European Reference Framework (CERF) โดยมีกรอบโครงสร้างไวยากรณ์ ดังตัวอย่างที่แสดงในตารางที่ 2.4

ตารางที่ 2.4 ตัวอย่างโครงสร้างไวยากรณ์ของแบบวัดความสามารถทางภาษาของเด็กของ Cambridge ESOL (2007)

Language (grammar and structure)	Language use (communication)	Language Items (Example)
The alphabet	writing down spelling	that's W-H-I-T-E
Nouns singular and plural, including limited, specified, irregular plural forms (Proper nouns) (common nouns) possessive forms: /'s/'s	Asking who people are and Identifying people Responding to request for Information about objects Talking about ownership	Are you Bill? It's Pat. They're oranges, not lemons. That's Ann's bike.
Adjectives size, age, colour	describing and identifying objects, people and animals identifying colours	He's a small boy. Your face is very dirty. It's a red car.
Determiners a, an, the, some my, your, his, her, our, their	Identifying objects' animals, fruit, vegetables, etc. (with countables and uncountables) Talking about possessions and relationship	It's a banana. Who's eating an egg? Put the tomato on the table. He's got some apples. It's my brother's birthday.

ซึ่งสำหรับนักเรียนอายุ 12 ปี (Flyer level) มีโครงสร้างไวยากรณ์ (Cambridge ESOL, 2007) ดังนี้

- 1) interrogatives: what (time); what is 'X' like?
- 2) what does 'X' look/ sound/ taste/ feel like?
- 3) past continuous
- 4) present perfect
- 5) future: to be going to
- 6) will, might, may, shall, let's / could/ should
- 7) present simple passive voice
- 8) verbs of sensation (looks like; feels like; sounds like; smells like)
- 9) adverbs of degree: enough; too; a little

- 10) adverb of time: then
- 11) conjunction: so
- 12) cardinal numbers 20 – 1,000
- 13) ordinal numbers up to 31st
- 14) question tag
- 15) adverbial clause with ‘when’
- 16) conditional sentences (present)

1.2.4 กิจกรรมที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษ

ความสามารถทางภาษารวมทั้งทักษะการพูดนั้นเป็นความสามารถในการสื่อสารทั้งรับสารและส่งสาร ซึ่งการที่จะประเมินความสามารถของผู้เรียนได้นั้นครูหรือผู้ประเมินควรวัดความสามารถของนักเรียนโดยให้นักเรียนได้แสดงความสามารถออกมาด้วยการปฏิบัติ เครื่องมือที่ใช้จึงควรเป็นภารกิจต่างๆ (tasks) ที่ให้นักเรียนได้พูด หรือ สนทนา แสดง และกิจกรรมที่ประเมินควรเป็นสิ่งที่สมจริง (authentic) และเป็นชีวิตจริง (real life) ที่มีในประจำวัน ดังนั้นการประเมินควรเป็นการประเมินทักษะบูรณาการ (integrated skill) (Bachman & Palmer, 2010) การประเมินทักษะการพูดนั้นสำหรับภารกิจ หรือกิจกรรม (tasks) ที่ใช้ในการประเมินในระดับประถมศึกษาชั้น Mckay (2006) ได้แบ่งการประเมินออกเป็น 2 ประเภท ดังนี้

1. กิจกรรมที่เป็นการพูดอย่างเดียว ได้แก่ 1) การเล่าเรื่อง (new telling) 2) การเล่านิทาน (storytelling) 3) การเล่าเรื่องจากภาพ (picture talks) 4) การพูดเกี่ยวกับการจัดหมวดหมู่ (categorization tasks) 5) การพูดนำเสนอเรื่องต่างๆ (oral presentation) และ 6) การพูดประเภทอื่นๆ (other speaking-only genres)

2. กิจกรรมที่รวมทักษะการฟังและการพูดเข้าด้วยกัน ได้แก่ 1) การสัมภาษณ์ (interview) และ 2) การถามและตอบคำถาม (question-and-answer tasks)

ส่วนการประเมินทักษะการพูดของ Efthymiou (2012) ที่ทำการประเมินทักษะการพูดกับนักเรียนระดับประถมศึกษาโดยใช้กิจกรรมการประเมิน 4 กิจกรรม ดังนี้ 1) การนำเสนอเกี่ยวกับบุคคลที่เป็นที่รู้จัก 2) การเล่านิทาน และการเล่าเรื่องจากหนังสือ 3) การอธิบายขั้นตอนการปฏิบัติบางสิ่งบางอย่าง และ 4) การถ่ายโอนข้อมูล การอธิบายการจัดวางสิ่งของในตำแหน่งต่างๆ อย่างไรก็ตาม O’ Sullivan (2008) ได้สรุปกิจกรรมต่างๆ ที่ใช้ประเมินทักษะการพูดไว้ 8 กิจกรรม ดังนี้

1. การอ่านออกเสียง (reading aloud) 2. การจำลอง (mimicry) 3. การสนทนาแลกเปลี่ยน (conversational exchanges) 4. การพูดนำเสนอ (Oral presentation) 5. การถ่ายโอนข้อมูล (information transfer) 6. กิจกรรมที่มีการปฏิสัมพันธ์ระหว่างบุคคล (interaction tasks) เช่น

ระหว่างนักเรียนกับนักเรียน นักเรียนกับผู้ประเมิน 7. การสัมภาษณ์ (interview) และ 8. การอภิปราย (discussion) ในประเด็นที่กำหนดให้

สำหรับการประเมินทักษะการพูดในระดับอุดมศึกษาของ Sato (2011) พบว่ามีการใช้กิจกรรมการประเมิน (tasks) จำนวน 3 กิจกรรม ดังนี้ 1) การอธิบายเกี่ยวกับงานอดิเรกที่ชอบ และอธิบายเหตุผลที่ชอบ โดยให้บอกรายละเอียดและยกตัวอย่างประกอบคำอธิบาย 2) การนำเสนอเกี่ยวกับกิจกรรมที่จะทำหากมีเวลาว่างเป็นเวลา 1 เดือน โดยอธิบายและให้เหตุผลประกอบ และ 3) การให้แสดงความคิดเห็นในประเด็นที่กำหนดให้ ซึ่งในการวิจัยครั้งนี้เป็นเรื่องเกี่ยวกับการให้แยกโรงเรียนสำหรับนักเรียนหญิงและนักเรียนชาย

นอกจากนี้แล้ว Educational testing service: ETS (2006) ได้แบ่งแบบสอบหรือกิจกรรม (tasks) ที่ใช้ในการประเมินทักษะการพูดออกเป็น 2 ประเภท ดังนี้

1. กิจกรรมอิสระ (independent tasks) เป็นแบบสอบที่ใช้สื่อประกอบการสอบชนิดเดียว เน้นให้ผู้สอบใช้ประสบการณ์ของตนเองในพูด

2. กิจกรรมบูรณาการ (integrated tasks) เป็นแบบสอบที่ใช้สื่อมากกว่า 1 สื่อ เช่น ให้อ่านข้อความและฟังประกอบกันแล้วจึงให้พูด เป็นต้น

นอกจากนี้ในการนำกิจกรรมการประเมินการพูดมาใช้ควรต้องคำนึงถึงความเหมาะสม จุดเด่น และจุดอ่อนของกิจกรรมการประเมินแต่ละประเภท ซึ่ง O' Sullivan (2008) ได้เสนอรายละเอียด จุดเด่น จุดด้อยของแต่ละกิจกรรมไว้ดังตารางที่ 2.5

ตารางที่ 2.5 รายละเอียดกิจกรรมการประเมินการพูด จุดเด่น และจุดด้อย

ประเภท	คำอธิบาย	จุดเด่น	จุดอ่อน
1. reading aloud	ให้นักเรียนอ่านเอกสารที่กำหนดให้ (อ่านในใจ) แล้วให้นักเรียนอ่านออกเสียงให้ผู้ประเมินฟัง	นักเรียนทุกคนต้องอ่าน เอกสารที่เหมือนกัน ดังนั้น คาดหวังได้ว่าจะได้รับ ระดับความสามารถเดียวกัน ซึ่งง่ายต่อการเปรียบเทียบ ง่ายต่อการควบคุม	- มีความแตกต่างอย่างชัดเจน ต่อความสามารถของผู้ที่เป็นเจ้าของภาษา - มีการแทรกแซงระหว่างทักษะการอ่านและการพูด - ไม่มีความน่าเชื่อถือ (เป็นการประเมินที่ใช้ความรู้สึก) ไม่เป็นที่ยอมรับในตำราหลายๆ เล่ม

ตารางที่ 2.5 (ต่อ)

ประเภท	คำอธิบาย	จุดเด่น	จุดอ่อน
2. mimicry	นักเรียนต้องพูดซ้ำหรือทวนซ้ำตามผู้ประเมินและบันทึกเสียงเก็บไว้เพื่อทำการวิเคราะห์	สามารถดำเนินการในห้องปฏิบัติการทางภาษากับนักเรียนในจำนวนมากได้ในเวลาเดียวกัน นักเรียนได้รับความตั้งต้นที่เท่าเทียมกัน ง่ายต่อการควบคุม จากการวิจัยพบว่ามีความคลาดเคลื่อนประเภทเดียวกับการพูดแบบอิสระ (free talking)	- ยากต่อการแปลความหมายของผลคะแนน - ไม่เป็นธรรมชาติ - ไม่ใช่การสื่อสาร - เป็นการประเมินสิ่งอื่น เช่น ความจำระยะสั้น และการฟัง - ตอบสนองต่อผลการสะท้อนกลับ (backwash effect)
3. conversational exchanges	กำหนดชุดของสถานการณ์ให้แก่ นักเรียน (จากการอ่านหรือได้ยิน) จากที่คาดหวังให้สร้างประโยคโดยใช้รูปแบบที่มีความเฉพาะเจาะจง โมเดลของภาษาที่คาดหวังอาจจะถูกกำหนดหรือไม่ถูกกำหนดให้ในครั้งแรก	- เหมาะสำหรับใช้กับนักเรียนจำนวนมากหรือในห้องปฏิบัติการทางภาษา - ภาษาถูกควบคุม ดังนั้นจึงมีความเป็นไปได้สำหรับการเปรียบเทียบ และมีความเที่ยงสูง - ความตรงเชิงเนื้อหาในการทดสอบมีความสัมพันธ์โดยตรงกับการเรียนในห้องเรียนของนักเรียน	- ไม่มีปฏิสัมพันธ์ตามจริง ดังนั้นการทดสอบจึงไม่มีการสื่อสารเกิดขึ้นได้ - ทักษะการอ่าน หรือการฟัง จะถูกแทรกแซงด้วยความสามารถที่จะตอบสนองต่อสถานการณ์ - ในกรณีที่ดีที่สุดในการทดสอบความสามารถของนักเรียนในการทำซ้ำรูปแบบที่เลือกภายใต้เงื่อนไขที่ถูกจำกัดอย่างสุดโต่ง

ตารางที่ 2.5 (ต่อ)

ประเภท	คำอธิบาย	จุดเด่น	จุดอ่อน
4. oral presentation			
4.1 verbal essay	<ul style="list-style-type: none"> - ให้นักเรียนพูด โดยไม่เตรียมตัว (ส่วนใหญ่จะเป็นการพูดสดกับเทพ) ภายใต้อเวลาที่กำหนด เช่น 3 นาที ในเนื้อหาที่เฉพาะเจาะจง 1 หัวข้อ หรือมากกว่า 1 หัวข้อ - หรือบางครั้งอาจยอมให้มีการนำเสนอโดยเตรียมตัวก่อน (เช่น 30 วินาที หรือ 1 นาที) 	<ul style="list-style-type: none"> - นักเรียนจะต้องพูดต่อ - เกณฑ์การให้คะแนนที่มีความหลากหลายและกว้างขวาง ที่อาจรวมอยู่ในการประเมินใดๆ ก็ตาม เช่น ความคล่องในการพูด ความถูกต้อง เป็นต้น 	<ul style="list-style-type: none"> - หัวข้ออาจไม่น่าสนใจต่อนักเรียน - ไม่เป็นตามสภาพจริงของชีวิตจริง - การนำเสนอทางเลือกโดยเลือกหัวข้อเองจะยากต่อการเปรียบเทียบ - เกิดหัวข้อปลายเปิดมากมาย และขาดเวลาเตรียมพร้อม อาจหมายถึงความสามารถขึ้นอยู่กับความรู้เดิมของนักเรียน (ไม่ใช่ความสามารถทางภาษา) - การบันทึกเทปอาจทำให้นักเรียนเกิดความเครียดได้
4.2 monologue	<ul style="list-style-type: none"> - เหมือนกันกับ verbal essay แต่ให้เวลานักเรียนในการเตรียมตัว 	<ul style="list-style-type: none"> - ง่ายในการจัดเตรียมและบริหารจัดการ - ทำให้เกิดกิจกรรมการประเมินการสนทนา - ปรากฏออกมา 	<ul style="list-style-type: none"> - มีความแตกต่างเล็กน้อยกับเจ้าของภาษา ทำให้นักเรียนไม่น่าเชื่อถือ และไม่มีมาตรฐานของขั้นตอน - นอกจากบทพูดเดี่ยวที่เหมือนกันกำหนดให้นักเรียนทุกคน ผลไม่สามารถเปรียบเทียบได้ - ความรู้หรือความสนใจในหัวข้อจะมีผลต่อความสามารถของนักเรียน - ระยะการเตรียมตัวที่ไม่เพียงพอจะทำให้นักเรียนถูกทดสอบความรู้แต่ไม่ใช่ความสามารถภาษาของนักเรียน

ตารางที่ 2.5 (ต่อ)

ประเภท	คำอธิบาย	จุดเด่น	จุดอ่อน
5. information transfer			
5.1 description of picture sequence	นักเรียนใช้ชุดของรูปภาพและพยายามเล่าเรื่องโดยใช้ tense ที่กำหนดให้ (past) โดยให้เวลาศึกษาภาพเล็กน้อย	<ul style="list-style-type: none"> - เป็นกิจกรรมที่มีความชัดเจน - ถ้าหลีกเลี่ยงความลำเอียงทางวัฒนธรรมหรือการศึกษาจากภาพการวัดจะไม่มีอุปสรรค - เป็นตัวอย่างของการตั้งเรื่องต่างๆ กับการเชื่อมโยงกับการพูด - เป็นการตรวจสอบความสามารถของนักเรียนในการใช้รูปแบบไวยากรณ์ - นักเรียนได้รับสิ่งกระตุ้นที่เหมือนกัน ดังนั้นจึงมีความถูกต้องในการเปรียบเทียบความสามารถ 	<ul style="list-style-type: none"> - มีความจำกัดตามสภาพจริง - บอกความสามารถของนักเรียนได้เพียงเล็กน้อยต่อการปฏิสัมพันธ์ในการพูด - รูปภาพที่มีคุณภาพต่ำมีผลต่อความสามารถของนักเรียน - ความเที่ยงของคะแนนอาจมาจากความแตกต่างในการตีความรูปภาพ
5.2 question on a single picture	ผู้ประเมินถามนักเรียนหลายๆ คำถามเกี่ยวกับเนื้อหาของรูปภาพ โดยให้ศึกษารูปภาพก่อนการทดสอบ	<ul style="list-style-type: none"> - เป็นวัสดุที่มีความเป็นจริง โดยเฉพาะเนื้อหาที่จัดเตรียมให้ตรงกับ ความสนใจของนักเรียน 	<ul style="list-style-type: none"> - นักเรียนสามารถตอบคำถามตามที่ถูกถามเท่านั้น - รูปภาพจะต้องเป็นภาพที่มีความชัดเจนและไม่คลุมเครือ - ถ้านักเรียนมีจำนวนมากจะยากต่อการเปรียบเทียบและและยากต่อการเป็นความลับของการทดสอบ

ตารางที่ 2.5 (ต่อ)

ประเภท	คำอธิบาย	จุดเด่น	จุดอ่อน
5.3 alternative visual stimuli	ใช้วัตถุที่เป็นของจริงแทนรูปภาพเป็นสิ่งกระตุ้นนักเรียน	- เป็นข้อได้เปรียบที่คล้ายกันกับการใช้รูปภาพที่ทำให้ได้สัมผัสกิจกรรมที่เป็นจริงมากขึ้น	- มีจุดอ่อนที่คล้ายกันกับการใช้ภาพ - ความรู้เกี่ยวกับวัสดุในคำถามอาจแทรกแซงการสร้างภาษา
6. interaction tasks			
6.1 information gap: student-student	โดยทั่วไปจะทำเป็นคู่ โดยผู้สอบแต่ละคนได้รับส่วนประกอบทั้งหมดของข้อมูล นักเรียนจะต้องทำงานด้วยกันเพื่อทำงานให้สมบูรณ์	- เมื่อนักเรียนได้รับอิสระในการเลือกคู่จะทำให้เกิดประสิทธิผลในกิจกรรมการสนทนา - สร้างความหลากหลายของเกณฑ์ในการให้คะแนนที่เกิดขึ้น - มีปฏิสัมพันธ์กันสูง	- ผู้สอบอาจถูกครอบงำได้ - ความสามารถที่ต่างกันมาก อาจมีผลต่อความสามารถได้ - ปัญหาในทางปฏิบัติ ได้แก่ ความยากในการบริหารเวลา การบริหารจัดการ และการรักษาความลับของการทดสอบ
6.2 information gap: student-examiner	เหมือนข้อ 6.1 แต่ นักเรียนที่ขาดข้อมูลที่จำเป็นในการทำกิจกรรม และจะต้องขอจากผู้ประเมิน ซึ่งทำหน้าที่เป็นคู่สนทนา	- ผู้สนทนาอาจทำหน้าที่ในวิธีที่คล้ายกับผู้สอบทั้งหมดเพื่อทำการเปรียบเทียบความสามารถที่มีความถูกต้อง	- ทำให้นักเรียนมีความกังวลได้มาก - ผู้ประเมินอาจมีการประเมินความสามารถของนักเรียนนอกเหนือที่กำหนด (ผู้ประเมินอาจไม่ปฏิสัมพันธ์ในลักษณะเดียวกันกับนักเรียนทุกคน)

ตารางที่ 2.5 (ต่อ)

ประเภท	คำอธิบาย	จุดเด่น	จุดอ่อน
6.3 role play (open)	ให้นักเรียนแสดงอย่างใด อย่างหนึ่งจากบทบาทที่มี การปฏิสัมพันธ์ในการใช้ ภาษาจริง อาจเป็นการแสดงระหว่าง นักเรียน-นักเรียน หรือ ผู้ประเมิน-นักเรียน	- มีความตรงเชิงประจักษ์ และความตรงเชิงเนื้อหา ในสถานการณ์ต่างๆ - อาจจะเป็นวิธีที่เชื่อถือ ได้จากการสังเกตและ การวัดความสามารถ ของนักเรียนในการ ดำเนินการตาม สถานการณ์ที่กำหนดให้	- นักเรียนที่แสดงออกมา (histrionic) อาจได้คะแนนสูง กว่านักเรียนที่มีบุคลิกเก็บตัว - ความคุ้นเคยในบทบาทที่ ได้รับอาจมีผลต่อการแสดง ความสามารถ - บางครั้งนักเรียนใช้ภาษา การรายงานแทนการแสดง บทบาทสมมติ - เมื่อทำการทดสอบกับกลุ่ม ใหญ่การแสดงบทบาทสมมติ ที่แตกต่างกันเป็นสาเหตุที่ให้ เกิดปัญหาเกี่ยวกับการ เปรียบเทียบและความลับ ในการทดสอบ
6.4 role play (guided)	ผู้ประเมิน (หรือ อาสาสมัคร) กำหนด บทบาท (บท) ที่แน่นอน ในการแสดง นักเรียนจะ แสดงตามบทบาทที่ กำหนดให้	- ผู้ประเมินมีการควบคุม ที่ดีตลอดทั้งการตั้ง การใช้ภาษาออกมา - สถานการณ์อาจจะถูก ควบคุมเพื่อสะท้อนการ นำเสนอความต้องการ ของการทดสอบหรือ วัตถุประสงค์การ ทดสอบ	- การใช้หัวข้อที่แตกต่างกัน อาจเพิ่มความเป็นกันเอง แต่ผลการประเมินอาจไม่ สามารถนำมาเปรียบเทียบ ได้ - ไม่มีการปฏิสัมพันธ์ที่เป็นจริง /ไม่สามารถขยายหัวเรื่องได้ ดังนั้นจึงไม่ใช้การทดสอบ การสนทนาที่แท้จริง

ตารางที่ 2.5 (ต่อ)

ประเภท	คำอธิบาย	จุดเด่น	จุดอ่อน
7. interview			
7.1 free	ไม่มีการกำหนดขั้นตอนไว้ล่วงหน้า การสนทนาดำเนินไปแบบไม่มีโครงสร้าง	- มีความตรงเชิงประจักษ์และความตรงเชิงเนื้อหาสูง	- ความสามารถที่หลากหลายในหัวข้อที่แตกต่างกันเนื่องจากวิธีการสัมภาษณ์ - ใช้เวลานานและยากในการบริหารจัดการ
7.2 structured	เครื่องมือที่ใช้เพื่อถึงความสามารถออกมาจะเป็นชุดของคำถามและ/หรือกำหนดไว้เพื่อเป็นแนวทางในการสัมภาษณ์	- มีความเป็นไปได้ที่นักเรียนทุกคนจะถูกถามคำถามเดียวกัน ดังนั้นจึงมีความถูกต้องในการเปรียบเทียบมากขึ้น - มีความตรงเชิงประจักษ์และความตรงเชิงเนื้อหาในระดับสูง - มีความเที่ยงของผู้ประเมินทั้งภายในและภายนอกสูง	- มีความจำกัดของช่วงของสถานการณ์ที่ครอบคลุม - ผู้ประเมินอาจไม่ยึดติดกับคำถามที่กำหนดไว้ล่วงหน้าตลอดเวลา
8. discussion			
8.1 student-student	ให้นักเรียนอภิปรายเป็นคู่หรือเป็นกลุ่มในหัวข้อต่างๆ	- มีความตรงเชิงประจักษ์และความตรงเชิงเนื้อหาในระดับสูง - หัวข้อในการสนทนาอาจจะเป็นครูหรือนักเรียนเป็นผู้กำหนด	- นักเรียนที่มีความสามารถอาจครอบงำนักเรียนคนอื่น ดังนั้นจึงต้องระมัดระวังในการจับคู่ - การแทรกแซงอาจเกิดจากการจัดการการสนทนาและทักษะการอภิปรายได้ - ถ้ามีนักเรียนจำนวนมากนักเรียนอาจถูกทดสอบเพียงครั้งเดียว

ตารางที่ 2.5 (ต่อ)

ประเภท	คำอธิบาย	จุดเด่น	จุดอ่อน
8.2 student-examiner	ผู้ประเมินจะกำหนดหัวข้อ (อาจจะกำหนดร่วมกับนักเรียน) เพื่อเป็นแนวทางในการอภิปราย	<ul style="list-style-type: none"> - มีความตรงเชิงประจักษ์ - ในระดับสูง หากนักเรียนมีบทบาทในหัวข้อที่เลือก - รูปแบบอาจสนับสนุนให้ผู้ประเมินขยายการสนทนาในทิศทางที่เลือก 	<ul style="list-style-type: none"> - หัวข้อที่ใช้ในการอภิปรายต้องเป็นหัวข้อเดียวกันจึงจะสามารถเปรียบเทียบคะแนนได้ - แม้ว่าจะใช้หัวข้อเดียวกันก็มีโอกาสน้อยมากที่นักเรียนจะพูดเหมือนกัน - ยากต่อการได้ความเที่ยงในระดับสูงจากผู้ประเมิน

จากการศึกษาเอกสารและงานวิจัยเกี่ยวกับการประเมินทักษะการพูดภาษาอังกฤษดังกล่าวข้างต้น พบว่า ภาระงาน (tasks) ที่นำมาใช้ในการประเมินนักเรียนมีหลากหลายประเภท และแต่ละภาระงานมีจุดเด่นจุดอ่อนแตกต่างกัน ซึ่งสำหรับการวิจัยในครั้งนี้ ผู้วิจัยได้คัดเลือกภาระงานที่จะใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 4 ภาระงาน คือ การสัมภาษณ์ (interview) การนำเสนอปากเปล่า (oral presentation) การเล่านิทาน (storytelling) และการบรรยายภาพ (picture description) เนื่องจากเป็นภาระงานที่เหมาะสมกับความสามารถของนักเรียนในระดับชั้นประถมศึกษาปีที่ 6 เพราะเป็นการพูดในเนื้อหาที่ใกล้ตัว เป็นเรื่องเกี่ยวกับตนเองที่นักเรียนมีความคุ้นเคยตามกรอบความสามารถ ในระดับ A1-B1 (CEFR, 2001) การสัมภาษณ์จะเป็นการสัมภาษณ์แบบมีโครงสร้าง ซึ่งมีจุดเด่นคือ ครูมีกรอบในการถามที่ชัดเจน ทำให้ดึงความสามารถของนักเรียนออกมาได้ตามเป้าหมายการเรียนรู้ มีความเที่ยงในการให้คะแนนที่สูง การนำเสนอปากเปล่าเป็นการพูดในเรื่องราวเกี่ยวกับตนเองซึ่งเป็นการพูดต่อเนื่องกับการสัมภาษณ์ กล่าวคือนักเรียนจะนำเรื่องราวที่ครูสัมภาษณ์มาเรียงร้อยเป็นเรื่องราวด้วยตนเอง ในขณะที่การเล่านิทาน เป็นการเล่าเรื่องที่นักเรียนได้เตรียมตัวมาก่อนทำให้เกิดการฝึกฝนก่อนการประเมิน และครูสามารถควบคุมความครบถ้วนของเนื้อหาในการพูดได้ นักเรียนสามารถเลือกเรื่องที่น่าสนใจอยากพูดได้ และการบรรยายภาพเป็นภาระงานที่นักเรียนจะได้สร้างประโยคและเรื่องราวจากสื่อหรือภาพที่นักเรียนจะต้องค้นคว้าคำศัพท์ สร้างประโยค และเชื่อมโยงเนื้อหาเป็นเรื่องราวให้สอดคล้องกัน

1.2.5 เกณฑ์การให้คะแนน

เกณฑ์การให้คะแนนเกี่ยวกับความสามารถทางภาษาอังกฤษ มีนักภาษาศาสตร์แบ่งไว้ ดังนี้ Weigle (2002) ได้เสนอแนวคิดเกี่ยวกับรูปแบบเกณฑ์การวัดและประเมินผลการเขียนนั้น ประกอบด้วยเกณฑ์ 2 แบบ คือ 1) เกณฑ์การให้คะแนนแบบรวม (holistic rating scales) และ 2) แบบแยกส่วน (analytic rating scale) ในขณะที่ McKay (2006) ได้แบ่งวิธีการให้คะแนนความสามารถทางภาษาในเด็ก (young learner language) เป็น 4 แบบ คือ 1) แบบตรวจสอบรายการ การสังเกต (observation checklist) 2) แบบอิงเกณฑ์ (criteria sheets) 3) เกณฑ์แบบรวม (holistic rating scale) และ 4) เกณฑ์แบบแยกส่วน (analytic rating scale) ซึ่งการศึกษาครั้งผู้วิจัยขอ นำเสนอรายละเอียดเกณฑ์การให้คะแนน 2 แบบ คือ 1) เกณฑ์การให้คะแนนแบบรวม (holistic rating scales) และ 2) แบบแยกส่วน (analytic rating scale) ซึ่งมีรายละเอียดดังนี้

1. เกณฑ์แบบรวม (holistic rating scale)

เกณฑ์แบบรวมเป็นการให้คะแนนโดยอธิบายความสามารถด้วยการบอกคะแนนที่แตกต่างกันในแต่ละระดับ ซึ่งระดับความสามารถนี้เป็นมาตรฐานเดียว ซึ่งแยกออกจากแต่ละช่วง (bands) หรือระดับที่จัดประเภทไว้ (label) เช่น จาก “ต้องปรับปรุง” “ดี” “ดีเยี่ยม” หรือ ระดับ 1 ถึง ระดับ 5 เป็นต้น เกณฑ์แบบรวมสามารถสร้างได้ทั้งจากหลักสูตรหรือทฤษฎีความสามารถทางภาษา (McKay, 2006) แนวทางการให้คะแนนแบบภาพรวมโดยองค์ประกอบสำคัญที่บ่งบอกถึงผลงานที่คาดหวัง มาจัดทำเป็นรายการระบุคำอธิบายที่บรรยายลักษณะของเกณฑ์ในแต่ละรายการเรียงต่อกัน มีระดับคะแนนที่ชัดเจนที่กำหนดเพียงคะแนนเดียวและมีความสะดวกในการตรวจให้คะแนน (Weigle, 2002) ซึ่งการให้คะแนนแบบองค์รวมนี้อัจฉรา วงศ์โสธร (2543) ได้เสนอเกณฑ์การให้คะแนนการพูดแบบอิงเกณฑ์ตามเนื้อหาไว้ 7 ระดับ ดังนี้

- ระดับที่ 7 พูดคล่องและชัดเหมือนเจ้าของภาษา มีความสามารถใช้ภาษาเพื่อวัตถุประสงค์ต่างๆ และในสถานการณ์ต่างๆ เหมือนเจ้าของภาษาพูดทุกอย่าง ความสามารถระดับนี้ มักจะเป็นผู้ที่เคยพักอาศัยในประเทศที่ใช้ภาษาอังกฤษมานาน
- ระดับที่ 6 สามารถพูดสื่อความหมายในทุกสถานการณ์ได้อย่างคล่องแคล่ว ชัดเจน สามารถพูดโต้แย้ง พูดเน้น และพูดขยายความ โดยเปลี่ยนเรื่องทีพูดอยู่ได้ทุกขณะ ผู้ที่ฟังอยู่ก็ฟังด้วยความเข้าใจ และไม่ต้องช่วยผู้พูดในเรื่องการใช้ภาษา ระดับนี้ผู้พูดยังมีการใช้ไวยากรณ์ผิด ใช้คำศัพท์ผิด หรือยังติดใช้เสียงแบบไทยๆ และใช้สำนวนผิด เพราะอิทธิพลภาษาไทยน้อยมากจนแทบสังเกตไม่ได้

- ระดับที่ 5 สามารถพูดสื่อความหมายในสถานการณ์ต่างๆ ได้ สามารถแสดงข้อคิดเห็นเกี่ยวกับเรื่องใดเรื่องหนึ่งได้ ถึงแม้จะยังขาดความคล่องเหมือนเจ้าของภาษาเมื่อพูดเรื่องซับซ้อนลึกซึ้ง มีการใช้ไวยากรณ์ผิดน้อยมาก ที่ผิดมักจะเป็นสำนวนหรือวิธีการพูดมากกว่าผู้สนทนาด้วย มักต้องซักถามเป็นครั้งคราวเพื่อให้ได้ใจความที่กระจ่างขึ้น
- ระดับที่ 4 สามารถที่จะพูดเพื่อสื่อความหมาย โดยให้ผู้ฟังเกิดความสับสนไม่เข้าใจน้อยมาก สามารถเล่าเรื่องราวที่ต่อเนื่องตามลำดับเหตุการณ์ได้ เช่น เล่าเรื่องประวัติบุคคลสำคัญๆ ได้ แต่ยังไม่สามารถพูดต่อเนื่องเพื่อแสดงเหตุผลโต้แย้งและมักจะนิ่งเฉยไม่พูดขณะที่เจ้าของภาษาอภิปรายกัน ยกเว้นเมื่อมีผู้พูดด้วยแบบตัวต่อตัว และมักจะขอใช้ภาษาไทยถ้าเป็นไปได้
- ระดับที่ 3 สามารถพูดเรื่องต่างๆ ไปได้โดยใช้ความรู้ทางไวยากรณ์สร้างประโยคที่ไม่ซับซ้อนได้ เจ้าของภาษาพอฟังเข้าใจได้ ยังมีการใช้ไวยากรณ์อย่างผิดๆ และเสียงที่พูดก็ยังไม่แปร่งอยู่ มีการพูดซ้ำๆ อย่างไม่แน่ใจ และเมื่อตอบคำถามต้องใช้เวลาคิดก่อน เพื่อไม่ให้พูดผิดพลาดจนเกินไป ผู้ฟังมีความลำบากในการฟังให้เข้าใจได้
- ระดับที่ 2 มีความสามารถอ่านออกเสียงได้พอที่จะพูดภาษาอังกฤษเพื่อบอกความต้องการของตนเองได้ โดยใช้การท่องจำ แทบจะไม่มีการเปลี่ยนแปลงคำที่ท่องจำไว้ ไม่สามารถใช้ภาษาเพื่อการสื่อความหมายหรือสร้างสรรค์ได้ มีการพูดติดตะกุกตะกัก มีที่ผิดไวยากรณ์ และการออกเสียงบ่อยครั้ง ต้องพูดไปแก้ไป และเมื่อมีผู้มาพูดด้วยจะมีการตอบที่ซ้ำมาก ผู้ที่พูดด้วยต้องคอยพยายามช่วยๆ ให้พูดต่อไปได้
- ระดับที่ 1 ไม่มีความรู้กฎไวยากรณ์ภาษาอังกฤษ สำเนียงพูดเป็นสำเนียงแบบไทยล้วนๆ สามารถที่จะพูดข้อความบางข้อความ หรือบางคำได้ในวงจำกัด เช่น good morning, goodbye ได้โดยที่เจ้าของภาษาผู้คุ้นเคยกับการใช้ภาษาของต่างชาติพอฟังเข้าใจบ้าง และการใช้ภาษาอังกฤษสั้นๆ นี้จำเป็นต้องใช้ท่าทาง มือไม้ประกอบเพื่อให้ผู้ฟังเข้าใจ

นอกจากนี้แล้ว ศูนย์การสอนภาษาต่างประเทศของสภาสหรัฐอเมริกา (ACTFL: American Council on the Teaching Foreign Language) (Sander, Pardee Lowe, Miles, & Swender, 2000) ได้แบ่งระดับความสามารถในการพูดภาษาต่างประเทศออกเป็น 4 ระดับใหญ่ๆ คือ ระดับความสามารถขั้นต่ำสุด (novice-level) ระดับความสามารถขั้นกลาง (intermediate-level) ระดับความสามารถขั้นสูง (advanced-level) และระดับความสามารถขั้นสูงสุด (superior-level) และ ACTFL ยังได้ทำการแบ่งระดับความสามารถในการพูดออกเป็นระดับย่อยออกไปอีกเป็น 10 ระดับ (American Council on the Teaching of Foreign Languages, 2012) ดังนี้

1. novice low ผู้พูดไม่สามารถใช้ภาษาต่างประเทศนั้นๆ ได้เลย โดยอยู่ในระดับของผู้ที่เริ่มต้นเรียน ผู้พูดอาจสามารถพูดทักทาย บอกชื่อสิ่งของที่คุ้นเคย บอกจำนวนได้บ้าง แต่ไม่สามารถสนทนาในสถานการณ์จริงได้
2. novice mid ผู้พูดสามารถใช้ภาษาเกี่ยวกับชีวิตประจำวันอย่างไม่เป็นทางการได้น้อยมาก ในการพูดจะมีการใช้ภาษาของแม่ร่วมด้วย
3. novice high ผู้พูดสามารถพูดเป็นคำๆ หรือวลีที่ได้ท่องจำมาเท่านั้นเกี่ยวกับชีวิตประจำวัน
4. intermediate low ผู้พูดสามารถใช้ภาษาพื้นฐานต่างๆ ไปในชีวิตประจำวันได้ แต่อยู่ในวงจำกัด
5. intermediate mid ผู้พูดสามารถใช้ภาษาต่างๆ ไปในหัวข้อเรื่องที่คุ้นเคยกับการดำเนินชีวิตประจำวันอย่างไม่เป็นทางการได้และยังมีข้อจำกัดในการใช้ภาษาที่ใช้กันในสังคม
6. intermediate high ผู้พูดสามารถใช้ภาษาต่างๆ ไปในชีวิตประจำวันได้อย่างมีประสิทธิภาพแต่มีข้อจำกัดในการใช้ภาษาที่ใช้กันในสังคม
7. advanced low ผู้พูดที่สามารถใช้ภาษาในสังคมที่ไม่เป็นทางการได้ และอาจสนทนาแบบเป็นทางการได้บ้าง ในหัวข้อที่เคยเรียน ที่บ้าน และกิจกรรมที่ง่ายๆ แต่ยังมีข้อจำกัดในการใช้ภาษาให้ตรงกับความต้องการทางอาชีพ แต่ยังคงอยู่ในข้อจำกัดในการใช้โครงสร้างและคำศัพท์ในสถานการณ์ที่เป็นทางการบ้าง
8. advanced mid ผู้พูดที่สามารถใช้ภาษาในสังคมได้ พูดเกี่ยวกับการทำงานที่ไม่เป็นทางการได้ และการทำงานที่เกี่ยวกับโรงเรียน บ้าน และกิจกรรมที่ง่ายๆ ได้ และความสนใจของตนเองและความสัมพันธ์ส่วนบุคคลได้
9. advanced high ผู้พูดที่สามารถใช้ภาษาในสังคมได้ พูดด้วยความมั่นใจ สามารถอธิบายรายละเอียดต่างๆ ได้อย่างถูกต้องภายในเวลาที่กำหนด อภิปรายประเด็นที่เป็นนามธรรมได้ และตรงกับความต้องการทางอาชีพ แต่ยังคงอยู่ในข้อจำกัดในการใช้โครงสร้างและคำศัพท์ในสถานการณ์ที่เป็นทางการบ้าง
10. superior ผู้พูดสามารถสนทนาโต้ตอบได้อย่างถูกต้องตามโครงสร้างและใช้คำศัพท์ได้อย่างมีประสิทธิภาพ ไม่ว่าจะ เป็นสถานการณ์ที่เป็นทางการหรือไม่เป็นทางการ สามารถโต้ตอบในเรื่องที่มีความสนใจต่างๆ ไป และตรงกับความต้องการทางอาชีพได้

2. เกณฑ์แบบแยกส่วน (analytic rating scale)

เกณฑ์แบบแยกส่วนเป็นแนวทางการให้คะแนนแบบแยกส่วน คือ พิจารณาเป็นประเด็น แต่ละส่วนของงานที่มีลักษณะการตอบจำกัด ซึ่งแต่ละส่วนต้องกำหนดแนวทางการให้คะแนนโดยมี นิยามศัพท์ หรือคำอธิบายของงานแต่ละส่วนไว้อย่างชัดเจน และเกณฑ์ส่วนใหญ่จะไม่เกิน 4 ระดับ เกณฑ์การให้คะแนนแบบแยกส่วนนี้ง่ายต่อการประเมินค่าและมีความน่าเชื่อถือในการประเมิน (Weigle, 2002) ซึ่งตัวอย่างการประเมินที่ใช้การประเมินแบบแยกส่วน เช่น การประเมินความสามารถ ในทักษะการพูดของ CEFR ดังตารางที่ 2.6

ตารางที่ 2.6 เกณฑ์การให้คะแนนทักษะการพูดของ CEFR (2001)

ระดับ	ขอบเขต (range)	ความถูกต้อง (accuracy)	ความคล่องแคล่ว (fluency)	ปฏิสัมพันธ์ (interaction)	การเชื่อมโยง (coherence)
A1	มีพื้นฐานของ องค์ประกอบคำ เบื้องต้น และวลี ง่ายๆ กับข้อมูล ส่วนบุคคล และ สถานการณ์ที่ เฉพาะเจาะจง แน่ชัด	แสดงถึงการ ควบคุมได้อย่าง จำกัดของ โครงสร้าง ไวยากรณ์ง่ายๆ และรูปแบบ ประโยคใน a memorized repertoire	สามารถจัดการ การแยกออก สั้นๆ โดยเฉพาะ การออกเสียง/ การพูดที่มีการ จัดชุดไว้ด้วยการ หยุดหลายครั้ง เพื่อค้นหาการ แสดงออกถึง การเชื่อมต่อของ คำที่มีความ เชื่อมโยงน้อย กว่า และแก้ไข การสื่อสารได้	สามารถถามและ ตอบคำถาม เกี่ยวกับข้อมูล บุคคลได้ สามารถ ปฏิสัมพันธ์ใน วิธีการง่ายๆ แต่การสื่อสารมี ความสัมพันธ์ อย่างเต็มที่กับ repertoire การปรับลีและ การแก้ไข	สามารถเชื่อมโยง คำหรือกลุ่มของคำ ด้วยตัวเชื่อมโยง เชิงเส้น เช่น and หรือ then ได้
A2	ใช้รูปแบบประโยค ง่ายๆ ด้วยวลีที่ จดจำ กลุ่มของคำ 2-3 คำ และ จัดเรียงรูปแบบการ สื่อสารสารสนเทศ อย่างจำกัด ในสถานการณ์ ประจำวัน	ใช้โครงสร้าง ประโยคง่ายๆ บางประโยคได้ อย่างถูกต้อง แต่ยังมีความ ผิดพลาด ในระดับพื้นฐาน อย่างเป็นระบบ	สามารถทำให้ ผู้ฟังเข้าใจในการ ออกเสียงสั้นๆ แม้ต้องมีการ หยุด ชะงัก มี ความผิดพลาด และการจัดเรียง ใหม่ ที่เด่นชัด	สามารถตอบ คำถามและ ตอบสนองต่อ ข้อความง่ายๆ ได้ สามารถชี้ให้เห็น ว่าผู้สนทนา (ผู้ ถูกประเมิน) ปฏิบัติตามแต่ สามารถทำได้ น้อยมาก	สามารถเชื่อมโยง กลุ่มของคำด้วย ตัวเชื่อมต่อต่างๆ เช่น and, but และ because ได้

ตารางที่ 2.6 (ต่อ)

ระดับ	ขอบเขต (range)	ความถูกต้อง (accuracy)	ความคล่องแคล่ว (fluency)	ปฏิสัมพันธ์ (interaction)	การเชื่อมโยง (coherence)
B1	มีภาษาเพียงพอที่จะใช้การได้ด้วย คำศัพท์ที่มากพอต่อการแสดงต่อผู้สนทนา (ผู้ประเมิน) ด้วยความลึกลับ และพูดวาทะ ในหัวข้อที่พูด เช่น ครอบครัว งานอดิเรกและความสนใจ การทำงาน การท่องเที่ยว และสถานการณ์ปัจจุบัน	ใช้ repertoire ง่ายๆ ได้อย่าง สมเหตุสมผลและ ถูกต้องต่อการใช้ เกี่ยวกับกิจวัตร ประจำวัน และ รูปแบบที่ เชื่อมโยงใน สถานการณ์ที่ ทำนายได้มากขึ้น	สามารถ ดำเนินการได้ อย่างคงไว้ซึ่ง ความรอบรู้ แม้จะมี หยุดชะงักต่อ ไวยากรณ์ การ วางแผน วิเคราะห์ศัพท์ และแก้ไขได้ อย่างชัดเจน โดยเฉพาะอย่างยิ่งการขยายออก ให้ยาวขึ้นของ การพูดอย่าง อิสระ	- สามารถเริ่มต้น คงไว้ และปิด การสนทนาแบบ เฉลียวหน้าง่ายๆ ได้ในหัวข้อที่มี ความคุ้นเคย หรือเป็นความ สนใจส่วนบุคคล ได้ - สามารถพูด ทวนซ้ำ ข้อความ บางส่วนเพื่อ ยืนยันความ เข้าใจร่วมกันได้	สามารถเชื่อมโยง ชุดขององค์ประกอบ ที่แยกแยะง่ายๆ สั้นๆ ให้เชื่อมต่อ เป็นจุดๆ ในการ เชื่อมต่อเป็นลำดับ เชิงเส้น

ที่มา: CEFR (2001)

1.2.6 งานวิจัยเกี่ยวกับการประเมินทักษะการพูด

งานวิจัยที่เกี่ยวข้องกับการประเมินทักษะการพูดภาษาอังกฤษที่ทำการศึกษาค้นคว้า มีผู้ทำวิจัยกับกลุ่มตัวอย่างทั้งที่เป็นนักเรียนในระดับประถมศึกษาและระดับอุดมศึกษา ดังนี้

Sawaki (2007) ได้ทำวิจัยเกี่ยวกับการตรวจสอบโครงสร้างของการประเมินความสามารถทางการพูดภาษาอังกฤษที่เป็นภาษาที่สอง โดยการให้คะแนนแบบแยกส่วน (analytic rating scales) และคะแนนรวมของการประเมิน (composite score) กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักศึกษาระดับปริญญาตรีที่เรียนในโปรแกรมการศึกษาต่างประเทศ จำนวน 214 คน เครื่องมือที่ใช้ในการประเมินความสามารถของนักศึกษาใช้การประเมินโดยใช้กิจกรรม (tasks) คือ กิจกรรมการแสดงบทบาทสมมติในการพูด (role-play speaking tasks) โดยกิจกรรมการประเมินนี้มีวัตถุประสงค์เพื่อเป็นการประเมินเพื่อจัดวางตำแหน่งผู้เรียนและเพื่อวินิจฉัยผู้เรียน การให้คะแนนเป็นแบบแยกองค์ประกอบ (analytic rating scales) ซึ่งองค์ประกอบที่ใช้ในการประเมินการพูดประกอบด้วย 5 องค์ประกอบ ดังนี้ 1) การออกเสียง (pronunciation) 2) คำศัพท์ (vocabulary) 3) การเชื่อมโยงหรือความสัมพันธ์

เนื้อหา (cohesion) 4) การเรียบเรียง (organization) และ 5) ไวยากรณ์ (grammar) โดยใช้ผู้ประเมินจำนวน 15 คน ซึ่งทุกคนเป็นนักศึกษาระดับบัณฑิตศึกษาที่เป็นเจ้าของภาษา (native speaker) และเรียนในภาควิชา TESL/Applied linguistics และ Spanish and Portuguese ในมหาวิทยาลัย UCLA ด้วยวิธีการบันทึกเสียงของนักศึกษาด้วยเครื่องบันทึกเสียง และผู้ประเมินทำการประเมินอย่างอิสระจากกัน แล้วจึงนำผลการประเมินมาประชุมร่วมกัน การวิเคราะห์ข้อมูลเพื่อตอบวัตถุประสงค์การวิจัย ได้แก่ การศึกษาองค์ประกอบของการประเมินโดยใช้การวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis) และวิเคราะห์ความน่าเชื่อถือในการวัด (multivariate G-theory) ผลการวิจัยพบว่า องค์ประกอบของการประเมินทั้ง 5 องค์ประกอบมีความสัมพันธ์กันสูง ความน่าเชื่อถือของการวัดสำหรับการตัดสินใจของการทดสอบระดับชาติโดยใช้เกณฑ์การให้คะแนนแบบแยกส่วน และการให้คะแนนรวม พบว่า ความแปรปรวนของคะแนนของการประเมินในโปรแกรมการศึกษาต่างประเทศนั้น องค์ประกอบด้านไวยากรณ์มีความแปรปรวนสูงที่สุด หรือมีความเกี่ยวข้องกับความสามารถทางภาษาอังกฤษของนักศึกษาโปรแกรมต่างประเทศมากกว่าองค์ประกอบอื่นๆ

Westerveld and Moran (2010) ได้ทำการวิจัยเกี่ยวกับทักษะการอธิบายด้วยการพูดของนักเรียนระดับประถมศึกษา (อายุ 6 และ 7 ปี) และเปรียบเทียบความสามารถในการพูดเพื่ออธิบายของนักเรียนที่มีอายุ 6 และ 7 ปี กับนักเรียนอายุ 11 ปี ซึ่งการวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาความสามารถทางภาษาของนักเรียนเกี่ยวกับการสร้างบทสนทนา (produce) กลุ่มตัวอย่างที่ใช้ในการวิจัย ประกอบด้วยนักเรียนระดับประถมศึกษาที่แบ่งออกเป็น 2 กลุ่ม คือ นักเรียนที่มีอายุ 6 และ 7 ปี จำนวน 65 คน ที่เรียนในโรงเรียนของรัฐบาลโดยมีที่ตั้งอยู่รอบนอกของเมืองไอค์แลนด์ และนักเรียนอายุ 11 ปี จำนวน 20 คน ที่มีโรงเรียนตั้งอยู่ในเมืองโครเชริท ประเทศนิวซีแลนด์ ซึ่งเครื่องมือที่ใช้ในการประเมินทักษะการพูดของนักเรียนโดยใช้กิจกรรม The favorite game (FGS) ของ Nippold et al. (2005) ซึ่ง FGS จะให้นักเรียนพูดเกี่ยวกับกีฬาที่นักเรียนชอบ และบอกเหตุผลประกอบด้วยการอธิบายวิธีการเล่นกีฬาชนิดดังกล่าว และบอกวิธีการที่จะเล่นให้ชนะ โดยการประเมินนี้จะไม่มีการกำหนดเวลาในการพูดของนักเรียนหรือนักเรียนจะใช้เวลาในการพูดเท่าใดก็ได้ บทสนทนาของนักเรียนจะถูกประเมินทักษะการพูดของนักเรียนจาก 4 องค์ประกอบ ได้แก่ การพูด (verbal productivity) ความซับซ้อนของประโยค (syntactic complexity) ความถูกต้องทางไวยากรณ์ (grammatical accuracy) และ ความคล่องทางภาษา (verbal fluency) ผู้ประเมินเป็นนักศึกษาระดับอุดมศึกษา จำนวน 3 คน โดยการอัดเสียง และผู้ประเมินจะต้องทำการประเมินนักเรียนคนเดิมตลอดระยะเวลา 4 สัปดาห์ ในการทดลอง และทำการประเมินซ้ำ จำนวน 2 ครั้ง ผลการวิจัยพบว่า นักเรียนกลุ่มอายุมากกว่า (11 ปี) สามารถสร้างบทสนทนาในการพูดได้มากกว่า และมีความคล่องทางภาษา (verbal fluency) ดีกว่า แสดงว่ามีความรู้ทางไวยากรณ์ที่ดีกว่า แต่ค่าเฉลี่ยของความยาว

ในการพูด (mean length of utterance in T-units) และการใช้อนุประโยค (dependent clauses) ไม่ต่างกัน

Sato (2011) ได้ทำการวิจัยเกี่ยวกับการให้คะแนนทักษะการพูดภาษาอังกฤษของนักศึกษา ซึ่งกลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักศึกษาระดับปริญญาตรีในประเทศญี่ปุ่น จำนวน 156 คน ประกอบด้วยนักศึกษาชั้นปีที่ 1, 2 และ 4 (69, 86 และ 1 คน ตามลำดับ) ที่เรียนหลักสูตรภาษาอังกฤษระดับกลาง และระดับสูง ผู้ประเมินที่ทำการประเมินมีจำนวน 9 คน ประกอบด้วยผู้ประเมินที่เป็นเจ้าของภาษา (native speaker) จำนวน 4 คน และไม่ใช่เจ้าของภาษาจำนวน 5 คน แต่มีความสามารถทางภาษาอังกฤษในระดับสูง ได้แก่ IELTS Band 6 EIKEN Grade 1 2 EIKEN Grade Pre-1 TOEIC 920 และ TOEFL iBT 105 ซึ่งผู้ประเมินแต่ละคนจะให้คะแนนนักศึกษาจำนวน 30 คน เครื่องมือที่ใช้ในการประเมินเป็นกิจกรรมที่กระตุ้นให้นักศึกษาได้พูดภาษาอังกฤษ 3 ฉบับ โดยมีเนื้อหาเกี่ยวกับความคิดเห็นเกี่ยวกับบุคคล และเจตคติเกี่ยวกับสถานการณ์ต่างๆ ไป เนื้อหาที่ใช้ประเมินคัดเลือกมาจากแบบสอบของ TOEIC ส่วนการให้คะแนนเป็นการให้คะแนนในภาพรวม และการให้คะแนนแบบแยกส่วน ได้แก่ 1) ประสิทธิภาพการพูดโดยรวม 2) ความถูกต้องของไวยากรณ์ 3) ความคล่องในการพูด 4) ขอบเขตของศัพท์ 5) การออกเสียง และ 6) เนื้อหา โดยองค์ประกอบที่ 2-6 จะให้คะแนนแบบมาตรฐานประเมินค่า 5 อันดับ จาก 1 คะแนน (ไม่เป็นที่น่าพอใจ) ถึง 5 คะแนน (เป็นที่น่าพอใจ) ส่วนด้านที่ 1 ประสิทธิภาพการพูดโดยรวม จะเป็นการให้คะแนนตามความรู้สึกของผู้ประเมิน การวิเคราะห์ข้อมูลประกอบด้วย Rasch measurement เพื่อประเมินความน่าเชื่อถือภายในของผู้ประเมิน และความเป็นอิสระของเกณฑ์การวิเคราะห์ multiple regression เพื่อศึกษาปัจจัยที่ทำนายต่อความสามารถในการพูดของนักศึกษาจากทั้ง 5 องค์ประกอบ และ multivariate G-theory เพื่อศึกษาสถานการณ์ในการตัดสินใจในการประเมินในน้ำหนักรวมของแต่ละองค์ประกอบที่ใช้ในการประเมินไปยังคะแนนรวม และศึกษาแหล่งความคลาดเคลื่อน 2 แหล่ง คือ นักศึกษา (person) และกิจกรรมการประเมิน (tasks) ผลการวิจัย พบว่า ผู้ประเมินมีความเที่ยงในระดับที่สูง .99 ส่วนความเป็นอิสระของเกณฑ์พบว่า ทั้งเกณฑ์แบบแยกส่วนและเกณฑ์แบบรวมมีความเป็นอิสระจากกัน และเกณฑ์ทั้งสองประเภทมีความเป็นเอกมิติ (unidimensionality) และเป็นอิสระ (local independence) ส่วนการศึกษาอำนาจการทำนายประสิทธิภาพการพูดโดยรวมพบว่า องค์ประกอบทั้ง 5 ที่สามารถทำนายประสิทธิภาพการพูดโดยรวมได้อย่างมีนัยสำคัญทางสถิติ โดยสามารถทำนายได้ร้อยละ 67 และการวิเคราะห์ด้วย multivariate G-theory พบว่า เกณฑ์การให้คะแนนแบบแยกส่วนนั้น ความคล่องในการพูด และเนื้อหาในการพูดมีน้ำหนักที่สำคัญมากกว่าองค์ประกอบอื่นๆ

Efthymiou (2012) เป็นการศึกษาเพื่อศึกษาการประเมินทักษะการพูดของนักเรียนระดับประถมศึกษาโดยใช้แฟ้มสะสมงาน ซึ่งกลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้เป็นนักเรียนประถมศึกษาที่กำลังเรียนอยู่ในเกรด 5 ของโรงเรียนรัฐบาล ซึ่งกรอบในการประเมินทักษะการพูดของนักเรียน

โดยการวัดความสามารถทางภาษาอังกฤษของนักเรียนในระดับ A1 และ A2 ตามกรอบของ CEFR (Council of Europe, 2001) ซึ่งมีขอบเขตความสามารถในการพูดในสิ่งที่ใกล้ตัวนักเรียน สิ่งที่คุณเคย และกิจกรรมต่างๆ ที่ทำในชีวิตประจำวัน โดยนักเรียนสามารถสร้างประโยคง่ายๆ ได้ จากหัวข้อที่มีความคุ้นเคยที่กล่าวมาข้างต้น ซึ่งกิจกรรมที่ใช้ในการประเมินการพูดในการวิจัยครั้งนี้ประกอบด้วย 4 กิจกรรม ได้แก่ 1) การอธิบายเกี่ยวกับบุคคล (describing people) 2) การบรรยายหรือการเล่าเรื่อง (narration (story-telling)) 3) การอธิบายขั้นตอน (explaining a procedure) และ 4) การถ่ายทอดข้อมูล (information transfer) ซึ่งกิจกรรมการประเมินการพูดทั้ง 4 กิจกรรมเป็นกิจกรรมที่มีลักษณะเกิดขึ้นจริงในชีวิต ทำการประเมินโดยให้ครูผู้สอนประเมินนักเรียน และนักเรียนประเมินตนเอง การประเมินกิจกรรมที่ 1-3 ใช้การบันทึกเสียงของนักเรียนไว้แล้วครูจึงทำการให้คะแนนภายหลัง ส่วนกิจกรรมสุดท้ายไม่ได้บันทึกเสียง เนื่องจากเป็นกิจกรรมที่ให้นักเรียนสนทนาเป็นคู่ ซึ่งกิจกรรมการประเมินที่ 1-3 นั้นนักเรียนจะได้รับสื่อประกอบการพูด เช่น รูปภาพที่นักเรียนต้องการเพื่อใช้ประกอบการอธิบาย ผลการวิจัยพบว่าการประเมินด้วยกิจกรรมทั้ง 4 กิจกรรม ที่อยู่ในรูปแฟ้มสะสมงาน (portfolios) นั้นครูมีความคิดเห็นว่าสามารถพัฒนาความสามารถของนักเรียนได้จริงและพัฒนา นักเรียนทุกคนได้ เนื่องจากเป็นกิจกรรมที่ทำให้นักเรียนทุกคนมีโอกาสได้ฝึกปฏิบัติจริง การประเมินในลักษณะนี้จากการสังเกตนักเรียนขณะประเมิน พบว่า นักเรียนสามารถปฏิบัติกิจกรรมต่างๆ ได้ทุกกิจกรรม แต่หากมีการบันทึกเสียงด้วยเครื่องบันทึกเสียงจะทำให้นักเรียนบางคนเครียดและวิตกกังวล และการประเมินจากกิจกรรมโดยใช้แฟ้มสะสมงานจะทำให้เกิดผลสะท้อนไปยังการเรียนและการทดสอบหลังการเรียนได้ (washforward effect) แต่การประเมินในลักษณะนี้จะใช้เวลาในการประเมินค่อนข้างมากและเป็นงานที่หนักสำหรับครู

จากการศึกษาวิจัยที่เกี่ยวข้องกับการประเมินทักษะการพูดหรือความสามารถทางการพูด พบว่าการประเมินทักษะการพูดทั้งในระดับระดับอุดมศึกษา และนักเรียนระดับประถมศึกษาจะใช้กิจกรรม (tasks) ในการประเมินเนื่องจากการประเมินที่ต้องการให้ผู้ถูกประเมินได้แสดงความสามารถทางการพูดออกมา และพบว่าเกณฑ์ที่ใช้จะเป็นเกณฑ์แบบแยกส่วน (analytic rating scale) โดยจะใช้ผู้ประเมินที่มีความสามารถและมีความเชี่ยวชาญทางภาษาอังกฤษเป็นผู้ประเมินและมีจำนวนมากกว่า 1 คน ซึ่งเกณฑ์ที่ใช้ในการประเมินจะมีองค์ประกอบคล้ายๆ กัน ได้แก่ การออกเสียง (pronunciation) 2) คำศัพท์ (vocabulary) 3) การเชื่อมโยงหรือความสัมพันธ์เนื้อหา (cohesion) 4) การเรียบเรียง (organization) และ 5) ไวยากรณ์ (grammar) และสำหรับการประเมินของนักเรียนระดับประถมศึกษาจะมีขอบเขตความสามารถอยู่ในระดับ A1 และ A2 ตามกรอบของ CEFR (2001) ซึ่งสามารถสรุปผลการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องได้ ดังตารางที่ 2.7

ตารางที่ 2.7 สรุปผลจากการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องกับการประเมินทักษะการพูดภาษาอังกฤษ

ผู้วิจัย	กลุ่มตัวอย่าง	องค์ประกอบ	ภาระงาน	เกณฑ์	จำนวน	การวิเคราะห์
Sawaki (2007)	นักศึกษา ระดับ ปริญญาตรี	- pronunciation - vocabulary - cohesion - organization - grammar	- role-play speaking tasks	- analytic rating scales	15 คน	- CFA - multivariate G-theory
Westerveld and Moran (2010)	นักเรียน ระดับประถม ศึกษา	- verbal productivity - syntactic complexity - grammatical accuracy - verbal fluency	- The favorite game: FGS (การพูดอธิบาย กีฬาที่นักเรียน ชอบ)	- analytic rating scales	3 คน	- eta squared
Sato (2011)	นักศึกษา ระดับ ปริญญาตรี	- overall communicative effectiveness - grammatical accuracy - fluency - vocabulary range - pronunciation - content elaboration/ development.	- การพูดความ คิดเห็นเกี่ยวกับ บุคคล และเจต คติเกี่ยวกับ สถานการณ์ ต่างๆ ไป	- analytic rating scales	9 คน	- Rasch measurement - multiple regression - multivariate G-theory
Efthymiou (2012)	นักเรียน ระดับประถม ศึกษา		- describing people - narration - explaining a procedure - information transfer	- analytic rating scales	-	- ความถี่ - ร้อยละ

ตอนที่ 2 การประเมินโดยแฟ้มสะสมงาน

แฟ้มสะสมงานเป็นแฟ้มที่ใช้ในการรวบรวมผลงานต่างๆ ที่นักเรียนได้จัดทำขึ้นระหว่างเรียน ซึ่งแฟ้มสะสมงานได้เข้ามามีบทบาทในการเรียนการสอนตั้งแต่ปี ค.ศ.1988 โดยเป็นเครื่องมือชนิดหนึ่งในการประเมินความสามารถที่แท้จริงของนักเรียน (อุทุมพร (ทองอุไทย) จามรมาน, 2540) และมีการดำเนินการที่เป็นระบบ ซึ่งมีรายละเอียด ดังนี้

2.1 ความหมายของแฟ้มสะสมงาน

แฟ้มสะสมงาน (portfolio) เป็นการรวบรวมผลงานนักเรียนจากการเรียนรู้ในชั้นเรียน และนำไปสู่การประเมินการเรียนของนักเรียนตามสภาพจริงวิธีหนึ่ง ซึ่งเป็นวิธีการที่ทำให้ผู้เรียนได้ลงมือปฏิบัติงานที่เป็นจริง และผู้เรียนได้รับข้อมูลป้อนกลับจากครู การสะท้อนคิด ได้ประเมินตนเอง และประเมินเพื่อนในชั้นเรียนด้วยทำให้มีความรับผิดชอบต่อการเรียนของตน ซึ่งแฟ้มสะสมงานมีผู้ให้ความหมายไว้มากมาย ดังนี้

Popham (2012) ให้ความหมายของแฟ้มสะสมงานไว้ว่า เป็นการรวบรวมตัวอย่างงานของแต่ละบุคคล เช่น ศิลปินที่รวบรวมงานศิลปะของตน สถาปนิกใช้รวบรวมภาพถ่ายหรือกล่าวได้ว่าแฟ้มสะสมงานเป็นการรวบรวมผลงานของบุคคลต่างๆ ที่ได้สร้างสรรค์ขึ้นเพื่อแสดงถึงความสามารถของตนในสาขานั้นๆ ดังนั้นการประเมินโดยแฟ้มสะสมงานจึงหมายถึงการประเมินผลงานอย่างมีระบบจากตัวอย่างผลงานนักเรียนที่เกิดจากการรวบรวมมาตลอดระยะเวลาหนึ่งเพื่อแสดงถึงความสามารถของเจ้าของแฟ้ม

Seitz and Bartholomew (2008) กล่าวว่าแฟ้มสะสมงาน คือ การรวบรวมผลงานอันน่ายกย่องของนักเรียน ซึ่งแฟ้มสะสมงานรวมถึงผลงานต่างๆ เอกสาร และการสะท้อนคิด (reflection) ต่อผลงานที่ได้รับการพัฒนาอย่างเหมาะสมของนักเรียน และประกอบด้วยหลักฐาน (evidence) ความเข้าใจและความสามารถของนักเรียน

Nitko (2004) แฟ้มสะสมงาน หมายถึง การรวบรวมงานที่ดีที่สุดของนักเรียนไว้ในแฟ้ม ซึ่งผลงานที่ดีที่สุดนี้เป็นผลงานที่ผ่านการเสนอแนะจากครูผู้สอนจนได้ผลงานที่ดีที่สุด หรือแสดงถึงพัฒนาการในการเรียนของนักเรียน

Weigle (2007 cited in Lam and Lee, 2010) แฟ้มสะสมงานการเขียน เป็นแฟ้มสะสมงานการรวบรวมงานอย่างมีวัตถุประสงค์ และเป็นงานที่แสดงให้เห็นถึงความพยายาม ความก้าวหน้า และผลสัมฤทธิ์ในการเขียนตลอดช่วงระยะเวลาที่เรียน

Trim (1997 cited in Efthymiou, 2012) ให้ความหมายของแฟ้มสะสมงานทางภาษาไว้ว่าเป็นเอกสารที่นักเรียนแต่ละคนรวบรวมไว้ในช่วงเวลาที่เรียน และมีจัดแสดงอย่างเป็นระบบ

มีการบันทึกคุณสมบัติ ผลสัมฤทธิ์ และประสบการณ์ในการเรียนรู้ทางภาษาและรวบรวมตัวอย่าง ผลงานต่างๆ ของนักเรียน (Cole, Ryan, Kick, & Mathies, 2000)

Northwest Evaluation association (1991 cited in Weigle, 2002) แฟ้มสะสมงาน หมายถึง การรวบรวมงานของนักเรียนที่มีจุดมุ่งหมาย ใช้แสดงผลงานของนักเรียนต่อผู้อื่น แสดงถึงความพยายาม ความก้าวหน้า หรือผลสัมฤทธิ์ในงานนั้นๆ

Vavrus (1990 cited in Cole, Ryan, Kick, & Mathies, 2000) ให้ความหมายของ แฟ้มสะสมงานไว้ว่าเป็นการรวบรวมหลักฐานอย่างเป็นระบบ ครูและนักเรียนใช้ในการกำกับติดตาม ความความก้าวหน้า ความรู้ ทักษะและเจตคติของนักเรียน ซึ่งองค์ประกอบที่สำคัญของแฟ้มสะสมงาน จะต้องประกอบด้วย หลักการและเหตุผล (วัตถุประสงค์ของแฟ้มสะสมงาน) เจตนา (เป้าหมายของ แฟ้มสะสมงาน) เนื้อหา (หลักฐานการเรียนรู้) เกณฑ์ (ลักษณะของผลงานที่ยอมรับและไม่ยอมรับ) และการตัดสินใจ (เนื้อหาในแฟ้มบอกอะไรเกี่ยวกับการเรียนรู้ของนักเรียน)

Arter and Spandel (1992 cited in Payne, 2003) ให้ความหมายแฟ้มสะสมงานของ นักเรียนไว้ว่าเป็นการรวบรวมงานอย่างมีจุดมุ่งหมาย แสดงถึงความพยายาม ความก้าวหน้าหรือ สัมฤทธิ์ผลในงานที่เรียนซึ่งจะต้องเป็นงานที่นักเรียนมีส่วนร่วมในการคัดเลือก กำหนดเกณฑ์การเลือก กำหนดเกณฑ์การให้คะแนน และแสดงถึงหลักฐานในการสะท้อนคิด (reflection) ของนักเรียน

อุทุมพร จามรมาน (2540ก) กล่าวว่า แฟ้มสะสมงานเป็นแฟ้มที่สะสมผลงานเพื่อ ชี้ให้เห็นว่าบุคคลได้ทำอะไรบ้าง ทำแล้วมีผลดีมาน้อยอย่างไร ซึ่งแฟ้มสะสมงานที่ได้ผลจะต้องเป็น แฟ้มที่ทำให้เจ้าของแฟ้มมีพัฒนาการทั้งทางร่างกาย จิตใจ สังคม สติปัญญาไปพร้อมกัน และทำให้ เจ้าของแฟ้มมีพัฒนาการในงานที่ทำด้วย

สุวิทย์ มูลคำ (2540) กล่าวว่า แฟ้มสะสมงาน เป็นสิ่งที่เก็บรวบรวมตัวอย่างงาน หรือ หลักฐานที่แสดงถึงผลสัมฤทธิ์ ความพยายาม ความถนัดของบุคคลหรือประเด็นที่ต้องการจัดทำแฟ้ม สะสมงานไว้อย่างเป็นระบบ โดยความรับผิดชอบร่วมกันของเจ้าของแฟ้มสะสมงานและผู้ที่เกี่ยวข้อง

ชมรมศึกษานิเทศก์ (2541) กล่าวว่า แฟ้มสะสมงานเป็นสิ่งบรรจุ รวบรวมผลงาน เอกสารชิ้นงาน ผลงานหลักฐานความสัมฤทธิ์ผลของประสบการณ์การเรียนรู้ของนักเรียนอย่างมี จุดมุ่งหมายและเป็นระบบ มีการสะท้อนภาพความคิด ความรู้สึก ทักษะ และกระบวนการพัฒนา ผลงานที่ผ่านการคัดเลือกตามเกณฑ์ โดยจัดเก็บไว้ในกล่อง หรือสมุด และเป็นเครื่องมือในการประเมิน ตามสภาพจริง

จากเอกสารและงานวิจัยข้างต้นกล่าวสรุปได้ว่า แฟ้มสะสมงาน หมายถึง สิ่งที่ใช้ในการ บรรจุเอกสาร ชิ้นงาน หลักฐานความก้าวหน้าของนักเรียน และเป็นสิ่งที่เก็บรวบรวมผลงานที่มี

จุดมุ่งหมาย ซึ่งสะท้อนให้เห็นถึงความคิด ความรู้สึก และทักษะของนักเรียน และเป็นเครื่องมือการประเมินที่สามารถประเมินได้ตามสภาพจริง

2.2 ลักษณะของแฟ้มสะสมงาน

Bittnger (1996) กล่าวถึงลักษณะที่สำคัญของแฟ้มสะสมงานว่าต้องมีการดำเนินการที่สำคัญใน 3 ประเด็น ได้แก่

1. ตัวอย่างงานนักเรียน (sample of student work) ตัวอย่างงานนักเรียนเป็นสิ่งที่แสดงให้เห็นถึงพัฒนาการของนักเรียนในสิ่งที่เรียนนั้นๆ ตัวอย่างงานสามารถเป็นได้ทั้งตัวอย่างงานเขียน เสียง หรือวิดีโอเทป ปัญหาทางคณิตศาสตร์ รายงานสังคมศึกษา เป็นต้น การรวบรวมชิ้นงานจำเป็นจะต้องมีการวางแผนที่ดี คือต้องวางแผนว่าต้องการหลักฐานอะไรบ้างให้ตรงกับเป้าหมายของการเรียน โดยครูต้องวางแผนร่วมกันกับนักเรียนในการรวบรวมตัวอย่างชิ้นงาน ซึ่งตัวอย่างนี้จะต้องร่วมแสดงความคิดระหว่างครูผู้สอน ครูคนอื่นๆ ผู้ปกครอง ผู้บริหาร

2. การประเมินตนเองของนักเรียน (sample self-assessment) หากไม่มีการประเมินตนเองของนักเรียนและการสะท้อนคิดของนักเรียนในงานของตนเองแล้วจะไม่เรียกว่าเป็นแฟ้มสะสมงาน (Paulson, Paulson, and Meyer, 1991 cited in Bittnger, 1996) การประเมินตนเองนั้นมี 3 ประเภท คือ 1) การตัดสินใจหรือตัดสินใจในการคัดเลือกผลงานตนเองใส่ในแฟ้มสะสมงาน โดยเลือกชิ้นงานที่ดีที่สุด 2) การเปรียบเทียบ นักเรียนเปรียบเทียบผลงานในปัจจุบันกับผลงานที่ผ่านมาเพื่อค้นหาแนวทางในการพัฒนาผลงานของตนให้ดีขึ้น 3) การบูรณาการ นักเรียนใช้แฟ้มสะสมงานเพื่อหาตัวอย่างของความก้าวหน้าในงานเขียนหรือการพูดของนักเรียนเป็นรายคน

3. เกณฑ์การให้คะแนนที่ชัดเจน (clearly stated criteria) นักเรียนจำเป็นต้องรู้ว่าผลงานของตนจะถูกประเมินอย่างไร และใช้มาตรฐานอย่างไรบ้าง เกณฑ์และมาตรฐานที่มีความเฉพาะเจาะจงที่มีการเตรียมไว้สำหรับการตัดสินผลงานนักเรียนจะช่วยให้เด็กกำหนดเป้าหมายและการทำงานของตนได้

Hamp-Lyond and Condon (2000 cited in Weigle, 2002) ได้กล่าวถึงลักษณะที่สำคัญของแฟ้มสะสมงานเขียนไว้ว่ามี 9 ลักษณะ ดังนี้

1. การรวบรวม (collection) แฟ้มสะสมงานคือการรวบรวมงานเขียนหลายชิ้น แทนที่จะเป็นตัวอย่างงานเขียนเพียงชิ้นเดียว

2. ขอบเขต (range) แฟ้มสะสมงานช่วยให้ผู้เขียนสามารถแสดงขอบเขตของความสามารถในการเขียนของผู้อ่านที่แตกต่างกันและวัตถุประสงค์แต่ละประเภท

3. ความสมบูรณ์ด้วยบริบท (context richness) แฟ้มสะสมงานที่มีบริบทที่มากมายจะสะท้อนสถานการณ์การเรียนรู้และแสดงถึงผู้เขียนได้บรรลุความสำเร็จภายในบริบทนั้น

4. การประเมินที่ช้าออกไป (delayed evaluation) คุณลักษณะที่สำคัญของโปรแกรม แฟ้มสะสมงานส่วนใหญ่ถูกประเมินอย่างช้าๆ (delayed evaluation) โดยให้ทั้งโอกาสและแรงจูงใจแก่นักเรียนในการทบทวนงานก่อนที่การประเมินครั้งสุดท้ายจะถูกกำหนด

5. การคัดเลือก (selection) แฟ้มสะสมงานปกติแล้วจะเกี่ยวข้องกับการคัดเลือกชิ้นงานเข้าใส่ในแฟ้มสะสมงาน โดยทั่วไปแล้วนักเรียนจะเป็นผู้คัดเลือกโดยได้รับการแนะนำจากครู

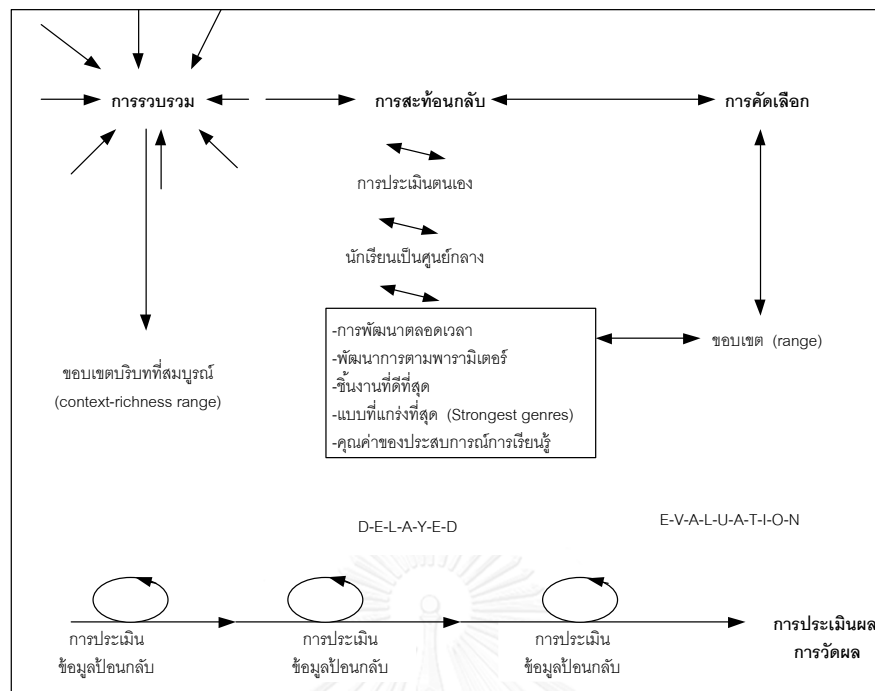
6. การยึดนักเรียนเป็นศูนย์กลาง (student centered control) การประเมินที่ถูกเลื่อนออกไป (delayed evaluation) และเสนอโอกาสให้กับการควบคุมที่มีศูนย์กลางจากนักเรียน ซึ่งนักเรียนสามารถเลือกชิ้นงานที่ดีที่สุดเพื่อเติมเต็มในเงื่อนไขการประเมินที่สร้างขึ้นมา และสามารถทบทวนได้ก่อนจัดเก็บเข้าในแฟ้มงานของเขา

7. การสะท้อนคิดและการประเมินตนเอง (reflection and self-assessment) แฟ้มสะสมงานโดยทั่วไปแล้วเกี่ยวข้องกับการสะท้อนคิด และการประเมินตนเองซึ่งนักเรียนต้องสะท้อนคิดงานของตนเองในการตัดสินใจว่าจะมีการจัดเรียงแฟ้มสะสมงานของตนได้อย่างไร และมักถูกขอให้เขียนบรรยายสะท้อนคิดเกี่ยวกับการพัฒนาในฐานะที่เป็นนักเขียน และการนำเสนอชิ้นงานจะแสดงการพัฒนาได้อย่างไร

8. พัฒนาการ (growth along specific parameters) แฟ้มสะสมงานสามารถทำให้ทราบความหมายของการวัดผลในพัฒนาการตามค่าตัวแปรที่กำหนด (parameters) เช่น ความถูกต้องทางภาษาศาสตร์ หรือความสามารถในการรวบรวมและพัฒนาหลักฐาน

9. การพัฒนาไปตามเวลา (development over time) แฟ้มสะสมงานทำให้ทราบการวัดผลของการพัฒนาไปตามเวลาในแนวทางที่ทั้งครูและนักเรียนจะสามารถมีส่วนร่วมได้

ลักษณะพื้นฐานของแฟ้มสะสมงานทั้ง 9 นี้มีความเกี่ยวข้องกันดังแสดงดังภาพที่ 2.5



ภาพที่ 2.5 ลักษณะทั่วไปของแฟ้มสะสมงาน

ที่มา: Hamp-Lyond and Condon (2000 cited in Weigle, 2002)

Cole et al. (2000) กล่าวถึงเนื้อหาในแฟ้มสะสมงานจะต้องประกอบด้วย 5 องค์ประกอบ ดังนี้

1. ผลงานที่นักเรียนคัดเลือก
2. ความคิดเห็นของนักเรียนที่มีต่อผลงานที่ตนคัดเลือก
3. วัตถุประสงค์มีการระบุชัดเจน
4. ตัวอย่างผลงานที่ทำและผลงานเมื่อทำเสร็จ
5. ตัวอย่างความก้าวหน้า

จากแนวคิดของนักวิชาการที่กล่าวมาข้างต้นสามารถสรุปได้ว่าลักษณะของแฟ้มสะสมงานของนักเรียนประกอบด้วย

1. การกำหนดวัตถุประสงค์ของแฟ้มสะสมงาน
2. ตัวอย่างผลงานที่นักเรียนคัดเลือกเองใส่ในแฟ้มสะสมงาน
3. การประเมินตนเองของนักเรียน
4. การสะท้อนคิดต่อผลงานของตนเอง
5. การประเมินผลของครู
6. เกณฑ์การให้คะแนนที่มีความชัดเจน และมีการกำหนดร่วมกันระหว่างครูผู้สอน

2.3 ประเภทของแฟ้มสะสมงาน

แฟ้มสะสมงานถูกนำมาใช้ในทางการศึกษาอย่างแพร่หลายทั้งระดับประถมศึกษา มัธยมศึกษา และมหาวิทยาลัย นอกจากนี้ยังมีการนำไปใช้กับการประเมินครูด้วย จึงทำให้มีนักการศึกษาได้แบ่งประเภทของแฟ้มสะสมงานออกเป็นหลายประเภทตามลักษณะต่างๆ ดังนี้

อุทุมพร จามรมาน (2540ก) ได้แบ่งแฟ้มสะสมงานที่ใช้ในโรงเรียน ไว้ 2 ประเภท ได้แก่ แฟ้มสะสมงานที่แสดงผลสัมฤทธิ์ทางการเรียน และแฟ้มสะสมงานที่แสดงผลสัมฤทธิ์ทางการเรียนบางวิชา ซึ่งแต่ละประเภทยังมีรายละเอียด ดังนี้

1. แฟ้มสะสมงานที่แสดงผลสัมฤทธิ์ทางการเรียน ซึ่งเป็นแฟ้มที่บรรจุสาระจำนวนมากที่ครอบคลุมงานทั้งปีการศึกษา/ภาคการศึกษา แฟ้มสะสมงานประเภทนี้อาจมี 1 แฟ้มที่บรรจุสาระจำนวนมากหรืออาจมีหลายแฟ้มที่บรรจุสาระที่เกี่ยวข้อง และแฟ้มทั้งหลายดังกล่าวเมื่อรวมกันแล้วจะแสดงผลงานที่คลุมเวลาทั้งหมด

2. แฟ้มสะสมงานที่แสดงผลสัมฤทธิ์ทางการเรียนบางวิชา/หน่วย การสอนบางวิชา/หน่วย แฟ้มดังกล่าวจะบรรจุสาระจำนวนน้อยกว่าแฟ้มที่แสดงผลสัมฤทธิ์ทางการเรียน

ชัยพฤกษ์ เสรีรักษ์ และคณะ (ม.ป.ป.) สุวิทย์ มูลคำ (2540) และ ชมรมศึกษานิเทศก์ (2541) แบ่งได้เป็น 4 ประเภท ซึ่งมีรายละเอียดดังนี้

1. แฟ้มสะสมงานส่วนบุคคล (personal portfolio) เป็นแฟ้มที่แสดงข้อมูลเกี่ยวกับเจ้าของแฟ้มในด้านความสามารถพิเศษหรือพรสวรรค์ ความสนใจในด้านกีฬา ศิลปะ กิจกรรมอาสาพัฒนาชุมชน

2. แฟ้มสะสมงานเชิงวิชาการ (academic portfolio) เป็นแฟ้มที่ใช้แสดงถึงความสามารถเชิงวิชาการด้านใดด้านหนึ่ง แบ่งออกเป็น 2 ประเภท คือ แฟ้มสะสมงานที่แสดงความก้าวหน้า (progress portfolio) และแฟ้มสะสมงานที่แสดงความรอบรู้ (Mastery portfolio) เช่น แฟ้มสะสมงานนักเรียน (student portfolio) แฟ้มสะสมงานการสอน (teaching portfolio)

3. แฟ้มสะสมงานเชิงวิชาชีพ (professional portfolio) เป็นแฟ้มสะสมงานที่แสดงถึงความเชี่ยวชาญในเรื่องใดเรื่องหนึ่งด้านอาชีพ วัตถุประสงค์ที่ทำ เพื่อใช้เป็นหลักฐานประกอบการพิจารณาความดีความชอบ การประชาสัมพันธ์ตัวเอง การสมัครงาน การเลื่อนระดับทางวิชาชีพ

4. แฟ้มสะสมงานของโครงการ (project portfolio) เป็นแฟ้มสะสมงานที่แสดงขั้นตอนการทำงานในโครงการ ซึ่งเป็นความพยายามในการแก้ปัญหาหรือเป็นส่วนหนึ่งของการศึกษาอิสระ มีลักษณะคล้ายภาพยนตร์ สารคดี เช่น โครงการงานวิทยาศาสตร์ โครงการพัฒนาสิ่งแวดล้อมในโรงเรียน เป็นต้น

Cole et al. (2000) ได้แบ่งแฟ้มสะสมงานออกเป็น 2 ประเภท ได้แก่

1. แฟ้มสะสมงานกระบวนการ (process portfolio) เป็นแฟ้มสะสมงานที่ดำเนินการระหว่างปีการศึกษาโดยใช้แฟ้มสะสมงานเพื่อเป็นเครื่องมือวัดความก้าวหน้าทางการเรียนของผู้เรียน ซึ่งจะต้องแสดงผลงานของนักเรียนตั้งแต่เริ่มต้นการเรียน โดยนักเรียนต้องตอบคำถามว่าทำไมนักเรียนจึงเริ่มเรียนในระดับนั้น นักเรียนคาดหวังว่าจะก้าวหน้าไปในทิศทางใด มีการวางแผนที่จะไปในทิศทางนั้นอย่างไร เมื่อไร เมื่อนักเรียนเรียนไปเรื่อยๆ ครูจึงค่อยๆ สะสมหลักฐานที่แสดงถึงความก้าวหน้าในการเรียนของนักเรียนไปเป็นระยะๆ เรื่อยๆ จนถึงขั้นสุดท้ายก็จะเก็บไว้ในแฟ้มสะสมงาน โดยเมื่อนักเรียนเรียนเรื่องใดจบนักเรียนต้องสรุปวิธีการเรียนที่ทำให้ตนเรียนได้สำเร็จ และสิ่งที่นักเรียนจำเป็นต้องทำในอนาคต รวมทั้งงานที่ไม่ประสบความสำเร็จนักเรียนก็สามารถเก็บในแฟ้มได้ และระบุถึงปัญหาที่พบ จากนั้นพิจารณาว่าเหตุใดเรื่องนั้นจึงเป็นปัญหา และสามารถจัดการกับปัญหานั้นได้อย่างไร ซึ่งแฟ้มสะสมงานกระบวนการต้องใช้แรงงานและเวลามาก

2. แฟ้มสะสมงานผลผลิต (product portfolio) เป็นแฟ้มสะสมงานที่แสดงผลงานของนักเรียนตอนปลายปีการศึกษา หรือจบโปรแกรม ซึ่งแฟ้มสะสมงานผลผลิตนี้เป็นการแสดงถึงความสำเร็จของนักเรียนในการเรียนตามหลักสูตร ซึ่งครูควรจะสรุปความคิดเห็นของครูที่มีต่อกระบวนการเรียนรู้ของนักเรียนด้วย เพื่อช่วยนักเรียนในการเรียนระดับสูงต่อไป

2.4 องค์ประกอบของแฟ้มสะสมงาน

Burke, Fogart and Begrad (1994) กำหนดองค์ประกอบของแฟ้มสะสมงานไว้ 3 ส่วน ดังนี้

ส่วนที่ 1 เป็นส่วนนำ ประกอบด้วย ปก ประวัติผู้ทำ รายการจุดประสงค์การเรียนรู้ แผนการศึกษาส่วนบุคคล สารบัญชิ้นงาน ตัวชี้บ่งประกอบงาน รายการเอาออกและนำเข้าชิ้นงาน

ส่วนที่ 2 เป็นที่เก็บชิ้นงานที่สร้างขึ้นจากการเรียนการสอน การสะท้อนความคิดเห็น แสดงประวัติของงาน จำนวนหนังสือที่อ่าน เวลาที่ใช้ทำงาน คะแนนจากการสอบ แบบสำรวจรายการของครู บันทึกความคิดเห็นเกี่ยวกับงานหรือวิชาที่เรียน

ส่วนที่ 3 เป็นที่เก็บเกณฑ์การตัดสินแฟ้มสะสมงาน และข้อมูลการประเมินของครู เพื่อน และผู้ปกครอง รวมทั้งหลักฐานการประเมินตนเองของนักเรียน แผนการและแนวคิดในการประชุมแฟ้มผลงาน

Council of Europe (2004) ได้พัฒนา European language portfolio (ELP) ขึ้นเพื่อใช้ในกลุ่มประเทศยุโรป โดย ELP ได้เริ่มดำเนินการอย่างเป็นทางการตั้งปี ค.ศ. 2001 เพื่อสนับสนุนการเรียนการสอนแก่ผู้เรียนในกลุ่มประเทศยุโรปและเป็นเครื่องมือประกอบการตัดสินการเรียนรู้

ของ Common European Framework of Reference for Language (CFR) มีการศึกษานำร่องใน 15 ประเทศในยุโรป โดย ELP ประกอบด้วย 3 ส่วน ได้แก่

1. The language passport เป็นส่วนที่อธิบายสรุปรวมความสามารถทางภาษาของนักเรียน โดยอธิบายในเทอมของทักษะและระดับความสามารถทางภาษาตาม CFR ที่มี 6 ระดับ ตั้งแต่ A1-C2 การบันทึกจะบันทึกทั้งคุณวุฒิ สมรรถนะ และลักษณะที่สำคัญทางภาษาของนักเรียน ที่เป็นผลการประเมินตนเอง การประเมินจากครู และการประเมินจากสถาบันการศึกษา คณะกรรมการทดสอบ

2. The language biography เป็นส่วนที่สนับสนุนการเข้าร่วมของนักเรียนในการวางแผนการสะท้อนคิด (reflection) และการประเมินกระบวนการเรียนรู้และความก้าวหน้าของนักเรียน กระตุ้นให้นักเรียนได้ทราบถึงสิ่งที่ตนสามารถทำได้ในแต่ละภาษา ทั้งการศึกษาในระบบและการศึกษานอกระบบ

3. The dossier เป็นส่วนที่ช่วยให้นักเรียนเลือกสื่อการเรียนเพื่อแสดงถึงผลสัมฤทธิ์หรือประสบการณ์การเรียนรู้ที่ได้บันทึกไว้ในส่วนของ The language passport และ The language biography

Cole et.al. (2000) และชัยพลกฤษ์ เสรีรักษ์ และคณะ (ม.ป.ป.) กล่าวว่าเพิ่มสะสมงานควรมีองค์ประกอบ ดังนี้

1. จุดมุ่งหมาย คือ วัตถุประสงค์ของเพิ่มสะสมงาน ที่ใช้เพื่ออธิบายว่าต้องการวัดอะไร ซึ่งถ้ามีจุดมุ่งหมายที่ชัดเจนจะเป็นแนวทางที่สำคัญที่แสดงว่านักเรียนประสบความสำเร็จอย่างไร

2. การกำหนดเวลา เป็นการวางแผนว่าจะให้นักเรียนประเมินเพิ่มสะสมงานเมื่อไร และเป็นไปตามจุดมุ่งหมายที่วางไว้หรือไม่ ซึ่งมีกระบวนการดังนี้

2.1 การประเมินตนเอง คือการให้นักเรียนประเมินเพิ่มสะสมงานของตนเองว่าเป็นไปตามจุดมุ่งหมายหรือไม่

2.2 การสะท้อนคิด คือการให้นักเรียนได้แสดงความรู้สึก และการให้คะแนนผลงาน เขียนแสดงความคิดเห็นต่อการเรียนและผลงานของตนเองว่าจุดเด่นและจุดด้อยอย่างไร ซึ่งจะทำให้ นักเรียนพัฒนาความสามารถในการเรียนได้ อันเป็นการให้นักเรียนได้ใช้ความคิดขั้นสูง

3. เนื้อหา ประกอบด้วยหลักฐานการเรียนรู้ ได้แก่ ผลงานที่นักเรียนเลือก ความคิดเห็นของนักเรียนที่มีต่อผลงานที่เลือกหรือการสะท้อนคิด วัตถุประสงค์ต้องมีความชัดเจน ตัวอย่างผลงานระหว่างที่ทำและผลงานสุดท้ายเมื่อทำแล้ว และตัวอย่างหลักฐานที่แสดงถึงความก้าวหน้า

4. เกณฑ์ คือ ลักษณะของผลงานที่เป็นที่ยอมรับ และผลงานที่ไม่เป็นที่ยอมรับ

5. การตัดสิน คือ เนื้อหาในเพิ่มสะสมงานบอกอะไรเกี่ยวกับการเรียนรู้ของนักเรียน

อุทุมพร จามรมาน (2540ข) ได้เสนอแนะเกี่ยวกับองค์ประกอบของแฟ้มสะสมงานไว้ดังนี้

1. สารบัญเนื้อหา สารระโนแฟ้ม
2. วัตถุประสงค์ของการนำเสนอเนื้อหาสารระโนแฟ้ม
3. สารระต่างๆ (อาจอยู่ในรูปงานเขียน ผลิตผล ชิ้นงาน)
4. ผลการวัดประเมินสารระต่างๆ และแหล่งการวัดประเมิน พร้อมหลักฐานการวัดประเมิน

เป็นระยะๆ

5. ภาคผนวก แหล่งอ้างอิง

กรมวิชาการ (2542) ได้เสนอส่วนประกอบของแฟ้มสะสมงานนักเรียนไว้ 3 ส่วน คือ

ส่วนที่ 1 ส่วนนำ ประกอบด้วย ปก ประวัติผู้ทำ รายการจุดประสงค์การเรียนรู้ แผนการศึกษาส่วนบุคคล สารบัญชิ้นงาน ตัวชี้บ่งประกอบงาน รายการเอาออกและนำเข้าชิ้นงาน

ส่วนที่ 2 ส่วนบรรจุหลักฐานชิ้นงานที่สร้างขึ้นจากการเรียนการสอน การสะท้อนคิด แสดงประวัติของงาน จำนวนหนังสือที่อ่าน เวลาที่ใช้ทำงาน คะแนนจากการทดสอบ แบบสำรวจ รายการครู บันทึกความคิดเห็นเกี่ยวกับงาน/วิชาของนักเรียน

ส่วนที่ 3 ส่วนบรรจุเกณฑ์การตัดสินแฟ้มสะสมงาน และข้อมูลการประเมินของครู เพื่อน และผู้ปกครอง รวมทั้งหลักฐานการประเมินตนเองของนักเรียน แผนการและแนวคิดในการประชุม แฟ้มสะสมงาน

นอกจากนี้แล้ว ชัยฤทธิ์ ศิลาเดช (2540) ได้เสนอส่วนประกอบของแฟ้มสะสมงานไว้ 4 ส่วน ดังนี้

1. ส่วนนำ ประกอบด้วย
 - 1.1 ข้อมูลนักเรียน
 - 1.2 สารบัญ
 - 1.3 จุดประสงค์การเรียนรู้ ข้อตกลง เกณฑ์การประเมินผลงาน
2. ส่วนเนื้อหา ประกอบด้วย ผลงานหรือหลักฐานต่างๆ ที่ผู้เรียนคัดเลือก
3. ส่วนสนับสนุน ประกอบด้วย
 - 3.1 แผนพัฒนาและปรับปรุงการเรียนรู้
 - 3.2 สรุปผลความก้าวหน้าในการเรียนเป็นรายเดือน
 - 3.3 ผลการสอบ ผลการสังเกต และประกาศนียบัตร
4. สรุปผล

2.5 การพัฒนาและการดำเนินการของแฟ้มสะสมงาน

การพัฒนาแฟ้มสะสมงานและขั้นตอนในการดำเนินการของแฟ้มสะสมงานเป็นการดำเนินการที่มีความต่อเนื่องและมีความสำคัญที่จะทำให้ครูและนักเรียนบรรลุเป้าหมายของแฟ้มสะสมงาน ซึ่งการพัฒนาและการดำเนินการของแฟ้มสะสมงานทางภาษาและแฟ้มสะสมงานของนักเรียน มีนักวิชาการเสนอไว้ดังนี้

European language portfolio: ELP การดำเนินการของแฟ้มสะสมงานทางภาษาของ ELP มีจุดเน้นใน 2 ประเด็นคือ การประเมินตนเอง (self-assessment) และการสะท้อนคิด (reflection) (Yilmaz & Akcan, 2011) ซึ่งในการดำเนินการมีขั้นตอนที่สำคัญ 5 ขั้นตอน ดังนี้

1. ให้ความสำคัญในการตระหนักรู้ (raising awareness) ในขั้นแรกผู้ประเมิน ครู หรือ ผู้เกี่ยวข้องต้องให้นักเรียนตระหนักรู้หรือเข้าใจเกี่ยวกับเป้าหมายในการเรียนและความสามารถที่คาดว่าจะได้รับจากการเรียน โดยการเขียนบรรยายให้นักเรียนได้รับทราบทั่วกัน ครูต้องเริ่มต้นด้วยการบรรยายในส่วนของ biography ในแฟ้มสะสมงานของนักเรียน (ELP)

2. การเกาะติดกับเป้าหมาย (goal tracking) การเกาะติดหรือยึดติดกับเป้าหมายการเรียน รวมถึงการวางแผนกลยุทธ์ที่จะช่วยให้นักเรียนบรรลุตามเป้าหมายที่ตั้งไว้ ซึ่งใน ELP นักเรียนไม่ได้เลือกเป้าหมายเอง แต่ครูจะเป็นผู้กำหนดให้

3. การสร้างทางเลือก (making choices) ใน ELP นักเรียนจะมีอิสระในการวางแผน และทำการงานให้สำเร็จ ซึ่งหมายความว่านักเรียนมีอิสระในการเลือกเนื้อหาในการเรียนภาษา ทั้งภาระงานการพูดและการเขียน ในกรณีการเขียนนักเรียนจะเลือกสถานที่ในการเขียนที่ตนสนใจเอง เป็นต้น

4. การสะท้อนคิด (reflection) เป็นการสะท้อนคิดเมื่อนักเรียนได้ประเมินตนเองว่าสิ่งใด ตนทำได้ และทำไม่ได้ และสะท้อนคิดว่่านักเรียนรู้สึกอย่างไรกับความสามารถในการปฏิบัติของตน โดยนักเรียนจะเขียนบรรยายและใช้แบบตรวจสอบรายงานของ ELP (ELP checklists)

5. การประเมินตนเอง (self-assessment) การประเมินตนเองจะทำในส่วนของ biography โดยทำการประเมินทั้งระหว่างเรียนและหลังเรียนเสร็จแล้ว ขั้นตอนการประเมินคือเริ่มจากนักเรียนประเมินตนเองก่อน และครูจึงจะประเมินนักเรียน นักเรียนจะเปรียบเทียบผลการประเมินระหว่างการประเมินตนเองและครูประเมิน

Barton and Collins (1997) ได้เสนอวิธีการออกแบบและการพัฒนาการจัดทำแฟ้มสะสมงานไว้ 3 ขั้นตอน ได้แก่ การกำหนดวัตถุประสงค์ การรวบรวมหลักฐาน และการกำหนดเกณฑ์การให้คะแนน ซึ่งมีรายละเอียด ดังนี้

1. การกำหนดวัตถุประสงค์ (purpose) เป็นการกำหนดว่าแฟ้มสะสมงานจะใช้เพื่ออธิบายความสามารถใดของนักเรียน ซึ่งจุดมุ่งหมายควรมีความชัดเจนเพื่อแสดงให้เห็นถึงความสำคัญที่แท้จริงของนักเรียน โดยวัตถุประสงค์ของแฟ้มสะสมงานเพื่อให้การเรียนการสอนบรรลุตามเป้าหมายของหลักสูตร เพื่อให้นักเรียนสามารถประเมินตนเองได้จากการสะท้อนคิดงานของตนเองและการพัฒนา และเพื่อประเมินความรับผิดชอบของครูจากความสามารถในการสอนให้นักเรียนบรรลุตามเป้าหมายของหลักสูตร

2. การรวบรวมหลักฐาน (evidence) เป็นหลักฐานต่างๆ ตามความเป็นจริงที่สะท้อนให้เห็นถึงความสำเร็จตามจุดมุ่งหมายที่วางไว้ ซึ่งหลักฐานสำหรับการประเมินของครูประกอบด้วย ผลงานที่นักเรียนสร้างขึ้น (artifacts) งานเสริมภายนอก (reproduction) การรับรองความถูกต้อง (attestation) และ ผลงาน (productions)

2.1 ผลงานที่นักเรียนสร้างขึ้น (artifacts) ได้แก่ เอกสารต่างๆ เช่น การบ้าน รายงานของนักเรียนที่สร้างขึ้นในชั้นเรียน

2.2 งานเสริมภายนอก (reproduction) ได้แก่งานต่างๆ ที่นักเรียนได้ทำนอกห้องเรียน เช่น โครงการพิเศษ การสัมภาษณ์กับสมาชิกครอบครัว

2.3 การรับรองความถูกต้อง (attestation) เป็นเอกสารที่ครูผู้สอนและผู้เกี่ยวข้องสร้างขึ้นเพื่อบันทึกความก้าวหน้าของนักเรียน เช่น การบันทึกการสังเกตนักเรียนในระหว่างที่นักเรียนนำเสนอผลงาน

2.4 ผลงาน (productions) เป็นเอกสารต่างๆ ที่นักเรียนทำและจัดเตรียมเพื่อจะใส่ในแฟ้มสะสมงาน ผลงานประกอบด้วย 3 ส่วน ได้แก่ 1) การบรรยายเป้าหมาย (goal statements) 2) การสะท้อนคิดของนักเรียน (reflections) และ 3) การรับรองคุณค่า (captions) ของผลงาน

โดยนักเรียนควรรวบรวมหลักฐานไม่น้อยเกินไปเพราะจะทำให้ไม่มีประสิทธิภาพเพียงพอในการประเมินผล แต่ก็ไม่มีความจำเป็นที่จะต้องเก็บทุกชิ้นเพราะมากเกินไป Haertel (1991 cited in Barton and Collins, 1997) ได้ให้แนวทางในการเพิ่มคุณค่า (The value-added principle) ไว้ดังนี้

1. การเลือกชิ้นงานที่ดีที่สุดที่เป็นหลักฐานของการบรรลุตามวัตถุประสงค์ของแฟ้มสะสมงาน

2. การเลือกชิ้นงานที่เป็นหลักฐานที่ดีอันดับที่สอง พร้อมตอบคำถามว่าหากรวบรวมผลงานชิ้นนี้ไว้จะทำให้เกิดคุณค่าอะไรเพิ่มขึ้นบ้าง หากประเมินแล้วพบว่ามีความสำคัญก็เก็บไว้ หากไม่มีความสำคัญก็ไม่ควรเก็บ ให้ปฏิบัติเช่นนี้ทุกชิ้นงาน จนกว่าจะเห็นว่าหลักฐานเพียงพอแล้ว

3. การกำหนดเกณฑ์การประเมิน (assessment criteria) การประเมินแฟ้มสะสมงานนั้น สิ่งที่ต้องตัดสินใจคือการให้คะแนนแฟ้มสะสมงานโดยต้องคำนึงถึงความตรง (validity) และความเที่ยง

(reliability) ในประเด็นเรื่องความตรงและความเที่ยงนี้ Elbow (1991 cited in Barton & Collins (1997) กล่าวว่า การประเมินผลด้วยแฟ้มสะสมผลงานมีความตรงสูงอย่างยิ่ง (external valid) เมื่อเทียบกับการวัดผลในปัจจุบัน โดยให้เหตุผลว่าการประเมินด้วยแฟ้มสะสมผลงานได้วัดตัวแปรต่างๆ ที่ซับซ้อน (complex variables) อย่างละเอียดรอบด้าน จึงส่งผลให้เห็นความสามารถที่แท้จริงของนักเรียนอย่างมีความแม่นยำ (accurately) และการวัดตัวแปรที่มีความซับซ้อนนี้เอง ทำให้ความเที่ยง (reliability) ในการลงความเห็นในการตัดสินให้คะแนนของครูหรือผู้ให้เกรด (graders) แต่ละคนค่อนข้างต่ำ

Burke, Fogart and Begrad (1994) ได้กำหนดแนวทางในการดำเนินการของแฟ้มสะสมงานไว้ 3 แนวทาง ดังนี้

แนวทางที่ 1 ประกอบด้วยกระบวนการหลัก 3 ขั้นตอน ได้แก่ 1) การเก็บรวบรวมและจัดระบบผลงาน 2) การคัดเลือกผลงาน และ 3) การแสดงความคิดเห็นหรือแสดงความรู้สึกต่อผลงาน

แนวทางที่ 2 มีการขยายขั้นตอนเพิ่มจากแนวทางที่ 1 เป็น 6 ขั้นตอน ได้แก่ 1) การกำหนดวัตถุประสงค์และประเภทของแฟ้มสะสมงาน 2) การเก็บรวบรวมและจัดระบบผลงาน 3) การคัดเลือกผลงาน 4) การแสดงความคิดเห็นหรือแสดงความรู้สึกต่อผลงาน 5) การประเมินค่าผลงาน และ 6) การสร้างความสัมพันธ์

แนวทางที่ 3 ประกอบด้วยกระบวนการที่สมบูรณ์ 10 ขั้นตอน ซึ่งมีรายละเอียดในแต่ละขั้นตอน ดังนี้

Burke, Fogart and Begrad (1994) สุวิทย์ มูลคำ (2540) และชัยพฤกษ์ เสรีรักษ์ และคณะ (ม.ป.ป.) ได้เสนอขั้นตอนการพัฒนาและดำเนินการจัดทำแฟ้มสะสมงานไว้ 10 ขั้นตอน ดังนี้

1. ขั้นกำหนดวัตถุประสงค์และประเภทของแฟ้มสะสมงาน (project purpose) เป็นขั้นตอนที่ครูผู้สอนร่วมกับนักเรียนในการกำหนดข้อตกลงร่วมกันถึงวัตถุประสงค์ของแฟ้มสะสมงาน และการตัดสินใจเลือกประเภทของแฟ้มสะสมงานซึ่งต้องสอดคล้องกับวัตถุประสงค์ของการเรียนรู้ ซึ่งต้องยึดหลักแห่งการเรียนรู้ กระบวนการเรียนรู้ และความแตกต่างระหว่างบุคคลในการกำหนดวัตถุประสงค์ของแฟ้มสะสมงาน โดยต้องให้นักเรียนมีส่วนร่วมในการตัดสินใจ

2. ขั้นรวบรวมและจัดระบบผลงาน (collect and organize) ครูสอนต้องวางแผนร่วมกับนักเรียนว่าจะเก็บผลงานและรวบรวมชิ้นงานอย่างไร จะออกแบบเครื่องมือและวิธีการที่ช่วยให้นักเรียนจัดระบบผลงานอย่างไร การรวบรวมชิ้นงานระหว่างดำเนินการต้องรวบรวมผลงานทุกชิ้นในระยะเวลาที่กำหนด ขั้นตอนนี้ผู้เรียนจะสามารถเห็นพัฒนาการและความก้าวหน้าในผลงานของตนเอง ซึ่งชิ้นงานที่รวบรวมมี 2 ชนิด คือ งานหลัก และงานที่ผู้เรียนสามารถคิดสร้างสรรค์เองได้โดยอิสระ

3. ขั้นคัดเลือกผลงาน (select) การคัดเลือกผลงานเพื่อนำมาใส่แฟ้มสะสมงานเป็นกระบวนการของการพัฒนาแฟ้มสะสมงาน โดยนักเรียนจะคัดเลือกผลงานเองและตัดสินใจในเชิงวิชาการเกี่ยวกับ

เนื้อหาว่าเป็นชิ้นงานที่มีหลักฐานเพียงพอ สอดคล้องกับวัตถุประสงค์ของแฟ้มสะสมงาน กระบวนการคัดเลือกจะเกี่ยวข้องกับเกณฑ์และมาตรฐานที่หากมีการกำหนดไว้ล่วงหน้าแล้วจะช่วยให้การดำเนินการสะดวกรวดเร็วขึ้น และยังสามารถเลือกผลงานได้ตรงกับเป้าหมายที่ตั้งไว้

4. ขั้นสร้างสรรค์ผลงานให้มีลักษณะเฉพาะ (interject personality) ขั้นตอนนี้เป็นขั้นที่ต้องใช้ความคิดสร้างสรรค์ของนักเรียน การประดิษฐ์ และตกแต่งแฟ้มสะสมงานของนักเรียนแต่ละคน เพื่อให้แต่ละชิ้นงานมีความแตกต่าง ซึ่งจะสะท้อนให้เห็นความแตกต่างระหว่างบุคคล บุคลิกภาพ ทักษะและความสามารถพิเศษของนักเรียน

5. ขั้นให้ข้อมูลเกี่ยวกับชิ้นงาน (reflect metacognitively) การสะท้อนคิดในขั้นตอนนี้เป็นการถ่ายทอดความรู้สึกเกี่ยวกับผลงานที่เลือกของตน เป็นการย้อนคิดเกี่ยวกับผลงานของตนเอง การไตร่ตรอง กำกับติดตาม และประเมินคุณภาพของผลงาน และยังเป็นการให้นักเรียนได้ใช้ความคิดขั้นสูง รวมทั้งเป็นการหาเหตุผลว่าทำไมงานชิ้นนี้ถึงถูกเลือก เป็นการฝึกการพิพากษ์ของนักเรียน

6. ขั้นตรวจสอบผลงานตนเอง (inspect to self-assess) ขั้นนี้เป็นการให้นักเรียนตั้งเป้าหมาย และตรวจสอบตนเอง ซึ่งเป็นการเสริมสร้างความคิดสร้างสรรค์ ทำให้ผู้เรียนได้ใช้ความรู้ความสามารถ สติปัญญา รวมทั้งความรับผิดชอบต่อการเรียนรู้ของตนเอง สนับสนุนให้นักเรียนเป็นอิสระสามารถประเมินตนเองได้ รวมทั้งตั้งเป้าหมายความสำเร็จได้ นอกจากนี้แล้วจะต้องมีจุดมุ่งหมายเชิงวิชาการและส่งเสริมทักษะทางสังคมในการอยู่ร่วมกับผู้อื่น ปลูกฝังการยอมรับความบกพร่องของตนเองและเสริมสร้างจุดแข็งให้ดียิ่งขึ้น และนักเรียนต้องระบุแนวทางในการปรับปรุงผลงาน และเสนอแนะเพื่อนได้

7. ขั้นทำให้สมบูรณ์และประเมินค่าผลงาน (perfect and evaluate) เมื่อนักเรียนได้ตรวจสอบความสามารถและกำหนดแผนการปรับปรุงผลงานแล้ว จะเข้าสู่ระยะทำให้ผลงานมีความสมบูรณ์ เพื่อเตรียมพร้อมที่จะได้รับการให้ระดับคะแนน ซึ่งขั้นนี้จะช่วยขัดเกลาทำให้ผลงานมีความสมบูรณ์ การให้คะแนนต้องพิจารณาจากเกณฑ์การให้คะแนนที่ครูและนักเรียนร่วมกันพัฒนาขึ้น และการประเมินจะให้ความสำคัญกับการประเมินความก้าวหน้ามากกว่าการประเมินเพื่อการเปรียบเทียบ

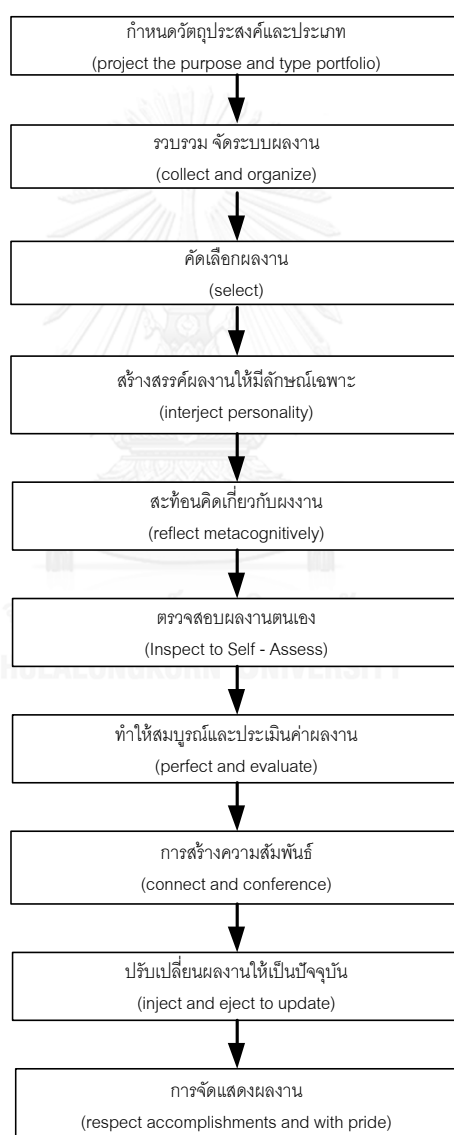
8. ขั้นการสร้างความสัมพันธ์ (connect conference) เป็นการเปิดโอกาสให้นักเรียนนำเสนอความสำเร็จของผลงาน มีคนอื่น ๆ ได้รับความสำเร็จนั้น และร่วมสนทนากันแลกเปลี่ยนซึ่งกันและกันมากขึ้น นักเรียนได้พัฒนาตนเอง ส่งเสริมให้มีการสื่อสาร การแสดงออกที่เหมาะสม อาจมีการสนทนาแบบตัวต่อตัว ตลอดจนการแสดงออกที่เป็นทางการเพื่อเกิดประโยชน์ต่อการพัฒนาผลงานและแฟ้มสะสมงาน

9. ขั้นปรับเปลี่ยนผลงานให้เป็นปัจจุบัน (inject/eject to update) เป็นขั้นตอนที่มีการหมุนเวียนชิ้นงานที่เป็นผลงานโดยการนำบางชิ้นเข้ามาและบางชิ้นเอาออกเพื่อให้แฟ้มสะสม

งานมีความทันสมัยอยู่เสมอ ช่วยให้นักเรียนสนใจความเปลี่ยนแปลงของสิ่งแวดล้อม ศึกษาชิ้นงานให้เหมาะสมกับสภาพการณ์และสอดคล้องกับเป้าหมายของแฟ้มสะสมงานที่กำหนดไว้ในตอนต้น

10. ขั้นการจัดแสดงผลงาน (respect accomplishment) เป็นขั้นที่มีการจัดแสดงผลงานของนักเรียนต่อสาธารณชน เพื่อนักเรียนคนอื่นๆ ผู้บริหาร และผู้ปกครอง เพื่อให้นักเรียนได้ยอมรับและชื่นชมในผลงานของตน ซึ่งนักเรียนจะต้องจัดเตรียมการแสดงผลนิทรรศการด้วยตนเอง

จากแนวทางการพัฒนาแฟ้มสะสมงานของ Burke, Fogart and Begrad (1994) สุวิทย์ มูลคำ (2540) และชัยพฤกษ์ เสรีรักษ์ และคณะ (ม.ป.ป.) สามารถสรุปขั้นตอนการพัฒนาได้ดังภาพที่ 2.6



ภาพที่ 2.6 ขั้นตอนการพัฒนาแฟ้มสะสมงาน

Hamp-Lyons and Condon (2000, cited in Lam & Lee, 2010) กล่าวว่า การดำเนินการพัฒนาแฟ้มสะสมงานประกอบด้วย 3 ขั้นตอน ดังนี้

1. การรวบรวมผลงาน (collection) เป็นการรวบรวมผลงานต่างๆ ที่ครูมอบหมายให้ ผู้เรียนทำระหว่างการเรียน โดยครูจะให้ผู้เรียนทำงานต่างๆ ตามวัตถุประสงค์ของการเรียนในแต่ละ หัวเรื่องจะมีการมอบหมายภาระงานหลายภาระงาน และแต่ละงานจะต้องทำหลายครั้งเพื่อปรับปรุง ให้ผลงานออกมามีคุณภาพเพื่อนำผลงานที่ตนทำเก็บไว้ในแฟ้มสะสมงาน

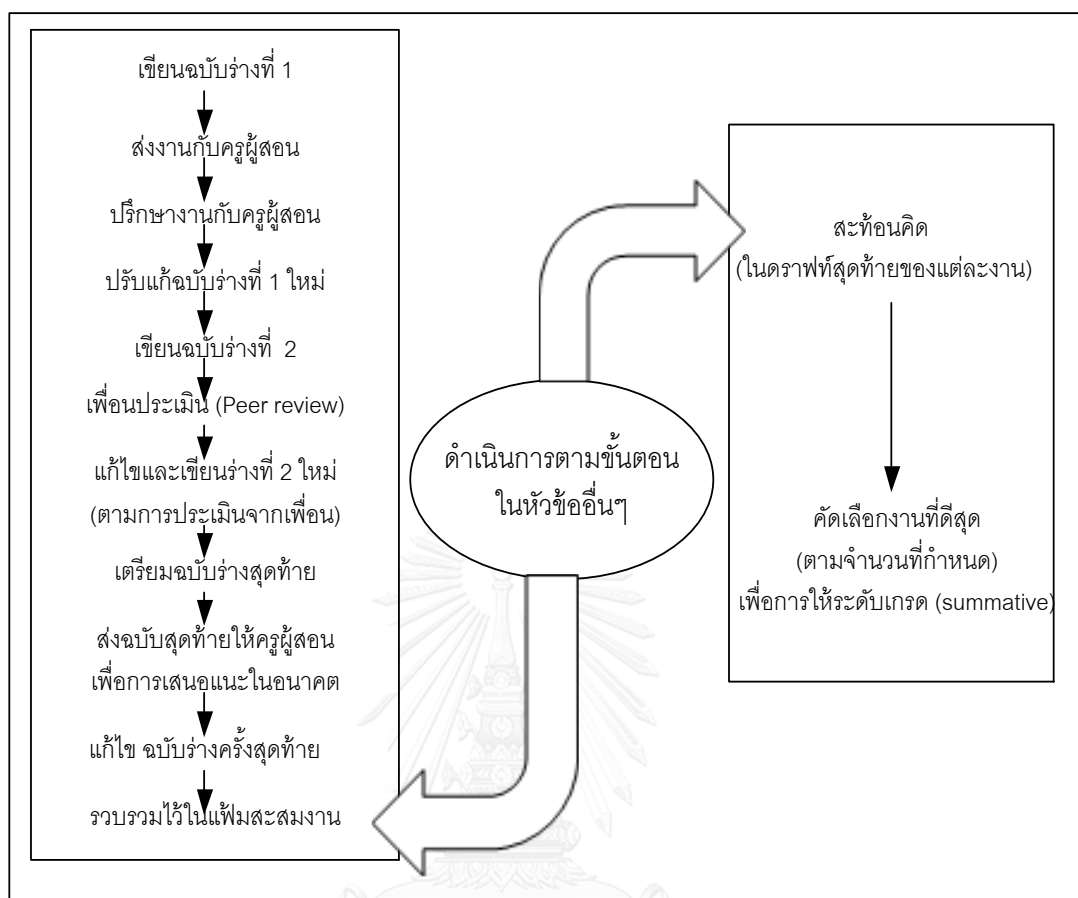
2. การคัดเลือกผลงาน (selection) เมื่อรวบรวมงานในขั้นตอนการรวบรวมงานและ ผ่านการให้คำปรึกษาจากครูผู้สอน และการได้รับข้อมูลป้อนกลับจากครูผู้สอนและนำไปปรับแก้และ ได้รับการประเมินผลงานจากเพื่อนร่วมชั้น (peer review) จะต้องคัดเลือกผลงานที่คิดว่าดีที่สุด มาจำนวนหนึ่งตามที่ร่วมกันกับครูผู้สอนไว้เข้ามาเก็บในแฟ้มงานของตนเอง และผลงานที่ถูกคัดเลือก ไว้นี้จะนำมาจัดอันดับหรือให้คะแนนในเรื่องที่เรียนเพื่อสรุปรวมผลการเรียน (summative)

3. การสะท้อนคิด (reflection) ลักษณะการประเมินโดยแฟ้มสะสมงานสิ่งที่สำคัญ ประการหนึ่งคือการสะท้อนคิดของผู้เรียน โดยเมื่อนักเรียนทำงานและมีการปรับแก้จนเสร็จสิ้นในแต่ละ ชิ้นงานแล้วนักเรียนจะต้องมีการสะท้อนคิดเกี่ยวกับการดำเนินงานของตนเอง จุดเด่นจุดด้อยของผลงาน ของตน

Lam and Lee (2009) ได้พัฒนาแฟ้มสะสมงานทางภาษาโดยใช้แฟ้มสะสมงานในการประเมินการเขียนภาษาอังกฤษของนักศึกษาที่เรียนภาษาอังกฤษในฐานะภาษาต่างประเทศ (EFL) โดยนำแฟ้มสะสมงานมาใช้ในการประเมินความก้าวหน้าหรือประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) ซึ่งมีการดำเนินการตามแนวคิดของ Hamp-Lyons and Condon (2000) และมีการดำเนินการระหว่างการประเมินระหว่างเรียน และ ประเมินสรุปรวม ดังภาพที่ 2.7 และแต่ละขั้นมีการดำเนินการ ดังนี้ (Hamp-Lyons & Condon, 2000)

1. ขั้นการประเมินระหว่างเรียน (formative assessment) มีสิ่งที่สำคัญและจำเป็น ที่ต้องดำเนินการ ดังนี้ 1) ครูให้ข้อมูลป้อนกลับ (feedback) แก่นักเรียนอย่างต่อเนื่อง 2) ครูและ นักเรียนมีการปรึกษาเกี่ยวกับผลงานของนักเรียน และ 3) มีการประเมินผลงานโดยเพื่อนนักเรียน ในชั้นเรียน (peer review)

2. ขั้นการประเมินสรุปรวม (summative assessment) มีการดำเนินการ ดังนี้ 1) เมื่อสิ้นสุด การเรียนนักเรียนสะท้อนคิดในงานของตนเอง วิเคราะห์จุดแข็ง จุดด้อยในผลงานของตนเอง 2) ครูทำการ ประเมินแฟ้มสะสมงานของนักเรียนตามเกณฑ์ 2) ครูให้เกรด (grading) แฟ้มสะสมงานของนักเรียน แต่ละคน โดยประเมินจากผลงานที่นักเรียนเลือกใส่ในแฟ้มสะสมงาน และผลงานฉบับร่างระหว่างการ ปรับปรุงจะไม่มีนำมาพิจารณาให้คะแนน



ภาพที่ 2.7 กระบวนการพัฒนาแฟ้มสะสมงานทางภาษาของ Lam and Lee (2010)

ที่มา: Lam and Lee (2010) ภาลงกรณ์มหาวิทยาลัย

Barabouti (2012) ได้เสนอขั้นตอนการดำเนินการประเมินแฟ้มสะสมงานการเขียน (ภาษาอังกฤษ) ในงานวิจัยเกี่ยวกับการพัฒนาการเขียนของนักเรียนระดับประถมศึกษา มีการดำเนินการ 6 ขั้นตอน ดังนี้

1. กำหนดวัตถุประสงค์ของแฟ้มสะสมงาน

แฟ้มสะสมงานที่ใช้ในการวิจัยครั้งนี้ผู้วิจัยระบุไว้ว่าเป็นการประเมินทักษะการเขียน (ภาษาอังกฤษ) จึงเน้นกระบวนการเขียนมากกว่าผลงานสุดท้ายของนักเรียน

2. กำหนดข้อมูลที่ใช้ในการประเมิน

เป็นการระบุเกี่ยวกับข้อมูลที่จะรวบรวมจากแฟ้มสะสมงานเพื่อ 1) พัฒนาการเขียน ซึ่งเน้นทั้งเนื้อหา รูปแบบ 2) วินิจฉัยจุดแข็งของนักเรียน ดังนั้นจึงเป็นให้ข้อคิดเห็นเป็นทางบวกเพื่อให้ นักเรียนตระหนักว่าอะไรที่ตนสามารถทำได้ (can do) 3) วินิจฉัยจุดด้อยของนักเรียน ดังนั้นจึงมีการปรับใช้สื่อการสอนที่สอดคล้องตามเนื้อหาอื่นๆ

3. จับคู่กิจกรรมกับวัตถุประสงค์

แฟ้มสะสมงานจะต้องมีการพิจารณาในขั้นตอนนี้ว่าเป้าหมายของการเรียนการสอน และจุดประสงค์ของการเรียนที่กำหนดโดยหลักสูตรเพื่อมีการจัดกิจกรรมให้สอดคล้องกัน ดังตัวอย่างที่แสดงในตารางที่ 2.8

ตารางที่ 2.8 ตัวอย่างการจับคู่กิจกรรมกับวัตถุประสงค์

กิจกรรม	หน้าที่ทางภาษา (language function)
1. ให้นักเรียนดูภาพและเขียนบรรยายเกี่ยวกับ Nick's daily activities	การเขียนเกี่ยวกับกิจกรรมที่ผู้คนทำในชีวิตประจำวัน
2. ให้นักเรียนเขียนโปสการ์ดถึงเพื่อนเกี่ยวกับสถานที่ที่ไปเที่ยวในวันหยุด	การเขียนเพื่อนำเสนอ

4. บันทึกความก้าวหน้าของนักเรียน

การบันทึกความก้าวหน้าของนักเรียนเป็นระยะที่สำคัญในกระบวนการของพัฒนาแฟ้มสะสมงาน ซึ่งเป็นการบันทึกเกี่ยวกับความก้าวหน้าของนักเรียนเป็นรายบุคคลเกี่ยวกับสิ่งที่นักเรียนได้เรียนรู้ และช่วยการเรียนรู้ในอนาคตแก่นักเรียน ซึ่งการใช้แฟ้มสะสมงานนั้นทำให้ครูได้สะท้อนคิดแก่ผลงานของนักเรียนด้วย

5. การให้นักเรียนมีส่วนร่วม

ในขั้นตอนนี้เป็นการให้นักเรียนได้เข้าร่วมกิจกรรมการเรียนรู้โดยใช้แฟ้มสะสมงาน ซึ่งครูจะต้องแจ้งให้นักเรียนทราบในการเข้าร่วมโครงการ และให้นักเรียนเชื่อว่าวัตถุประสงค์ของแฟ้มสะสมงานมาจากการกำหนดร่วมกันระหว่างครูและนักเรียนหรือเป็นการตัดสินใจร่วมกัน โดยครูสำรวจประเด็นที่นักเรียนสนใจที่นักเรียนอยากจะเขียน กำหนดเกณฑ์ในการให้คะแนน นักเรียนต้องประเมินแฟ้มสะสมงานของเพื่อน ในการประเมินนักเรียนมีบทบาทที่สำคัญ 2 ประการ ดังนี้ 1) นักเรียนประเมินตนเอง (self-assessment) โดยนักเรียนสะท้อนคิดเกี่ยวกับข้อผิดพลาดของตนเองในการเขียนให้เหตุผลว่าทำไมถึงทำผิดพลาด 2) การประเมินเพื่อน (peer-assessment) ดำเนินการดังนี้ ประเมินผลงานเพื่อน 1 คน สะท้อนคิดว่าได้เรียนรู้อะไรจากงานเขียนของเพื่อนและประเมินเพื่อนโดยเสนอแนะการทำให้งานเขียนดีขึ้นได้อย่างไร นักเรียนต้องประเมินแฟ้มสะสมงานของเพื่อนในตอนสุดท้ายของการเรียน และบันทึกจุดแข็งจุดอ่อนของเพื่อน

6. สื่อสารเกี่ยวกับวัตถุประสงค์และผลที่ได้รับ

ขั้นตอนนี้เป็นการร่วมตัดสินใจในทุกขั้นตอนระหว่างครูและนักเรียนในการดำเนินการเรียนการสอนของโครงการ และให้ความสำคัญกับผู้ปกครองนักเรียนคือมีการนำเสนอแฟ้มสะสมงานแก่ผู้ปกครอง

Lo (2010) ได้เสนอการพัฒนาและการดำเนินการของแฟ้มสะสมงานในการพัฒนา นักศึกษาระดับอุดมศึกษา ในการจัดการเรียนการสอนเพื่อให้นักศึกษาเกิดการเรียนรู้ด้วยตนเองในวิชา ภาษาอังกฤษในฐานะเป็นภาษาต่างประเทศ (EFL) ประกอบด้วย 7 ขั้นตอน ดังนี้

1. วิเคราะห์ข้อมูลจากแบบสอบถามก่อนเรียน (สอบถามผู้เรียนเกี่ยวกับพฤติกรรมการอ่าน หนังสือพิมพ์ภาษาอังกฤษ)

2. ตัดสินใจเลือกรูปแบบแฟ้มสะสมงาน

3. พัฒนาจุดประสงค์การเรียนรู้

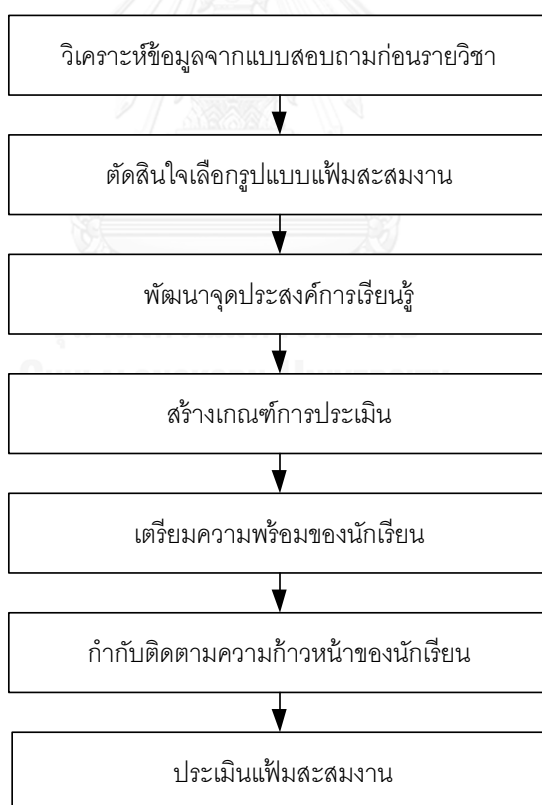
4. สร้างเกณฑ์การให้คะแนน

5. เตรียมความพร้อมของนักเรียน

6. กำกับติดตามความก้าวหน้าของนักเรียน

7. ประเมินแฟ้มสะสมงาน

จากขั้นตอนการพัฒนาและดำเนินการแฟ้มสะสมงานทางภาษาอังกฤษของนักศึกษา สามารถสรุปได้ ดังภาพที่ 2.8



ภาพที่ 2.8 กระบวนการพัฒนาและดำเนินการโครงการแฟ้มสะสมงานของ Lo (2010)

ที่มา: Lo (2010)

อุทุมพร (ทองอุไทย) จามรमान (2540) ได้เสนอขั้นตอนการดำเนินงานจัดทำแฟ้มสะสมงาน เพื่อให้ผู้เรียนได้แสดงความสามารถที่แท้จริง ประกอบด้วย 15 ขั้นตอน ดังนี้

1. ระบุหลัก ภาระงาน บทบาท หน้าที่ของเจ้าของแฟ้มสะสมงานว่ามีอะไรบ้าง
2. มีความรู้ภาคทฤษฎี และค้นคว้าเพิ่มเติม
3. การระบุประเด็น/เรื่องที่จะบรรจุไว้ในแฟ้มสะสมงาน
4. การวางแผนดำเนินงานเพื่อหาเข้ามาในแฟ้มสะสมงาน
5. การดำเนินงานศึกษา ค้นคว้า ทดลอง ปฏิบัติงาน
6. การแสดงผลประเมินเป็นระยะๆ ของเจ้าของแฟ้มสะสมงานกับผู้เกี่ยวข้อง เช่น นักเรียนเจ้าของแฟ้มสะสมงานกับคณะครูผู้เกี่ยวข้อง ครูเจ้าของแฟ้มสะสมงานกับเพื่อนครู-ผู้บริหารที่เกี่ยวข้อง เป็นต้น

7. การคัดเลือกสาระมาใส่แฟ้ม
8. การเขียนสารบัญเนื้อหาสาระ ความนำ และวัตถุประสงค์
9. การสรุปผลประเมินเป็นระยะๆ จากผู้เกี่ยวข้อง
10. การเขียนสรุปผลที่ศึกษาในประเด็นที่ระบุในข้อ 2
11. การจัดเรียงสาระในแฟ้มสะสมงาน พร้อมหลักฐานแหล่งค้นคว้าอ้างอิง
12. การจัดทำแฟ้มสะสมงานให้สวยงาม นำเปิดอ่าน ตรวจสอบแฟ้มว่ามีการจัดเรียงอย่างเป็นระบบ และครบถ้วนตามสาระที่ต้องการหรือไม่
13. การนำเสนอ (presentation) สู่คณะบุคคลผู้เกี่ยวข้องภายนอก ซึ่งอาจทำในรูปแบบ การบรรยาย การจัดนิทรรศการ การใช้สื่อโสตประกอบการแสดง เป็นต้น
14. การโต้ตอบทางบวกต่อคณะบุคคลในข้อ 13
15. การสรุปผลการประเมินจากคณะบุคคลภายนอก และการวางแผนการแก้ไขจุดอ่อน เสริมจุดแข็งต่อไป

2.6 จุดเด่นและจุดอ่อนของแฟ้มสะสมงาน

2.6.1 จุดเด่นและประโยชน์ของแฟ้มสะสมงาน

แฟ้มสะสมงานให้ข้อมูลเกี่ยวกับผู้ที่เป็นเจ้าของแฟ้มอย่างครบถ้วนมากกว่าการทดสอบด้วยแบบทดสอบ เนื่องจากการจัดทำแฟ้มสะสมงานต้องใช้ความคิด ความรู้ ความอดทน วิจัยรณญาณ ความพยายามและทักษะต่างๆ ซึ่งครอบคลุมทั้งทางสมองและจิตใจ ซึ่งเป็นการพัฒนาผู้เรียนแต่ละคนในหลายๆ ด้านไปพร้อมๆ กัน (อุทุมพร (ทองอุไทย) จามรमान, 2540ก)

Payne (2003) กล่าวถึงจุดเด่นของแฟ้มสะสมงานที่เป็นเหตุผลสำคัญที่ครูควรนำแฟ้มสะสมงานมาใช้ในการประเมินนักเรียนไว้ดังนี้

1. เป็นการรวบรวมผลงานด้วยนักเรียนเองที่เรียนในชั้นเรียน ทำให้แฟ้มสะสมงานมีความเป็นธรรมชาติ
2. แฟ้มสะสมงานเป็นสื่อกลางที่นำนักเรียนไปสู่การสร้างผลสัมฤทธิ์ในการเรียนที่มีความสัมพันธ์ระหว่างนักเรียน ครู และผู้ปกครอง
3. เป็นการเกาะติดความก้าวหน้าและพัฒนาการตลอดทั้งภาคเรียนหรือทั้งปีการศึกษา และเป็นโอกาสในการแสดงความสามารถในระดับสุดท้าย
4. เป็นการเรียนการสอนที่เน้นนักเรียนเป็นศูนย์กลาง
5. กระตุ้นครูให้ได้พบปะพูดคุยทั้งกับนักเรียนและผู้ปกครองนักเรียน
6. ส่งเสริมทักษะการประเมินตนเองของนักเรียนและการประเมินที่เป็นจริง (realistic of assessment)
7. ทำให้เห็นภาพของการปฏิบัติงานที่ต่อเนื่องที่ถูกตัดแบ่งให้เห็นในแต่ละช่วงๆ
8. ภาระงานเป็นตัวแทนของหลักฐานที่มีอยู่จริง
9. ทำให้นักเรียนได้มีโอกาสในการเป็นเจ้าของการเรียนรู้

2.6.2 จุดอ่อนของแฟ้มสะสมงาน

แม้แฟ้มสะสมงานจะมีประโยชน์หลายประการแต่ก็ยังมีจุดด้อยบางประการ ได้แก่ ความสอดคล้องระหว่างผู้ประเมินงาน/แฟ้มของผู้เรียน มีมากน้อยเพียงใด ค่าที่ได้จากงาน/แฟ้มในช่วงระยะเวลาต่างๆ (ต้นปี ถึง ปลายปี) ควรมีค่าคงที่หรือไม่ ความเป็นมาตรฐานของค่าที่แสดงความสามารถของนักเรียนอยู่ที่ใคร และจำนวนงาน/แฟ้มที่ต้องพิจารณา มีปริมาณมากจะเกิดความคุ้มค่า แรงงานหรือไม่ และส่งผลต่อ คุณภาพของค่าที่ได้หรือไม่ (อุทุมพร (ทองอุไทย) จามรมาน, 2540ก)

(Payne, 2003) กล่าวถึงจุดด้อยของแฟ้มสะสมงานไว้ดังนี้

1. ทำให้เกิดความยุ่งยากในการเก็บเอกสารและชิ้นงานต่างๆ อาจจะยับยั้งประโยชน์ต่อการใช้งาน
2. นักเรียน ครู และผู้ปกครองบางคนมีการต่อต้านการดำเนินงานเนื่องจากเป็นวิธีการใหม่
3. ยากที่จะได้รับการยอมรับความน่าเชื่อถือของผู้อ่าน ผู้ให้คะแนน และผู้ประเมิน
4. มีความจำเป็นสำหรับการพัฒนาเจ้าหน้าที่โดยอาจบังคับด้วยเวลาและทรัพยากร
5. การขาดโอกาสและการไม่ให้ความสำคัญของนักเรียนในการแสดงออกอาจนำไปสู่การสรุปอ้างอิงที่ไม่ถูกต้อง
6. ความล้มเหลวในการรวบรวมผลงานต่างๆ นำไปสู่การสรุปอ้างอิงที่ไม่ถูกต้อง

7. อาจเกิดความล้มเหลวในการสร้างการให้คะแนนแบบรูบริก
8. เสียเวลาในการสร้างการให้คะแนนแบบรูบริกเพื่อการทดสอบภาคสนาม

2.7 การประเมินแฟ้มสะสมงาน

การประเมินหรือการให้ค่าแก่แฟ้มสะสมงานเป็นการให้ค่าหรือประเมินจากหลักฐานที่แสดงการดำเนินงาน ซึ่งถูกรวบรวมไว้ในแฟ้มสะสมงาน ซึ่งรวมทั้งกระบวนการผลิต และผลผลิตเข้าไว้ด้วยกัน เพื่อให้ผู้อ่าน เช่น ครู อาจารย์ ผู้บริหาร ผู้ปกครอง ได้เห็นภาพความพยายาม ตลอดจนขั้นตอนการดำเนินงานตั้งแต่ต้นจนจบ (อุทุมพร (ทองอุไทย) จามรมาน, 2540) ซึ่งการให้คะแนนแฟ้มสะสมงานหรือการประเมินแฟ้มสะสมงานจะนิยมใช้วิธีการประเมิน 3 วิธี (กมลวรรณ ดังชันกานนท์, 2557; Randal et al., 1987 อ้างถึงใน ชัยฤทธิ์ ศิลาเดช, 2540; Kuhs, 1994 อ้างถึงใน สมศักดิ์ ภูวิภาดารวรรณ, 2544) ได้แก่ 1) การประเมินชิ้นแต่ละชิ้นแล้วหาค่าเฉลี่ย 2) การประเมินตามองค์ประกอบ และ 3) การประเมินชิ้นงานโดยรวม ซึ่งแต่ละวิธีมีรายละเอียด ดังนี้

1) การประเมินชิ้นงานแต่ละชิ้นแล้วหาค่าเฉลี่ย โดยทำการประเมินชิ้นงานแต่ละชิ้นในแฟ้มสะสมงานด้วยเกณฑ์การให้คะแนน (scoring rubric) หาคะแนนรวมของชิ้นงานทั้งแฟ้มสะสมงานแล้วทำการหาค่าเฉลี่ยของชิ้นงานเพื่อเป็นคะแนนของแฟ้มสะสมงานทั้งแฟ้ม

2) การประเมินตามองค์ประกอบ หรือสมรรถภาพ หรือตามมิติของงาน โดยทำการแยกวิเคราะห์ชิ้นงานตามมิติ ตามองค์ประกอบ หรือตามวัตถุประสงค์ของแฟ้มสะสมงาน ทำการให้คะแนนตามมิติของงานหรือตามสมรรถภาพที่ต้องการวัด แยกชิ้นงานตามมิติหรือสมรรถภาพ แล้วทำการให้คะแนนหรือประเมินชิ้นงานโดยแยกเกณฑ์ตามแต่ละสมรรถภาพหรือตามมิติโดยใช้เกณฑ์การให้คะแนนที่สร้างขึ้นมาเฉพาะของสมรรถภาพนั้นๆ ทั้งนี้ 1 ชิ้นงานอาจมีการประเมินได้มากกว่า 1 สมรรถภาพได้เมื่อทำครบทุกชิ้นงานแล้วนำคะแนนแต่ละสมรรถภาพมารวมกันและหาค่าเฉลี่ยในแต่ละสมรรถภาพ

3) การประเมินชิ้นงานโดยรวม เป็นการประเมินแฟ้มสะสมงานที่มีวิธีการที่แตกต่างจาก 2 วิธีแรก กล่าวคือ เป็นการประเมินแฟ้มสะสมงานที่มองภาพรวมของแฟ้มสะสมงาน ไม่ให้คะแนนชิ้นงานเป็นรายชิ้น แต่จะให้คะแนนตามความสมบูรณ์ของของแฟ้มสะสมงานทั้งฉบับ ความครบถ้วนของชิ้นงาน คุณภาพของชิ้นงาน ความสร้างสรรค์

อย่างไรก็ตามก่อนการประเมินแฟ้มสะสมงานครูผู้สอนควรตรวจสอบคุณภาพของหลักฐานแต่ละชิ้น โดยพิจารณาถึงคุณลักษณะ 3 ประการ ได้แก่ มีการตอบสนองจุดประสงค์ มีการสะท้อน และมีการรับรอง (ประเมิน) และพิจารณาเนื้อหาของแฟ้มสะสมงานว่าบรรลุจุดมุ่งหมาย ผู้เรียนมีความก้าวหน้าตามจุดมุ่งหมายของแฟ้มสะสมงานที่กำหนดไว้ หากไม่บรรลุตามจุดมุ่งหมายครูผู้สอนจะต้องเพิ่มหลักฐานหรือชิ้นงานอีกโดยควรทำเป็นรายบุคคลเพื่อให้ให้นักเรียนได้มีการปรับปรุง

และมีความก้าวหน้าและบรรลุจุดมุ่งหมายทางการเรียนต่อไป (กมลวรรณ ตั้งธนกานนท์, 2557)
การประเมินแฟ้มสะสมงานตัวอย่างดังที่แสดงในตารางที่ 2.9

ตารางที่ 2.9 ตัวอย่างเกณฑ์การประเมินแฟ้มสะสมงานเพื่อให้เกรด (grading)

ชื่อนักเรียน.....วันที่.....

ผู้ประเมิน.....

งานในแฟ้มสะสมงานนี้มีความสำเร็จอย่างไรในด้านต่างๆ ต่อไปนี้ (ให้คะแนนจาก 1 ถึง 5 คะแนน)

	สูงที่สุด				ต่ำที่สุด
1. แสดงถึงความคิดสร้างสรรค์	1	2	3	4	5
2. แสดงถึงความหลากหลายของผลงาน	1	2	3	4	5
3. แสดงถึงความรู้ของการทำงานในรูปแบบที่ถูกต้อง	1	2	3	4	5
4. มีการจัดระเบียบอย่างดี	1	2	3	4	5
5. แสดงถึงความชัดเจนของผลงาน	1	2	3	4	5
6. แสดงถึงผลงาน (งานเขียน) ที่กระชับ	1	2	3	4	5
7. มีการปรับให้เหมาะสมกับผู้อ่าน	1	2	3	4	5
8. โนม่น่าผู้อ่านได้	1	2	3	4	5
9. ใช้หลักไวยากรณ์ได้ถูกต้อง	1	2	3	4	5
10. สะกดคำได้อย่างถูกต้อง	1	2	3	4	5
11. เขียนย่อหน้าได้ดี	1	2	3	4	5
12. มีความประณีตและเป็นมืออาชีพ	1	2	3	4	5

ที่มา: Payne (2003)

ในการวิจัยครั้งนี้จะขอกล่าวถึงการประเมินที่เป็นแบบเกณฑ์การให้คะแนน (scoring rubric) เนื่องจากเป็นการประเมินที่มีการแยกประเมินตามองค์ประกอบและเป็นวิธีการประเมินที่ได้รับการนำมาใช้อย่างแพร่หลาย ซึ่งมีรายละเอียดดังต่อไปนี้

2.7.1 การประเมินแบบรูบริก (scoring rubric)

การประเมินแฟ้มสะสมงาน เป็นการประเมินแบบตามแนวคิดของการประเมินตามสภาพจริง (authentic assessment) ซึ่งในการประเมินตามสภาพจริงนั้นมักนิยมใช้รูบริก (rubric) เพื่ออธิบายความสัมฤทธิ์ผลของนักเรียน (สมศักดิ์ ภูวิภาดารัตน์, 2544; อุทุมพร (ทองอุไทย) จามรมาน, 2540) การประเมินที่ใช้ในการประเมินแฟ้มสะสมงานนั้น เป็นการให้คะแนนแบบรูบริก (scoring rubric)

เป็นเกณฑ์ที่ถูกลำเอียงไปอย่างแพร่หลาย การให้คะแนนแบบรูบริก คือ การให้ค่ากับลักษณะที่กำหนดขึ้นระหว่างเจ้าของแฟ้มสะสมงานกับผู้เกี่ยวข้องในช่วงการจัดทำแฟ้มสะสมงานและการนำเสนอแฟ้มสะสมงาน (อุทุมพร (ทองอุไทย) จามรมาน, 2540ข) เมื่อนักเรียนคัดเลือกงานใส่ในแฟ้มสะสมงานครบตามที่กำหนดแล้วและเสร็จสิ้นการเรียนหรือโครงการครูและผู้เกี่ยวข้องจะทำการประเมินแฟ้มสะสมงานตามเกณฑ์ที่สร้าง

2.7.2 ประโยชน์ของรูบริก

รูบริกมีประโยชน์มากสำหรับการประเมิน ซึ่งประโยชน์ของรูบริกมีดังนี้ (Wiener & Cohen, 1994; อ้างอิงใน สมศักดิ์ ภูวิภาตววรรณ, 2544)

1. ช่วยให้การคาดหวังของครูที่มีต่อผลงานของนักเรียนบรรลุผลสำเร็จได้ โดยนักเรียนจะเกิดความสามารถ และสามารถใช้อุปกรณ์ต่อการประเมินและพัฒนาชิ้นงานของตน
2. ช่วยให้ครูเกิดความชัดเจนมากขึ้นเกี่ยวกับสิ่งที่ต้องการให้นักเรียนเกิดการเรียนรู้หรือพัฒนาการ
3. ช่วยให้นักเรียนบรรลุคุณลักษณะจากงานที่เป็นตัวอย่างได้ โดยรูบริกตรวจสอบ
4. ช่วยให้นักเรียนสามารถควบคุมตนเองในการปฏิบัติงานเพื่อไปสู่ความสำเร็จได้
5. เป็นเครื่องมือในการเชื่อมโยงความสัมพันธ์ระหว่างกิจกรรมการปฏิบัติงานต่างๆ ของนักเรียนได้เป็นอย่างดี
6. ช่วยให้ผู้เกี่ยวข้อง เช่น ผู้ปกครอง ผู้สนับสนุน ผู้นิเทศ ได้เกิดความเข้าใจเกณฑ์การตัดสินผลงานนักเรียนที่ครูใช้
7. ช่วยในการให้เหตุผลประกอบการให้เกรดนักเรียนได้
8. ช่วยเพิ่มคุณภาพผลงานของนักเรียน

2.7.3 ประเภทของรูบริก

การประเมินแบบรูบริกสามารถแบ่งออกได้เป็น 2 ประเภท ได้แก่ รูบริกแบบองค์รวม (holistic rubric) และรูบริกแบบแยกองค์ประกอบ (analytic rubric) (อุทุมพร จามรมาน, 2540ข และ สมศักดิ์ ภูวิภาตววรรณ, 2544) ซึ่งแต่ละประเภทมีรายละเอียด ดังนี้

1. รูบริกแบบองค์รวม (holistic rubric)

รูบริกแบบองค์รวมเป็นเครื่องมือในการให้คะแนนที่บรรยายคุณภาพของงานในแต่ละระดับคะแนน (grad) ที่ควรได้โดยพิจารณาจากภาพรวม ในการประเมินจะต้องระบุได้ว่าการให้คะแนนในแต่ละระดับหมายถึงอะไร เช่น ในงานเขียน มีการกำหนดคะแนน 4 3 2 1 ผู้ประเมินต้องระบุว่า 4 คะแนน นักเรียนจะต้องมีลักษณะหรือคุณภาพอย่างไร (สมศักดิ์ ภูวิภาตววรรณ, 2544) ดังตัวอย่างในตารางที่ 2.10

ตารางที่ 2.10 ตัวอย่างรูบริกแบบองค์รวมเพื่อประเมินการนำเสนอปากเปล่า

ชื่อ-สกุล นักเรียน.....	ชั้น.....วิชา.....
ระดับ	คำอธิบายความสามารถ
5 - ยอดเยี่ยม	นักเรียนอธิบายเรื่องที่ศึกษาได้อย่างชัดเจน พร้อมให้เหตุผลที่หนักแน่น ชี้ให้เห็นความสำคัญของเรื่องที่ศึกษา มีข้อมูลที่เฉพาะเจาะจงสนับสนุนข้อสรุป การนำเสนอมีความต่อเนื่อง มีการสบสายตาผู้ฟังโดยตลอด มีการเตรียมการอย่างดี มีการจัดลำดับระบบระเบียบในการนำเสนออย่างดี มีความกระตือรือร้น มีการใช้สื่อในการนำเสนออย่างมีประสิทธิภาพ ตอบคำถามของผู้ฟังได้ชัดเจน ตรงประเด็น และเหมาะสม
4 - ดีมาก	นักเรียนอธิบายเรื่องที่ศึกษาให้เหตุผลที่ชี้ให้เห็นความสำคัญของเรื่องที่ศึกษา มีข้อมูลเพียงพอที่สนับสนุนข้อสรุป การนำเสนอถูกต้อง มีการเตรียมการก่อนนำเสนอ และมีการจัดลำดับการนำเสนอ มีการใช้สื่อ และตอบคำถามผู้ฟังได้ชัดเจน
3 - ดี	นักเรียนอธิบายเรื่องที่ศึกษามาพร้อมกล่าวถึงข้อสรุป แต่ข้อมูลสนับสนุนข้อสรุปไม่มากพอเหมือน 4 กับ 5 การนำเสนอถูกต้อง มีการเตรียมการก่อนนำเสนอ และมีการจัดลำดับการนำเสนอ มีการใช้สื่อและตอบคำถามผู้ฟัง
2 - พอใช้	นักเรียนอธิบายเรื่องที่ศึกษามาแต่ไม่ค่อยสมบูรณ์ ไม่มีการกล่าวถึงข้อสรุป ลำดับการนำเสนอที่ทำให้เข้าใจได้ แต่มีข้อผิดพลาดบางจุด เตรียมการไม่ดี จัดระบบระเบียบไม่ดี มีการตอบคำถามได้บางคำถาม
1 - ควรปรับปรุง	นักเรียนนำเสนอโดยไม่ได้เริ่มที่คำถามหรือความสำคัญ หัวข้อที่นำเสนอไม่ชัดเจนและมีข้อสรุปที่ไม่เพียงพอ การนำเสนอเน้นย้ำต่อการติดตามฟัง ขาดการเตรียมตัวหรือขาดการจัดระบบระเบียบ มีการตอบคำถามเฉพาะบางคำถามหรือไม่ตอบ
0	ไม่มีการนำเสนอปากเปล่า

ที่มา Wiggins (1998; อ้างอิงใน สมศักดิ์ ภาวภูตานาวรรณ, 2544)

2. รูบริกแบบแยกส่วน (analytic rubric)

รูบริกแบบแยกส่วน หรือแยกมิติเป็นเครื่องมือในการให้คะแนนที่มีลักษณะแยกสิ่งที่ต้องการประเมินออกเป็นองค์ประกอบหรือหลายมิติ (dimension of performance) และแต่ละองค์ประกอบ/มิติจะมีการบรรยายคุณภาพออกเป็นระดับต่างๆ จากสูงสุดไปต่ำสุดหรือควรปรับปรุง เช่นงานที่จะได้รับการประเมินงานเขียน การประเมินจะประเมินแยกตามองค์ประกอบ ได้แก่ การสร้างประโยค การเรียบเรียง สไตลการเขียน เป็นต้น ในแต่ละมิตินั้นจะต้องบรรยายคุณภาพเป็นลำดับสูงสุดถึงต่ำสุด ซึ่งการกำหนดระดับว่ามีกี่ระดับแล้วแต่ความเหมาะสมและการให้คะแนนก็ต้องพิจารณาว่าในองค์ประกอบหรือมิติที่จะประเมินนั้น งานชิ้นนั้นมีคุณภาพในระดับใด โดยเทียบกับรายละเอียดที่เขียนบรรยายไว้ในรูบริก (สมศักดิ์ ภาวภูตานาวรรณ, 2544) ดังตัวอย่างในตารางที่ 2.11

ตารางที่ 2.11 การให้คะแนนแบบรูบริกเพื่อประเมินแฟ้มสะสมงานแบบแยกส่วน

ด้าน	4	3	2	1
รูปร่างหน้าแฟ้ม	สะอาด	ค่อนข้างสะอาด	ค่อนข้างสะอาด	ไม่เรียบร้อย
	ชัดเจน	ชัดเจนพอควร	ชัดเจนพอสมควร	เขียนจําหน้าแฟ้ม
	สวยงาม	ตัวพิมพ์ธรรมดา	เขียนด้วยลายมือ	อ่านยาก ไม่สะดวก
	สะดวกในการใช้	สะดวกในการใช้	สะดวกในการใช้	ในการใช้
เอกสารในแฟ้ม	จัดเรียงเป็นระบบ	จัดเรียงเป็นระบบ	จัดเรียงอ่านพอเข้าใจ	ไม่ได้จัดเรียง
	อ่านเข้าใจง่าย	อ่านเข้าใจ ขาด	ขาดสาระสำคัญ	อ่านพอเข้าใจ
	เขียนชัดเจน	สาระบางตอนที่ไม่	บางอย่าง สะท้อน	ขาดสาระหลายอย่าง
	ได้สาระประเด็น	สำคัญ สะท้อน	พัฒนาการด้าน	สะท้อนพัฒนาการ
	สะท้อนพัฒนาการ	พัฒนาการด้าน	ความคิดบ้าง	ด้านความคิด
	ด้านความคิด	ความคิด		
หลักฐานสนับสนุน	สมบูรณ์ครบถ้วนมี	สมบูรณ์ครบถ้วน	ไม่สมบูรณ์	ไม่มีหลักฐาน
	ความเป็นไปได้	มีความเป็นไปได้	ขาดหลักฐาน	
	หลักฐานถูกต้อง	หลักฐานส่วนใหญ่	บางอย่าง	
		ถูกต้อง		
รวม				

ที่มา: (อุทุมพร (ทองอุไทย) จามรมาน, 2540ข)

นอกจากรูบริกแบบองค์รวม และรูบริกแบบแยกองค์ประกอบแล้วแฟ้มสะสมงานอาจใช้รูบริกแบบอื่น คือ รูบริกแบบกำหนดน้ำหนัก (weighted scoring rubric) โดยแยกเกณฑ์การประเมินออกเป็นองค์ประกอบย่อยและหากองค์ประกอบเหล่านั้นมีน้ำหนักไม่เท่ากันก็สามารถกำหนดน้ำหนักให้แก่เกณฑ์เหล่านั้นได้ตามความเหมาะสม และเป็นที่ยอมรับร่วมกันระหว่างผู้เกี่ยวข้อง (สมศักดิ์ ภูริภาตาวรรณ, 2544) ดังตัวอย่างในตารางที่ 2.12

ตารางที่ 2.12 ตัวอย่างรูปกรีกแบบกำหนดน้ำหนัก

ชื่อ.....		วัน-เดือน-ปี.....			
ชื่อวิชา.....		ชื่อ-ผู้สอน.....			
เกณฑ์	1	2	3	น้ำหนัก	คะแนน
ภาพปรากฏโดยรวม	<input type="checkbox"/> ไม่สะอาดเรียบร้อยและ ไม่มีความเป็น เอกลักษณ์ของตนเอง	<input type="checkbox"/> สะอาดเรียบร้อย มีบางอย่างแสดง ความเป็นเอกลักษณ์ ของตนเอง	<input type="checkbox"/> มีความสร้างสรรค์ แสดงความเป็น เอกลักษณ์ของตนเอง	X 1	3
การจัดระบบ	<input type="checkbox"/> วางโครงร่างของแฟ้ม สะสมงานไม่ดี ไม่สามารถติดตาม และ งานหลายชิ้นขาด หายไป	<input type="checkbox"/> วางลำดับโครงร่าง ของแฟ้มสะสมงาน ใช้ได้ แต่ยากในการ ติดตามชิ้นงานบางชิ้น ขาดหายไป	<input type="checkbox"/> วางลำดับโครงร่าง ของแฟ้มสะสมงานไว้ ชัดเจน เรียงลำดับ ชิ้นงานอย่างเป็นระบบ ง่ายต่อการติดตาม	X 3	9
รูปฟอร์ม/ลักษณะ การเขียน	<input type="checkbox"/> รูปฟอร์มของชิ้นงาน ไม่มีความหลากหลาย ลักษณะการเขียน มีข้อผิดพลาดมาก	<input type="checkbox"/> รูปฟอร์มของชิ้นงาน บางชิ้นมีความ หลากหลาย ลักษณะ การเขียนมีความ ผิดพลาดบ้าง	<input type="checkbox"/> รูปฟอร์มของชิ้นงาน มีความหลากหลาย ลักษณะการเขียนไม่มี ข้อผิดพลาด	X 3	9
ความเข้าใจใน เนื้อหา	<input type="checkbox"/> ขาดความแม่นยำ ใช้ข้อมูลพื้นฐาน ขาดความริเริ่ม	<input type="checkbox"/> ขาดความแม่นยำ บางเรื่อง จำกัดการใช้ ความคิด มีความริเริ่ม บางเรื่อง	<input type="checkbox"/> มีความแม่นยำ ประยุกต์ความคิดใหม่ ได้ แสดงความริเริ่ม	X 4	12
คุณภาพชิ้นงาน	<input type="checkbox"/> ขาดการสะท้อน ความรู้สึกที่ลุ่มลึก ขาดความช่างคิด มีการ แสดงออกไม่มาก	<input type="checkbox"/> สะท้อนความรู้สึก บางครั้งไม่ชัดเจน ความคิดไม่ต่อเนื่อง การแสดงออก ค่อนข้างจำกัด	<input type="checkbox"/> สะท้อนความรู้สึก อย่างใช้ความคิด มีความช่างคิด มีการ แสดงออกอย่างมาก	X 5	15
ข้อเสนอแนะ	สเกล			คะแนนรวม :
	A = 43-48			เกรด
	B = 38-42				
	C = 33-37				
	D = 28-32				
	F = ตั้งแต่ 27 ลงมา				

ที่มา: สมศักดิ์ ภู่วิภาตาวรรณ (2544)

2.7.4 การพัฒนาและการสร้างรูบริกการประเมิน

การพัฒนาการประเมินที่สอดคล้องกับสมรรถนะการเรียนรู้ของผู้เรียน การพัฒนาการประเมินที่สอดคล้องกับสมรรถนะการเรียนรู้ของผู้เรียน และการสอนของแต่ละคน ทั้งนี้เพื่อให้รูบริกมีความสัมพันธ์กับการเรียนรู้มากขึ้น การออกแบบรูบริกต้องทำให้นักเรียนมีส่วนร่วมในแต่ละขั้นตอน ซึ่งการพัฒนาการประเมินมี 7 ขั้น (สมศักดิ์ ภูวิภาดาบรรณ, 2544) ดังนี้

ขั้นที่ 1 เห็นรูปแบบต่างๆ (look at models) ขั้นแรกนี้ครูให้นักเรียนเห็นตัวอย่างชิ้นงานที่ดีและไม่ดีนัก ระบุคุณลักษณะที่จะทำให้ชิ้นงานดีและลักษณะที่ทำให้ชิ้นงานไม่ดี

ขั้นที่ 2 ระบุรายการที่เป็นเกณฑ์ (list criteria) ขั้นนี้เป็นการอภิปรายชิ้นงาน แล้วนำความเห็นมาสรุปเป็นเกณฑ์ที่บอกว่าชิ้นงานที่ดีเป็นอย่างไร

ขั้นที่ 3 ระบุระดับของคุณภาพ (articulate gradation of quality) ขั้นนี้เป็นการบรรยายลักษณะของชิ้นงานที่ถือว่ามีความดีที่สุด และบรรยายลักษณะชิ้นงานที่มีคุณภาพต่ำสุด จากนั้นบรรยายลักษณะที่อยู่ระหว่างกลาง

ขั้นที่ 4 ฝึกใช้เกณฑ์ (practice on models) ขั้นนี้ให้นักเรียนฝึกใช้รูบริกที่สร้างขึ้นในการประเมินชิ้นงานที่นำเสนอเป็นตัวอย่างในขั้นที่ 1

ขั้นที่ 5 ประเมินตนเองและเพื่อน (use self-and peer-assessment) ขั้นนี้ให้นักเรียนผลิตชิ้นงาน ขณะทำงานให้หยุดบางช่วงเพื่อให้นักเรียนใช้รูบริกประเมินชิ้นงานของตนเอง และของเพื่อน

ขั้นที่ 6 แก้ไข ปรับปรุง (revise) ขั้นนี้เปิดโอกาสให้นักเรียนแก้ไข ปรับปรุงชิ้นงานของตนเองจากข้อเสนอแนะที่ได้จากขั้นที่ 5

ขั้นที่ 7 ครูใช้รูบริกที่นักเรียนพัฒนาขึ้นในการประเมิน (use teacher assessment) ขั้นนี้ครูต้องใช้รูบริกที่นักเรียนพัฒนาขึ้นมาแล้วประเมินชิ้นงานของนักเรียนต่อไป

2.8 แนวคิดเกี่ยวกับการให้ข้อมูลย้อนกลับ

การให้ข้อมูลย้อนกลับของครูต่อนักเรียนเป็นกระบวนการที่มีความสำคัญต่อการประเมินระหว่างเรียน (formative assessment) ที่ทำให้เกิดความก้าวหน้าในการเรียนรู้ของนักเรียน (Tunstall & Gipps, 1996) การให้ข้อมูลย้อนกลับของครูแก่นักเรียนในการเรียนภาษาที่สองเป็นสิ่งจำเป็นเพื่อให้นักเรียนปรับปรุงและแก้ไขข้อผิดพลาดให้ถูกต้อง (corrective feedback) ต่อไป ซึ่งการให้ข้อมูลย้อนกลับเพื่อให้นักเรียนนำไปปรับปรุงนี้ TESOL (Teachers of English to Speakers of Other Languages) ได้ให้ความสำคัญมาอย่างยาวนานแล้ว (Mori, 2011)

2.8.1 ความหมายของการให้ข้อมูลย้อนกลับ

การให้ข้อมูลย้อนกลับเป็นการบอกสารสนเทศเกี่ยวกับช่องว่างระหว่างระดับของสิ่งที่เป็นอยู่จริง (actual level) และระดับที่ใช้อ้างอิง (reference level) หรือต้องการให้เป็น และเพื่อการนำข้อมูลนั้นไปใช้ในการปรับเปลี่ยนหรือแก้ไข (Ramprasad, 1983 cited in Plakht, Shiyovich, Nusbaum, & Raizer, 2012)

2.8.2 ประเภทของการให้ข้อมูลย้อนกลับ

Tunstall and Gipps (1996) ได้เสนอการให้ข้อมูลย้อนกลับไว้ 8 ประเภท ได้แก่

แบบ A1 การให้รางวัล เป็นการให้ข้อมูลย้อนกลับจากครูในเชิงที่เป็นบวกมากที่สุด เช่น การแสดงออกด้วยใบหน้าที่ยิ้มแย้มของครูต่อนักเรียน การให้สติ๊กเกอร์ แสตมป์ ดาว เป็นต้น

แบบ A2 การลงโทษ เป็นการให้ข้อมูลย้อนกลับจากครูในเชิงที่เป็นลบมากที่สุด แสดงถึงการไม่ยอมรับ เช่น ครูยึดสติ๊กเกอร์ แสตมป์ ดาว ที่นักเรียนเคยได้คืน ให้นักเรียนออกไปอยู่นอกห้องเรียน เป็นต้น

แบบ B1 การยอมรับ เป็นการให้ข้อมูลย้อนกลับจากครูในเชิงบวกที่แสดงออกถึงความเห็นอย่างอบอุ่น ให้การยอมรับแก่นักเรียน เช่น การสัมผัส การแสดงออกทางสีหน้าในเชิงยิ้มแย้ม การทำเครื่องหมายถูก (✓) ในชิ้นงาน และการกล่าวชมเชย เช่น ดีมาก ดี ถูกต้อง เป็นต้น

แบบ B2 การไม่ยอมรับ เป็นการให้ข้อมูลย้อนกลับจากครูในเชิงที่เป็นลบทั้งการแสดงออกและคำพูด เช่น นักเรียนควรทำให้ดีกว่านี้ ครูผิดหวังมากๆ กับงานชิ้นนี้ เป็นต้น

แบบ C1 การระบุความสำเร็จ เป็นการให้ข้อมูลย้อนกลับจากครูในเชิงบรรยายและเฉพาะเจาะจง โดยระบุเกณฑ์ที่เฉพาะเจาะจงสำหรับการได้มาซึ่งความสำเร็จ ซึ่งจะสนับสนุนความสำเร็จของนักเรียน เช่น นักเรียนทำได้ดีมาก เพราะนักเรียน..... เป็นต้น

แบบ C2 การระบุสิ่งที่ควรปรับปรุง เป็นการให้ข้อมูลย้อนกลับจากครูในเชิงบรรยาย โดยการระบุข้อผิดพลาดที่เฉพาะเจาะจง ว่าข้อผิดพลาดคืออะไร ควรแก้ไขอย่างไร ซึ่งจะสนับสนุนให้นักเรียนบรรลุในการเรียน โดยจะเน้นความผิดพลาดเกี่ยวกับผลสัมฤทธิ์ของนักเรียนมากกว่าความผิดพลาดด้านบุคลิกภาพ เช่น นักเรียนยังออกเสียงบางคำไม่ดี นักเรียนควรแก้ไขเป็น..... เป็นต้น

แบบ D1 การสร้างความสำเร็จ เป็นการให้ข้อมูลย้อนกลับจากครูและนักเรียนร่วมกัน โดยจะแทรกในบทสนทนาหรือแลกเปลี่ยนความคิดเห็นของครูและนักเรียน เพื่อสะท้อนความก้าวหน้าในการทำงาน ทั้งนี้ผู้สอนจะเสริมให้เกิดกระบวนการเรียนรู้ และชี้แนะผู้เรียนให้สามารถแสดงความสำเร็จจากการทำงานได้ด้วยตนเอง

แบบ D2 การสร้างวิธีการที่ก้าวหน้าเป็นการให้ข้อมูลย้อนกลับจากครูและนักเรียนร่วมกันโดยพิจารณาว่าร่วมกัน ครูมีลักษณะเป็นหุ้นส่วนทางการเรียนรู้กับนักเรียน

การให้ข้อมูลย้อนกลับทั้ง 8 วิธี Tunstall and Gipps (1996) ได้แบ่งออกเป็น 6 มิติ คือ การให้ข้อมูลย้อนกลับเชิงบวก และเชิงลบ การให้ข้อมูลย้อนกลับเชิงความสำเร็จและข้อมูลย้อนกลับที่บอกการปรับปรุง การให้ข้อมูลย้อนกลับเกี่ยวกับการประเมิน และการบรรยาย สามารถสรุปได้ดังตารางที่ 2.13

ตารางที่ 2.13 ประเภทของการให้ข้อมูลย้อนกลับ

	แบบ A	แบบ B	แบบ C	แบบ D	
1 เชิงบวก	การให้รางวัล	การยอมรับ	การระบุความสำเร็จ	การสร้างความสำเร็จ	1 เชิงความสำเร็จ
2 เชิงลบ	การลงโทษ	การไม่ยอมรับ	การระบุสิ่งที่ควรปรับปรุง	การสร้างวิธีการที่ก้าวหน้า	2 การปรับปรุง
	การประเมิน		การบรรยาย		

นอกจากนี้แล้วยังพบว่าการให้ข้อมูลย้อนกลับตามเป้าหมายของการสะท้อนกลับสามารถแบ่งการให้ข้อมูลย้อนกลับได้ 2 วิธีการ ได้แก่ 1) การให้ข้อมูลย้อนกลับต่องานปัจจุบัน (feedback) และการให้ข้อมูลย้อนกลับเพื่อปรับปรุงงานครั้งต่อไป (feedforward) (Hounsell, 2007 and Garcia-Sanpedro, 2012 อ้างอิงใน โชติกา ภาชีผล และคณะ, 2558)

ซึ่งสำหรับการวิจัยครั้งนี้ผู้วิจัยใช้การให้ข้อมูลย้อนกลับแบบเชิงบรรยายที่ให้ข้อมูลย้อนกลับแบบการระบุความสำเร็จ (achievement feedback) และการระบุสิ่งที่ควรปรับปรุง (improvement feedback) และแก้ไขให้ถูกต้อง (corrective feedback) เพื่อเป็นการเสริมแรงในการเรียน รวมทั้งรู้เกี่ยวกับสิ่งที่จะนำตนเองไปสู่ความสำเร็จในการเรียน และรู้ถึงข้อผิดพลาดและแก้ไขให้ถูกต้อง (correction of errors)

2.9 งานวิจัยที่เกี่ยวกับการประเมินด้วยแฟ้มสะสมงาน

Tasdemir Tasdemid and Yildirim (2008) ได้ทำวิจัยโดยการใช้แฟ้มสะสมงานกับการเรียนรู้แบบร่วมมือ กลุ่มตัวอย่างที่ใช้ในการทดลอง เป็นนักศึกษาคณะครุศาสตร์ ในมหาวิทยาลัย Ahi Evran University ชั้นปีที่ 2 จำนวน 81 คน การดำเนินการทดลองแบ่งนักศึกษาออกเป็น 3 กลุ่ม กลุ่มที่ 1 เรียนโดยใช้วิธีการประเมินด้วยแฟ้มสะสมงานกับกระบวนการเรียนแบบร่วมมือ กลุ่มที่ 2 เรียนโดยใช้วิธีการประเมินด้วยแฟ้มสะสมงานเพียงอย่างเดียว และกลุ่มที่ 3 เรียนโดยวิธีการปกติ (กลุ่มควบคุม) วิธีการดำเนินการในการประเมินด้วยแฟ้มสะสมงานมี 6 ชั้น ได้แก่ 1) การกำหนดจุดประสงค์การเรียนรู้ก่อนเริ่มเรียน โดยนักศึกษาร่วมกันกำหนด 2) ให้ความรู้แก่นักศึกษาเกี่ยวกับ

การเรียนรู้แบบร่วมมือ และตัวอย่างผลงาน การให้รางวัลกลุ่ม การประเมินเป็นรายบุคคล การปฏิสัมพันธ์แบบเผชิญหน้า และการประเมินกระบวนการกลุ่ม 3) ให้ความรู้นักศึกษาเกี่ยวกับขั้นตอนการประเมินในแต่ละขั้นตอน 4) ฝึกปฏิบัติ 10 สัปดาห์ โดยครูเป็นผู้ให้คำแนะนำ เกี่ยวกับกิจกรรมการเรียนรู้ การประเมินตนเอง 5) นักศึกษารวบรวมผลงาน ทำรายงาน และรายงานผลงานในชั้นเรียน 6) นักศึกษาทำกิจกรรมตามที่ฝึก เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล ได้แก่ แบบสอบวัดผลสัมฤทธิ์ทางการเรียน

ผลการวิจัยพบว่า 1) นักศึกษาที่เรียนโดยใช้วิธีการประเมินด้วยแฟ้มสะสมงานกับกระบวนการเรียนแบบร่วมมือมีผลสัมฤทธิ์ทางการเรียนสูงกว่านักศึกษาที่เรียนด้วยวิธีการปกติ 2) นักศึกษาที่เรียนโดยใช้วิธีการประเมินด้วยแฟ้มสะสมงานเพียงอย่างเดียวมีผลสัมฤทธิ์ทางการเรียนสูงกว่ากลุ่มที่เรียนโดยวิธีการปกติ (กลุ่มควบคุม) และ 3) และนักศึกษาที่เรียนโดยใช้วิธีการประเมินด้วยแฟ้มสะสมงานกับกระบวนการเรียนแบบร่วมมือ และนักศึกษาที่เรียนโดยใช้วิธีการประเมินด้วยแฟ้มสะสมงานเพียงอย่างเดียวมีผลสัมฤทธิ์ทางการเรียนไม่แตกต่างกัน

Lam และ Lee (2010) ได้ทำวิจัยเพื่อพัฒนาแฟ้มสะสมงานการเขียนภาษาอังกฤษ กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นนักศึกษาที่เรียนในระดับ post-secondary ของ Hong Kong University วิชาเอกภาษาอังกฤษ อายุ 18-20 ปี จำนวน 31 คน โดยนักศึกษาทุกคนใช้ภาษาจีนกวางตุ้งเป็นภาษาหลัก มีความสามารถทางภาษาอังกฤษอยู่ในระดับ intermediate (IELTS 5.5) และมีความสามารถในทักษะการเขียนอยู่ในระดับอ่อน การดำเนินการของแฟ้มสะสมงานใช้แนวคิดของ Hamp-Lyons and Condon (2000) ที่ประกอบด้วย 3 ขั้นตอน คือ การรวบรวมผลงาน การคัดเลือกผลงาน และการสะท้อนคิด การพัฒนาแฟ้มสะสมงานครั้งนี้ผู้วิจัยเน้นการประเมินผู้เรียนให้สมดุลระหว่างการประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) โดยในขั้นตอนของการประเมินระหว่างเรียนได้เพิ่มขั้นตอนบางขั้นตอนเข้าไป คือ การให้ข้อมูลย้อนกลับของครู (ongoing teacher feedback) การปรึกษาระหว่างครูและผู้เรียน (conferencing) และการประเมินจากเพื่อนในชั้นเรียน (peer review) การดำเนินการวิจัยมีดังนี้ 1) ภาระงานที่ใช้ให้นักศึกษาเขียน ได้แก่ narration illustration และ persuasion 2) นักศึกษาคัดเลือกผลงานที่ดีที่สุด 2 ชิ้นใส่ในแฟ้มสะสมงาน ซึ่งเป็นฉบับร่าง (draft) สุดท้ายของแต่ละชิ้น 3) นักศึกษาสะท้อนคิดต่องานของตน 4) ครูทำการประเมินแฟ้มสะสมงานในภาพรวมเพื่อให้ระดับคะแนน (grading) โดยในการเขียนผลงานทั้ง 3 ชิ้น นักศึกษาจะต้องเขียนฉบับละหลายครั้ง มีการปรับแก้ตามคำแนะนำจากผู้สอน และเพื่อน (peer review)

ผลการวิจัย พบว่า นักศึกษามีความคิดเห็นต่อการประเมินระหว่างเรียนโดยใช้กระบวนการของแฟ้มสะสมงานในทางบวก โดยนักศึกษาและครูเห็นว่าแฟ้มสะสมงานที่ใช้การประเมินระหว่างเรียนมีประโยชน์มากและทำให้นักศึกษาสามารถพัฒนางานเขียนขึ้นได้ และยังคงเห็นว่าการประเมินในส่วน

ของการสรุปรวมเป็นสิ่งมีประโยชน์เช่นกันเพราะเชื่อว่าระดับคะแนนเป็นสิ่งที่บอถึงความสามารถในการเขียนของตนในปัจจุบันได้

Lo (2010) ได้วิจัยเพื่อศึกษาผลการใช้แฟ้มสะสมงานเพื่อพัฒนาการเรียนการสอนในหลักสูตร Journalistic English กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้คือ นักศึกษาระดับปริญญาตรีที่เข้าร่วมโครงการวิจัย ที่เรียนวิชาเอกภาษาอังกฤษ จำนวน 101 คน ของ National University ในประเทศไต้หวัน ขั้นตอนของแฟ้มสะสมงานที่พัฒนาขึ้นประกอบด้วย 7 ขั้นตอน ดังนี้ 1) วิเคราะห์ข้อมูลจากแบบสอบถามก่อนเรียน (สอบถามผู้เรียนเกี่ยวกับพฤติกรรมการอ่านหนังสือพิมพ์ภาษาอังกฤษ) 2) ตัดสินใจเลือกรูปแบบแฟ้มสะสมงาน 3) พัฒนาจุดประสงค์การเรียนรู้ 4) สร้างเกณฑ์การประเมิน 5) เตรียมความพร้อมของนักเรียน 6) กำกับติดตามความก้าวหน้าของนักเรียน และ 7) ประเมินแฟ้มสะสมงาน

ผลการวิจัยพบว่า นักศึกษายังไม่มีประสบการณ์ในการรวบรวมงานที่เป็นแฟ้มสะสมงาน และยังไม่เกิดการเรียนรู้ที่เกิดขึ้นเอง (learning autonomous) ส่วนครูพบว่าการดำเนินการยังมีข้อบกพร่องหลายประการ ได้แก่ นักเรียนไม่มีบทบาทในการเรียนเท่าที่ควร ครูไม่ได้ให้นักศึกษามีส่วนร่วมในการตัดสินใจในการเรียน ดังนั้นผู้วิจัยจึงมีข้อเสนอแนะบทบาทของครู ดังนี้ เป็นผู้อำนวยความสะดวกให้การสนับสนุนแหล่งข้อมูลเพื่อให้นักเรียนเกิดการเรียนรู้ขึ้นเอง นักเรียนควรได้บริหารเวลาในการทำภาระงาน (tasks) ด้วยตนเอง

Barabouti (2012) ได้วิจัยผลการใช้แฟ้มสะสมงานเพื่อการประเมินทักษะการเขียนภาษาอังกฤษของนักเรียนระดับประถมศึกษา กลุ่มตัวอย่างในการวิจัยเป็นนักเรียนที่กำลังเรียนในระดับประถมศึกษา อายุ 11-12 ปี ทักษะที่ประเมินคือทักษะการเขียนภาษาอังกฤษ วิธีดำเนินการของแฟ้มสะสมงานเพื่อการประเมินทักษะการเขียนภาษาอังกฤษ มีการดำเนินการ 6 ขั้นตอน ได้แก่ 1) การกำหนดวัตถุประสงค์ของแฟ้มสะสมงาน 2) กำหนดหรือระบุข้อมูลที่จะใช้ในการประเมินในแฟ้มสะสมงาน คือ เนื้อหา และรูปแบบ การวินิจฉัยจุดแข็งจุดอ่อนนักเรียน 3) การจับคู่เนื้อหาเกี่ยวกับกิจกรรม 4) บันทึกความก้าวหน้าในการเขียนของนักเรียน 5) การให้นักเรียนเข้าร่วมการดำเนินการ ได้แก่ การประเมินตนเอง การประเมินเพื่อน และการสะท้อนคิด 6) การพูดคุยเกี่ยวกับวัตถุประสงค์และผลการเรียน มีการนำเสนอผลงานของนักเรียนแก่ผู้ปกครองนักเรียน

ผลการวิจัยพบว่า นักเรียนมีประสิทธิผลในทักษะการพูดในทางที่ดีคือมีการพัฒนาขึ้น นักเรียนมีการวางแผนการเรียนและจัดการกับการเรียนของตน และนักเรียนมีการปรับปรุงการเขียนของตนเองตามข้อเสนอแนะที่ได้รับจากข้อมูลย้อนกลับ (feedback) จากครู นักเรียนมีเจตคติที่ดีต่อการเขียนและการประเมินตนเอง

Yilmaz and Akcan (2012) ได้ทำการวิจัยการใช้แฟ้มสะสมงานทางภาษาของยุโรป (the European Language Portfolio: ELP) ในบริบทของประเทศตุรกี กลุ่มตัวอย่างที่ใช้ในการวิจัย

เป็นนักเรียนเกรด 4 และ 5 อายุ 9-11 ปี ระดับชั้นละ 22 คน ครูผู้สอน จำนวน 2 คน ที่ไม่ใช่เจ้าของภาษา (native speaker) โดยเพิ่มสะสมงานภาษาอังกฤษมีการดำเนินการ 5 ขั้นตอน ได้แก่ 1) การให้นักเรียนตระหนักในเป้าหมายการเรียนรู้ (raising awareness) 2) การเกาะติดกับเป้าหมาย (goal tracking) 3) การสร้างตัวเลือก (making choice) 4) การสะท้อนคิดต่อผลงานของตนเอง (reflection) และ 5) การประเมินตนเอง (self-assessment)

ผลการวิจัย พบว่า นักเรียนและครู ทั้ง 2 ระดับ มีเจตคติทางบวกต่อ ELP โดยครูกล่าวว่า ELP เป็นนวัตกรรมสำหรับการเรียนรู้ทางภาษาที่ทำให้นักเรียนระดับประถมศึกษาที่มีประสบการณ์ทางภาษาที่ดีและช่วยพัฒนานักเรียนในระดับประเทศได้ด้วย ส่วนนักเรียนพบว่าทำให้นักเรียนมีทักษะการสะท้อนคิดและการประเมินตนเอง

Kouzouli (2012) ได้ทำวิจัยเพื่อศึกษาผลการใช้เพิ่มสะสมงานเพื่อการประเมินการเรียนรู้สำหรับนักเรียนเกรด 3 ในโรงเรียนประถมศึกษาของรัฐบาล ประเทศกรีซ กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นโรงเรียนประถมศึกษาของรัฐบาล แนวคิดของเพิ่มสะสมงานที่ใช้ในการวิจัยใช้แนวคิดของ Kemp and Toperoff (1998) and Tsagari (2005) มีขั้นตอนดังนี้ 1) การสร้างวัตถุประสงค์การเรียนรู้ 2) การสร้างเนื้อหาของเพิ่มสะสมงานและรูปแบบ 3) การสร้างความเป็นเจ้าของ 4) การแนะนำแนวคิดของเพิ่มสะสมงาน 5) แจกผู้สนใจอื่นๆ (ผู้บริหารโรงเรียน ผู้ปกครองนักเรียนให้เข้าร่วม) 6) กำหนดแนวทางของเพิ่มสะสมงาน 7) จัดเตรียมความพร้อมแก่นักเรียน 8) การประเมินเพิ่มสะสมงาน 9) การติดตามผล (follow-up) และ 10) การใช้ผลของเพิ่มสะสมงาน

ผลการวิจัยพบว่า การประเมินโดยเพิ่มสะสมงานเป็นวิธีการที่เหมาะสมกับนักเรียนในวัยเด็ก (young learner) นักเรียนมีเจตคติที่ดีต่อการเรียน สามารถประเมินตนเองและประเมินเพื่อนในชั้นได้ มีการสะท้อนคิดต่อผลงานของตนเองได้ ทำให้เกิดการเรียนรู้ที่เป็นอิสระ มีความเข้าใจในปัญหาที่เกิดขึ้นในการเรียนได้อย่างถ่องแท้และสามารถจัดการกับปัญหาได้

ชัยฤทธิ์ ศิลาเดช (2540) ได้พัฒนาเพิ่มสะสมงานในการประเมินผลการเรียนวิชาภาษาอังกฤษระดับชั้นมัธยมศึกษาปีที่สาม กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 43 คน ทักษะที่มุ่งประเมินได้แก่ ความสามารถในการใช้ภาษาอังกฤษ และความก้าวหน้าในการเรียนภาษาอังกฤษในทักษะการฟัง ทักษะการพูด ทักษะการอ่าน และทักษะการเขียน เครื่องมือที่ใช้ในการวิจัยประกอบด้วย แบบสังเกตและวิเคราะห์พฤติกรรมการสอนของครู พฤติกรรมการเรียนรู้ของนักเรียนแบบตรวจสอบรายการใช้กระบวนการทักษะการฟังและการพูด ทักษะการอ่านและเขียน แบบบันทึกการสังเกตกระบวนการอ่านและการเขียน แบบบันทึกการสนทนาเกี่ยวกับเพิ่มสะสมงานรายบุคคล บันทึกประจำวัน บันทึกเสียงการสอน แบบสอบถามความคิดเห็นผู้ปกครอง แบบประเมินความก้าวหน้าทักษะการฟัง ทักษะการพูด ทักษะการอ่าน และทักษะการเขียน รูปแบบของเพิ่มสะสมงานประกอบด้วย 4 ส่วน คือ 1) ส่วนนำ ประกอบด้วย ข้อมูลนักเรียน สารบัญ จุดประสงค์

การเรียนรู้ ข้อตกลง เกณฑ์การประเมินผลงาน 2) ส่วนเนื้อหา ประกอบด้วย ผลงานหรือหลักฐาน ต่างๆ ที่ผู้เรียนคัดเลือก 3) ส่วนสนับสนุน ประกอบด้วย แผนพัฒนาและปรับปรุงการเรียนรู้สรุปผล ความก้าวหน้าในการเรียนเป็นรายเดือน ผลการสอบ ผลการสังเกต และประกาศนียบัตร และ 4) สรุปผลการวิเคราะห์ความเชื่อมั่นของคะแนนนักเรียนแต่ละคนใช้วิธีการสรุปอ้างอิงความน่าเชื่อถือของการวัดผล (generalization theory) โดยใช้ $p \times e \times r$ Design กำหนดชิ้นงานเป็น 4 8 12 16 20 24 และ 28 ชิ้น กำหนดผู้ประเมินเป็น 1-5 คน ผลการวิจัยพบว่าการประเมินโดยเพิ่มสะสมงานวิชา ภาษาอังกฤษ ประกอบด้วย 1) สภาวะแวดล้อม 2) จุดมุ่งหมายของเพิ่มสะสมงาน 3) โครงสร้างของ เพิ่มสะสมงาน 4) การพัฒนาเพิ่มสะสมงาน 5) การประเมินเพิ่มสะสมงาน และ 6) การประชาสัมพันธ์ร่วม ชมผลงานดีเด่น ความเชื่อมั่นของคะแนนนักเรียนตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของการ วัดผล จากผลการศึกษา D (D-study) พบว่าเมื่อใช้ผู้ประเมินจำนวน 2 คน และจำนวนชิ้นงานทักษะละ 4 ชิ้น รวมทั้งสิ้น 16 ชิ้น มีค่าเท่ากับ .6517 และถ้าใช้ผู้ประเมินเพียงคนเดียวความน่าเชื่อถือจะต่ำ และ จำนวนผู้ประเมินที่เหมาะสมคือ 2-3 คน ส่วนชิ้นงานพบว่าเมื่อใช้ชิ้นงานมากขึ้นเกิน 16 ชิ้นงาน มีผลทำให้ค่าความน่าเชื่อถือเพิ่มขึ้นเพียงเล็กน้อย

กนกพงศ์ จิตต์ปลื้ม (2549) ได้ศึกษาผลการใช้เพิ่มสะสมงานอิเล็กทรอนิกส์ประกอบการ เขียนภาษาอังกฤษเพื่อการสื่อสารที่มีต่อผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนจุฬารัตน์ราชวิทยาลัย เพชรบุรี ปีการศึกษา 2549 จำนวน 55 คน โดยแบ่งเป็นกลุ่มทดลอง 27 คน และกลุ่มควบคุม 28 คน เครื่องมือที่ใช้ ในการวิจัย ได้แก่เพิ่มสะสมงานอิเล็กทรอนิกส์ โปรแกรมการสร้างเพิ่มสะสมงานอิเล็กทรอนิกส์ แผนการสอน แบบประเมินผลงานดีเด่น และแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน โดยกระบวนการ ในเพิ่มสะสมงานอิเล็กทรอนิกส์ประกอบด้วย 1) การกำหนดเป้าหมาย 2) การสร้างผลงานและรวบรวม ผลงาน 3) การคัดเลือกผลงาน 4) การสะท้อนคิด 5) การประเมินงาน และ 6) การประชาสัมพันธ์ ผู้วิจัยร่วมกับนักเรียนเพื่อสร้างเกณฑ์ประเมินคุณภาพผลงานโดยใช้มาตราส่วนประเมินค่า (rating scale) 4 ระดับ คือ ดีมาก ดี พอใช้ และปรับปรุง เกณฑ์ที่ใช้ในการประเมินการเขียนมี 5 รายการ คือ เนื้อหา (content) คำศัพท์ (vocabulary) การเรียบเรียง (organization) ไวยากรณ์ (grammar) และการใช้เครื่องหมายวรรคตอน (mechanics) ผลการวิจัยพบว่า นักเรียนที่เรียนโดยใช้เพิ่มสะสมงาน อิเล็กทรอนิกส์ประกอบการเขียนภาษาอังกฤษเพื่อการสื่อสารและนักเรียนที่ไม่ได้เรียนโดยใช้ เพิ่มสะสมงานอิเล็กทรอนิกส์ประกอบการเขียนภาษาอังกฤษเพื่อการสื่อสารมีผลสัมฤทธิ์ทางการเรียน ไม่แตกต่างกัน ส่วนนักเรียนที่เรียนโดยใช้เพิ่มสะสมงานอิเล็กทรอนิกส์ประกอบการเขียนภาษาอังกฤษ เพื่อการสื่อสารมีผลสัมฤทธิ์ทางการเรียนสูงขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

โชติกา ภาชีผล และคณะ (2558) ได้ทำการวิจัยเพื่อพัฒนารูปแบบเพิ่มสะสมงาน อิเล็กทรอนิกส์ที่สะท้อนข้อมูลย้อนกลับในการประเมินพัฒนาการเรียนรู้ของนักเรียนในสังกัด

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน โดยให้การสะท้อนข้อมูลย้อนกลับแบบทั่วไป (feedback) และการสะท้อนข้อมูลชี้แนะเพื่อการปรับปรุง (feedforward) ผลวิจัยพบว่า แฟ้มสะสมงานอิเล็กทรอนิกส์ที่สะท้อนข้อมูลย้อนกลับสำหรับนักเรียนฯ ประกอบด้วย 6 องค์ประกอบ ดังนี้

- 1) จุดมุ่งหมายของแฟ้มสะสมงานอิเล็กทรอนิกส์
- 2) เนื้อหาของแฟ้มสะสมงานอิเล็กทรอนิกส์
- 3) บทบาทผู้เกี่ยวข้องกับแฟ้มสะสมงานอิเล็กทรอนิกส์
- 4) เครื่องมือในการพัฒนาแฟ้มสะสมงานอิเล็กทรอนิกส์
- 5) การให้ข้อมูลย้อนกลับ และ
- 6) การประเมินผลแฟ้มสะสมงานอิเล็กทรอนิกส์

และขั้นตอนการพัฒนาแฟ้มสะสมงานประกอบด้วย 7 ขั้นตอน ดังนี้

- 1) กำหนดจุดมุ่งหมาย เนื้อหา และเกณฑ์การประเมินแฟ้มสะสมงาน
- 2) เก็บรวบรวมผลงานหรือหลักฐานในแฟ้มสะสมงานระหว่างดำเนินการ
- 3) สะท้อนความคิดต่อผลงานประเมินตนเอง และให้ข้อมูลย้อนกลับ
- 4) ประเมินและให้ข้อมูลย้อนกลับโดยเพื่อน ครู และผู้ปกครอง
- 5) คัดเลือกผลงานเพื่อจัดทำแฟ้มสะสมงานอิเล็กทรอนิกส์สำหรับนำเสนอ
- 6) เผยแพร่แฟ้มสะสมงานอิเล็กทรอนิกส์สำหรับนำเสนอ และ
- 7) ประเมินแฟ้มสะสมงานอิเล็กทรอนิกส์สำหรับนำเสนอ

ผลด้านพัฒนาการเรียนรู้ของนักเรียนพบว่านักเรียนระดับประถมศึกษา และมัธยมศึกษาที่เป็นกลุ่มตัวอย่างมีพัฒนาการเรียนรู้ในวิชาภาษาไทยและวิทยาศาสตร์สูงขึ้น โดยสำหรับนักเรียนประถมศึกษาพบว่าการใช้แฟ้มสะสมงานอิเล็กทรอนิกส์ที่สะท้อนข้อมูลย้อนกลับแบบทั่วไป ทำให้นักเรียนส่วนใหญ่ของนักเรียนที่มีพัฒนาการต่ำกว่าไม่มีพัฒนาการ ในขณะที่การใช้แฟ้มสะสมงานอิเล็กทรอนิกส์ที่สะท้อนข้อมูลชี้แนะเพื่อการปรับปรุงร้อยละของนักเรียนที่มีพัฒนาการสูงกว่าไม่มีพัฒนาการ ส่วนนักเรียนระดับมัธยมศึกษาพบว่าการใช้แฟ้มสะสมงานอิเล็กทรอนิกส์ที่สะท้อนข้อมูลย้อนกลับแบบทั่วไปและแฟ้มสะสมงานอิเล็กทรอนิกส์ที่สะท้อนข้อมูลชี้แนะเพื่อการปรับปรุงร้อยละของนักเรียนที่มีพัฒนาสูงกว่าไม่มีพัฒนาการ

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการประเมินโดยแฟ้มสะสมงานพบว่ามีกรนำมาใช้ในการจัดการเรียนรู้ในหลายระดับตั้งแต่ระดับประถมศึกษาจนถึงระดับอุดมศึกษา และพบว่ามีกรนำมาใช้ในการประเมินทางภาษาอย่างแพร่หลาย โดยพบว่าองค์ประกอบและขั้นตอนที่นำมาใช้มีลักษณะที่คล้ายกัน ได้แก่ การรวบรวมผลงาน การคัดเลือกงานตัวอย่าง การประเมินตนเอง การประเมินจากครู การประเมินจากเพื่อน และการสะท้อนคิด นอกจากนี้ยังพบว่าแฟ้มสะสมงานมีการนำกระบวนการประเมินระหว่างเรียน (formative assessment) เข้ามาทำให้นักเรียนมีพัฒนาการและความก้าวหน้าในการเรียนตามเป้าหมายที่วางไว้ รวมทั้งมีความสามารถในการประเมินตนเอง ประเมินเพื่อนในชั้นเรียน ได้ดียิ่งขึ้นด้วย ส่วนการประเมินในภาพรวม (summative assessment) มีความสำคัญคือทำให้นักเรียนทราบถึงระดับความสามารถของตนเองนั้นอยู่ในระดับใดและจะมีการพัฒนาได้อย่างไร

ตอนที่ 3 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ (G theory) มีกรอบแนวคิดที่กว้างขวางและเป็นเครื่องมือทางสถิติที่มีประสิทธิภาพสำหรับประเด็นการวัดจำนวนมาก ซึ่ง G theory มีที่มาจากทฤษฎีการวัดผลแบบดั้งเดิม (classical test theory: CTT) และการวิเคราะห์ของความแปรปรวน (analysis of variance) (Robert L. Brennan, 2011) แต่เนื่องจาก CTT เป็นการประมาณค่าความเที่ยงภายใต้เงื่อนไขหรือสถานการณ์ที่เฉพาะ หากสถานการณ์แตกต่างกันไปจะไม่สามารถสรุปอ้างอิงได้ รวมทั้ง CTT ถือว่าความคลาดเคลื่อนของการวัดเป็นคะแนนความคลาดเคลื่อนรวมอันเดียวกันไม่สามารถแบ่งแยกได้ (unique error) ส่วน G theory นั้นได้เสนอโมเดลการศึกษาการวิเคราะห์ความน่าเชื่อถือหรือความเที่ยงให้ครอบคลุมและแบ่งแยกความคลาดเคลื่อนออกเป็นหลายแหล่ง ทำให้ผู้ออกแบบการวัดได้ผลการวัดที่น่าเชื่อถือมากขึ้น และมีความเที่ยงสูงในระดับที่ต้องการ (Shavelson & Webb, 2005; Brennan, 2011) แต่เนื่องจากข้อมูลที่มีอยู่จริงในธรรมชาติไม่ได้เป็นเพียงตัวแปรเดียว (univariate) เท่านั้น แต่ยังเป็นข้อมูลที่เป็นแบบพหุตัวแปร (multivariate) ดังนั้นจึงได้มีการขยายแนวคิดของการศึกษา G-theory ออกไปในรูปแบบพหุตัวแปรหรือทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร (multivariate generalizability theory: MGT) Cronbach Gleser Nanda and Rajaratnum ได้เป็นผู้เสนอแนวคิดนี้ไว้ในปี ค.ศ. 1972 และมีผู้พัฒนาวิธีการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิง (G-coefficient) และพัฒนาโปรแกรมสำหรับวิเคราะห์ข้อมูลให้มีความสะดวกในการวิเคราะห์ข้อมูลยิ่งขึ้นจึงทำให้การนำทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรมาใช้ในการวัดและประเมินผลต่างๆ แพร่หลายมากขึ้น

การศึกษาทบทวนเอกสารเกี่ยวกับการวิเคราะห์พหุตัวแปรของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ เพื่อทำความเข้าใจในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือทั้งในกรณีตัวแปรเดียว (G-theory) และหลายตัวแปร (multivariate G- theory) เพื่อที่จะนำทฤษฎีไปประยุกต์ใช้ โดยเนื้อหาสาระนำเสนอเป็น 3 ส่วน ได้แก่ (1) ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ (2) การวิเคราะห์พหุตัวแปรของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ และ (3) งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์พหุตัวแปรของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ รายละเอียดของเนื้อหาสาระแต่ละส่วนเป็นดังนี้

3.1 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ

การศึกษาเกี่ยวกับ G-theory ผู้ศึกษาได้แบ่งออกเป็น 5 ประเด็น ได้แก่ (1) นิยามศัพท์ที่เกี่ยวข้องกับ G-theory (2) ข้อตกลงเบื้องต้นของ G-theory (3) ขั้นตอนการศึกษา ทั้งการศึกษาเชิงสรุปอ้างอิง (generalizability study, G-study) และการศึกษาเชิงตัดสินใจ (decision study, D-study) (4) สัมประสิทธิ์การสรุปอ้างอิง (generalizability coefficient, G-coefficient) และ (5) การออกแบบการวัดใน G-theory

3.1.1 นิยามศัพท์ที่เกี่ยวข้องกับ G-theory (ศิริชัย กาญจนวาสี, 2550; Brennan, 2001, Shavelson & Webb, 2004) (ศิริชัย กาญจนวาสี, 2552)

ในการศึกษา G-theory มีคำศัพท์ที่เป็นศัพท์เฉพาะ และคำศัพท์ที่เกี่ยวข้อง ดังนี้ ประชากร (population) หมายถึง สิ่งที่มีงวัดทั้งหมด ในสถานการณ์ของการทดสอบหรือการประเมิน ซึ่งโดยทั่วไปสิ่งที่มีงวัดหรือประเมินมักได้แก่ คน หรือ ผู้ที่ถูกประเมิน

เอกภพ (universe) หมายถึง เงื่อนไขของการวัดที่สนใจทั้งหมดในสถานการณ์ของการทดสอบหรือการประเมิน

ฟาเซท (facet) หมายถึง กลุ่มเงื่อนไขของการวัดหรือการประเมินที่คาดว่าจะมีผลต่อความคลาดเคลื่อนของการวัด ซึ่งฟาเซทที่ต้องการศึกษาอาจเป็นองค์ประกอบสุ่ม (random) หรือองค์ประกอบเจาะจง (fixed)

เอกภพของการสรุปอ้างอิง (universe of generalizability) หมายถึง เงื่อนไขหรือสถานการณ์ที่ครอบคลุมสิ่งที่สนใจของการวัดทั้งหมดอันเป็นเป้าหมายของการสรุปอ้างอิงของแบบสอบ

รูปแบบความสัมพันธ์ระหว่างองค์ประกอบ หมายถึง องค์ประกอบที่เป็นสิ่งที่ชี้ถึงแหล่งความแปรปรวนของรูปแบบการวัด ในแบบจำลองการวิเคราะห์ความแปรปรวน ซึ่งรูปแบบความสัมพันธ์นี้แบ่งออกเป็น 3 ลักษณะ คือ

1. ความสัมพันธ์แบบไขว้ (crossed) หมายถึง ความสัมพันธ์แต่ละลักษณะที่แต่ละระดับของสิ่งที่ถูกวัด ถูกวัดภายใต้เงื่อนไขเดียวกันทั้งหมด สัญลักษณ์คือ "X" อ่านว่า "crossed with" เช่น กำหนดรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น $p \times i \times r$ หมายถึง นักเรียน (p) ทำข้อสอบ (i) ทุกข้อ และผู้ตรวจ (r) ตรวจข้อสอบของนักเรียนทุกคน ทุก ๆ ข้อ

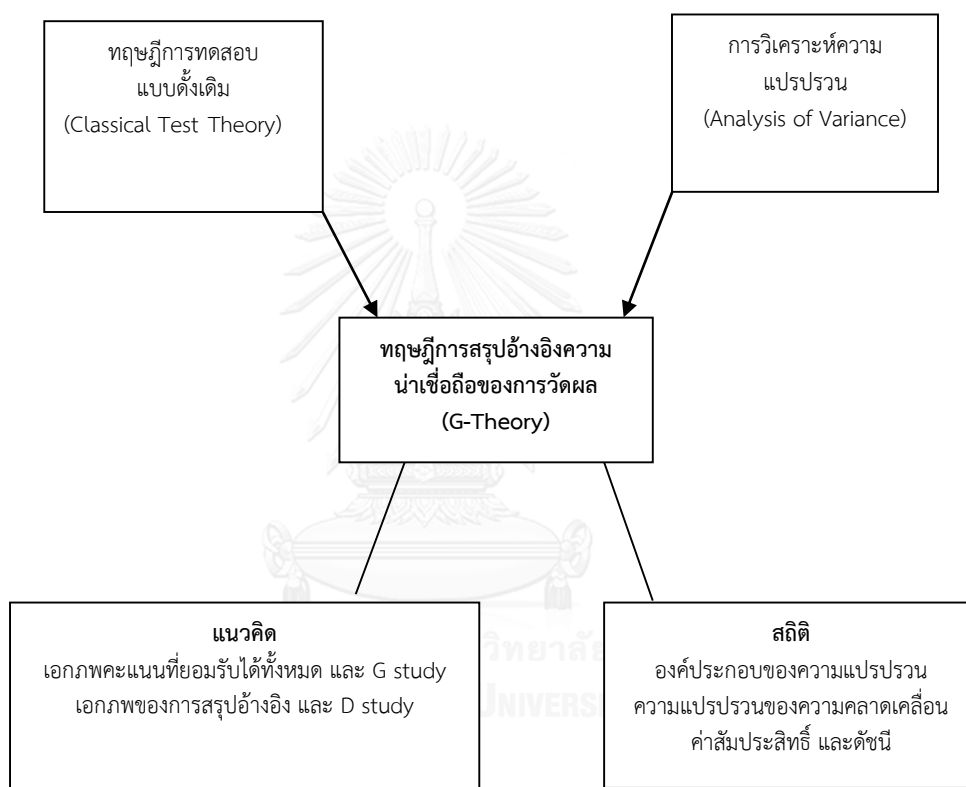
2. ความสัมพันธ์แบบแฝง (nested) หมายถึง ความสัมพันธ์ในลักษณะที่แต่ละระดับของสิ่งที่ถูกวัด ถูกวัดภายใต้เงื่อนไขที่แตกต่างกัน สัญลักษณ์คือ " : " อ่านว่า "nested within" เช่น รูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น $i:r$ หมายถึง ผู้ตรวจข้อสอบ (r) ตรวจข้อสอบ (i) ต่างข้อกัน

3. ความสัมพันธ์แบบผสม (confounded) หมายถึง ความสัมพันธ์ที่มีทั้งความสัมพันธ์แบบไขว้และความสัมพันธ์แบบแฝงปนกันอยู่ เช่น $p \times (i:t)$ หมายถึง นักเรียน (p) ทำข้อสอบ (i) บางข้อที่อยู่ในแบบทดสอบ (t)

3.1.2 แนวคิดพื้นฐาน

G-theory มีแนวคิดมาจาก CTT และ การวิเคราะห์ความแปรปรวน (Analysis of Variance) (Brennan, 2011; Brennan, 2001) ซึ่ง CTT เป็นการศึกษาความเที่ยงของการวัดโดยไม่ได้แบ่งแยกแหล่งของความคลาดเคลื่อนนั้น ในขณะที่ G-theory ให้ความสำคัญกับองค์ประกอบของความแปรปรวน

และมีการประมาณค่าความคลาดเคลื่อนเหล่านั้น ซึ่งเป็นแนวคิดการวิเคราะห์ความแปรปรวน โดยในการศึกษา G-theory นั้นจะมีขั้นตอนในการศึกษาเพื่อนำไปสู่การตัดสินใจของผู้ศึกษา โดยการศึกษา เอกภพที่ยอมรับได้ทั้งหมดและทำการประมาณค่าเงื่อนไขที่เป็นแหล่งความคลาดเคลื่อนใน G-study เพื่อนำไปสู่การตัดสินใจเลือกเงื่อนไขที่มีความคลาดเคลื่อนน้อยที่สุดและมีความเที่ยงสูงที่สุด ใน D-study ส่วนในการประมาณค่านั้น G-theory เป็นการประมาณค่าองค์ประกอบของความแปรปรวนความแปรปรวนของความคลาดเคลื่อนที่เกิดจากแต่ละแหล่ง และเป็นประมาณค่าสัมประสิทธิ์การสรุปร่างของการวัด ซึ่งมีลักษณะคล้ายกันกับสัมประสิทธิ์ความเที่ยงใน CTT สามารถสรุปได้ดัง ภาพที่ 2.9



ภาพที่ 2.9 กรอบแนวคิดของ G-Theory

ที่มา: Brennan, 2001; ใน Generalizability Theory

3.1.3 ข้อตกลงเบื้องต้น

G-theory มีข้อตกลงเบื้องต้น ดังนี้ (ศิริชัย กาญจนวาสี, 2550)

1. คุณลักษณะที่มุ่งวัดของบุคคลไม่ว่าจะเป็น เจตคติ ทักษะ หรือคุณลักษณะอื่นๆ เป็นค่าที่อยู่ในสภาวะคงที่ (Steady state)

2. ผู้สอบคนเดียวกันได้คะแนนสอบต่างกันจากการวัดในแต่ละสถานการณ์ หรือเงื่อนไขของการวัด เนื่องมาจากความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง โดยองค์ประกอบด้านวุฒิภาวะ (maturation) และการเรียนรู้ (learning) ระหว่างการวัดไม่เป็นแหล่งความคลาดเคลื่อนของคะแนนที่ได้จากการวัด

3. เมื่อพิจารณาผู้สอบทั้งกลุ่ม ความแปรปรวนของคะแนนที่สังเกตได้ ประกอบด้วย ความแปรปรวนของคะแนนจริง ซึ่งเป็นความแตกต่างที่แท้จริงของบุคคลความแปรปรวนของคะแนนความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง และความแปรปรวนของคะแนนความคลาดเคลื่อนสุ่ม

3.1.4 ขั้นตอนการศึกษา

G-theory มีขั้นตอนการศึกษาที่สำคัญ 2 ขั้นตอน (ศิริชัย กาญจนวาสี, 2550; Brennan, 2001; Brennan, 2009) คือ การศึกษาเชิงสรุปอ้างอิงและการศึกษาเชิงตัดสินใจ ดังนี้

1. การศึกษาเชิงสรุปอ้างอิง (G-study)

G-study เป็นการสรุปอ้างอิงผลจากการศึกษาตัวอย่างการวัดตามเงื่อนไขการวัดที่สนใจไปยังเอกภพของการวัด เพื่อทำนายและจัดเตรียมสารสนเทศเกี่ยวกับแหล่งความแปรปรวนจากการวัดให้ได้มากที่สุดเท่าที่จะเป็นไปได้ การเก็บรวบรวมข้อมูลควรให้ครอบคลุมการทดสอบภายใต้เงื่อนไขการวัดที่สนใจทั้งหมดหรือที่เรียกว่า เอกภพคะแนนที่ยอมรับได้ทั้งหมด (universe of admissible observation) ซึ่ง G-theory เป็นการประมาณค่าขององค์ประกอบความผันแปร (variance components) ที่ให้ความสำคัญกับการสรุปอ้างอิงที่เน้นกับการประมาณค่าและการแปลความหมายขององค์ประกอบของความผันแปรเพื่อตัดสินใจ ใน D-study (decisions) โดยขึ้นอยู่กับค่าเฉลี่ยของคะแนนสังเกตได้ (observed mean score)

2. การศึกษาเชิงตัดสินใจ (D-study)

D-study เป็นการใช้ข้อมูลจาก G-study ในการสรุปอ้างอิงที่เน้นกับการประมาณค่าและการแปลความหมายขององค์ประกอบของความผันแปร เพื่อการตัดสินใจ โดยใน D-study ผู้วิเคราะห์จะต้องระบุและนิยามเอกภพของการสรุปอ้างอิงว่าจะใช้การตัดสินใจแบบใด ระหว่างการตัดสินใจแบบสัมบูรณ์ (absolute) ซึ่งเป็นการตัดสินใจโดยการแปลผลเปรียบเทียบกับเกณฑ์มาตรฐาน เช่น การเปรียบเทียบว่าผู้สอบคนหนึ่งจะมีความสามารถในการออกแบบผ่านเกณฑ์มาตรฐานของคณะหรือไม่ โดยไม่สนใจเปรียบเทียบความสามารถกับผู้เรียนคนอื่น ๆ หรือ แบบสัมพัทธ์ (relative) ซึ่งเป็นการตัดสินใจโดยการแปลผลเปรียบเทียบระหว่างบุคคล เช่น การเปรียบเทียบว่าผู้สอบคนหนึ่งจะมีความสามารถในการออกแบบสูงหรือต่ำกว่าผู้สอบคนอื่น ๆ ที่เรียนในกลุ่มเดียวกัน แล้วจึงใช้ข้อมูลจากที่ได้จากการวิเคราะห์ G-study ที่ให้ค่าความคลาดเคลื่อนจากแหล่งความแปรปรวนต่าง ๆ เพื่อประเมินว่าจะเลือกรูปแบบ (design) การวัดแบบใดที่ให้ค่าความคลาดเคลื่อนต่ำสุด ในขณะที่มีความเที่ยงสูงสุด

3.1.5 สัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือ (G-coefficient)

Brennan (2001) และ Webb et al. (2007) ได้เสนอรายละเอียดการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบสอบที่มีหลายแหล่งของความคลาดเคลื่อน ซึ่ง Johnson and Johnson (2009) ได้ค้นพบความสามารถของ G-theory เพื่อศึกษาความเที่ยงของผลการวัดและความคลาดเคลื่อนของการวัดจากแหล่งต่างๆ รวมทั้งความสัมพันธ์ระหว่างข้อสอบ (Item-related) ความสัมพันธ์ระหว่างผู้สอบ (marker-related) และแหล่งอื่นๆ การวัดแบบสมบูรณ์ และสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ (absolute generalizability coefficients) นั้น G-theory สามารถใช้ตรวจสอบ (investigate) ความเที่ยงของการประเมินได้ซึ่งในความเป็นจริงแล้ว G-theory study design มีความจำเป็นต่อการพิจารณาจำนวนของฟาเซทหลักเพื่อการสำรวจขนาดของชุดข้อมูลที่จำเป็นต่อการประมาณค่าที่มีความเสถียรภาพสำหรับองค์ประกอบของความแปรปรวน ซึ่งในการวิเคราะห์ G-theory เป็นทฤษฎีการวัดที่ใช้การวิเคราะห์แหล่งความคลาดเคลื่อนจากหลายแหล่ง ดังนั้นการวิเคราะห์จึงเป็นการประยุกต์ใช้การวิเคราะห์ความแปรปรวน (Analysis of Variance, ANOVA) ค่าสัมประสิทธิ์การสรุปอ้างอิงความเชื่อถือของผลการวัด (G-coefficient) มีลักษณะเหมือนกับสัมประสิทธิ์ความเที่ยง (reliability coefficient) ใน CTT ซึ่งสัมประสิทธิ์การสรุปความน่าเชื่อถือใน G-theory ที่ใช้สำหรับการตัดสินใจใน D-study มี 2 ประเภท คือ สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (E_{ρ^2}) และสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (Φ)

1. สัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ (E_{ρ^2})

เป็นสัมประสิทธิ์ที่ใช้ในการตัดสินใจสำหรับสถานการณ์ของการวัดที่เป็น การเปรียบเทียบระหว่างผู้สอบ ซึ่ง E_{ρ^2} มีความหมายเดียวกับความเที่ยง ในการทดสอบแบบดั้งเดิม โดยมีสูตรในการคำนวณ (Cronbach et al, 1972; cited in Brennan, 2001) เป็นดังนี้

$$E_{\rho^2} = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\delta)}$$

เมื่อ $\sigma^2(p)$ คือ ความแปรปรวนของคะแนนของผู้สอบ

$\sigma^2(\delta)$ คือ ความแปรปรวนของแหล่งความคลาดเคลื่อน
ในการตัดสินใจเชิงสัมพัทธ์

2. สัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ (Φ)

เป็นสัมประสิทธิ์ที่ใช้ในตัดสินใจสำหรับสถานการณ์การวัดที่เป็นการเปรียบเทียบกับเกณฑ์มาตรฐานหรือเรียกว่า Phi Coefficient หรือดัชนีของความน่าเชื่อถือ (Index of dependability) โดยมีสูตรในการคำนวณ (Brennan and Kane, 1977a,b; cited in Brennan, 2001) เป็นดังนี้

$$\Phi = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\Delta)}$$

เมื่อ $\sigma^2(p)$ คือ ความแปรปรวนของคะแนนของผู้สอบ

$\sigma^2(\Delta)$ คือ ความแปรปรวนของแหล่งความคลาดเคลื่อนในการตัดสินใจเชิงสัมบูรณ์

3.1.6 การออกแบบการวัดใน G-theory

การออกแบบการวัดใน G-theory เป็นการออกแบบสถานการณ์ในการวัดให้มีเงื่อนไขหรือ ฟาเซท ตามที่ผู้ประเมินสนใจ เช่น ฟาเซทเดียว สองฟาเซทหรือสามฟาเซท เป็นต้น เพื่อทำการเก็บรวบรวมข้อมูลที่สนใจและนำไปใช้ในขั้น D-study ต่อไป ซึ่งในที่นี้จะกล่าวถึงลักษณะการออกแบบการวัดในกรณีฟาเซทเดียว (single-facet) และกรณีสองฟาเซท (two-facet) ดังต่อไปนี้

1. การออกแบบกรณีฟาเซทเดียว (Single-Facet Design)

เป็นการออกแบบการวัดที่เอกภพเป็นแบบเอกภพที่มีหนึ่งองค์ประกอบหรือมีองค์ประกอบของความคลาดเคลื่อนเพียง 1 แหล่ง เช่น ในการสอบการเขียนเรียงความภาษาอังกฤษ ผู้บริหารการสอบสนใจความคลาดเคลื่อนที่เกิดจากผู้ตรวจให้คะแนนหรือความยาวของแบบสอบ อย่างไรก็ตามหนึ่ง เป็นต้น สถานการณ์ของการวัดในเอกภพของหนึ่งฟาเซทนี้ ความคลาดเคลื่อน (error) ที่เกิดขึ้นผู้บริหารการสอบหรือผู้ตัดสินใจต้องการสรุปอ้างอิงไปยังความน่าเชื่อถือของคะแนนที่ได้จากข้อสอบที่มีจำนวนของผู้ตรวจต่างกัน หรือความยาวข้อสอบต่างกัน แหล่งความคลาดเคลื่อนของคะแนนสอบในตัวอย่างนี้ คือ จำนวนของผู้ตรวจ หรือความยาวของข้อสอบ เป็นต้น

2. การออกแบบกรณีสองฟาเซท (Two-facet design)

เป็นการออกแบบการวัดที่เอกภพเป็นแบบเอกภพที่มีหนึ่งองค์ประกอบหรือมีองค์ประกอบของความคลาดเคลื่อน 2 แหล่ง ตัวอย่างเช่น การสอบเขียนเรียงความภาษาอังกฤษที่ผู้ตรวจจำนวนหนึ่งทำการประเมินผลการเขียนเรียงความของผู้เรียน โดยมีแหล่งของผู้ประเมินจากหลายแหล่ง เช่น ผู้เชี่ยวชาญภายใน ผู้เชี่ยวชาญภายนอก เพื่อสรุปอ้างอิงไปยังความน่าเชื่อถือได้ของคะแนนจากการสอบที่มีผู้ประเมินจำนวนต่างกัน และจำนวนแหล่งของผู้ประเมินต่างกัน แหล่งความคลาดเคลื่อนของคะแนนสอบในที่นี้ คือ จำนวนผู้ประเมิน และจำนวนแหล่งผู้ประเมิน เป็นต้น

3.2 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร (Multivariate generalizability theory, MGT)

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบตัวแปรเดียว (Univariate G-theory) สามารถใช้ในการประมาณค่าองค์ประกอบของความคลาดเคลื่อน และความน่าเชื่อถือของคะแนนสำหรับคะแนนในภาพรวม ส่วนทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร (Multivariate G-theory) สามารถใช้ในการตรวจสอบความสัมพันธ์ของคะแนนเอกภพระหว่างองค์ประกอบของแบบสอบ (หรือการทดสอบองค์ประกอบ (component tests)) และเงื่อนไขของการจัดกลุ่มเพื่อสร้างความเที่ยงสูงสุดสำหรับคะแนนรวม (Brennan, 2001; Lee, 2006; Webb et al, 2007; Ofqual, 2010) ให้สารสนเทศเกี่ยวกับความสัมพันธ์ระหว่างส่วนประกอบของแบบสอบซึ่งเป็นสิ่งสำคัญเพื่อตรวจสอบความเหมาะสมของความสัมพันธ์ขององค์ประกอบของแบบสอบที่นำไปสู่รูปแบบองค์ประกอบ (Lee, 2006; Brennan, 2011) องค์ประกอบความแปรปรวนร่วมเป็นสิ่งที่ทำให้ได้สารสนเทศของผู้สอบเกี่ยวกับคะแนนเอกภพและความแตกต่างร่วมของความคลาดเคลื่อน (errors co-vary) เป็นอย่างไร ซึ่งองค์ประกอบความแปรปรวนร่วมสำหรับผลที่แตกต่างกันนั้นสามารถที่จะแปลความหมายความสัมพันธ์ไปยังองค์ประกอบความแปรปรวนที่สอดคล้องกัน (corresponding variance components) สำหรับคะแนนเอกภพหลายเอกภพได้ ดังนั้นทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรจึงมีความเหมาะสมสำหรับการวิเคราะห์ความน่าเชื่อถือของคะแนนรวม (composite scores) ในการสอบที่มีเนื้อหาหลายส่วน (multiple content domains) หรือหลายคุณลักษณะ (traits) มีความสัมพันธ์กัน โดยมีความสำคัญเกี่ยวกับความเหมาะสมในสองเรื่อง คือ 1) การประมาณค่าความน่าเชื่อถือของคะแนนรวมที่ใช้รูปแบบการถ่วงน้ำหนักที่แตกต่างกัน (different weighting schemes) (Brennan, 2001) และ 2) ผลอื่นๆ ที่พบบ่อย คือ คะแนนเอกภพที่แสดงให้เห็นว่ามีความสัมพันธ์ที่แท้จริงระหว่างคะแนนของการวัดที่หลากหลาย (multiple measures)

ผลการวิเคราะห์จากทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร นอกจากให้สารสนเทศที่มีอยู่ใน univariate G theory แล้วยังรวมทั้ง การประมาณค่าองค์ประกอบความคลาดเคลื่อนสำหรับแหล่งความแตกต่างของความผันแปรของคะแนนและผลสรุปของดัชนีต่างๆ ของความน่าเชื่อถือของคะแนน ยิ่งไปกว่านั้นสารสนเทศที่แสดงในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรที่เกี่ยวกับความสัมพันธ์ที่เฉพาะเพื่อการศึกษาคือ ความสัมพันธ์ (interrelationships) ระหว่างชุดของมาตรประมาณค่า และการวิเคราะห์คะแนนรวมที่ครอบคลุม (comprehensive composite score) (Yasuyo Sawaki, 2007 cited in Brennan, 2011)

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรเป็นทฤษฎีการวัดที่นำเสนอขึ้นโดย Cronbach Gleser Nanda and Rajaratnum ในปี ค.ศ.1972 ในบทความที่มีชื่อว่า “The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles”

(Webb and Shavelson, 1981; cited in Chauser, et al, 2006) โดยแบ่งเมตริกซ์ ความแปรปรวน-ความแปรปรวนร่วม (variance-covariance matrix) ของคะแนนที่สังเกตได้ให้อยู่ในรูปของเมตริกซ์ ความแปรปรวน-ความแปรปรวนร่วมย่อยของแต่ละแหล่ง ความแปรปรวน ซึ่งจะทำได้รายละเอียดมากกว่า univariate G-theory ซึ่งแม้ว่า Cronbach และคณะจะเป็นผู้ที่ได้เสนอแนวคิดของทฤษฎี การสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรไว้ แต่ Joe และ Woodward ในปี ค.ศ. 1976 ได้วิเคราะห์ ค่าสัมประสิทธิ์ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร (multivariate generalizability coefficient) โดยขยายวิธีการประมาณค่าจาก G-theory ของ Cronbach และคณะ โดยได้แบ่ง การศึกษาทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรออกเป็น 2 ขั้นตอน คือ G-study และ D-study เช่นกัน (Webb and Shevelson, 1981 cited in Chauser, et al, 2006)

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรเป็นทฤษฎีที่พิเศษและมีความยืดหยุ่น ซึ่ง MGT เป็นพหุตัวแปรตั้งต้น ซึ่งในที่นี้พหุเอกภพ (multiple universe) โดยทั่วไปแล้วคือคะแนน พหุเอกภพ (multiple universe scores) ของนักเรียนแต่ละคน ซึ่งเป็นพหุเอกภพที่สอดคล้อง ของการสังเกตที่ยอมรับได้ทั้งหมด แต่ละพหุเอกภพมีความเกี่ยวข้องกับเงื่อนไขแบบคงที่ (a single fixed condition) ของการวัด การวิเคราะห์ใน Multivariate G theory ต่างจากใน univariate G theory คือ ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรนั้น แต่ละเป้าหมายของการวัด จะมีคะแนนพหุเอกภพ (multiple universe scores) ดังนั้น จึงมีการวิเคราะห์ ทั้งความแปรปรวน และความแปรปรวนร่วม ระหว่าง คะแนนเอกภพ และระหว่างความคลาดเคลื่อนในองค์ประกอบ (Brennan, 2001) ซึ่งในทางสถิติแล้วการวิเคราะห์ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร จะมีการวิเคราะห์ทั้งองค์ประกอบของความแปรปรวน (variance components) และองค์ประกอบ ของความแปรปรวนร่วม (covariance components) (Brennan, 2011) สถิติที่กล่าวถึงใน D study ประกอบด้วย ความแปรปรวนของคะแนนเอกภพ (universe score variances) ความแปรปรวน ของความคลาดเคลื่อนเชิงสัมพัทธ์และเชิงสมบูรณ์ (relative and absolute error variances) ความคลาดเคลื่อนมาตรฐานของการวัด (standard errors of measurement) และสัมประสิทธิ์ การสรุปอ้างอิง (G-coefficient) และสัมประสิทธิ์ความน่าเชื่อถือ (dependability coefficients) (Yin & Sconing, 2008) ซึ่งในรายละเอียดต่างๆ ที่เกี่ยวข้องกับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ แบบพหุตัวแปรมีดังนี้

3.2.1 G study designs

การออกแบบ G-study ในการออกแบบสำหรับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร (multivariate G theory) นั้น มีหลายรูปแบบซึ่งในที่นี่จะกล่าวถึงเฉพาะแบบฟาเซตเดี่ยว (single facet) และสองฟาเซต (two facet) โดยสรุปรูปแบบไว้ดังตารางที่ 2.14 การศึกษาครั้งนี้ได้ใช้สัญลักษณ์ตามที่ Brennan (2011) ใช้เพื่อความเข้าใจที่ง่ายขึ้นในการอธิบาย โดยให้สัญลักษณ์ “•” หมายถึง เงื่อนไขที่ถูกเชื่อมโยง (linked conditions) หรือเป็นฟาเซตที่มีการเชื่อมโยง (linked facet) หรือ การสุ่มตัวอย่างร่วมกัน (joint sampling) และให้สัญลักษณ์ “◦” เงื่อนไขที่เป็นอิสระ (Independent conditions) หรือ การสุ่มตัวอย่างที่เป็นอิสระต่อกัน (independent sampling) หรือเป็นฟาเซตที่เป็นอิสระ (independent facet) ซึ่งสามารถสรุปได้ดังตารางที่ 2.14 และรายละเอียดของแต่ละรูปแบบจะกล่าวในหัวข้อการออกแบบ G-study แบบฟาเซตเดี่ยว และการออกแบบ G-study แบบสองฟาเซต ต่อไป

ตารางที่ 2.14 ตัวอย่างการออกแบบแบบพหุตัวแปร (multivariate designs) และการออกแบบ univariate

Multivariate Designs	Variance components Design	Covariance components Design	การออกแบบ Univariate
$p^{\bullet} \times i^{\bullet}$	$p \times i$	$p \times i$	$p \times i \times v$
$p^{\bullet} \times i^{\circ}$	$p \times i$	p	$p \times (i : v)$
$i^{\circ} : p^{\bullet}$	$i : p$	$i : p$	$(i : p) \times v$
$i^{\circ} : p^{\circ}$	$i : p$	p	$i : (p \times v)$
$p^{\bullet} \times i^{\bullet} \times h^{\bullet}$	$p \times i \times h$	$p \times i \times h$	$p \times i \times h \times v$
$p^{\bullet} \times i^{\bullet} \times h^{\circ}$	$p \times i \times h$	$p \times i$	$p \times i \times (h : v)$
$p^{\bullet} \times i^{\circ} \times h^{\bullet}$	$p \times i \times h$	$p \times h$	$p \times (i : v) \times h$
$p^{\bullet} \times i^{\circ} \times h^{\circ}$	$p \times i \times h$	p	$p \times [(i \times h) : v]$
$p^{\bullet} \times (i^{\circ} : h^{\bullet})$	$p \times (i : h)$	$p \times (i : h)$	$p \times (i : h) \times v$
$p^{\bullet} \times (i^{\circ} : h^{\circ})$	$p \times (i : h)$	$p \times h$	$p \times [(i \times h) : v]$
$p^{\bullet} \times (i^{\circ} : h^{\circ})$	$p \times (i : h)$	p	$p \times (i : h : v)$
$(p^{\bullet} : c^{\bullet}) \times i^{\bullet}$	$(p \times c) \times i$	$(p \times c) \times i$	$(p : c) \times i \times v$
$(p^{\bullet} : c^{\bullet}) \times i^{\circ}$	$(p \times c) \times i$	$p : c$	$(p : c) \times (i \times v)$
$(p^{\circ} : c^{\bullet}) \times i^{\circ}$	$(p \times c) \times i$	c	$(p \times i) : (c : v)$

ที่มา: Brennan (2011)

จากตารางที่ 2.14 การออกแบบ G-study ทั้งแบบฟาเซตเดี่ยว และสองฟาเซตในแต่ละรูปแบบของการออกแบบของ MGT โดยให้ v แทน การออกแบบระดับของ fixed facet หรือตัวแปรที่ศึกษา หรือ มิติ p แทน ผู้สอบ, i แทน ข้อสอบ, h แทน ผู้ตรวจ หรือผู้ให้คะแนน โดยที่ทั้ง i และ h เป็นฟาเซตสุ่ม (random facets) และ c แทน การรวมของผู้สอบ (เช่น classes)

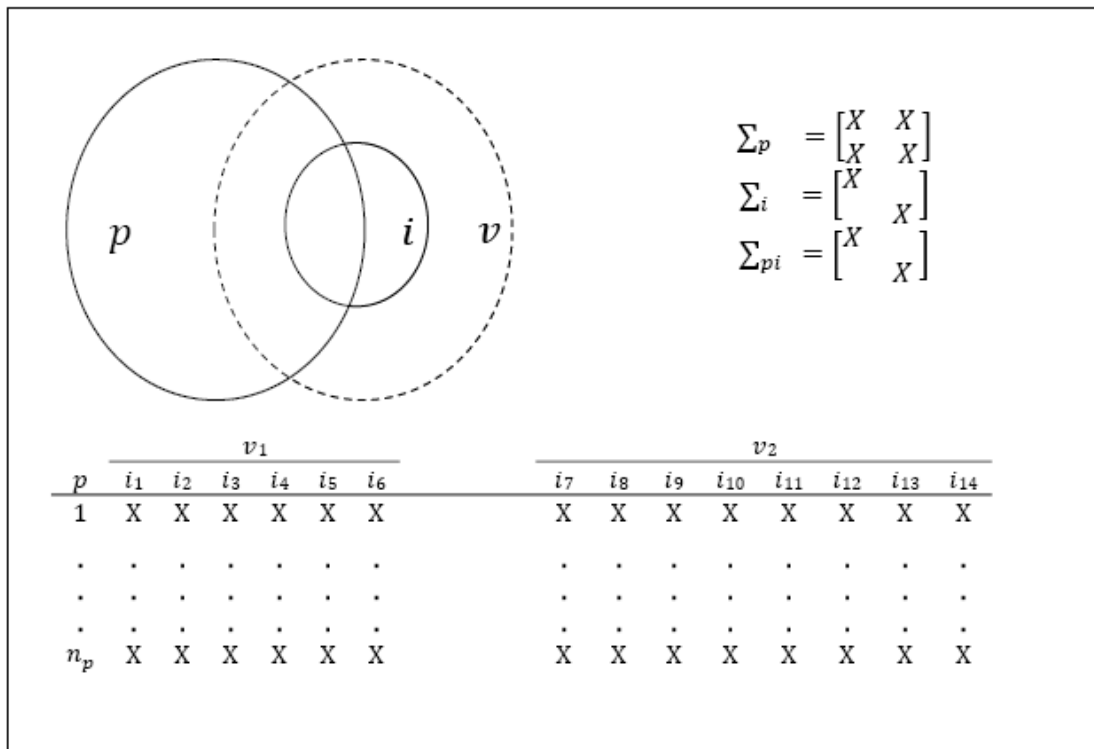
ซึ่งความสัมพันธ์ที่พบในการออกแบบองค์ประกอบความแปรปรวน (variance components design) กับ univariate G-theory คือ มีความเกี่ยวข้องในทุกตัวแปรคงที่ (fixed variable) ส่วนการออกแบบองค์ประกอบความแปรปรวนร่วม (covariance components design) มีความสัมพันธ์กับ univariate G-theory คือ มีความเกี่ยวข้องกันคู่ของตัวแปรคงที่ (pair of fixed variable)

3.2.1.1 การออกแบบแบบ Single-Facet Designs

สำหรับการออกแบบของพหุตัวแปร ในส่วนของ Single-Facet Designs มีทั้งหมด 4 แบบ 2 แบบแรก คือ เงื่อนไขที่เป็นการไขว้ (crossed) กับผู้สอบ ส่วน อีก 2 แบบ คือ เงื่อนไขที่เป็นแบบแฝง (nested) ภายในผู้สอบ ดังนี้ (Office of Qualifications and Examinations Regulation, 2010) (Office of Qualifications and Examinations Regulation, 2010)

1. $p \times i$ Design

การออกแบบ $p \times i$ Design นี้มีความแตกต่างในส่วนของชุดของข้อสอบ (i) ที่ใช้ในแต่ละมิติของแบบสอบ หรือในแต่ละระดับของ v สำหรับการออกแบบนี้ สามารถอธิบายเกี่ยวกับการดำเนินการคือ ผู้สอบ (p) ทำข้อสอบทุกข้อ และเป็นข้อสอบชุดเดียวกัน แต่ข้อสอบในแต่ละมิติของแบบสอบต่างกัน ดังนี้ องค์ประกอบความแปรปรวนร่วมจึงมีความเกี่ยวข้องกับ p เท่านั้น ส่วน univariate design นั้นคือ $p \times (i: v)$ แต่การออกแบบในลักษณะของ $p \times i$ Design นี้เป็นการออกแบบที่มีสมรรถภาพสูงกว่า $p \times (i: v)$ ในสองประการ ประการแรก คือ การออกแบบในลักษณะนี้มักจะมีจำนวนข้อสอบไม่เท่ากันในแต่ละมิติของแบบสอบ หรือในแต่ละระดับ ซึ่งทำให้เกิดความซับซ้อนและมีความกำกวมสำหรับการวิเคราะห์ในการศึกษาแบบ ตัวแปรเดียว (univariate design) ประการที่สอง คือ การศึกษาแบบตัวแปรเดียว ไม่มีการรองรับสำหรับการประมาณค่าแบบแยกส่วนขององค์ประกอบความแปรปรวนและความแปรปรวนร่วมในแต่ละระดับของฟาเซทแบบจำกัด (fixed facet) ดังภาพที่ 2.10



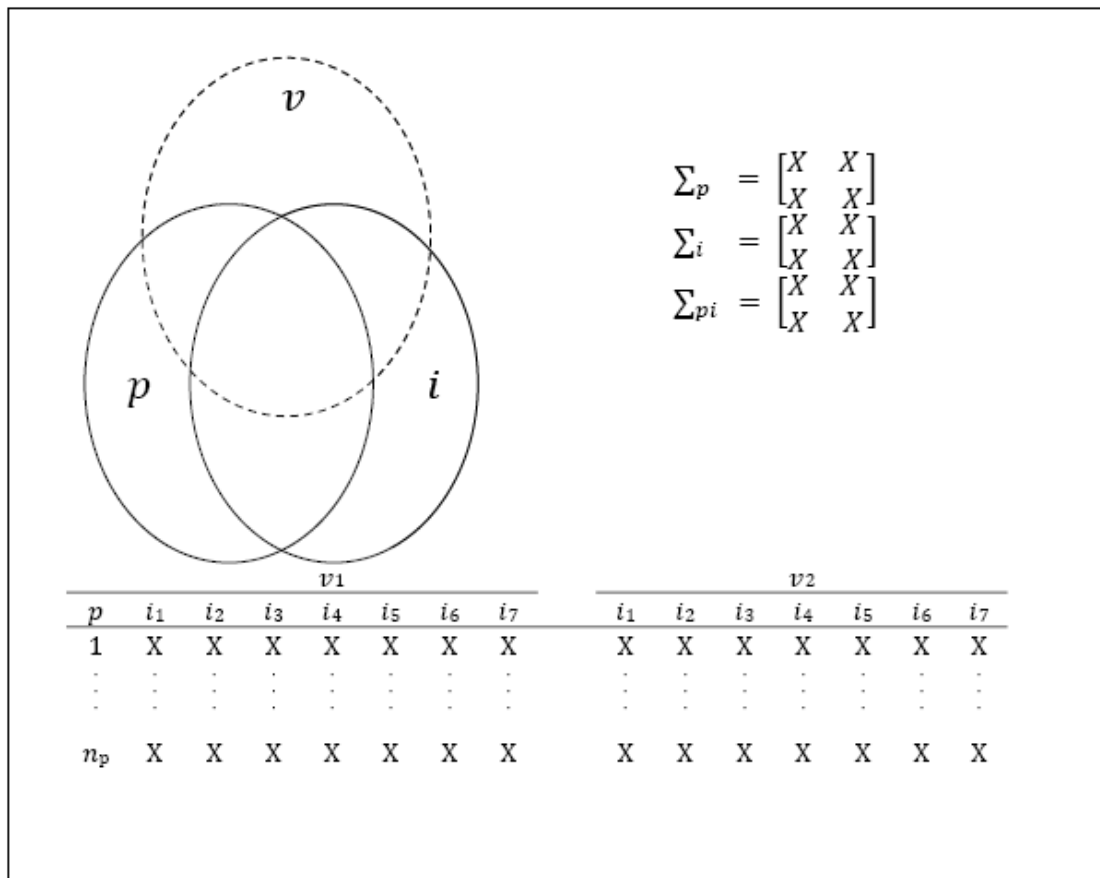
ภาพที่ 2.10 การออกแบบ แบบ $p \times i$ Design

ที่มา: Brennan (2011)

จากภาพที่ 2.10 เป็นการออกแบบการบันทึกข้อมูลสำหรับ $p \times i$ Design โดยแสดงให้เห็นว่า ข้อสอบในแต่ละมิติของแบบสอบนั้นเป็นข้อสอบที่ต่างชุดกัน และส่วนขององค์ประกอบความแปรปรวนร่วมจะมีเฉพาะในส่วนของผู้สอบเท่านั้น

3. $p \times i$ Design

การออกแบบ $p \times i$ Design เป็นการออกแบบที่มีลักษณะคล้ายกับ $p \times i$ Design แต่แตกต่างกันส่วนของข้อสอบ กล่าวคือ สำหรับชุดของข้อสอบในการออกแบบ $p \times i$ Design นั้นในแต่ละมิติของแบบสอบ หรือ ในแต่ละระดับของ v เป็นข้อสอบชุดเดียวกัน ดังนั้น องค์ประกอบความแปรปรวนร่วมจึงมีทั้งจาก ผู้สอบ (p) ข้อสอบ (i) และปฏิสัมพันธ์ระหว่าง ผู้สอบ และ ข้อสอบ ดังภาพที่ 2.11



ภาพที่ 2.11 การออกแบบ แบบ $p^* \times i^*$ Design

ที่มา: Brennan (2011)

จากภาพที่ 2.11 แสดงให้เห็นว่า ในการออกแบบ $p^* \times i^*$ Design ผู้สอบแต่ละคน (p) ต้องทำข้อสอบ (i) ทุกข้อ และข้อสอบเป็นชุดเดียวกันในทุกมิติ ตัวอย่าง เช่น ผู้สอบทำข้อสอบชุดเดียวกันทั้งสองครั้ง (การทำข้อสอบ Pretest posttest)

4. $i^{\circ}:p^*$ และ $i^{\circ}:p^{\circ}$ Design

สำหรับการออกแบบทั้ง $i^{\circ}:p^*$ และ $i^{\circ}:p^{\circ}$ Design นี้ มีความแตกต่างในส่วนของ i เกี่ยวกับการเชื่อมโยงในแต่ละมิติของแบบสอบหรือระดับของ v กล่าวคือ $i^{\circ}:p^*$ Design ในข้อสอบ (i) ในแต่ละมิติของแบบสอบมีความเกี่ยวข้องกันกับทุกมิติ ดังนั้น องค์ประกอบความแปรปรวนร่วมจึงมีทั้ง ของผู้สอบ และข้อสอบ และการออกแบบ univariate คือ $i^{\circ}:(p \times v)$

ในขณะที่ $i^{\circ}:p^{\circ}$ Design ข้อสอบ (i) ในแต่ละมิติของแบบสอบจะไม่มี ความเกี่ยวข้องกันหรือกล่าวได้ว่า ข้อสอบในแต่ละมิติเป็นข้อสอบที่ต่างชุดกัน ดังนั้นองค์ประกอบความแปรปรวนร่วมจึงมีเฉพาะของผู้สอบเท่านั้น และการออกแบบ univariate คือ $(i^{\circ}:p) \times v$

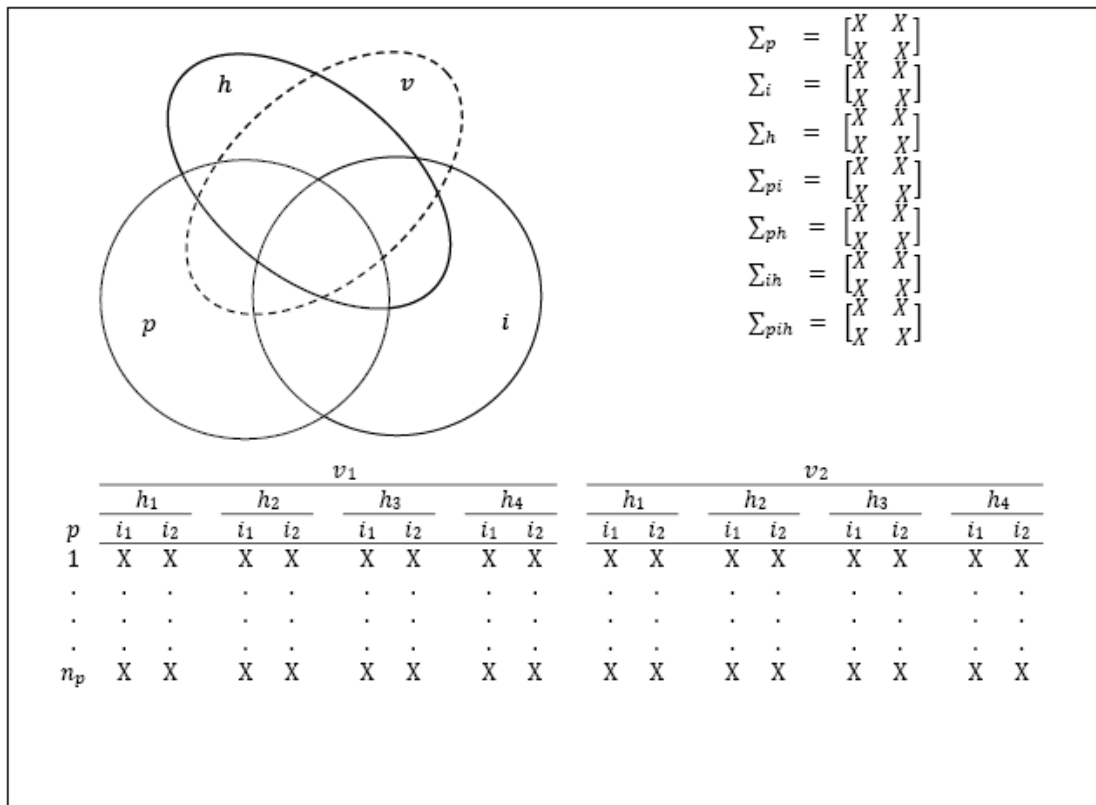
3.2.1.2 การออกแบบแบบ Two-Facet Crossed Designs

การออกแบบแบบ Two-Facet Crossed Designs เป็นการออกแบบที่มีแหล่งของความคลาดเคลื่อนที่นำมาพิจารณา 2 แหล่ง ซึ่งในการนำเสนอครั้งนี้ ได้นำเสนอใน 4 รูปแบบ โดยทุกแบบนั้นมีแหล่งขององค์ประกอบความแปรปรวน และองค์ประกอบความแปรปรวนร่วมจำนวน 7 แหล่ง ซึ่งแต่ละรูปแบบมีรายละเอียดดังนี้

1. $p \times i \times h$ Design

การออกแบบแบบ $p \times i \times h$ Design เป็นการออกแบบที่ทุกเมตริกขององค์ประกอบความแปรปรวนร่วมมีความสมบูรณ์ (full) ดังรูปที่ 12 ซึ่งการออกแบบ univariate คือ $p \times i \times h \times v$ ซึ่งมีจุดตัด (intersecting) ของวงรีใน 4 จุด ตัวอย่างสำหรับการออกแบบใน $p \times i \times h$ Design เช่น ผู้สอบ (p) ทุกคน ทำชิ้นงาน (i) จำนวน 16 ชิ้นงาน และทุกๆ ชิ้นงานถูกตรวจโดยผู้ตรวจ (h) ชุดเดียวกัน ในข้อสอบที่มี 2 มิติ

สำหรับการออกแบบด้วย $p \times i \times h$ Design นี้ นับว่าเป็นการออกแบบที่มีสมรรถภาพสูงมากเพราะมีการประมาณค่าองค์ประกอบความแปรปรวน และองค์ประกอบความแปรปรวนร่วมในทุกเมตริกซ์ แต่ในความเป็นจริงการออกแบบในลักษณะนี้ถูกนำมาใช้น้อยเนื่องมาจากการเก็บรวบรวมข้อมูลที่ได้ยากทั้งเรื่องจำนวนครั้งที่มาก ค่าใช้จ่ายที่สูง เช่น การประเมินความสามารถทางศิลปะของนักเรียนในชั้นประถมศึกษาปีที่ 5 (p) จำนวน 60 คน โดยให้ทำชิ้นงาน (i) การวาดภาพสีน้ำจำนวน 4 ชิ้น และถูกตรวจให้คะแนนด้วยผู้ตรวจ (h) จำนวน 25 คน ดังนั้นผู้ตรวจแต่ละคนต้องตรวจให้คะแนน จำนวน 240 ภาพวาด และมีการให้คะแนนใน 3 ประเด็นที่ต่างกัน โดยมีเกณฑ์การให้คะแนนเป็นมาตร 10 ระดับ ได้แก่ ความเป็นตัวแทนของวัตถุประสงค์ของภาพ ภูมิหลังของภาพ และความสัมพันธ์ระหว่างวัตถุประสงค์ที่มีความชัดเจนในการนำเสนอแก่ผู้ชมได้เข้าใจ



ภาพที่ 2.12 การออกแบบ $p^* \times i^* \times h^*$ Design

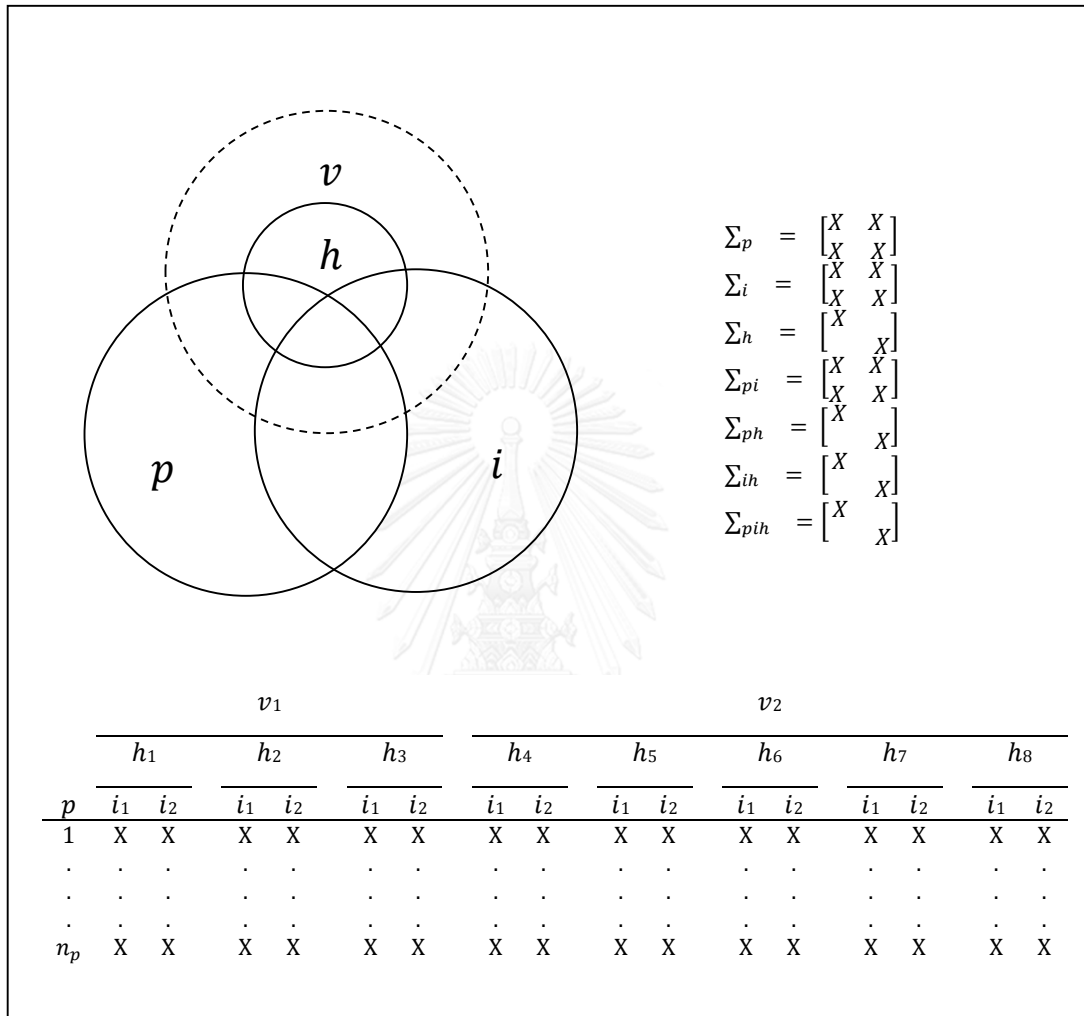
ที่มา: Brennan (2011)

จากภาพที่ 2.12 แสดงให้เห็นว่า เมตริกองค์ประกอบความแปรปรวนและองค์ประกอบความแปรปรวนร่วมถูกนำมาประมาณค่าครบทั้ง 7 แห่ง และในส่วนของมิติของการสอบพบว่าในทั้งสองมิติมีจำนวนชุดของข้อสอบที่เป็นชุดเดียวกัน

2. $p^* \times i^* \times h^*$ Design

การออกแบบ $p^* \times i^* \times h^*$ Design มีความแตกต่างกับ $p^* \times i^* \times h^*$ Design ในส่วนของการเกี่ยวข้องของ h ในแต่ละมิติของแบบสอบ ดังนั้นในแต่ละระดับของ v $p \times i \times h$ คือ องค์ประกอบความแปรปรวน และ $p \times i$ คือองค์ประกอบความแปรปรวนร่วมที่อยู่ในเมตริกซ์แบบ diagonal ดังภาพประกอบ 17 โดยมี univariate counterpart คือ $p \times i \times (h: v)$ ตัวอย่างการออกแบบนี้ เช่น นักเรียนจำนวน 50 คน (p) สอบทักษะการฟังจากเทป จำนวน 12 ชิ้นงาน (i) ทำการบันทึกระหว่างการฟัง ด้วยการเขียน (writing) การตรวจให้คะแนนเป็นแบบองค์รวม (holistic scale) สำหรับการตรวจให้คะแนนทั้งทักษะการเขียนและทักษะการฟัง ผู้ตรวจ (h) ใช้เกณฑ์การให้คะแนนเป็นแบบ 5 ระดับ ทั้งนี้มีชุดของผู้ตรวจ 2 ชุด ที่ต่างกัน แต่ละชุดมีผู้ตรวจจำนวน 3 คน ชุดแรกให้คะแนนทักษะการฟัง และอีกชุดให้คะแนนทักษะการเขียน การดำเนินการคือ

นักเรียนทุกคนทำข้อสอบการฟังและการเขียนทุกชิ้นงาน ผู้ตรวจชุดที่ 1 จำนวน 3 คน ตรวจให้คะแนนทักษะการฟัง จากชิ้นงานทุกชิ้นงานที่นักเรียนเขียนจากการฟังเทป และผู้ตรวจชุดที่ 2 จำนวน 3 คน ตรวจให้คะแนนทักษะการเขียน จากชิ้นงานทุกชิ้นงานที่นักเรียนเขียนจากการฟังเทป ดังภาพที่ 2.13



ภาพที่ 2.13 การออกแบบ $p^\bullet \times i^\bullet \times h^\circ$ Design

ที่มา: Brennan (2011)

3. $p^\bullet \times i^\circ \times h^\bullet$ และ $p^\bullet \times i^\circ \times h^\circ$ Design

การออกแบบของ $p^\bullet \times i^\circ \times h^\bullet$ Design เป็นการออกแบบที่ใช้ชิ้นงาน (i) ต่างกันสำหรับแต่ละมิติของแบบสอบ และผู้สอบ (p) และผู้ตรวจ (h) ยังคงเหมือนกันในทุกมิติของแบบสอบ

ในขณะที่การออกแบบของ $p^* \times i^* \times h^*$ Design เป็นการออกแบบที่แตกต่างกันส่วนของการไม่เชื่อมโยงของผู้ตรวจในทุกมิติของแบบสอบ กล่าวคือ ในแต่ละมิติของแบบสอบนั้นนอกจากจะใช้ชุดของข้อสอบที่ต่างกันแล้ว ยังใช้ชุดของผู้ตรวจที่ต่างกันด้วย

3.2.1.3 การออกแบบแบบ Two-Facet Nested Designs

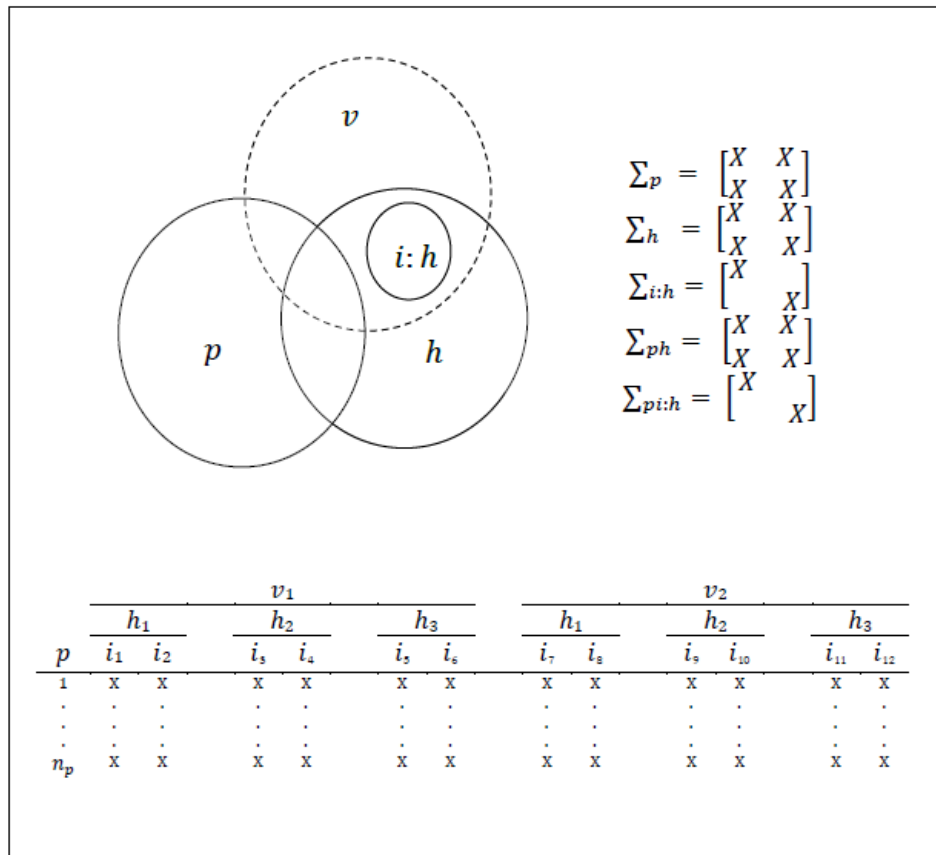
การออกแบบที่เป็น Two-Facet Nested Designs ในการนำเสนอครั้งนี้ได้นำเสนอไว้ 3 รูปแบบ คือ $p^* \times (i^*:h^*)$ Design $p^* \times (i^*:h^*)$ Design และ $p^* \times (i^*:h^*)$ Design ซึ่งแต่ละรูปแบบของการออกแบบมีรายละเอียดดังนี้

1. $p^* \times (i^*:h^*)$ Design

สำหรับการออกแบบ $p^* \times (i^*:h^*)$ Design องค์ประกอบความแปรปรวนคือ $p \times (i:h)$ และองค์ประกอบความแปรปรวนร่วมคือ $p \times (i:h)$ เช่นกัน ตัวอย่างเช่น ในการทักษะการฟัง และทักษะการเขียน ให้ผู้ตรวจ (h) ทุกๆ 3 คน ตรวจชิ้นงาน (i) เพียง 4 ชิ้น จากทั้งหมด 12 ชิ้น แต่ผู้ตรวจทุกคนตรวจทั้งสองทักษะ (มิติของการวัด มี 2 มิติ) ในกรณีนี้ชิ้นงาน Nested อยู่ในผู้ตรวจ ซึ่งทั้งชิ้นงาน และผู้ตรวจ Crossed อยู่ในมิติของ v

2. $p^* \times (i^*:h^*)$ Design

การออกแบบ $p^* \times (i^*:h^*)$ Design เป็นการออกแบบที่มีองค์ประกอบความแปรปรวนคือ $p \times (i:h)$ และองค์ประกอบความแปรปรวนร่วมคือ $p \times h$ ส่วน univariate counterpart คือ $p \times [i:(h \times v)]$ หรือกล่าวได้ว่า i Nested ในทั้ง h และ v ตัวอย่างของการออกแบบ เช่น การวัดความสามารถในการอ่านของนักเรียน โดยการสะท้อนจากการอ่านข้อความ (passage) ดังนี้ ในการวัดทักษะการอ่านของนักเรียนมีการใช้ชิ้นงานการอ่าน (i) จำนวน 5 ชุด ในแต่ละชุดมีจำนวน ข้อ 5, 3, 6, 9 และ 7 ตามลำดับ และสำหรับข้อสอบอีกชุด ที่มีจำนวน 5 ชุดเช่นกัน มีจำนวนข้อสอบ 4, 5, 6, 7 และ 8 ตามลำดับ ทั้งนี้ ผู้ตรวจ (h) มีจำนวน 3 คน โดยเป็นผู้ตรวจให้คะแนนทั้ง 2 ชุดข้อสอบ หรือกล่าวได้ว่า ผู้สอบทุกคนทำข้อสอบในสองมิติ และผู้ตรวจตรวจข้อสอบทั้งสองมิติเป็นผู้ตรวจชุดเดียวกัน แต่ผู้ตรวจแต่ละคนตรวจข้อสอบไม่ซ้ำซ้อนกันเลย ดังภาพที่ 2.14

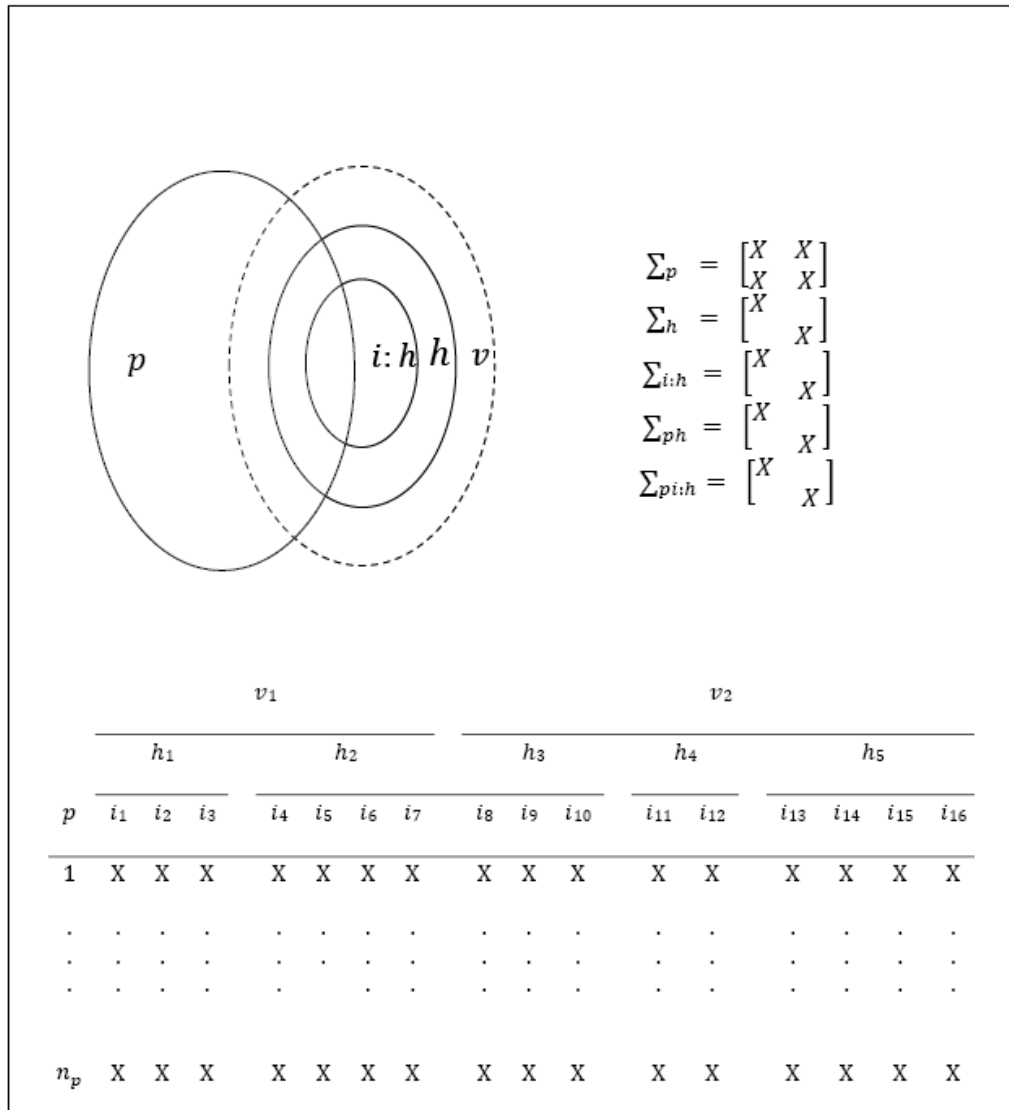


ภาพที่ 2.14 การออกแบบ $p \times (i:h)$ Design

ที่มา: Brennan (2011)

3. $p \times (i:h)$ Design

สำหรับการออกแบบ $p \times (i:h)$ Design เป็นการออกแบบที่มีองค์ประกอบ ความแปรปรวนคือ $p \times (i:h)$ และองค์ประกอบความแปรปรวนร่วมคือ p ส่วน การออกแบบ univariate คือ $p \times (i:h:v)$ หรือกล่าวได้ว่า i Nested ใน h และ h Nested อยู่ใน v ตัวอย่างของการออกแบบ เช่น ผู้ตรวจให้คะแนน (h) แต่ละคนตรวจให้คะแนนข้อสอบต่างชุดกัน โดยที่นักเรียน (p) ทำข้อสอบทุกข้อ และนักเรียนนั้นทำข้อสอบในทุกมิติ ในขณะที่ ข้อสอบ และ ผู้ตรวจนั้น Nested อยู่ในแต่ละมิติ หรือ ไม่มีการใช้ข้อสอบ และผู้ตรวจร่วมกันในแต่ละมิติของแบบสอบ หรือในแต่ละระดับของ v ดังภาพที่ 2.15



ภาพที่ 2.15 การออกแบบ $p^{\circ} \times (i^{\circ}:h^{\circ})$ Design

ที่มา: Brennan (2011)

3.2.1.4 การออกแบบแบบ Class-Means Design

การออกแบบแบบ Class-Means Design ได้แก่ $(p^{\circ}:c^{\circ}) \times i^{\circ}$, $(p^{\circ}:c^{\circ}) \times i^{\circ}$ และ $(p^{\circ}:c^{\circ}) \times i^{\circ}$ การออกแบบแบบ Class-Means Design เป็นการออกแบบที่ในแต่ละระดับชั้น หรือกลุ่ม เช่น ในชั้นเรียน ในโรงเรียน มีความแตกต่างกันไป ซึ่งในที่นี้ได้นำเสนอไว้ 3 รูปแบบ ได้แก่ $(p^{\circ}:c^{\circ}) \times i^{\circ}$, $(p^{\circ}:c^{\circ}) \times i^{\circ}$ และ $(p^{\circ}:c^{\circ}) \times i^{\circ}$ ซึ่งการออกแบบในลักษณะนี้เป็นเหมือนการออกแบบระดับชั้น-ระดับค่าเฉลี่ย เรียกว่า “Class-Means” โดยการออกแบบนี้เป็น การออกแบบที่ผู้สอบ (p) Nested ในชั้นเรียน หรือ โรงเรียน แต่ยังคงทำข้อสอบ (i) เหมือนกัน

แต่แตกต่างกันในส่วนของมิติของแบบสอบโดย $(p^\bullet : c^\bullet) \times i^\bullet$ เป็นการออกแบบที่เหมือนกับ $p^\bullet \times i^\bullet$, $(p^\bullet : c^\bullet) \times i^\circ$ เป็นการออกแบบที่เหมือนกับ $p^\bullet \times i^\circ$ และ $(p^\circ : c^\bullet) \times i^\circ$ เป็นการออกแบบที่เหมือนกับ $p^\circ \times i^\circ$ และ $p^\bullet \times i^\circ$ ความแตกต่างคือ นักเรียนที่สอบนั้น เป็นนักเรียนที่อยู่ในชั้นเรียนบางคน หรือในโรงเรียนบางคนที่ทำกรสอบ ในการออกแบบ $p^\bullet \times i^\bullet$ นั้น นักเรียนทำการสอบทุกข้อเหมือนกัน และข้อสอบเป็นชุดเดียวกันทุกมิติ ส่วน $p^\bullet \times i^\circ$ มีความแตกต่างจากแบบ $p^\bullet \times i^\bullet$ ในส่วนของข้อสอบที่ไม่ซ้ำกันในแต่ละมิติเท่านั้น แต่นักเรียนยังคงทำข้อสอบทุกข้อและทุกมิติ แต่เนื่องจากวิธีนี้ มีความยุ่งยากและเป็นไปได้ยากในปฏิบัติ เพราะการนำนักเรียนจากแต่ละกลุ่มที่ต่างกัน และทำข้อสอบที่มีเนื้อหาต่างกันด้วย จึงทำให้วิธีนี้ไม่ค่อยถูกนำมาใช้

สำหรับการวิจัยครั้งนี้การออกแบบที่ผู้วิจัยเลือกใช้มี 4 แบบ ได้แก่ เป็น 1) $p^\bullet \times c^\circ$ 2) $p^\bullet \times t^\circ$ 3) $p^\bullet \times t^\circ \times r^\bullet$ และ 4) $p^\bullet \times t^\circ \times r^\circ$ โดยแต่ละแบบมีเหตุผลประกอบดังนี้

1. $p^\bullet \times c^\circ$ Design

1.1 เพื่อศึกษาความคลาดเคลื่อนและความน่าเชื่อถือของเกณฑ์การให้คะแนนของภาระงาน และเพื่อศึกษาถึงความน่าเชื่อถือของเกณฑ์ที่ได้ค่าความเที่ยงเหมาะสมที่สุด เกณฑ์ในการให้คะแนนเป็นแบบแยกส่วน ดังนี้ คำศัพท์ การสร้างประโยค การออกเสียง และการเชื่อมโยงเนื้อหา ซึ่งจะทำให้ทราบว่าเกณฑ์การให้คะแนนด้านใดมีความคลาดเคลื่อนมาน้อยเพียงใด เพื่อนำไปสู่การปรับปรุงในการพัฒนาและการนำเกณฑ์ไปใช้

1.2 เพื่อศึกษาความคลาดเคลื่อนและความน่าเชื่อถือของเกณฑ์การให้คะแนนของการประเมินแฟ้มสะสมงานฯ ที่ได้ค่าความเที่ยงเหมาะสมที่สุด เกณฑ์ในการให้คะแนนเป็นแบบแยกส่วน ซึ่งจะทำให้ทราบว่าเกณฑ์การให้คะแนนด้านใด มีความคลาดเคลื่อนมาน้อยเพียงใดเพื่อนำไปสู่การปรับปรุงในการพัฒนาและการนำเกณฑ์ไปใช้

2. $p^\bullet \times t^\circ$ Design การออกแบบในลักษณะนี้เพื่อศึกษาคลาดเคลื่อนและความน่าเชื่อถือของจำนวนภาระงานว่าควรใช้จำนวนเท่าใดในการประเมินแต่ละภาระงาน และในการประเมินแฟ้มสะสมงานในภาพรวมจึงจะได้ความน่าเชื่อถือในระดับที่ยอมรับได้และมีความเป็นไปได้ในทางปฏิบัติ และเพื่อศึกษาถึงความน่าเชื่อถือของจำนวนภาระงานที่ได้ค่าความเที่ยงเหมาะสมที่สุด

3. $p^\bullet \times t^\circ \times r^\bullet$ Design และ $p^\bullet \times t^\circ \times r^\circ$ Design การออกแบบในลักษณะนี้เพื่อศึกษาความคลาดเคลื่อนและความน่าเชื่อถือของรูปแบบการให้คะแนนของผู้ประเมินที่มีการให้คะแนนทุกภาระงาน กับการให้คะแนนภาระงานจากผู้ประเมินที่ต่างคนกัน เพื่อเปรียบเทียบความน่าเชื่อถือในการประเมินจากทั้งสองรูปแบบและผู้บริหารการสอบสามารถเลือกรูปแบบการให้คะแนนนี้ไปปรับใช้ได้ อย่างเหมาะสมต่อไป

องค์ประกอบความแปรปรวนร่วม (Covariance Components)

องค์ประกอบความแปรปรวนร่วม (covariance components) เป็นค่าคาดหวังของผลผลิตของผลจากสองแหล่ง (Brennan, 2001) เช่น ถ้ากำหนดให้ \mathbf{v} เป็น Random effects สำหรับตัวแปร \mathbf{v} และให้ ξ เป็น Random effects สำหรับตัวแปร \mathbf{v}' (โดยที่ $\mathbf{v} \neq \mathbf{v}'$) องค์ประกอบความแปรปรวนร่วมสำหรับ Effects α คือ

$$\sigma_{vv'} \equiv E v_{\alpha} \xi_{\alpha} \quad \text{สมการที่ 1}$$

ตัวอย่างสมการ สำหรับ \mathbf{v} และ \mathbf{v}' สำหรับ $p \cdot x(i : h \cdot)$ Design คือ

$$X_{pihv} = \mu_v + v_p + v_h + v_{i:h} + v_{ph} + v_{pi:h} \quad \text{สมการที่ 2}$$

และ

$$X_{pihv'} = \mu_{v'} + \xi_p + \xi_h + \xi_{i:h} + \xi_{ph} + \xi_{pi:h} \quad \text{สมการที่ 3}$$

เมื่อ i (Item) แตกต่างกันระหว่าง \mathbf{v} และ \mathbf{v}' ภายใน fixed facet ที่ไม่มีความสัมพันธ์กัน นั่นคือ

$$E v_{\alpha} v_{\beta} = E \xi_{\alpha} \xi_{\beta} = 0 \quad \text{สมการที่ 4}$$

ถ้า $\alpha \neq \beta$ แล้ว,

$$E v_{\alpha} \xi_{\beta} = 0 \quad \text{สมการที่ 5}$$

เมื่อ $\alpha \neq \beta$

สำหรับผลในสมการที่ 2 และ 3 คือ 3 องค์ประกอบความแปรปรวนร่วม ที่มีค่าไม่เท่ากับศูนย์ ได้แก่ p , h และ ph ซึ่ง องค์ประกอบความแปรปรวนร่วม คือ

$$\sigma_{vv'}(p) = E v_p \xi_p$$

$$\sigma_{vv'}(h) = E v_h \xi_h$$

และ

$$\sigma_{vv'}(ph) = E v_{ph} \xi_{ph}$$

เมื่อ $\sigma_{vv'}(p)$, $\sigma_{vv'}(h)$, $\sigma_{vv'}(h)$ คือ องค์ประกอบร่วมของ p (persons) h (raters) ตามลำดับ

การประมาณค่าองค์ประกอบความแปรปรวนร่วม

การประมาณค่าองค์ประกอบความแปรปรวนร่วมสามารถประมาณค่าได้ดังนี้

Variance-Covariance Matrices สำหรับ $p \times i$ Design คือ (Brennan, 2001)

$$\Sigma_p = \begin{bmatrix} \sigma_v^2(p) & \sigma_{vv'}(p) \\ \sigma_{vv'}(p) & \sigma_{v'}^2(p) \end{bmatrix}$$

$$\Sigma_i = \begin{bmatrix} \sigma_v^2(i) & \sigma_{vv'}(i) \\ \sigma_{vv'}(i) & \sigma_{v'}^2(i) \end{bmatrix}$$

$$\Sigma_{pi} = \begin{bmatrix} \sigma_v^2(pi) & \sigma_{vv'}(pi) \\ \sigma_{vv'}(pi) & \sigma_{v'}^2(pi) \end{bmatrix}$$

เมื่อ

Σ_p , Σ_i และ Σ_{pi}

คือ เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วม
ของผู้สอบ ข้อสอบ และปฏิสัมพันธ์
(Interaction) ระหว่างผู้สอบและข้อสอบ
ตามลำดับ

$\sigma_v^2(p)$, $\sigma_v^2(i)$ และ $\sigma_v^2(pi)$ คือ ความแปรปรวนของคะแนนจากตัวแปร **U**
ของผู้สอบข้อสอบและปฏิสัมพันธ์ (interaction)
ระหว่าง ผู้สอบและข้อสอบตามลำดับ

$\sigma_{v'}^2(p)$, $\sigma_{v'}^2(i)$ และ $\sigma_{v'}^2(pi)$ คือ ความแปรปรวนของคะแนนจากตัวแปร **U'**
ของผู้สอบข้อสอบและปฏิสัมพันธ์ (interaction)
ระหว่างผู้สอบและผู้ข้อสอบตามลำดับ

$\sigma_{vv'}(p)$, $\sigma_{vv'}(i)$ และ $\sigma_{vv'}(pi)$ คือ ความแปรปรวนร่วมของคะแนนจากตัวแปร
U และ **U'** ของผู้สอบ ข้อสอบและปฏิสัมพันธ์
(interaction) ระหว่าง ผู้สอบและข้อสอบ
ตามลำดับ

3.2.2 D study

Variance-Covariance Matrices for D-study Effects

สำหรับการออกแบบแบบ $p^* \times i^*$ Design นั้น สมการเชิงเส้น สำหรับ ตัวแปร u และ u' คือ

$$\bar{X}_{pv} = X_{pIv} = \mu_v + v_p + v_i + v_{pI}$$

และ

$$\bar{X}_{pv'} = X_{pIv'} = \mu_{v'} + \xi_p + \xi_i + \xi_{pI}$$

เมตริกซ์ของ Variance-Covariance ของ D-study ประกอบด้วย 3 เมตริกซ์ ดังนี้ (Brennan, 2001)

$$\Sigma_p = \begin{bmatrix} \sigma_v^2(p) & \sigma_{vv'}(p) \\ \sigma_{vv'}(p) & \sigma_{v'}^2(p) \end{bmatrix}$$

$$\Sigma_I = \begin{bmatrix} \sigma_v^2(I) & \sigma_{vv'}(I) \\ \sigma_{vv'}(I) & \sigma_{v'}^2(I) \end{bmatrix}$$

$$\Sigma_{pI} = \begin{bmatrix} \sigma_v^2(pI) & \sigma_{vv'}(pI) \\ \sigma_{vv'}(pI) & \sigma_{v'}^2(pI) \end{bmatrix}$$

Variance-Covariance Matrices for Universe Scores and Errors of Measurement

ความคลาดเคลื่อนในการวัดสำหรับ $p^* \times i^*$ Design คือ (Brennan, 2001)

$$\Sigma_\tau = \Sigma_p$$

$$\Sigma_\delta = \Sigma_{pI}$$

$$\Sigma_\Delta = \Sigma_I + \Sigma_{pI}$$

เมื่อ

Σ_τ	คือ	เมตริกซ์ของคะแนนเอกภพ (Universe Scores)
Σ_δ	คือ	เมตริกซ์ของความคลาดเคลื่อนเชิงสัมพัทธ์ (Relative Errors)
Σ_Δ	คือ	เมตริกซ์ของความคลาดเคลื่อนเชิงสัมบูรณ์ (Absolute Errors)
Σ_p	คือ	เมตริกซ์ของคะแนนของผู้สอบ
Σ_I	คือ	เมตริกซ์ของความคลาดเคลื่อนที่เกิดจากข้อสอบ
Σ_{pI}	คือ	เมตริกซ์ของความคลาดเคลื่อนเกิดจากปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบ

ดังนั้น เมตริกซ์ของ Variance-Covariance ของ D-study ประกอบด้วย 3 เมตริกซ์ ดังนี้ (Brennan, 2001)

$$\Sigma_{\tau} = \begin{bmatrix} \sigma_v^2(\tau) & \sigma_{vv'}(\tau) \\ \sigma_{vv'}(\tau) & \sigma_{v'}^2(\tau) \end{bmatrix}$$

$$\Sigma_{\delta} = \begin{bmatrix} \sigma_v^2(\delta) & \sigma_{vv'}(\delta) \\ \sigma_{vv'}(\delta) & \sigma_{v'}^2(\delta) \end{bmatrix}$$

$$\Sigma_{\Delta} = \begin{bmatrix} \sigma_v^2(\Delta) & \sigma_{vv'}(\Delta) \\ \sigma_{vv'}(\Delta) & \sigma_{v'}^2(\Delta) \end{bmatrix}$$

ซึ่ง ความแปรปรวนร่วมของ ความคลาดเคลื่อนเชิงสัมพัทธ์ (Relative Errors) ความคลาดเคลื่อนเชิงสัมบูรณ์ (Absolute Errors) สามารถประมาณค่าได้ดังนี้ (Brennan, 2001)

$$\sigma_{vv'}(\delta) = \sigma_{vv'}(pI) = \frac{\sigma_{vv'}(pi)}{n'_i}$$

และ

$$\begin{aligned} \sigma_{vv'}(\Delta) &= \sigma_{vv'}(I) = \sigma_{vv'}(pI) \\ &= \frac{\sigma_{vv'}(i)}{n'_i} + \frac{\sigma_{vv'}(pi)}{n'_i} \end{aligned}$$

จาก ความแปรปรวนร่วมของความคลาดเคลื่อนเชิงสัมพัทธ์ (relative errors) และความคลาดเคลื่อนเชิงสัมบูรณ์ (absolute errors) นั้นสามารถนำไปสู่ความคลาดเคลื่อนเชิงสัมพันธ์ (correlated error) ดังนี้ (Brennan, 2001)

$$\rho_{vv'}(\delta) = \frac{\sigma_{vv'}(\delta)}{\sqrt{\sigma_v^2(\delta)\sigma_{v'}^2(\delta)}}$$

และ

$$\rho_{vv'}(\Delta) = \frac{\sigma_{vv'}(\Delta)}{\sqrt{\sigma_v^2(\Delta)\sigma_{v'}^2(\Delta)}}$$

เมื่อ $\rho_{vv'}(\delta)$ และ $\rho_{vv'}(\Delta)$ คือ ความคลาดเคลื่อนเชิงสัมพันธ์ (correlated error) เชิงสัมพัทธ์ (Relative Errors) และเชิงสัมบูรณ์ (Absolute Errors) ตามลำดับ

3.2.3 สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจ

สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจ ประกอบด้วย 2 ส่วน ได้แก่ 1) สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (Multivariate Generalizability Coefficient) และ 2) สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (Multivariate phi Coefficient)

1. สัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ (Multivariate Generalizability Coefficient)

สัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ (Multivariate Generalizability Coefficient) สามารถหาค่าได้สัดส่วนของความแปรปรวนของคะแนนเอกภพกับผลบวกของความแปรปรวนของคะแนนเอกภพและความแปรปรวนของความคลาดเคลื่อนเชิงสัมพัทธ์ (Brennan, 2001) ดังนี้

$$E_{p^2} = \frac{\sigma_c^2(\tau)}{\sigma_c^2(\tau) + \sigma_c^2(\delta)}$$

2. สัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ (Multivariate phi Coefficient)

สัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ (Multivariate Phi Coefficient) สามารถหาค่าได้สัดส่วนของความแปรปรวนของคะแนนเอกภพกับผลบวกของความแปรปรวนของคะแนนเอกภพและความแปรปรวนของความคลาดเคลื่อนเชิงสัมบูรณ์ (Brennan, 2001) ดังนี้

$$\Phi = \frac{\sigma_c^2(\tau)}{\sigma_c^2(\tau) + \sigma_c^2(\Delta)}$$

3.3 การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลสำหรับการวิเคราะห์ MGT นั้น Brennan (2001) ได้พัฒนาโปรแกรมสำเร็จรูปขึ้น คือ โปรแกรม mGENOVA ซึ่งเป็นที่นิยมและมีการใช้อย่างแพร่หลาย

3.4 ขั้นตอนการวิเคราะห์การสรุปอ้างอิงแบบพหุตัวแปรสำหรับการตัดสินใจ (Multivariate Generalizability Decision Analysis)

ขั้นตอนในการดำเนินการวิเคราะห์การสรุปอ้างอิงแบบพหุตัวแปรสำหรับการตัดสินใจ (multivariate generalizability decision Analysis) นั้น Mao Cuiyum, Han Yuanhang, Mei Qiang and Liu Jianyi (2010) ได้สรุปไว้ดังนี้

1. กำหนดเอกภพของการสุรूपอ้างอิงตามวัตถุประสงค์ของการประเมิน
 2. ประเมินผลกระทบจากทุกฟาเซท หรือ ปฏิสัมพันธ์ ใน G-study ในที่นี้ หมายถึง จำนวนของระดับของกลุ่มตัวอย่างตามระดับของเงื่อนไขในแต่ละเอกภพการสุรूपอ้างอิง (generalizability universe) จากนั้นหาค่ากำลังสองของค่าเฉลี่ยของปัจจัยทั้งหมด
 3. ประมาณค่าความแปรปรวนคลาดเคลื่อนสัมพัทธ์ (relative error variation) และความแปรปรวนคลาดเคลื่อนสัมบูรณ์ (absolute error variation) ตามลำดับ
 4. ประมาณค่าสัมประสิทธิ์การสุรूपอ้างอิง (generalizability coefficient) หรือดัชนีความเที่ยง
 5. นิยามเอกภพการสุรूपอ้างอิงใหม่ และทำซ้ำในขั้นตอนที่ 4 อีกครั้ง
- ซึ่งในท้ายที่สุดจะได้ข้อสุรूपที่เชื่อถือได้และน่าพึงพอใจ และข้อสุรूपการอ้างอิงโดยการเปรียบเทียบการประมาณค่าความถูกต้องของผลการทดสอบในแต่ละเอกภพการสุรूपอ้างอิง

3.5 งานวิจัยที่เกี่ยวข้องกับทฤษฎีการสุรूपอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร

งานวิจัยที่นำแนวคิดทฤษฎีการสุรूपอ้างอิงความน่าเชื่อถือแบบตัวแปรเดียว และพหุตัวแปรมาใช้นั้นมีหลากหลาย ซึ่งจากการทบทวนงานวิจัยต่างๆ พบว่ามีการนำมาใช้ทั้งทางด้านการเรียนการสอนของแพทย์ คุณภาพการปฏิบัติงาน การวัดความสามารถทางภาษา การศึกษาความเที่ยงและความเป็นพหุมิติ การกำหนดคะแนนจุดตัดของแบบสอบ การออกแบบแบบสอบที่มีรูปแบบผสม โดยรายละเอียดของงานวิจัยแต่ละเรื่องสุรूपได้ดังนี้

1. การประเมินการปฏิบัติงาน

Lisa, Brian and David (2010) ได้ศึกษาเกี่ยวกับประเมินความสามารถในการปฏิบัติงาน (performance) ของนักศึกษาแพทย์ โดยมีวัตถุประสงค์ในการศึกษาคือ เพื่อศึกษาผลของรูปแบบการวิเคราะห์การประมาณค่าขององค์ประกอบความผันแปร ความเที่ยง และความคลาดเคลื่อนมาตรฐานของการวัด (standard error of measurement) เพื่อตรวจสอบความแตกต่างและความสัมพันธ์ระหว่างหมวดของเนื้อหาที่แตกต่างกัน และเพื่ออธิบายประโยชน์ของการใช้ทฤษฎีการสุรूपอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร ในการสร้างแบบสอบหรือการปรับตารางวิเคราะห์เนื้อหา (table of specifications) การศึกษาใช้การจำลองข้อมูลด้วยคอมพิวเตอร์เพื่อประเมินทักษะการบริหารจัดการผู้ป่วย ซึ่งระบบจะทำการบันทึกรายการที่ผู้สอบดำเนินการสอบ และรายการที่ถูกบันทึกจะถูกประเมินโดยผู้เชี่ยวชาญในกรณีที่มีความเหมาะสมหรือถูกให้คะแนนจากคอมพิวเตอร์ กลุ่มตัวอย่างที่ศึกษาเป็นนักเรียนแพทย์ จำนวน 179 คน ที่มาจากคนละรัฐในสหรัฐอเมริกา รายการที่จำลองมาจะถูกตรวจโดยแพทย์ผู้ดูแลหลัก 4 คน ซึ่งทั้งหมดเป็นแพทย์ในโรงเรียนแพทย์เดียวกันและเป็นผู้ที่มีประสบการณ์ในการให้คะแนนในกรณีเหล่านี้มาแล้ว รายการที่ถูกตรวจให้คะแนน เป็นแบบมาตรวัด 9 ระดับ การวิเคราะห์ข้อมูลประกอบด้วย โปรแกรม GENOVA urGENOVA และ mGENOVA

ผลการศึกษาพบดังนี้ ผลการวิเคราะห์ Case within strata (C:S) พบผลมีค่าใกล้เคียงกัน ทั้ง Univariate design and multivariate designs และพบความแตกต่างของ absolute SEM ใน Phi Coefficient เมื่อเปรียบเทียบ G-coefficients และ phi-coefficients (measures of reliability) พบว่า G-coefficients ของ MGT มีค่าสูงที่สุด ถึงแม้ว่าความแตกต่างระหว่างรูปแบบการวิเคราะห์จะมีความแตกต่างกันเล็กน้อยและความคลาดเคลื่อนมาตรฐานเชิงสัมพัทธ์มีค่าต่ำที่สุดด้วยเช่นกัน ส่วน Phi-coefficient พบว่า Univariate nested design มีค่าสูงที่สุด และ ความคลาดเคลื่อนมาตรฐานเชิงสัมบูรณ์ค่าน้อยที่สุด ซึ่งเป็นปัจจัยหนึ่งในความแปรปรวนของความคลาดเคลื่อนเชิงสัมบูรณ์ (absolute error variance) และ phi-coefficient แต่ไม่ใช่ในความแปรปรวนของความคลาดเคลื่อนเชิงสัมพัทธ์ (relative error variance) และ G-coefficient แสดงให้เห็นว่า ถึงแม้จะแตกต่างกันไม่มาก แต่ความแตกต่างระหว่างชั้น โดยเฉพาะอย่างยิ่งมีความหลากหลายภายใน strata ที่ต่างกันทำให้เห็นได้ว่า อาจมีความสำคัญในการเพิ่มจำนวนข้อสอบหรือกรณีการสอบจากทั้ง 4 strata ที่มีความแปรปรวนระหว่างกรณีที่มีมากที่สุดนี้ ซึ่งมีแนวโน้มที่จะมีค่าสัมประสิทธิ์สูงขึ้นซึ่งจะส่งผลกระทบต่อความเที่ยงของการประเมินด้วย ซึ่งข้อมูลนี้ไม่สามารถวิเคราะห์ได้จากการศึกษาแบบ Univariate และจากการวิเคราะห์ G-study นำไปสู่ D-study คือ มีการเพิ่มจำนวนกรณีใน ชั้นที่ 4 (strata) และลดจำนวนกรณีในชั้นที่ 1 ซึ่งจะทำให้ค่าความเที่ยงสูงที่สุด

Cuiyun, Yuanhang, Mei and Jianyi (2010) ศึกษาแบบวัดคุณภาพของผู้ประกอบการ (entrepreneur's qualities) ซึ่งเป็นคุณภาพที่เป็นลักษณะเฉพาะและเป็นความสามารถที่ทำให้ธุรกิจประสบความสำเร็จ ซึ่งประกอบด้วย ความรู้ คุณภาพของบุคลิกภาพ คุณภาพทางจิต คุณภาพความสามารถ ดังนั้น มาตรฐานของ Entrepreneur's qualities จึงไม่ได้เป็นแนวคิดที่มีมิติเดียว แต่เป็นแนวคิดที่เป็นระบบที่ประกอบด้วยหลายมิติ และมีเนื้อที่มีหลายระดับ การออกแบบ G-study Design เป็นแบบ $p \times r$ Design ในการศึกษาครั้งนี้การวิเคราะห์ข้อมูลการใช้โปรแกรม mGENOVA ผลการศึกษาพบว่า ดัชนีทุกตัวมีน้ำหนักเฉลี่ย 0.2 สำหรับระบบการประเมินดัชนีทั้ง 5 มิติ ความเที่ยงและความน่าเชื่อถือของคะแนนรวมของการประเมินจะมีค่ามากกว่า 0.59 ในกรณีของผู้เชี่ยวชาญในการประเมินมีจำนวน 3 คน ซึ่งผู้เชี่ยวชาญจำนวนเพียงเท่านี้เป็นที่ยอมรับโดยทั่วไป แต่ถ้าหากต้องการได้ความเที่ยงรวม (composite reliability) ของการทดสอบทั้งหมดไม่น้อยกว่า 0.8 จะต้องใช้ผู้เชี่ยวชาญอย่างน้อย 6 คน

2. การประเมินความสามารถทางภาษา

Lee (2006) ได้ศึกษาความน่าเชื่อถือของแบบประเมินการพูด ESL (English as a Second Language) ฉบับใหม่ของ TOEFL โดยแบบวัดเป็นแบบประเมินการพูดแบบพหู (multitask speaking) ซึ่งประกอบด้วย 2 ประเภท คือ แบบอิสระ (เป็นแบบประเมินที่ใช้สื่อเพียงอย่างเดียว เน้นให้ผู้พูดใช้ประสบการณ์ของตนเอง) และแบบบูรณาการ (integrated) มี 2 แบบ คือ listening-

speaking และ reading- speaking กลุ่มตัวอย่างที่ใช้ในการศึกษา มีจำนวน 216 คน ฟาเซทที่ศึกษาครั้งนี้เป็นแบบ 2 ฟาเซท ได้แก่ แบบประเมิน (tasks) มี 3 ประเภท และ 11 ฉบับ ส่วนผู้ตรวจ (raters) มีจำนวน 3 คน การออกแบบ G study มี 2 คือ 1) $(r:p) \times t$ และ $p \times t$ แบบที่ 2) $p \times t \times r$ และ $p \times t$ โปรแกรมที่ใช้วิเคราะห์ข้อมูล GENOVA และ mGENOVA ผลการศึกษาพบว่า ความแปรปรวนของคลาดเคลื่อนในแบบประเมินมีขนาดใหญ่กว่าในผู้ตรวจและผู้ประเมินปฏิสัมพันธ์กับแบบประเมินมีความแปรปรวนของความคลาดเคลื่อนมากที่สุด ส่วนการเพิ่มจำนวนของแบบประเมินมีผลต่อความน่าเชื่อถือทำให้ความน่าเชื่อถือเพิ่มขึ้น คะแนนเอกภพมีความสัมพันธ์ที่สูงระหว่างการรวมประเภทของแบบประเมิน และมีการเพิ่มขึ้นเล็กน้อยของคะแนนความเที่ยงรวมเมื่อเพิ่มจำนวนของแบบประเมิน listening-speaking แต่ความแตกต่างของความถูกต้องของความน่าเชื่อถือระหว่างการรวมของแบบประเมินชุดย่อยแบบต่างๆ สำหรับความยาวที่จำกัดมีขนาดไม่มาก

Brennan (2007) ได้ศึกษาความแตกต่างของสัมประสิทธิ์การสรุปอ้างอิงของการอ่านภาษาอังกฤษ และรูปแบบการวิเคราะห์ที่แตกต่างกันระหว่างกลุ่มผู้สอบสองกลุ่ม คือ นักเรียน native English และนักเรียน nonnative English การวิเคราะห์ที่เปรียบเทียบคือ การวิเคราะห์ระหว่าง univariate G-theory และ multivariate G-theory โดยรูปแบบที่ใช้ ได้แก่ $p \times i$, $p \times i^*$, $p \times (i:h)$ และ $p \times (i^*:h^*)$ กลุ่มตัวอย่างที่ใช้ในการศึกษา ประกอบด้วย นักเรียน native English และนักเรียน nonnative English กลุ่มละ 500 คน เครื่องมือที่ใช้ได้แก่ บทความ จำนวน 8 บทความ และข้อสอบ จำนวน 48 ข้อ ประกอบด้วยด้าน 3 ด้าน ที่แทรกอยู่ในแต่ละ passage ผลการศึกษาพบว่า เมื่อเปรียบเทียบระหว่างการใช้ crossed และ nested design พบว่า สำหรับการวิเคราะห์ด้วยทั้ง univariate G-theory และ multivariate G-theory จากทั้งสองกลุ่ม ค่าความคลาดเคลื่อนเชิงสัมพัทธ์และความคลาดเคลื่อนเชิงสัมบูรณ์ของ nested design สูงกว่า crossed design เพียงเล็กน้อย และส่งผลให้สัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์และเชิงสัมบูรณ์ใน nested design ต่ำกว่า crossed design ด้วย เมื่อเปรียบเทียบระหว่างการวิเคราะห์ univariate G-theory และ multivariate G-theory พบว่าสำหรับกลุ่ม nonnative ELL ความคลาดเคลื่อนเชิงสัมพัทธ์ที่วิเคราะห์ด้วย univariate G-theory สูงกว่า multivariate G-theory ทั้ง nested design และ crossed design ส่วนความคลาดเคลื่อนเชิงสัมบูรณ์ ใน crossed design พบว่า univariate G-theory สูงกว่า multivariate G-theory ส่วนกลุ่ม native ELL พบว่าความคลาดเคลื่อนเชิงสัมพัทธ์และเชิงสัมบูรณ์นั้น การวิเคราะห์ univariate G-theory ต่ำกว่า multivariate G-theory ทั้ง nested design และ crossed design เมื่อพิจารณาค่า g coefficient พบว่า ในกลุ่ม nonnative ELL univariate G-theory ต่ำกว่า multivariate G-theory ทั้ง nested design และ crossed design ส่วน phi coefficient พบว่า univariate G-theory สูงกว่า multivariate G-theory สำหรับ crossed

design และ พบว่า univariate G-theory ต่ำกว่า multivariate G-theory สำหรับ nested design ส่วนกลุ่ม native ELL พบว่า ค่า g coefficient และ phi coefficient สำหรับ crossed design นั้น univariate G-theory สูงกว่า multivariate G-theory ส่วน nested design พบว่า univariate G-theory ต่ำกว่า multivariate G-theory

Xiaoming (2007) ได้ทำการศึกษาการใช้ประโยชน์ในการวิเคราะห์การให้คะแนน สำหรับ TAST (TOEFL Academic Speaking Test) ในประโยชน์และสารสนเทศการวินิจฉัย ความเที่ยง เพื่อการนำไปใช้กับ 3 ลักษณะของความสามารถในการปฏิบัติของผู้สอบ คือ Delivery, Language use และ Topic development โดยใช้ G-theory ในการตรวจสอบความน่าเชื่อถือของการ วิเคราะห์คะแนน ความแตกต่างของการวิเคราะห์ขนาดความเที่ยงของการวิเคราะห์โปรไฟล์ของ คะแนน กลุ่มตัวอย่าง คือ นักเรียนที่ไม่ใช่เจ้าของภาษา (non-native speakers) จำนวน 140 คน จาก 7 มหาวิทยาลัยในสหรัฐอเมริกา ซึ่ง TAST ที่มีการพัฒนาการให้คะแนนแบบเกณฑ์แบบองค์รวม (holistic rubrics) ผู้ตรวจให้คะแนน 11 คน (จากทั้งหมด 14 คน) เป็นเจ้าของภาษา (native speakers) ของภาษาอังกฤษ และที่เหลือเป็นผู้ที่ใช้ภาษาอังกฤษเป็นภาษาที่สอง แต่สามารถใช้ภาษา ได้อย่างคล่องแคล่ว เครื่องมือที่ใช้เป็นแบบสอบที่มีรูปแบบเดียว ซึ่งมีทั้งหมด 6 ฉบับ การออกแบบ G study มี 2 ระยะ ดังนี้ ระยะที่ 1 $p \times i \times r$ ระยะที่ 2 $r: (p \times t)$ ผลการศึกษา ความ น่าเชื่อถือของการวิเคราะห์คะแนนเฉลี่ยทั้ง 6 งาน และการให้คะแนนเป็น 2 เท่า นั้นสามารถยอมรับ ได้สำหรับการปฏิบัติ และการตั้งค่าการปฏิบัติ แต่คะแนนเฉลี่ยทั้ง 2 ชิ้น และการให้คะแนนสองเท่ามี ความเที่ยงไม่เพียงพอสำหรับการนำไปปฏิบัติ ความสัมพันธ์ระหว่างการวิเคราะห์คะแนนกับงานพบว่า สูง แต่ระหว่าง delivery and topic development พบว่า มีค่าค่อนข้างต่ำ

Yasuyo (2007) ได้ศึกษาความตรงเชิงทฤษฎีของการประเมินการพูดของภาษาที่สอง จากโปรไฟล์ทางภาษาที่รายงานโดยเป็นการประเมินโดยใช้มาตราประมาณค่า (rating scale) และ มาตรรวม (composite scale) มุ่งศึกษา 3 ประเด็น คือ ความน่าเชื่อถือของมาตรวัดความตรงแบบคู่เข้า และแบบคู่ออก (convergent/discriminant validity) ของแบบวัด และการถ่วงน้ำหนักของการให้ คะแนนการวิเคราะห์ในคะแนนรวม (composite scale) โดยใช้ CFA และ MGT ในการศึกษา จากผู้สอบจำนวน 214 คน กับการศึกษาโปรแกรมการศึกษาต่างประเทศ กับแบบฝึกการพูดโดยการ แสดงบทบาทสมมติ (role-play speaking tasks) แบบฝึกที่เป็นการประเมินภาษาที่สองในประเทศ สเปนที่ออกแบบสำหรับการจัดตำแหน่งนักเรียน (student placement) และการวินิจฉัยผู้เรียน (diagnosis) การให้คะแนนความสามารถด้านการพูดนั้นผู้ประเมินจะให้คะแนนจากมาตรประมาณค่า ซึ่งประกอบด้วย การประเมิน 5 แบบวัด ได้แก่ การออกเสียง (pronunciation) คำศัพท์ (vocabulary) การเชื่อมโยงเนื้อหา (cohesion) การเรียบเรียง (organization) และไวยากรณ์ (grammar) โดย ทุกแบบวัดให้คะแนนเป็นมาตรวัด 4 ระดับ ให้คะแนน 1-4 คะแนน (จากไม่ปรากฏ หลักฐาน-ดี ตามลำดับ) ส่วน ไวยากรณ์ เป็นมาตร 7 ระดับ ให้คะแนน 1-7 คะแนน (จากไม่ปรากฏ หลักฐานอย่าง เป็นระบบ- สมบูรณ์หรือไม่มี ความคลาดเคลื่อนอย่างเป็นระบบ ตามลำดับ) ผู้ตรวจ

ให้คะแนน คือ นักศึกษาระดับบัณฑิตศึกษา จำนวน 15 คน และอาจารย์ภาควิชา of TESL/Applied Linguistics ของ Spanish และ Portuguese ที่เป็นเจ้าของภาษาและภาษาใกล้เคียงของสเปน (native or nearnative speakers) การออกแบบการศึกษา G-study เป็นแบบ 2 facets crossed design ($p \times r \times t$) โดย p คือ ผู้สอบ r คือ ผู้ประเมิน และ t คือแบบฝึกหัด การวิเคราะห์ข้อมูลประกอบด้วย โมเดลของ CFA สำหรับการวิเคราะห์ Multitrait-multimethod (MTMM) โดยใช้ EQS 6.0 Beta และ MGT ศึกษาการออกแบบวัดสอบเพื่อการตัดสินใจในการเลือกรูปแบบการสอบ โดยใช้โปรแกรม mGENOVA ผลการวิเคราะห์ข้อมูล พบว่า ความสัมพันธ์ระหว่างแบบวัดทั้ง 5 อยู่ในระดับสูง คือ อยู่ระหว่าง .85-.98 แสดงว่าแต่ละแบบวัดมีความสัมพันธ์กันสูงจึงทำการวิเคราะห์ CFA ต่อ ซึ่ง Multicomponential และความสัมพันธ์ที่สูงตามธรรมชาติของคุณลักษณะของแบบวัดการพูด LAAS ที่สามารถยืนยันได้ด้วย Higher-order Trait Factor Model ผลจาก MGT พบว่า phi coefficient เท่ากับ .95 หรือกล่าวได้ว่า มากกว่า 95 % ของความแปรปรวนในคะแนนรวมเป็นสัดส่วนจากความแปรปรวนของคะแนนเอกภพ หรือ ความแปรปรวนของคะแนนเป็นผลของความแตกต่างระหว่างคะแนนจริงของผู้สอบ การตรวจสอบการประมาณค่าของ phi-lambdas ($\Phi\lambda$) สำหรับแบบวัดแต่ละฉบับที่ได้จาก D-study ซึ่งเป็นดัชนีสำหรับความน่าเชื่อถือของการตัดสินใจ ก่อนการกำหนดคะแนนจุดตัด (cut score) ซึ่งพบว่า แบบวัดแต่ละฉบับมีค่า Phi-lambdas ที่สูง (.89-.94) แสดงว่ามีความน่าเชื่อถือของการตัดสินใจในคะแนนจุดตัดของแบบวัดได้

Jinyan (2008) ได้ศึกษาความเที่ยงของการให้คะแนน และความเที่ยงของแบบสอบการเขียน ของนักเรียน ESL (ใช้ภาษาอังกฤษเป็นภาษาที่สอง, English secondary language) ในประเทศแคนาดา คำถามวิจัย คือ มีความแตกต่างระหว่างความเที่ยงของการให้คะแนนและความเที่ยงของแบบสอบการเขียนของนักเรียน ESL students and to Native English (NE) students หรือไม่ โดยการออกแบบ G-study มีดังนี้ 1) $p:l \times t \times r'$ (p คือ นักเรียน l คือ กลุ่มภาษา t คือ ชี้นงาน และ r' คือ ผู้ตรวจให้คะแนน) แหล่งความแปรปรวนคือ $l, p:l, t, r', l \times t, l \times r', p:l \times t, p:l \times r', t \times r', l \times t \times r'$ และ $p:l \times t \times r'$ 2) G-study Design คือ $p \times t \times r'$ (p คือ นักเรียน t คือ ชี้นงาน และ r' คือ ผู้ตรวจให้คะแนน) แหล่งความแปรปรวน คือ $p, t, r', p \times t, p \times r', t \times r'$ และ $p \times t \times r'$ การวิเคราะห์ข้อมูล ใช้โปรแกรมสำเร็จรูป GENOVA ผลการศึกษาพบว่า มีอิทธิพลขนาดใหญ่สำหรับทั้งสองกลุ่มภาษา และผู้สอบมีปฏิสัมพันธ์ระหว่างภาษากับแบบฝึกหัดในส่วนขององค์ประกอบความแปรปรวนส่วนที่เหลือ (residual variance component) ที่ไม่พึงประสงค์ พบว่ามีนัยสำคัญสำหรับนักเรียน ESL มากกว่าในนักเรียน NE ทั้งสามปี และความแปรปรวนที่พึงประสงค์มีความสัมพันธ์กับเป้าหมายของการวัด พบว่านักเรียน ESL มีนัยสำคัญน้อยกว่านักเรียน NE ในหนึ่งปี นั่นคือ generalizability coefficient ที่สังเกตได้ของ นักเรียน ESL มีนัยสำคัญต่ำกว่า นักเรียน NE ในปีนั้น

Atilla and Ezel (2010) การศึกษาครั้งนี้เป็นศึกษา CTT และ MGT ซึ่งเป็นข้อมูลที่ได้จากแบบวัดการปฏิบัติงานด้านอาชีพ จากปี 2005-2006 โดยการเปรียบเทียบสัมประสิทธิ์ความเที่ยงของวิธีการ CTT และ MGT ซึ่งใน CTT ใช้การวัดซ้ำ (test retest and Cronbach's alpha reliability coefficient) และใน MGT ใช้ Kendall's concordance coefficient สำหรับความเที่ยงระหว่างผู้ประเมิน (interrater) ในสัมประสิทธิ์การสรุปอ้างอิง (G-coefficients) และสัมประสิทธิ์แบบ Phi (Phi-coefficients) เพื่อตรวจสอบความสอดคล้องของค่าสัมประสิทธิ์ ผลการศึกษาพบว่า สำหรับเงื่อนไขการวัดเดียวกัน สัมประสิทธิ์ความเที่ยงที่ได้จาก CTT และ MGT พบผลการศึกษาที่สอดคล้องกัน กลุ่มตัวอย่างที่ใช้ในการศึกษามีจำนวน 176 คน ในส่วนงานการให้บริการของบริษัท กลุ่มตัวอย่างประกอบด้วย ผู้จัดการ ผู้ช่วยผู้จัดการ หัวหน้างาน และผู้เชี่ยวชาญ แบบวัดการปฏิบัติงานที่ใช้ในการศึกษาครั้งนี้เป็นแบบวัดที่ขึ้นอยู่กับการสำรวจอย่างละเอียดของคุณลักษณะทางจิตมิติ มีจำนวน 32 ข้อ ประกอบด้วย 4 ด้าน ให้คะแนนเป็น 5 ระดับ จาก 1-5 คะแนน ผู้ที่ได้คะแนนมากกว่าแสดงว่ามีความสำเร็จในงานมากกว่า ซึ่งแบบวัดฉบับนี้ได้เคยมีการหาความเที่ยงจาก CTT มีค่าเท่ากับ สัมประสิทธิ์ของแอลฟา .94 (โดยค่าที่น้อยที่สุดมีสัมประสิทธิ์ของแอลฟา เท่ากับ .70) และแบบ test-retest correlation เท่ากับ .84 การวิเคราะห์ข้อมูลใช้โปรแกรมสำเร็จรูป SPSS สำหรับการวิเคราะห์ CTT และใช้โปรแกรมสำเร็จรูป GENOVA สำหรับการวิเคราะห์ข้อมูล MGT โดย MGT จะทำการวิเคราะห์แหล่งความคลาดเคลื่อนอย่างเป็นระบบและแหล่งความคลาดเคลื่อนจากการวัดพาเซทความผันแปรของคลาดเคลื่อน คือ เวลา จำนวนข้อ และปฏิสัมพันธ์ระหว่างเวลาและจำนวนข้อ โดยทำการศึกษา D-Study (decision Study) สำหรับการตัดสินใจการทำงาน การวิเคราะห์เพื่อนำไปสู่ผลการประมาณค่าความสมบูรณ์ของคะแนน สำหรับการเปลี่ยนแปลงที่คัดเลือกในการกำหนดข้อมูล (protocol) ของการวัด ผลการศึกษาพบว่า การตัดสินใจเชิงสัมพัทธ์ และค่าดัชนีความน่าเชื่อถือ (phi coefficients) มีความสัมพันธ์กับการตัดสินใจเชิงสัมบูรณ์

Gebriel (2010) ได้ศึกษาผลการเปรียบเทียบแบบแผนการให้คะแนนของแบบสอบการเขียนภาษาอังกฤษที่ต่างกันระหว่างรูปแบบการตรวจให้คะแนน ดังนี้ 1) ผู้ตรวจต่างกันในแต่ละประเภทของชิ้นงาน 2) ผู้ตรวจเหมือนกันทั้งสองประเภทของชิ้นงาน กลุ่มตัวอย่างที่ทำชิ้นงานการเขียนจำนวน 115 คน ที่ศึกษาในระดับปริญญาตรี ชิ้นงานการเขียนที่ใช้ในการศึกษามีทั้งหมด 4 ฉบับ แบ่งเป็น 2 ประเภท คือ 1) ชิ้นงานแบบอิสระ 2 ฉบับ (writing อย่างเดียว) 2) ชิ้นงานแบบบูรณาการ 2 ฉบับ (reading-to-writing) ผู้ตรวจให้คะแนนมีจำนวน 3 คน การออกแบบ G-study แบ่งออกเป็น 2 แบบ คือ 1) $p \times t \times r$ สำหรับ univariate G-theory และ 2) $p^\circ \times i^\circ \times r^\circ$ และ $p^\circ \times i^\circ \times r^\circ$ Multivariate G-theory โปรแกรมที่ใช้ในการวิเคราะห์ข้อมูล คือ mGENOVA ผลการศึกษาชิ้นงานทั้ง 2 ประเภท มีค่าความน่าเชื่อถือหรือความเที่ยงใกล้เคียงกัน ส่วนผลการเปรียบเทียบรูปแบบการให้คะแนน พบว่า การให้คะแนนทั้งแบบที่มีผู้ตรวจคนเดียวและผู้ตรวจต่างกัน ได้ผล

ในการทำงานเดียวกัน คือ เมื่อผู้ตรวจเพิ่มขึ้น จำนวนชิ้นงานเพิ่มขึ้น ค่า G-coefficient และ phi-coefficient จะเพิ่มขึ้นด้วย

3. การประเมินความเป็นมิติ

Yasuyo (2007) ได้ศึกษาความตรงเชิงทฤษฎีของการประเมินการพูดของภาษาที่สอง จากโปรไฟล์ทางภาษาที่รายงานโดยเป็นการประเมินโดยใช้มาตราประมาณค่า (rating scale) และ มาตรารวม (composite scale) มุ่งศึกษา 3 ประเด็น คือ ความน่าเชื่อถือของมาตรวัด ความตรงแบบลู่เข้า และแบบลู่ออก (convergent/discriminant validity) ของแบบวัด และการถ่วงน้ำหนักของการให้คะแนนการวิเคราะห์ในคะแนนรวม (composite scale) โดยใช้ CFA และ MGT ใน การศึกษา จากผู้สอบจำนวน 214 คน กับการศึกษาโปรแกรมการศึกษาต่างประเทศ กับแบบฝึกการพูดโดยการแสดงบทบาทสมมุติ (role-play speaking tasks) 2 แบบฝึก ที่เป็นการประเมินภาษาที่สองในประเทศสเปนที่ออกแบบสำหรับการจัดตำแหน่งนักเรียน (student placement) และการวินิจฉัยผู้เรียน (diagnosis) การให้คะแนนความสามารถด้านการพูดนั้นผู้ประเมินจะให้คะแนนจากมาตรประเมินค่า ซึ่งประกอบด้วยการประเมิน 5 แบบวัด ได้แก่ การออกเสียง (pronunciation) คำศัพท์ (vocabulary) การเชื่อมโยงเนื้อหา (cohesion) การเรียบเรียง (organization) และ ไวยากรณ์ (grammar) โดยทุกแบบวัดให้คะแนนเป็นมาตรวัด 4 ระดับ ให้คะแนน 1-4 คะแนน (จาก ไม่ปรากฏหลักฐาน-ดี ตามลำดับ) ส่วน Grammar เป็นมาตร 7 ระดับ ให้คะแนน 1-7 คะแนน (จากไม่ปรากฏหลักฐานอย่างเป็นทางการเป็นระบบ- สมบูรณ์หรือไม่มีความคลาดเคลื่อนอย่างเป็นทางการเป็นระบบตามลำดับ) ผู้ตรวจให้คะแนน คือ นักศึกษาระดับบัณฑิตศึกษา จำนวน 15 คน และอาจารย์ภาควิชา of TESL/Applied Linguistics ของ Spanish และ Portuguese ที่เป็นเจ้าของภาษาและภาษาใกล้เคียงของสเปน (native or nearnative speakers) การออกแบบการศึกษา G-study เป็นแบบ 2 facets crossed design ($p \times r \times t$) โดย p คือ ผู้สอบ r คือ ผู้ประเมิน และ t คือแบบฝึกหัด การวิเคราะห์ข้อมูลประกอบด้วย โมเดลของ CFA สำหรับการวิเคราะห์ Multitrait-multimethod (MTMM) โดยใช้ EQS 6.0 Beta และ MGT ศึกษาการออกแบบวัดสอบเพื่อการตัดสินใจในการเลือกรูปแบบการสอบ โดยใช้โปรแกรม mGENOVA ผลการวิเคราะห์ข้อมูล พบว่า ความสัมพันธ์ระหว่างแบบวัดทั้ง 5 พบว่าอยู่ในระดับสูง คือ อยู่ระหว่าง .85-.98 แสดงว่าแต่ละแบบวัดมีความสัมพันธ์กันสูง จึงทำการวิเคราะห์ CFA ต่อ ซึ่ง Multicomponential และความสัมพันธ์ที่สูงตามธรรมชาติของคุณลักษณะของแบบวัดการพูด LAAS ที่สามารถยืนยันได้ด้วย Higher-order Trait Factor Model ผลจาก MGT พบว่า phi coefficient เท่ากับ .95 หรือกล่าวได้ว่า มากกว่า 95 % ของความแปรปรวนในคะแนนรวมเป็นสัดส่วนจากความแปรปรวนของคะแนนเอกภพ หรือความแปรปรวนของคะแนนเป็นผลของความแตกต่างระหว่างคะแนนจริงของผู้สอบ การตรวจสอบการประมาณค่าของ phi-lambdas ($\Phi\lambda$) สำหรับแบบวัดแต่ละฉบับที่ได้จาก D-study ซึ่งเป็นดัชนีสำหรับความน่าเชื่อถือของการ

ตัดสินใจ ก่อนการกำหนดคะแนนจุดตัด (cut score) ซึ่งพบว่า แบบวัดแต่ละฉบับมีค่า Phi-lambdas ที่สูง (.89-.94) แสดงว่ามีความน่าเชื่อถือของการตัดสินใจในคะแนนจุดตัดของแบบวัดได้

4. การความเที่ยงของรูปแบบของแบบสอบแบบต่างๆ

Brennan (2009) ได้ศึกษาความเที่ยงของแบบที่มีการรวมรูปแบบของแบบสอบหลายแบบเข้าด้วยกัน หรือที่เรียกว่า แบบสอบที่ใช้รูปแบบผสม (mixed-formats exams) โดยการศึกษาครั้งนี้เป็นการตรวจสอบผลของการสร้างเครื่องมือ ส่วนของการถ่วงน้ำหนัก และจำนวนข้อต่อตอน ในความเที่ยง error variance และ standard errors of measurement ตามเงื่อนไข โดยการใช้ MGT ซึ่งแบบสอบที่มีรูปแบบผสม ประกอบด้วย แบบสอบ multiple-choice scoring และแบบสอบที่มีการตอบแบบอิสระ (free-response items) ในวิชาชีววิทยา และ AP ประวัติศาสตร์โลก ผลการศึกษามีความจำเป็นต่อการพิจารณาการสอบ การจัดวางตำแหน่งขั้นสูง (advanced placement, AP) ผลการศึกษาค้นคว้าชี้ให้เห็นว่า คะแนนจากแบบสอบ multiple-choice scoring (จำนวนข้อที่ถูกกับสูตรการให้คะแนน (number correct vs. formula scoring)) อาจมีผลกระทบต่อส่วนค่าความเที่ยงโดยรวม และผลในการถ่วงน้ำหนักที่แตกต่างกันที่ตีที่สุด ถึงแม้ว่าส่วนของ multiple-choice มีส่วนช่วย 50-60% ของคะแนนส่วนประกอบ (composite score) จุดที่สามารถใช้ในการถ่วงน้ำหนักที่ดีที่สุดจากความเที่ยงหรือความแปรปรวนคลาดเคลื่อน คือ ควรเพิ่ม multiple-choice เพื่อช่วยให้คะแนนรวม ถึง 80 หรือ 90% การเพิ่มจำนวนข้อของข้อสอบแบบตอบแบบอิสระ (free-response items) มักจะช่วยเพิ่มความเที่ยงโดยรวม แต่การบริหารเวลาจะเป็นการจำกัดจำนวนข้อสอบแบบตอบแบบอิสระ

5. การศึกษาจุดตัดของแบบสอบ

Ping and Sconing (2008) การศึกษาค้นคว้าครั้งนี้เป็นการศึกษาโดยใช้ G-theory ในการวัดความเที่ยงของผลของคะแนนจุดตัด (cut score) จากกระบวนการจัดแบบมาตรฐาน (standard-setting process) โดย Gt ใช้ในการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนจุดตัดจาก 2 วิธี คือ Item rating (Angoff-based) และ mapmark (bookmark-based) Methods วิธีการที่นิยมใช้ Standard-setting procedures คือ Angoff ซึ่งวิธีนี้เป็นวิธีดั้งเดิมที่ใช้ในการประเมินระดับชาติด้านความก้าวหน้าทางการศึกษา โดยใช้ผู้เชี่ยวชาญในการตัดสิน ประมาณค่าความน่าจะเป็นจากผู้ที่มีความผลสัมฤทธิ์น้อยที่สุดในแต่ละระดับ ส่วน Bookmark procedure เป็นวิธีที่ใช้ IRT โดยผู้เชี่ยวชาญจะเรียงข้อสอบตามความยาก เริ่มจากที่ง่ายที่สุด ผู้เชี่ยวชาญแต่ละที่จะเลือกข้อสอบที่รู้สึกว่ายากที่สุดใน 2 กลุ่ม คือ กลุ่มข้อสอบที่นักเรียนที่มีคุณภาพน้อยที่สุดจะสามารถทำได้ และกลุ่มข้อสอบที่ยากเกินไปสำหรับการสอบครั้งนั้น โดยศึกษา 2 แบบ คือ 1) มี 4 ฟาเซท คือ $(p: g: f) \times r$ design เมื่อ p คือ ผู้เชี่ยวชาญ g คือ ตารางกลุ่ม f คือ ข้อที่ง่าย และ r คือ conceptualization of a facet (round) 2) มี 3 ฟาเซท คือ $(p: g: f)$ design การวิเคราะห์

ข้อมูลใช้โปรแกรม urGENOVA ผลการศึกษาพบว่า คะแนนจุดตัดมีความสอดคล้องกันทั้งสองวิธี The first round standard setting สนับสนุนความเที่ยงโดยรวมมากที่สุดสำหรับวิธี Mapmark method และชัดเจนว่าไม่มีความคลาดเคลื่อนมาตรฐานใดที่มีความสัมพันธ์กับ certain cut score ตารางที่ 2.15 สรุปผลจากการทบทวนเอกสารงานวิจัยที่เกี่ยวข้องกับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร

ผู้วิจัย	ปัญหาวิจัย/ วัตถุประสงค์	บริบทการศึกษา	การออกแบบ การวิจัย	โปรแกรมการ วิเคราะห์ข้อมูล	ผลการวิจัย
Lee (2006)	ศึกษาความ น่าเชื่อถือของแบบ ประเมินการพูด	การวัด ความสามารถ ทางการพูด ภาษาอังกฤษ	$(r:p) \times t p \times t$ $p^* \times t^* \times r^*$ $p^* \times t^*$	GENOVA mGENOVA	- ความแปรปรวน ของปฏิสัมพันธ์ ระหว่างผู้ตรวจ และแบบ ประเมินมีขนาด ใหญ่ที่สุด - ความเที่ยง เพิ่มขึ้นเล็กน้อย เมื่อเพิ่มจำนวน แบบประเมิน
Xi (2007)	ศึกษาความเที่ยง ของแบบสอบที่มี ลักษณะต่างกัน 3 แบบ	การวัด ความสามารถ ทางการพูด ภาษาอังกฤษ	$p \times i \times r$ $r: (p \times t)$	GENOVA	ความน่าเชื่อถือของ แบบสอบทั้ง 6 ฉบับ เมื่อให้คะแนน แบบ 2 เท่า สามารถยอมรับได้ ส่วนการใช้แบบ สอบจำนวน 2 ฉบับ และการให้คะแนน แบบ 2 เท่าไม่มี ความเหมาะสม
Sawaki (2007)	- ศึกษาความ น่าเชื่อถือของ แบบสอบ - ความตรงแบบ คู่เข้า และแบบ คู่ออกของแบบสอบ - การถ่วงน้ำหนัก ของการให้คะแนน แบบสอบ	การวัด ความสามารถ ทางการพูด ภาษาอังกฤษ	$p \times r \times t$	EQS 6.0 Beta mGENOVA	phi-coefficient มีค่าเท่ากับ .95 และแบบวัดแต่ละ ฉบับมีความ น่าเชื่อถือสูง โดย phi-lambdas เท่ากับ .89-.94
Li and Brennan (2007)	เปรียบเทียบ ระหว่างการ ออกแบบ crossed และ nested design และ การ วิเคราะห์ GT และ MGT	การวัดการอ่าน	$p \times i,$ $p^* \times i^*,$ $p \times (i: h)$ $p^* \times (i^*: h^*)$	mGENOVA	ความคลาดเคลื่อน ของ MGT สูงกว่า GT และ ส่วน G และ phi coefficient ของ nested design สูงกว่า crossed design

ตารางที่ 2.15 (ต่อ)

ผู้วิจัย	ปัญหาวิจัย/ วัตถุประสงค์	บริบทการศึกษา	การออกแบบ การวิจัย	โปรแกรมการ วิเคราะห์ข้อมูล	ผลการวิจัย
Ping and Scoring (2008)	ศึกษาความเที่ยง ของแบบสอบจาก วิธีการทำคะแนน จุดตัดระหว่างวิธี Angoff-based และBookmark- based	การศึกษา	$(p: g: f) \times r$ $(p: g: f)$	urGENOVA	คะแนนจุดตัดมีความสอดคล้องกัน ทั้งสองวิธี และ พบว่า วิธี Bookmark- based มีความ เที่ยงสูงสุด
Huang (2008)	ศึกษาความเที่ยง ของแบบสอบ	การวัด ความสามารถ ทางการเขียน ภาษาอังกฤษ	$p: i \times t \times r'$ $p \times t \times r'$	GENOVA	- ปฏิสัมพันธ์ ระหว่างผู้สอบ กับแบบฝึก มีความ แปรปรวนของ ความ คลาดเคลื่อนสูง - G-coefficient ของนักเรียนกลุ่ม ESL สูงกว่า กลุ่ม NE
Powers and Brennan (2009)	ความเที่ยงของแบบ สอบที่มีรูปแบบ ผสม (Mixed- formats exams)	การศึกษา	$p^* \times i^o$	mGENOVA	- การเพิ่ม Multiple-choice ทำให้ความเที่ยง ของแบบสอบ อยู่ระหว่าง 80 หรือ 90 - การเพิ่มจำนวน ข้อสอบแบบ Free-response item ทำให้ความ เที่ยงเพิ่มขึ้น เช่นกัน แต่อาจมี ปัญหาเกี่ยวกับ การบริหารเวลา ในการสอบ
Atila and Ezel (2010)	เปรียบเทียบความ เที่ยงระหว่าง CTT และ MGT	การประเมิน ความสามารถใน การปฏิบัติงาน	$p \times t \times i$	- SPSS สำหรับ CTT - GENOVA สำหรับ GT - mGENOVA สำหรับ MGT	ค่าสัมประสิทธิ์เชิง สัมพันธ์มี ความสัมพันธ์กัน กับการตัดสินใจเชิง สัมบูรณ์

ตารางที่ 2.15 (ต่อ)

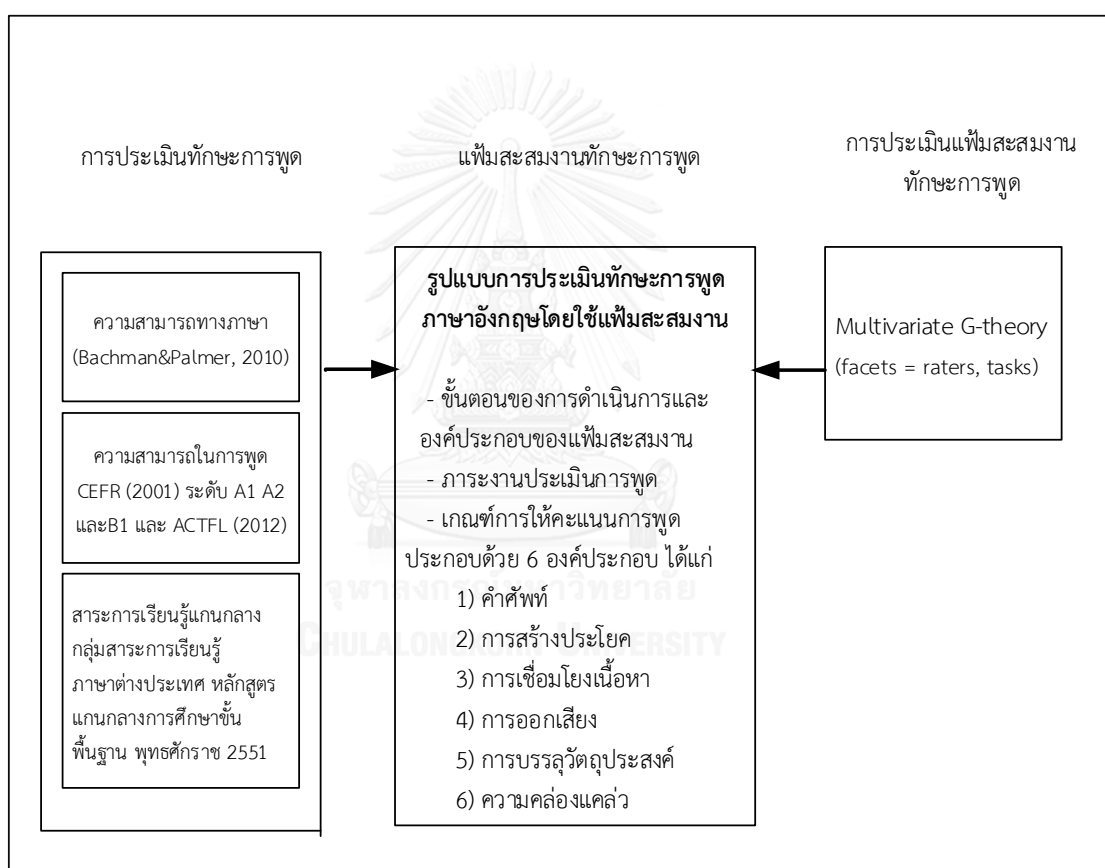
ผู้วิจัย	ปัญหาวิจัย/ วัตถุประสงค์	บริบทการศึกษา	การออกแบบ การวิจัย	โปรแกรมการ วิเคราะห์ข้อมูล	ผลการวิจัย
Lisa, et al (2010)	ศึกษาความแตกต่าง ความสัมพันธ์ ระหว่างหมวดของ เนื้อหาที่ต่างกัน	การประเมิน ความสามารถใน การปฏิบัติของ นักศึกษาแพทย์	$p \times (c:s)$	- GENOVA - urGENOVA - mGENOVA	G-coefficients ของ MGT มีขนาด ใหญ่ที่สุด และ ความคลาดเคลื่อน เชิงสัมพัทธ์ของ MGT มีขนาดเล็ก ที่สุดด้วยเช่นกัน
Cuiyun., et al (2010)	ศึกษาคุณภาพของ การประเมิน คุณภาพของ ผู้ประกอบการ	คุณภาพของ ผู้ประกอบการ	$p \times r$	mGENOVA	หากต้องการค่า ความเที่ยงไม่น้อย กว่า .80 ต้องใช้ ผู้เชี่ยวชาญสำหรับ การประเมินตั้งแต่ 6 คนขึ้นไป
Gebriel (2010)	เปรียบเทียบแบบ แผนการให้คะแนน ของแบบสอบถาม เขียนภาษาอังกฤษ ที่ต่างกันระหว่าง ของรูปแบบการ ตรวจให้คะแนน	การวัด ความสามารถ ทางการเขียน ภาษาอังกฤษ	$p \times t \times r$ $p^* \times i^* \times r^*$ $p^* \times i^* \times r^*$	mGENOVA	การให้คะแนน ทั้ง 2 แบบ ได้ผล ในทำนองเดียวกัน คือ เมื่อผู้ตรวจ เพิ่มขึ้น จำนวน ชิ้นงานเพิ่มขึ้น ค่า G-coefficient และ phi- coefficient จะ เพิ่มขึ้น

จากตารางที่ 2.15 แสดงให้เห็นถึงงานวิจัยที่ใช้ในการวัดและประเมินผลโดยใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรนั้นมีการนำมาใช้ในหลายสาขา ได้แก่ การประเมินการปฏิบัติงานของนักศึกษาการแพทย์ การประเมินการปฏิบัติงาน การออกแบบรูปแบบของแบบสอบ การศึกษาความเที่ยงของแบบสอบ การประเมินคุณภาพผู้ประกอบการ และการวัดและประเมินผลทางภาษา ซึ่งเป็นการศึกษาทั้ง univariate G-theory และ multivariate G-theory มีการออกแบบทั้งแบบ crossed design และ nested ส่วนโปรแกรมที่ใช้ในการวิเคราะห์ข้อมูล พบว่า ใน univariate G-theory ใช้ GENOVA และ urGENOVA ในการวิเคราะห์ข้อมูล ส่วน multivariate G-theory พบว่าโปรแกรมที่ใช้คือ mGENOVA และพบว่ามีมีการนำมาพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษ แต่ถึงอย่างไรก็ตามยังไม่พบว่ามีการใช้ multivariate G-theory มาประยุกต์ใช้ในการพัฒนารูปแบบแฟ้มสะสมงาน และการประเมินทักษะทางภาษาของเด็ก

ตอนที่ 4 กรอบแนวคิดการวิจัย

กรอบแนวคิดการวิจัยในการศึกษาค้นคว้าครั้งนี้เป็นการพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ที่มุ่งศึกษาการสร้างรูปแบบแฟ้มสะสมงานเพื่อการประเมินทักษะการพูดภาษาอังกฤษที่มีการให้คะแนนทักษะการพูดของนักเรียนทั้งระหว่างเรียน (formative assessment) และสรุปรวม (summative assessment) ตามกรอบแนวคิดของ Burke, Fogart and Begrad (1994), Lam and Lee (2010) เนื่องจากเป็นแนวคิดในการพัฒนาแฟ้มสะสมงานที่มีขั้นตอนการพัฒนาที่สมบูรณ์ ละเอียดชัดเจน และสามารถนำมาพัฒนาได้จริง นอกจากนี้ในการวิจัยครั้งนี้ยังเป็นการพัฒนาแฟ้มสะสมงานที่มีการประเมินทั้งระหว่างเรียน (formative assessment) เพื่อมุ่งเน้นพัฒนาความก้าวหน้าของนักเรียน ส่งเสริมให้นักเรียนมีการปรับปรุงวิธีการเรียนของตนเองตลอดเวลาและพัฒนาความสามารถทางภาษาได้ตามเป้าหมายที่กำหนดไว้ เมื่อสิ้นสุดการเรียนการสอนจะมีการประเมินการพูดภาษาอังกฤษของนักเรียนจากครูผู้สอนและผู้เชี่ยวชาญเพื่อประเมินระดับทักษะการพูดภาษาอังกฤษในภาพรวม (summative assessment) ทำให้นักเรียนทราบระดับความสามารถของตน และวางแผนการพัฒนาการเรียนในอนาคตได้ ส่วนการประเมินทักษะการพูด ผู้วิจัยนำกรอบแนวคิดการประเมินความสามารถทางภาษาของ Bachman and Palmer (2010) ในส่วนของความรู้ทางภาษา (language knowledge) ซึ่งประกอบด้วย 2 องค์ประกอบ ได้แก่ ความรู้ในการเรียบเรียง (organizational knowledge) และความรู้เกี่ยวกับการใช้ภาษาในทางปฏิบัติ (pragmatic knowledge) ความสามารถทางภาษาตามกรอบของ Common European Framework Reference ซึ่งมีอยู่ 6 ระดับ (A1 A2 B1 B2 C1 และ C2) แต่เนื่องจากการศึกษาค้นคว้าครั้งนี้กลุ่มตัวอย่างที่ศึกษาเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ดังนั้นจึงทำการประเมินทักษะการพูดของนักเรียนใน 3 ระดับเท่านั้น ได้แก่ A1 A2 และ B1 (อายุ 11-13 ปี) (Mckay, 2006) เกณฑ์การให้คะแนนการพูดของ ACTFL Oral Proficiency Interview (OPI) จาก American Council on the Teaching Foreign Language (ACTFL, 2012) และหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ซึ่งแนวคิดการประเมินความสามารถทางภาษาของ Bachman และ Palmer (2010) เป็นแนวคิดที่ได้รับการยอมรับอย่างแพร่หลายในการประเมินความสามารถทางภาษาที่สอง (second language) และภาษาต่างประเทศ (EFL) เพราะได้รับการพัฒนาองค์ประกอบความสามารถทางภาษามาอย่างต่อเนื่อง ส่วน CEFR (Common European Framework Reference, 2001) และ ACFTL (American Council on the Teaching Foreign Language, 1999, 2012) เป็นการประเมินเกี่ยวกับระดับความสามารถทางภาษาต่างประเทศที่เป็นที่ยอมรับทั่วโลก มีการนำมาใช้เป็นมาตรฐานในการจัดการเรียนการสอนในยุโรป และในระดับโลก และการประเมินทักษะการพูดภาษาอังกฤษ โดยใช้แฟ้มสะสมงานครั้งนี้เป็นการประยุกต์ใช้ทฤษฎี

ความน่าเชื่อถือแบบพหุตัวแปร เพื่อการวิเคราะห์สัมประสิทธิ์ความน่าเชื่อถือและความคลาดเคลื่อน ในการประเมินทักษะการพูดภาษาอังกฤษ ซึ่งเป็นแนวคิดทางการวัดที่สามารถแยกแหล่งความคลาดเคลื่อน ออกเป็นหลายแหล่ง และทำให้การประเมินสามารถเลือกสถานการณ์ที่มีความเหมาะสมและมีความ น่าเชื่อถือในระดับที่ยอมรับได้ ซึ่งการวิจัยครั้งนี้ มุ่งศึกษาแหล่งความคลาดเคลื่อนจำนวน 2 แหล่ง (facets) ได้แก่ ผู้ประเมิน (raters) และภาระงานการพูดภาษาอังกฤษ (tasks) เพราะเป็นสิ่งที่มีความ เกี่ยวข้องกับการประเมินความสามารถของนักเรียนที่ต้องแสดงหรือปฏิบัติ (performance) ดังภาพที่ 2.16



ภาพที่ 2.16 กรอบแนวคิดการวิจัย

บทที่ 3 วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนารูปแบบการประเมินแฟ้มสะสมงานทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษาปีที่ 6 ที่มุ่งศึกษาการสร้างรูปแบบแฟ้มสะสมงานของการประเมินทักษะการพูดภาษาอังกฤษ ที่มีการให้คะแนนทักษะการพูดของนักเรียนทั้งระหว่างเรียน (formative assessment) และสรุปรวม (summative assessment) ตามกรอบแนวคิดของ Burke, Fogart and Begrad (1994), Hamp-Lyons และ Condon (2000), Lam and Lee (2010) ส่วนการประเมินทักษะการพูดใช้แนวคิดการประเมินความสามารถทางภาษาของ Bachman และ Palmer (2010) Common European Framework Reference (2001) ACFTL (American Council on the Teaching Foreign Language, 2012) และหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 โดยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร โดยแบ่งการดำเนินการออกเป็น 4 ระยะ ดังนี้

ระยะที่ 1 การพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

ระยะที่ 2 การทดลองใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

ระยะที่ 3 การวิเคราะห์ผลการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

ระยะที่ 4 การพัฒนาคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ระยะที่ 1 การพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

ตอนที่ 1 การพัฒนาเกณฑ์การให้คะแนนการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ดำเนินการ ดังนี้

1.1 ศึกษาเอกสารเกี่ยวกับการทดสอบความสามารถทางภาษาและการประเมินทักษะการพูด ดังนี้

1.1.1 มาตรฐานการเรียนรู้และตัวชี้วัดสาระการเรียนรู้ภาษาต่างประเทศ (ภาษาอังกฤษ) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 องค์ประกอบความสามารถทางภาษาของ Bachman and Palmer (2010) มาตรฐานความสามารถทางภาษาอังกฤษของสหภาพยุโรป

(CEFR, 2001) และ ACFTL (American Council on the Teaching Foreign Language, 1999, 2012)

1.1.2 ระดับความสามารถทางภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 และภาระงาน (tasks) ที่จะใช้ในการประเมินความสามารถ และองค์ประกอบที่จะใช้ในการประเมินการพูดภาษาอังกฤษของนักเรียน

1.1.3 เกณฑ์การให้คะแนนที่ใช้ในการประเมินความสามารถทางการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ตามแนวคิดการประเมินความสามารถทางภาษาของ Bachman and Palmer (2010) CEFR (2001) และ ACFTL (2012) และการสังเคราะห์จากงานวิจัยที่เกี่ยวข้อง และทำการสัมภาษณ์ผู้ทรงคุณวุฒิ จำนวน 9 คน

เกณฑ์การคัดเลือกผู้ทรงคุณวุฒิ ดังนี้

1) ผู้ทรงคุณวุฒิด้านการสอนภาษาอังกฤษ ที่มีประสบการณ์ในการสอนวิชาภาษาอังกฤษไม่ต่ำกว่า 10 ปี จบสาขาการสอนภาษาอังกฤษหรือสาขาที่เกี่ยวข้อง หรือเป็นผู้ที่จบการศึกษาด้านการสอนภาษาอังกฤษในระดับปริญญาโทขึ้นไป และมีประสบการณ์ในการสอนวิชาภาษาอังกฤษไม่ต่ำกว่า 5 ปี จำนวน 3 คน

2) ผู้ทรงคุณวุฒิด้านแฟ้มสะสมงาน เป็นผู้ที่มีประสบการณ์ในการใช้แฟ้มสะสมงานในการจัดการเรียนการสอน จำนวน 3 คน

3) ผู้ทรงคุณวุฒิด้านการวัดและประเมินผลการศึกษา เป็นผู้ที่มีประสบการณ์ด้านการวัดและประเมินผลการศึกษาอย่างน้อย 5 ปี และจบการศึกษาด้านการวัดและประเมินผล การศึกษาระดับปริญญาโทขึ้นไป จำนวน 3 คน

โดยเมื่อทำการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้อง รวมทั้งการสัมภาษณ์ผู้ทรงคุณวุฒิแล้วได้ทำการร่างเกณฑ์การพูดภาษาอังกฤษให้ผู้ทรงคุณวุฒิทางภาษาอังกฤษ จำนวน 5 คน ตรวจสอบความตรงและความเหมาะสมของเกณฑ์การพูด และทำการปรับปรุงแก้ไขตามข้อเสนอแนะของผู้ทรงคุณวุฒิ พบผลดังนี้

ภาระงานการพูดภาษาอังกฤษตามระดับความสามารถทางภาษาอังกฤษของนักเรียนอยู่ในระดับต้น (A1 (breakthrough) A2 (way stage)) และระดับอิสระ (B1 (threshold)) โดยมีภาระงาน (tasks) ที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษครั้งนี้ประกอบด้วย 4 ภาระงาน โดยแบ่งออกเป็นภาระงานที่ใช้ประเมินระหว่างเรียน 3 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอ ปากเปล่า และการเล่านิทาน ส่วนภาระงานที่ใช้ในการประเมินสรุปรวม 1 ภาระงาน คือ การบรรยายภาพ ซึ่งแต่ละภาระงานมีรายละเอียดดังนี้

1) การสัมภาษณ์ (interview) เป็นกิจกรรมการพูดที่มีการสัมภาษณ์โดยครูเป็นผู้สัมภาษณ์เกี่ยวกับเรื่องใกล้ตัว เช่น กิจกรรมที่ทำในตอนเช้าก่อนมาโรงเรียน กิจกรรมที่ทำในห้องเรียน กีฬาที่ชอบ และเหตุผลที่ชอบ เป็นต้น

2) การนำเสนอปากเปล่า (oral presentation) เป็นกิจกรรมการพูดที่ให้นักเรียนได้พูดนำเสนอเรื่องราวเกี่ยวกับตนเอง โดยครูให้นักเรียนเตรียมตัวมาก่อนล่วงหน้า เช่น ให้ค้นหาคำศัพท์ที่เกี่ยวข้อง ลองแต่งประโยค ปรึกษากับครูในชั้นเรียน หรือผู้ปกครอง เนื้อหาที่พูด เช่น ประวัติโดยย่อ งานอดิเรก กิจกรรมที่สนใจ เป็นต้น

3) การเล่านิทาน (storytelling) เป็นกิจกรรมที่ครูให้นักเรียนเตรียมตัวก่อนล่วงหน้า โดยเตรียมนิทานมาเล่าคนละ 1 เรื่อง หรือเลือกเรื่องจากที่ครูเตรียมให้ อาจนำหนังสือนิทานหรืออุปกรณ์มาประกอบการเล่า โดยนิทานที่จะใช้ในการเล่าครูเป็นผู้เตรียมให้

4) การบรรยายภาพ (picture description) เป็นกิจกรรมที่ครูให้นักเรียนพูดบรรยายภาพ โดยนักเรียนจะได้เตรียมตัวก่อนล่วงหน้า คือ ศึกษารายละเอียดต่างๆ ของภาพ ชักถามรายละเอียดของภาพจากครูผู้สอน ค้นหาคำศัพท์ที่เกี่ยวข้อง ลองแต่งประโยค ปรึกษากับครูในชั้นเรียน หรือผู้ปกครอง เป็นต้น โดยภาพประกอบที่ใช้เป็นภาพที่ครูเตรียมไว้ให้ แล้วให้นักเรียนบรรยายเรื่องราวในภาพของตนเอง

ส่วนเกณฑ์การให้คะแนนที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ทั้ง 4 ภาระงานเป็นเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) มีขอบเขตเนื้อหาในการประเมิน ดังนี้ มีรายละเอียดดังนี้

1) คำศัพท์ (vocabulary) คือ การให้คะแนนด้านคำศัพท์ จะพิจารณาจากการพูดของนักเรียนที่สะท้อน 2 สิ่ง คือ ขอบเขต (range) ของคำศัพท์ที่ใช้ และความถูกต้อง (accuracy) กล่าวคือ ขอบเขต (range) ของการใช้คำศัพท์ที่มีความหลากหลาย ความสามารถในการใช้คำศัพท์ได้อย่างถูกต้อง (accuracy) และเหมาะสม (appropriation) ตรงตามความหมายที่ต้องการสื่อสาร

2) การสร้างประโยค (syntax) คือ การให้คะแนนด้านการสร้างประโยค พิจารณาจากความสามารถของนักเรียนในสร้างประโยคได้อย่างมีความถูกต้อง (accuracy) ตามที่ต้องการสื่อสาร เช่น มีประธาน กริยา กรรม ที่เหมาะสม เป็นต้น

3) การเชื่อมโยงเนื้อหา (cohesion) การให้คะแนนด้านการเชื่อมโยงเนื้อหา พิจารณาจากการที่นักเรียนสามารถพูดเชื่อมโยงประโยคแต่ละประโยคให้เป็นเรื่องราวที่มีความสัมพันธ์กันได้อย่างถูกต้องเหมาะสม

4) การออกเสียง (pronunciation) การให้คะแนนด้านการออกเสียง พิจารณาจากการที่นักเรียนสามารถเน้นเสียงหนักเบา (stress) เว้นวรรคการหยุด (segmental sound) และการใช้เสียงสูงต่ำ (intonation) ได้อย่างถูกต้อง เหมาะสม และทำให้ผู้ฟังเข้าใจในสิ่งที่ต้องการสื่อสารได้

5) บรรลุวัตถุประสงค์ (ideational function) คือ การให้คะแนนด้านบรรลุวัตถุประสงค์ พิจารณาจากการที่นักเรียนสามารถพูดสื่อสารได้บรรลุตามวัตถุประสงค์ที่ต้องการจะสื่อสาร โดยการพูดนั้นทำให้ผู้ฟังเข้าใจว่าต้องการสื่อสารอะไร เช่น พูดได้ตรงกับหัวเรื่องที่ระบุ ตอบตรงคำถาม ตอบคำถามได้อย่างเหมาะสมกับลักษณะคำถาม เป็นต้น

6) ความคล่องแคล่ว (fluency) คือ การใช้ภาษาที่มีความคล่องแคล่ว ลื่นไหล ซึ่งพิจารณาจากการที่นักเรียนสามารถพูดภาษาอังกฤษได้อย่างคล่องแคล่ว ราบรื่น และมีความลื่นไหลในระหว่างที่พูด และไม่สะดุดหรือกระตุก หยุดชะงัก ระหว่างที่พูด

โดยให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี ตามลำดับ ซึ่งมีผู้วิจัยพัฒนามาจาก Sapsirin (2006) และ Sawaki (2007) และการสัมภาษณ์ผู้ทรงคุณวุฒิทางภาษาอังกฤษ จำนวน 9 คน

1.2 การตรวจสอบความเหมาะสมของของภาระงานและเกณฑ์การให้คะแนน

การตรวจสอบความเหมาะสมของของภาระงานและเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 แบบแยกส่วน (analytic rating scale) โดยผู้ทรงคุณวุฒิจำนวน 9 คน และวิเคราะห์ค่า เพื่อตรวจสอบความเหมาะสมและความสอดคล้องของเกณฑ์การให้คะแนนกับวัตถุประสงค์การประเมินทักษะการพูดภาษาอังกฤษของนักเรียน เมื่อทำการปรับปรุงแก้ไขเกณฑ์การให้คะแนนตามที่ผู้ทรงคุณวุฒิเสนอแนะเรียบร้อยแล้ว ผู้วิจัยนำเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษไปทดลองใช้กับนักเรียนชั้นประถมศึกษาปีที่ 6 ที่ไม่ใช่กลุ่มตัวอย่างแต่มีลักษณะใกล้เคียงกับกลุ่มตัวอย่าง จำนวน 119 คน โดยมีวิธีดำเนินการ ดังนี้

1.2.1 การเก็บรวบรวมข้อมูลจากนักเรียนกลุ่มตัวอย่าง ให้นักเรียนทดสอบการพูดภาษาอังกฤษ โดยใช้ภาระงานการนำเสนอปากเปล่า จำนวน 2 เรื่อง ได้แก่ เรื่องราวเกี่ยวกับตนเอง และบุคคลที่ตนชื่นชอบ

1.2.2 ขั้นตอนการทดสอบ ดำเนินการดังนี้

- 1) ให้นักเรียนอ่านคำอธิบายภาระงานทั้ง 2 เรื่อง
- 2) ผู้วิจัยอธิบายภาระงานทั้ง 2 เรื่อง ให้นักเรียนทราบอีกครั้ง
- 3) กำหนดความยาวในการพูดไม่เกิน 3 นาที
- 4) ให้นักเรียนเตรียมตัว และซักถามผู้วิจัยเกี่ยวกับภาระงานที่จะทดสอบเป็นเวลา 30 นาที
- 5) ทำการบันทึกเสียงพูดของนักเรียนด้วยคอมพิวเตอร์

1.2.3 การตรวจให้คะแนนการพูดภาษาอังกฤษของนักเรียน ดำเนินการ ดังนี้

1.2.3.1 ผู้ประเมิน (raters) ผู้ประเมินที่ใช้ในการให้คะแนนการพูดภาษาอังกฤษครั้งนี้ เป็นนิสิตระดับปริญญาเอกสาขาภาษาอังกฤษเป็นภาษานานาชาติ (English as an International Language) จุฬาลงกรณ์มหาวิทยาลัย จำนวน 3 คน

1.2.3.2 การฝึกการให้คะแนนสำหรับผู้ประเมิน ดำเนินการดังนี้

1) ให้ผู้ประเมินศึกษาทำความเข้าใจกับเกณฑ์การให้คะแนนที่เป็นแบบแยกส่วน (analytic rating scale) ขั้นตอนการให้คะแนน การตัดสินใจให้คะแนน และร่วมกันอภิปรายการให้คะแนน

2) ให้ผู้ประเมินฟังเสียงพูดภาษาอังกฤษของนักเรียนที่บันทึกเสียงไว้ และลองให้คะแนนในแต่ละด้านทั้ง 6 ด้าน โดยต่างคนต่างให้คะแนน เมื่อให้คะแนนครบทุกด้าน ผู้ประเมิน อภิปรายผลคะแนนร่วมกันถึงความเหมือนความต่างของการให้คะแนนพร้อมอธิบายเหตุผล และทำการให้คะแนนจนกว่าจะมีการให้คะแนนที่สอดคล้องกัน

3) ให้ผู้ประเมินทำการให้คะแนนเสียงพูดของนักเรียนที่เป็นกลุ่มตัวอย่างที่บันทึกไว้ทั้งหมด จำนวน 119 คน อย่างอิสระ

4) เมื่อทำการตรวจให้คะแนนแล้ว ผู้ประเมินกลับมาทำการอภิปรายผลการให้คะแนนอีกครั้ง พร้อมทั้งให้ความคิดเห็น ข้อเสนอแนะเกี่ยวกับภาระงานที่ใช้ในการทดสอบนักเรียน และเกณฑ์การให้คะแนนที่ผู้วิจัยพัฒนาขึ้น ซึ่งหลังตรวจให้คะแนนแล้วผู้ประเมินได้ให้ข้อคิดเห็น และข้อที่ควรปรับปรุง ดังตารางที่ 3.1

ตารางที่ 3.1 ข้อเสนอแนะและข้อคิดเห็นจากผู้ประเมินจากการทดลองใช้เกณฑ์การให้คะแนนการพูดภาษาอังกฤษ

ข้อเสนอแนะ/ความคิดเห็นผู้ประเมิน	การปรับปรุงแก้ไขของผู้วิจัยในระยะทดลอง
<p>1. จำนวนประโยค หรือความยาวในการพูด ระยะเวลาในการพูด กล่าวคือ</p> <ul style="list-style-type: none"> - นักเรียนพูดเพียงประโยคเดียว หรือพูดน้อย แต่พูดถูกต้องตามโครงสร้างภาษาอังกฤษ กับนักเรียนที่พูดหลายประโยคหรือมีทั้งผิดและถูก จะให้คะแนนอย่างไร 	<ul style="list-style-type: none"> - มีการกำหนดจำนวนประโยคในการพูดแต่ละภาระงาน กล่าวคือ ให้พูดตั้งแต่ 5 ประโยคขึ้นไป
<p>2. เมื่อนักเรียนไม่พูดเลย</p> <ul style="list-style-type: none"> - จะให้คะแนนอย่างไร - ครูควรช่วยหรือไม่อย่างไร แต่หากเป็นการนำเสนอปากเปล่า ครูไม่ควรป้อนคำถามหรือกระตุ้น - ควรให้อิสระในการคิดแก่นักเรียน 	<ul style="list-style-type: none"> - ให้คะแนนในระดับ อ่อนมาก (1 คะแนน) และปรับคำอธิบายเกณฑ์ - ในระยะการทดลอง ผู้วิจัยได้อธิบายกับครูผู้สอนในขั้นตอนการเตรียมครู คือ ครูไม่ควรแทรกแซงระหว่างการพูดของนักเรียน แต่ควรให้ข้อเสนอแนะเมื่อพูดจบ เพื่อการปรับปรุงแก้ไขในการพูดครั้งต่อไป

1.3 การวิเคราะห์ข้อมูล ผู้วิจัยทำการวิเคราะห์ข้อมูลเพื่อเป็นการตรวจสอบคุณภาพของ เกณฑ์การให้คะแนนการพูด และเพื่อศึกษาจำนวนผู้ประเมินที่เหมาะสมในเบื้องต้น การนำไปใช้จริง ในระยะทดลองโดยใช้โปรแกรมสำเร็จรูป คือ mGENOVA (Brennan, 2001) โดยใช้การออกแบบ $p \times t \times r$ Design หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกข้อ และถูกประเมินจากผู้ประเมิน (r) ทุกคน โดยการให้คะแนนภาระงานจากผู้ประเมินคนเดียวกัน (r) ผลการวิเคราะห์ข้อมูลพบว่า เกณฑ์การให้คะแนนการพูดภาษาอังกฤษที่พัฒนาขึ้นมีค่าคะแนนรวม ของการประเมิน (composite score) โดยมีค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ (Generalizability coefficient) หรือความเที่ยงในทฤษฎีการวัดแบบดั้งเดิมอยู่ในระดับสูง คือมีค่า เท่ากับ 0.924 ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ (Phi coefficient) มีค่าอยู่ในระดับสูง เช่นกัน คือมีค่าเท่ากับ 0.885 และเมื่อพิจารณารายด้าน พบว่ามีความเที่ยง (Generalizability coefficient) ตั้งแต่ .700 ขึ้นไปทุกด้าน ยกเว้นด้านการออกเสียง (pronunciation) ที่มีค่าเท่ากับ .649 ดังตารางที่ 3.2

ตารางที่ 3.2 ผลการวิเคราะห์คุณภาพเกณฑ์การให้คะแนนแบบแยกส่วน

Contributions to	Variance and covariance components					
	vocabulary	syntax	cohesion	pronunciation	ideational function	fluency
A priori weight	0.166	0.166	0.166	0.166	0.166	0.166
Universe score variance (%)	17.10	13.71	18.44	11.14	18.79	20.82
Relative error variance (%)	25.57	14.12	17.55	20.08	15.42	7.26
Absolute error variance (%)	22.78	2.21	42.62	13.08	18.63	0.67
Gen Coefficient	0.776	0.716	0.764	0.649	0.872	0.821
Phi Coefficient	0.656	0.617	0.448	0.616	0.790	0.784
Composite						
Relative Error Variance				0.022		
Absolute Error Variance				0.034		
Generalizability coefficient				0.924		
Phi coefficient				0.885		

ตอนที่ 2 การพัฒนาแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ ดังนี้

การพัฒนาแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ ประกอบด้วยขั้นตอน การดำเนินการและองค์ประกอบของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ชั้นประถมศึกษาปีที่ 6 รวมถึงเครื่องมือที่ใช้ในแฟ้มสะสมงานฯ ได้แก่ การประเมินทักษะการพูด ภาษาอังกฤษ ประกอบด้วย 1) แบบประเมินตนเองของนักเรียน 2) แบบประเมินเพื่อน 3) แบบประเมิน ของครู 4) แบบการให้ข้อมูลย้อนกลับแก่นักเรียน (สำหรับครู) 5) แนวทางการเขียนการสะท้อนคิด ต่อผลงานของนักเรียน และ 6) แบบประเมินทักษะการพูดภาษาอังกฤษ (สำหรับครู)

2.1 การสร้างแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ

การพัฒนาแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ โดยเน้นกระบวนการให้ข้อมูลย้อนกลับของครู (teacher feedback) ที่เป็นการให้ข้อมูลย้อนกลับแบบบรรยายทั้งระบุความสำเร็จที่นักเรียนทำได้ดีแล้ว และระบุข้อผิดพลาดและบอกแนวทางแก้ไข (corrective feedback) มีขั้นตอนการดำเนินการ ดังนี้

2.1.1 สัมภาษณ์ผู้ทรงคุณวุฒิเกี่ยวกับรูปแบบแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยมีประเด็นที่สัมภาษณ์ ดังนี้

- 1) องค์ประกอบของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ควรมีองค์ประกอบอะไรบ้าง เพราะเหตุใด
- 2) ขั้นตอนการพัฒนาแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ควรมีขั้นตอนใดบ้าง
- 3) การประเมินระหว่างดำเนินการ (formative assessment) ภาระงานที่ใช้ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ควรมีภาระงานอะไรบ้าง เพราะเหตุใด
- 4) ภาระงานที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษฯ ควรมีจำนวนกี่ภาระงาน เพราะเหตุใด
- 5) เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษฯ แบบแยกส่วน (analytic rating scale) ควรมีองค์ประกอบอะไรบ้าง และแต่ละองค์ประกอบมีการให้คะแนนที่ระดับ
- 6) เครื่องมือที่ใช้ในกระบวนการแฟ้มสะสมงานทักษะการพูดภาษาอังกฤษฯ ควรมีอะไรบ้าง เช่น นักเรียนประเมินตนเอง ประเมินเพื่อน ครูประเมินนักเรียน การสะท้อนคิดต่อผลงานของนักเรียน การให้ข้อมูลย้อนกลับจากครู เป็นต้น และควรมีลักษณะการให้คะแนนแบบใดจึงจะเหมาะสม เช่น แบบตรวจสอบรายการ แบบมาตราประเมินค่า เป็นต้น
- 7) การให้ข้อมูลย้อนกลับของครู (teacher feedback) โดยใช้การให้ข้อมูลย้อนกลับแบบบรรยาย ทั้งระบุความสำเร็จที่นักเรียนทำได้ดีแล้ว และระบุข้อผิดพลาดและบอกแนวทางแก้ไข (correction of errors) มีข้อควรคำนึงถึงอะไรบ้าง และมีแนวทางในการดำเนินการจัดการเรียนการสอนอย่างไร
- 8) การคัดเลือกงานใส่ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษฯ ควรคัดเลือกกี่ชิ้นงานจึงจะเหมาะสม เพราะเหตุใด
- 9) จำนวนผู้ประเมินภาระงานการพูดในระหว่างเรียนควรมีจำนวนกี่คน และเป็นใครบ้าง
- 10) จำนวนผู้ประเมินแฟ้มสะสมงานฯ ควรมีจำนวนกี่คน และเป็นใครบ้าง

2.2 ร่างรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

ซึ่งจากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องผู้วิจัยได้ร่างขั้นตอนการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ดังนี้

ทำการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนใน 2 ระยะ คือ 1) ระยะที่ 1 การประเมินระหว่างเรียน (formative assessment) และ 2) ระยะที่ 2 การประเมินสรุปรวม (summative assessment) มีรายละเอียดการดำเนินการดังนี้

2.2.1 การประเมินระหว่างเรียน (formative assessment) ดำเนินการ ดังนี้

- 1) ประเมินทักษะการพูดของนักเรียน
- 2) ครูให้ข้อมูลย้อนกลับ (teacher feedback) โดยใช้การให้ข้อมูลย้อนกลับแบบบรรยาย ทั้งระบุความสำเร็จที่นักเรียนทำได้ดีแล้ว และระบุข้อผิดพลาดและบอกแนวทางแก้ไข (correction of errors)
- 3) นักเรียนปรับปรุงการพูด และประเมินการพูดอีกครั้ง
- 4) นักเรียนประเมินตนเอง (self- assessment) ครูประเมิน (teacher assessment) และเพื่อนประเมิน (peer review)
- 5) ครูให้ข้อมูลย้อนกลับ (teacher feedback)
- 6) ปรับปรุงตามคำแนะนำของครู และเพื่อน (peer review)
- 7) ประเมินการพูดอีกครั้ง
- 8) เก็บรายงานการพูด และบันทึกเสียงการพูดใส่แฟ้มสะสมงานฯ
- 9) ทำกระบวนการเติมในภาระงานการพูดอื่นๆ
- 10) ทำการประเมินการพูดของนักเรียนในแต่ละภาระงานการพูด โดยนำเสียงที่บันทึกไว้ให้ครูผู้สอนและผู้ทรงคุณวุฒิในกลุ่มสาระการเรียนรู้ภาษาต่างประเทศในโรงเรียนที่ผ่านการฝึกอบรมการเป็นผู้ให้คะแนนการพูด ทำการให้คะแนนการพูด ตามเกณฑ์การให้คะแนนการพูดที่ผู้วิจัยพัฒนาขึ้น และบันทึกในแบบการให้คะแนน

2.2.2 การประเมินสรุปรวม (summative assessment) ดำเนินการดังนี้

- 1) นักเรียนทำการทดสอบการพูดภาษาอังกฤษโดยใช้ภาระงานการพูดที่เป็นแบบบูรณาการรวมทั้ง 3 ภาระงานที่เป็นการประเมินระหว่างเรียน
- 2) ครูหรือผู้ประเมิน ทำการบันทึกเสียง
- 3) ครูหรือผู้ประเมิน ทำการประเมินทักษะการพูดภาษาอังกฤษเพื่อสรุประดับความสามารถในภาพรวม

ส่วนประกอบของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ ควรประกอบด้วย 3 ส่วน ดังนี้

ส่วนที่ 1 เป็นส่วนนำ ประกอบด้วย ปก ประวัติผู้ทำ รายการจุดประสงค์การเรียนรู้ แผนการศึกษาส่วนบุคคล

ส่วนที่ 2 เป็นที่เก็บชิ้นงานที่สร้างขึ้นจากการเรียนการสอน การสะท้อนความคิดเห็น คะแนนจากการสอบ แบบสำรวจรายการของครู บันทึกความคิดเห็นเกี่ยวกับงาน

ส่วนที่ 3 ภาคผนวก เป็นที่เก็บเกณฑ์การให้คะแนน คำอธิบายภาระงาน ซีดีที่บันทึกเสียงพูดของนักเรียน

ทั้งนี้จากการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้องผู้วิจัยได้ร่างเครื่องมือที่ใช้ในแฟ้มสะสมงานฯ ซึ่งประกอบด้วยเครื่องมือต่างๆ ดังนี้

1. แบบประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ดังนี้
 - 1.1 แบบประเมินตนเองของนักเรียน
 - 1.2 แบบประเมินเพื่อน
 - 1.3 แบบประเมินของครู
 2. แนวทางการเขียนการสะท้อนคิดต่อผลงานของนักเรียน
 3. แบบประเมินทักษะการพูดภาษาอังกฤษ (สำหรับครูและผู้ทรงคุณวุฒิ)
 4. แบบประเมินแฟ้มสะสมงาน (สำหรับครูและผู้เชี่ยวชาญ)
 5. เกณฑ์การให้คะแนนแฟ้มสะสมงาน (สำหรับครูและผู้ทรงคุณวุฒิ)
- 2.3 นำต้นแบบแฟ้มสะสมงานฯ ที่สร้างขึ้นเสนอผู้ทรงคุณวุฒิเพื่อตรวจสอบความเหมาะสม

ต้นแบบของแฟ้มสะสมงานประกอบด้วยรายละเอียดขององค์ประกอบและขั้นตอนของแฟ้มสะสมงานฯ ภาระงานการพูดภาษาอังกฤษและเกณฑ์การให้คะแนน ที่ผู้วิจัยสร้างขึ้นตามผลการสัมภาษณ์และการศึกษาเอกสารในตอนที่ 1 และ 2 โดยให้ผู้ทรงคุณวุฒิ จำนวน 9 คน ตรวจสอบความสอดคล้องระหว่างเกณฑ์การให้คะแนนกับวัตถุประสงค์/กิจกรรมการเรียนรู้ (Item Objective Congruence: IOC)

โดยลักษณะของแบบสอบถามซึ่งมีจุดมุ่งหมายเพื่อให้ผู้ทรงคุณวุฒิพิจารณาความเหมาะสมและความเป็นไปได้ของร่างแฟ้มสะสม มีลักษณะเป็นมาตราประมาณค่า 3 ระดับ ประกอบด้วย 2 ส่วน คือ

ส่วนที่ 1 ข้อมูลส่วนตัวของผู้ตอบแบบสอบถาม

ส่วนที่ 2 ความคิดเห็นเกี่ยวกับแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ เพื่อให้พิจารณาความเหมาะสมและความเป็นได้โดยผู้ทรงคุณวุฒิแสดงความคิดเห็นในแบบสอบถาม ดังนี้

1	หมายถึง	มีความเหมาะสม
0	หมายถึง	ไม่แน่ใจ
-1	หมายถึง	ควรปรับปรุง

เกณฑ์การคัดเลือกผู้ทรงคุณวุฒิ ดังนี้

1) ผู้ทรงคุณวุฒิด้านการสอนภาษาอังกฤษ ที่มีประสบการณ์ในการสอนวิชาภาษาอังกฤษ ไม่ต่ำกว่า 10 ปี จบสาขาการสอนภาษาอังกฤษหรือสาขาที่เกี่ยวข้อง หรือเป็นผู้ที่จบการศึกษา ด้านการสอนภาษาอังกฤษในระดับปริญญาโทขึ้นไป และมีประสบการณ์ในการสอนวิชาภาษาอังกฤษ ไม่ต่ำกว่า 5 ปี จำนวน 3 คน

2) ผู้ทรงคุณวุฒิด้านแฟ้มสะสมงาน เป็นผู้ที่มีประสบการณ์ในการใช้แฟ้มสะสมงาน ในการจัดการเรียนการสอน จำนวน 3 คน

3) ผู้ทรงคุณวุฒิด้านการวัดและประเมินผลการศึกษา เป็นผู้ที่มีประสบการณ์ด้านการวัด และประเมินผลการศึกษาอย่างน้อย 5 ปี และจบการศึกษาด้านการวัดและประเมินผลการศึกษาระดับ ปริญญาโทขึ้นไป จำนวน 3 คน

3. ปรับปรุงแก้ไขแฟ้มสะสมงานฯ ตามข้อเสนอแนะของผู้ทรงคุณวุฒิในข้อ 1 และนำร่าง แฟ้มสะสมงานที่ปรับแก้แล้วเสนอต่อผู้ทรงคุณวุฒิอีกครั้ง

กลุ่มตัวอย่างที่ใช้ในระยะที่ 1

1. ผู้ทรงคุณวุฒิด้านการจัดการเรียนการสอนภาษาอังกฤษ การวัดผลทางภาษา นักวิชาการ ด้านการวัดและประเมินผลการศึกษา จำนวน 9 คน

2. นักเรียนชั้นประถมศึกษาปีที่ 6 ที่เป็นกลุ่มตัวอย่างในการทดลองใช้เกณฑ์การให้คะแนน การพูด จำนวน 119 คน

3. ผู้ประเมินที่ทำหน้าที่ประเมินความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียน จำนวน 3 คน

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลในระยะที่ 1

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล 4 ฉบับ ได้แก่ แบบสัมภาษณ์สำหรับผู้ทรงคุณวุฒิ แบบมีโครงสร้าง แบบประเมินความเหมาะสมของรูปแบบแฟ้มสะสมงานการประเมินทักษะ ภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 แบบประเมินความเหมาะสมของเกณฑ์การให้คะแนน

ทักษะการพูดภาษาอังกฤษแบบแยกส่วน (analytic rating scale) แบบประเมินความเหมาะสมของเกณฑ์การให้คะแนนแฟ้มสะสมงานฯ แบบบูรณาการแบบแยกส่วน (analytic rubric)

การวิเคราะห์ข้อมูลในระยะที่ 1

1. การสังเคราะห์เอกสารและงานวิจัยเพื่อสร้างเกณฑ์การให้คะแนนพูดและองค์ประกอบและขั้นตอนของแฟ้มสะสมงาน ใช้การวิเคราะห์เอกสาร
2. ข้อมูลจากการสัมภาษณ์ใช้การวิเคราะห์เนื้อหา
3. การวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร (Multivariate generalizability theory) โดยใช้โปรแกรม mGENOVA version 2.1

ระยะที่ 2 การทดลองใช้รูปแบบการประเมินแฟ้มสะสมงานทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

การดำเนินในระยะทดลองผู้วิจัยมีการดำเนินการ ดังนี้

ตอนที่ 1 การเตรียมความพร้อมของครูผู้สอน ผู้ประเมิน และผู้เกี่ยวข้อง

1. การประชุมชี้แจงการดำเนินการประเมินทักษะการพูดภาษาอังกฤษโดยใช้แฟ้มสะสมงานให้แก่ ครูผู้สอนภาษาอังกฤษชั้นประถมศึกษาปีที่ 6 ครูกลุ่มสาระภาษาต่างประเทศ (ภาษาอังกฤษ) หัวหน้ากลุ่มสาระภาษาต่างประเทศ (ภาษาอังกฤษ) ซึ่งเป็นครูที่ทำการสอนวิชาภาษาอังกฤษ และจะทำหน้าที่เป็นผู้ประเมิน
2. ฝึกอบรมครูผู้สอนภาษาอังกฤษในโรงเรียนเกี่ยวกับการสอนและการประเมินโดยใช้แฟ้มสะสมงานฯ
3. ฝึกอบรมครูในการให้ข้อมูลย้อนกลับ (teacher feedback) แก่ครูผู้สอน ที่เป็นการให้ข้อมูลป้อนกลับแบบบรรยาย ทั้งระบุความสำเร็จที่นักเรียนทำได้ดีแล้ว และระบุข้อผิดพลาดและบอกแนวทางแก้ไข (correction of errors)
4. ฝึกอบรมครูผู้สอนและครูในกลุ่มสาระการเรียนรู้ภาษาอังกฤษในโรงเรียนที่จะทำหน้าที่เป็นผู้ประเมินในเรื่องเกี่ยวกับการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน และทำการให้คะแนนความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนตามเกณฑ์การให้คะแนนที่ผู้วิจัยพัฒนาขึ้น เพื่อให้ผู้ประเมินมีความเข้าใจวิธีการให้คะแนน และให้คะแนนที่มีความสอดคล้องกันโดยดำเนินการดังนี้
5. ตรวจสอบคุณภาพของผลการประเมินด้านความเที่ยงในการประเมินโดยการวิเคราะห์ค่าความสัมพันธ์ระหว่างผู้ประเมิน (inter-rater reliability) โดยการหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson' Product Moment Coefficient of Correlation)

โดยหลังจากเก็บรวบรวมคะแนนจากการจัดการเรียนสอนโดยใช้แฟ้มสะสมงาน การประเมินทักษะการพูดภาษาอังกฤษแล้ว ก่อนที่จะให้ผู้ประเมินที่เป็นครูผู้วิชาภาษาอังกฤษในโรงเรียน จำนวน 4 ท่าน ทำการประเมินทักษะการพูดของนักเรียนทั้ง 4 ภาระงานนั้น ผู้วิจัยได้ทำการฝึก การประเมินทักษะการพูดภาษาอังกฤษของผู้ประเมินโดยทบทวนความเข้าใจภาระงานทั้ง 4 เรื่อง และเกณฑ์การให้คะแนนอีกครั้ง และให้ครูที่เป็นผู้ประเมินทดลองให้คะแนนนักเรียน และอภิปราย ความเหมือนความต่างของการให้คะแนน และทดลองให้อีกครั้งจนกว่าจะสอดคล้องกัน หลังจากนั้น ให้ผู้ประเมินประเมินนักเรียนอย่างเป็นอิสระต่อกัน จาก 4 ภาระงาน จำนวนทั้งสิ้น 10 คน ผลการหา ความสอดคล้องของการประเมินของผู้ประเมินทั้ง 4 คน พบว่า การประเมินโดยใช้เกณฑ์การให้คะแนน แบบแยกส่วน ทั้ง 6 ด้าน มีค่าสัมประสิทธิ์ความสัมพันธ์อยู่ในระดับกลางถึงระดับสูง คือ มีค่าสัมประสิทธิ์ ความสัมพันธ์อยู่ระหว่าง .625-1.000 (Luoma, 2004) ดังตารางที่ 3.3



ตารางที่ 3.3 ค่าสหสัมพันธ์ระหว่างคะแนนของผู้ประเมินที่ใช้เกณฑ์การให้คะแนนแบบแยกส่วน

เกณฑ์การให้คะแนน	ผู้ประเมิน	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4
คำศัพท์	คนที่ 1	1	1.000**	.802**	1.000**
	คนที่ 2	1.000**	1	.802**	1.000**
	คนที่ 3	.802**	.802**	1	.802**
	คนที่ 4	1.000**	1.000**	.802**	1
การสร้างประโยค	คนที่ 1	1	1.000**	1.000**	1.000**
	คนที่ 2	1.000**	1	1.000**	1.000**
	คนที่ 3	1.000**	1.000**	1	1.000**
	คนที่ 4	1.000**	1.000**	1.000**	1
การเชื่อมโยงเนื้อหา	คนที่ 1	1	.820**	.873**	1.000**
	คนที่ 2	.820**	1	.678*	.820**
	คนที่ 3	.873**	.678*	1	.873**
	คนที่ 4	1.000**	.820**	.873**	1
การออกเสียง	คนที่ 1	1	.899**	.778**	1.000**
	คนที่ 2	.899**	1	.899**	.899**
	คนที่ 3	.778**	.899**	1	.778**
	คนที่ 4	1.000**	.899**	.778**	1
บรรลุนิติอุปประสงค์	คนที่ 1	1	1.000**	.830**	.867**
	คนที่ 2	1.000**	1	.830**	.867**
	คนที่ 3	.830**	.830**	1	.745*
	คนที่ 4	.867**	.867**	.745*	1
ความคล่องแคล่ว	คนที่ 1	1	.885**	.625	.815**
	คนที่ 2	.885**	1	.722*	.745*
	คนที่ 3	.625**	.722*	1	.745*
	คนที่ 4	.815**	.745*	.745*	1

** p < .01

ตอนที่ 2 การเตรียมความพร้อมของนักเรียน

1. จัดอบรมเกี่ยวกับการจัดการเรียนการสอนโดยใช้แฟ้มสะสมงานฯ โดยให้ความรู้เกี่ยวกับ วัตถุประสงค์ ขั้นตอนการเรียนรู้ องค์ประกอบของแฟ้มสะสมงานฯ วิธีการเรียน วิธีการประเมินผล

2. ฝึกอบรมนักเรียนเกี่ยวกับการประเมินตนเอง ประเมินเพื่อนร่วมชั้น โดยให้นักเรียนได้เรียนรู้เกี่ยวกับลักษณะของแบบประเมิน วิธีการประเมิน ความหมายของเกณฑ์การให้คะแนน ดำเนินการดังนี้

2.1 ให้นักเรียนจับกลุ่ม กลุ่มละ 3 คน

2.2 ให้นักเรียนแต่ละคนทดลองประเมินความสามารถในการพูดภาษาอังกฤษของตนเองโดยใช้แบบประเมินที่ผู้วิจัยสร้างขึ้น ครูและผู้วิจัยคอยให้คำแนะนำ

2.3 ให้นักเรียนแต่ละคนประเมินเพื่อนในกลุ่ม โดยใช้แบบประเมินที่ผู้วิจัยสร้างขึ้น ครูและผู้วิจัยคอยให้คำแนะนำ

2.4 ให้นักเรียน 2 คนในกลุ่ม นำคะแนนที่ตนประเมินเพื่อนอีกคนมาปรึกษาหารือถึงการให้คะแนนที่ตนให้ และที่เพื่อนให้ว่าเหมือนหรือต่างอย่างไร เพราะเหตุใด โดยมีครูและผู้วิจัยคอยให้คำแนะนำ

2.5 สลับคู่กันนักเรียนในกลุ่มเพื่อประเมินเพื่อนในกลุ่ม โดยดำเนินการตามข้อ 2.4 ทำจนครบทั้ง 3 คู่

3. ฝึกอบรมนักเรียนเกี่ยวกับการสะท้อนคิดต่อผลงานของตน โดยให้นักเรียนใช้แบบบันทึกการสะท้อนคิดที่ผู้วิจัยสร้างขึ้น ครูและผู้วิจัยคอยให้คำแนะนำ

ตอนที่ 3 การดำเนินการเรียนการสอนและการประเมินทักษะการพูดภาษาอังกฤษโดยใช้รูปแบบแฟ้มสะสมงานฯ ที่ผู้วิจัยพัฒนาขึ้น

1. ครูดำเนินการจัดการเรียนการสอนตามกระบวนการแฟ้มสะสมงานฯ และทำการประเมินทักษะการพูดภาษาอังกฤษตามภาระงานที่กำหนดตามเป้าหมายการเรียนรู้ที่ภาระงาน และทำการบันทึกเสียงนักเรียนในการพูดครั้งสุดท้ายของแต่ละภาระงาน

2. ครูผู้สอน และผู้ประเมินทำการประเมินความสามารถในการพูดภาษาอังกฤษของนักเรียนแต่ละคนโดยการฟังจากเสียงที่บันทึกไว้และทำการประเมินโดยต่างคนต่างประเมินอย่างอิสระ และใช้เกณฑ์การให้คะแนนที่ผู้วิจัยพัฒนาขึ้นและทำการบันทึกคะแนนในแบบบันทึกคะแนน

3. นักเรียนเลือกผลงานการพูดของตนเองที่ครูบันทึกไว้ใส่แฟ้มสะสมงานฯ

4. เมื่อสิ้นสุดการเรียน กลุ่มผู้ประเมินทำการประเมินความสามารถในทักษะการพูดของนักเรียนแต่ละคนตามเกณฑ์การให้คะแนนที่ผู้วิจัยพัฒนาขึ้น และทำการบันทึกคะแนนในแบบบันทึกคะแนน

การดำเนินการดังกล่าวมีการดำเนินกิจกรรมในช่วงเวลาต่างๆ ดังตารางที่ 3.4

ตารางที่ 3.4 ช่วงเวลาและการดำเนินกิจกรรม

ระยะที่ดำเนินการ	เวลาที่ดำเนินการ	กิจกรรมที่ดำเนินการ
<p>ระยะที่ 1 การเตรียมความพร้อมของครู</p> <p>1.1 ครูผู้สอนที่จะต้องทำการประเมินโดยใช้แฟ้มสะสมงาน การประเมินทักษะการพูดภาษาอังกฤษ</p> <p>1.2 ครูที่ทำหน้าที่เป็นผู้ประเมินทักษะการพูดภาษาอังกฤษของนักเรียน</p>	19-31 พฤษภาคม 2557	<ul style="list-style-type: none"> - ผู้วิจัยร่วมกับครูผู้สอนกลุ่มสาระการเรียนรู้ภาษาต่างประเทศในโรงเรียน ประชุมชี้แจงการดำเนินการประเมินฯ - ฝึกอบรมครูผู้สอนเกี่ยวกับเกณฑ์การให้คะแนน การรายงาน การให้ข้อมูลย้อนกลับ - ฝึกอบรมครูผู้สอนในกลุ่มสาระฯ เพื่อเป็นผู้ประเมินทักษะการพูดภาษาอังกฤษของนักเรียน
ระยะที่ 2 การเตรียมความพร้อมของนักเรียน	26-31 พฤษภาคม 2557	<ul style="list-style-type: none"> - ผู้วิจัยร่วมกับครูผู้สอนจัดอบรมเกี่ยวกับการจัดการเรียนการสอนโดยใช้แฟ้มสะสมงานฯ - ฝึกอบรมนักเรียนเกี่ยวกับการประเมินตนเอง ประเมินเพื่อนร่วมชั้น - ฝึกอบรมนักเรียนเกี่ยวกับการสะท้อนคิดต่อผลงานของตน
ระยะที่ 3 การดำเนินการสอนและการประเมินทักษะการพูดภาษาอังกฤษโดยใช้รูปแบบแฟ้มสะสมงานฯ	<p>2 มิถุนายน 2557</p> <p>4 มิถุนายน-30 กรกฎาคม 2557</p>	<ul style="list-style-type: none"> - ครูผู้สอนร่วมกับนักเรียนกำหนดวัตถุประสงค์ในการเรียน - ครูผู้สอนดำเนินการสอนและประเมินความสามารถของนักเรียนโดยใช้แฟ้มสะสมงานฯ

ตารางที่ 3.4 (ต่อ)

ระยะที่ดำเนินการ	เวลาที่ดำเนินการ	กิจกรรมที่ดำเนินการ
	6 สิงหาคม 2557	- จัดแสดงผลงานนักเรียน โดยโรงเรียนจัดเวทีในการแสดงความสามารถในการพูดภาษาอังกฤษภายในโรงเรียนและให้นักเรียนและครูในโรงเรียนได้เข้าร่วมชมการแสดง
ระยะที่ 4 การประเมินความสามารถในการการพูดภาษาอังกฤษ	10 กรกฎาคม 2557	- ผู้วิจัยทบทวนความเข้าใจกับคณะผู้ประเมินอีกครั้งเกี่ยวกับเกณฑ์การให้คะแนนภาระงาน - ฝึกการประเมินแก่คณะผู้ประเมิน - หาความเที่ยงโดยการวิเคราะห์ค่าความสัมพันธ์ระหว่างผู้ประเมิน (inter-rater reliability)
	11-30 กรกฎาคม 2557	ผู้ประเมินทำการประเมินความสามารถในการพูดภาษาอังกฤษของนักเรียน

กลุ่มตัวอย่างที่ใช้ในระยะที่ 2

1. ครูผู้สอนวิชาภาษาอังกฤษ ชั้นประถมศึกษาปีที่ 6 จำนวน 1 คน
2. นักเรียนชั้นประถมศึกษาปีที่ 6 ที่เป็นกลุ่มตัวอย่างในการทดลองใช้เกณฑ์การให้คะแนนการพูด จำนวน 34 คน
3. ผู้ประเมินที่ทำหน้าที่ประเมินความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียน จำนวน 4 คน

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลในระยะที่ 2

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลมี 7 ฉบับ ได้แก่ แบบประเมินตนเอง แบบประเมินเพื่อน แบบประเมินของครู แบบบันทึกการสะท้อนคิด แบบบันทึกการให้ข้อมูลย้อนกลับของครู (teacher feedback) แบบบันทึกพฤติกรรมการเรียนของนักเรียน แบบบันทึกการให้คะแนนการพูด (สำหรับครูและผู้ทรงคุณวุฒิ)

การวิเคราะห์ข้อมูลในระยะที่ 2

1. การสังเคราะห์เอกสารและงานวิจัยเพื่อสร้างเกณฑ์การให้คะแนนพูด และองค์ประกอบและขั้นตอนของแฟ้มสะสมงาน ใช้การวิเคราะห์เนื้อหา
2. ข้อมูลจากการสัมภาษณ์ ใช้การวิเคราะห์เนื้อหา

3. การวิเคราะห์ข้อมูลเพื่อทดลองใช้เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ วิเคราะห์ข้อมูลด้วย Multivariate G-theory ด้วยโปรแกรม mGENOVO

ระยะที่ 3 การวิเคราะห์ผลการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

การวิเคราะห์ผลการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับ ประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน แบ่งออกเป็น 3 ส่วน ดังนี้ 1) การศึกษาสัมประสิทธิ์ การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้น ประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือพหุตัวแปร โดยใช้ คะแนนความสามารถด้านทักษะการพูดภาษาอังกฤษที่ได้จากผู้ประเมิน จำนวน 4 คนที่ไม่ใช่ครูผู้สอน ทำการประเมินหลังจากการจัดการเรียนการสอนเสร็จสิ้นแล้ว โดยการประเมินจากเสียงที่บันทึกไว้ใน การประเมินครั้งที่ 3 2) การศึกษาความก้าวหน้าด้านการพูดภาษาอังกฤษของนักเรียน โดยใช้คะแนน ความสามารถด้านทักษะการพูดภาษาอังกฤษที่เก็บรวบรวมข้อมูลโดยครูผู้สอนระหว่างการจัดการเรียน การสอน ที่ทำการประเมินความก้าวหน้าของนักเรียนในแต่ละภาระงาน ภาระงานละ 3 ครั้ง และ 3) ผลการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษทั้ง 7 ขั้นตอน ซึ่งแต่ละส่วนมีรายละเอียด ดังนี้

3.1 การศึกษาสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูด ภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ด้วยทฤษฎีการสรุปอ้างอิง ความน่าเชื่อถือพหุตัวแปร

3.1.1 สัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของเกณฑ์การให้คะแนนและภาระงาน การประเมินทักษะการพูดภาษาอังกฤษ โดยใช้รูปแบบ $p^* \times t^* \times r^*$ และ $p^* \times t^* \times r^*$ Design

1) การออกแบบ $p^* \times t^* \times r^*$ เป็นการวิเคราะห์ค่าสัมประสิทธิ์ความน่าเชื่อถือ และความคลาดเคลื่อนของเกณฑ์การให้คะแนนแบบแยกส่วน (analytic scoring) โดยใช้รูปแบบ ความสัมพันธ์ระหว่างองค์ประกอบเป็น $p^* \times t^* \times r^*$ Design หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกภาระงาน และภาระงานการพูดภาษาอังกฤษทุกภาระงาน มีองค์ประกอบในการประเมินเหมือนกัน คือ ประกอบด้วย 6 เกณฑ์ ได้แก่ คำศัพท์ การสร้างประโยค การออกเสียง การเชื่อมโยงเนื้อหา การบรรลุเป้าหมาย ความคล่องแคล่ว และผู้ประเมิน (r) ทำการประเมินนักเรียนทุกคนและทุกภาระงาน

2) การออกแบบ $p^* \times t^* \times r^*$ เป็นการตรวจสอบคุณภาพของภาระ การประเมินทักษะการพูดภาษาอังกฤษแต่ละภาระงานด้วยการออกแบบ $p^* \times t^* \times r^*$ Design หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกภาระงาน โดยแต่ละภาระงาน

ไม่ซ้ำกัน ผู้ประเมินตรวจให้คะแนนนักเรียนทุกคน และทุกภาระงาน (r) ผลการตรวจสอบความเที่ยงของภาระงานที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ทั้ง 4 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ

3.1.2 การเปรียบเทียบรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือพหุตัวแปร โดยใช้การออกแบบ $p^\bullet \times t^\bullet \times r^\bullet$ Design $p^\bullet \times t^\circ \times r^\bullet$ Design และ $p^\bullet \times t^\circ \times r^\circ$ Design

1) การออกแบบ $p^\bullet \times t^\bullet \times r^\bullet$ Design เพื่อการเปรียบเทียบจำนวนผู้ประเมินและจำนวนภาระงานที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษของการใช้เกณฑ์การให้คะแนนแบบแยกส่วน ซึ่งการออกแบบในรูปแบบนี้ หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกภาระงาน โดยแต่ละภาระงานซ้ำกัน ผู้ประเมินตรวจให้คะแนนนักเรียนทุกคน และทุกภาระงาน (r)

2) การออกแบบ $p^\bullet \times t^\circ \times r^\bullet$ Design เพื่อการเปรียบเทียบจำนวนผู้ประเมินและจำนวนภาระงานที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษของภาระงานการประเมินทักษะการพูดภาษาอังกฤษ จำนวน 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และ 4) การบรรยายภาพ ซึ่งการออกแบบในรูปแบบนี้ หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกภาระงาน โดยแต่ละภาระงานไม่ซ้ำกัน ผู้ประเมินตรวจให้คะแนนนักเรียนทุกคน และทุกภาระงาน (r)

3) การออกแบบ $p^\bullet \times t^\circ \times r^\circ$ Design เพื่อการเปรียบเทียบรูปแบบการประเมินของผู้ประเมินที่ต่างกัน จำนวน 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และ 4) การบรรยายภาพ ซึ่งการออกแบบในรูปแบบนี้ หมายถึง นักเรียนทุกคน (p) ทำภาระงาน (t) ทุกภาระงาน โดยแต่ละภาระงานไม่ซ้ำกัน ผู้ประเมินตรวจให้คะแนนนักเรียนทุกคน แต่ต่างภาระงาน (r)

4) การออกแบบ $p^\bullet \times t^\circ \times r^\bullet$ Design เพื่อการศึกษาจำนวนผู้ประเมินและจำนวนภาระงานทักษะการพูดภาษาอังกฤษโดยการประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) โดยการประเมินในระหว่างเรียนใช้ภาระงานจำนวน 3 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า และ 3) การเล่านิทาน และการประเมินสรุปรวมใช้ภาระงาน จำนวน 1 ภาระงาน คือ 4) การบรรยายภาพ ซึ่งการออกแบบในรูปแบบนี้ หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกภาระงาน โดยแต่ละภาระงานไม่ซ้ำกัน ผู้ประเมินตรวจให้คะแนนนักเรียนทุกคน และทุกภาระงาน (r)

3.2 การศึกษาความก้าวหน้าด้านทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

เป็นการวิเคราะห์คะแนนที่ได้จากครูผู้สอนที่เป็นผู้ดำเนินการสอนโดยใช้รูปแบบการประเมินการพูดโดยใช้แฟ้มสะสมงานที่ผู้วิจัยพัฒนาขึ้น จากนักเรียนกลุ่มทดลอง จำนวน 34 คน โดยในแต่ละภาระงานทำการประเมินจำนวน 3 ครั้ง และทำการวิเคราะห์ความก้าวหน้าด้านทักษะการพูดภาษาอังกฤษของนักเรียนโดยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำชนิดทางเดียว (one-way repeated ANOVA measurement) ทั้งนี้เกณฑ์การให้คะแนนเพื่อกำหนดระดับคะแนนความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ข้างต้น พิจารณาจากคะแนนเฉลี่ยของคะแนนการพูดในแต่ละภาระงานและแปลความหมาย ดังนี้

คะแนนเฉลี่ย ต่ำกว่า 1.51	หมายถึง อ่อนมาก
คะแนนเฉลี่ย 1.51-2.50	หมายถึง อ่อน
คะแนนเฉลี่ย 2.51-3.50	หมายถึง พอใช้
คะแนนเฉลี่ย 3.51-4.00	หมายถึง ดี

3.3 การศึกษาผลการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

เป็นการวิเคราะห์ผลการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ทั้ง 7 ขั้นตอน ได้แก่ ขั้นวางแผน ขั้นเตรียมความพร้อมนักเรียน ขั้นการเก็บรวบรวมหลักฐาน ขั้นการติดตามความก้าวหน้า ขั้นการปรับปรุงผลงาน ขั้นการสะท้อนคิด และขั้นการจัดแสดงผลงาน เก็บรวบรวมข้อมูลการสังเกต การสัมภาษณ์นักเรียน และครูผู้สอน และวิเคราะห์ข้อมูลโดยการวิเคราะห์เนื้อหา

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลในระยะที่ 3

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล 2 ฉบับ ได้แก่ เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 และแบบประเมินคุณภาพและความเหมาะสมของเกณฑ์การให้คะแนนแฟ้มสะสมงานประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

การวิเคราะห์ข้อมูลในระยะที่ 3

1. การวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร โดยใช้โปรแกรม mGENOVA version 2.1
2. วิเคราะห์ความก้าวหน้าด้านทักษะการพูดภาษาอังกฤษของนักเรียนโดยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำชนิดทางเดียว (one-way repeated ANOVA measurement)

ระยะที่ 4 การพัฒนาคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ชั้นประถมศึกษาปีที่ 6

4.1 ผู้วิจัยร่างคู่มือการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน โดยนำข้อมูลวิจัยที่ได้จากระยะที่ 3 มาใช้ในการร่างคู่มือ เกี่ยวกับประเด็นองค์ประกอบของแฟ้มสะสมงาน เกณฑ์การให้คะแนนระหว่างเรียนและสรุปรวม จำนวนผู้ประเมิน และจำนวนภาระงาน

รายละเอียดของคู่มือประกอบด้วย 3 ส่วน ดังนี้

ส่วนที่ 1 ส่วนหน้า ประกอบด้วยรายละเอียดต่างๆ ดังนี้

1. คำชี้แจง
2. วัตถุประสงค์ของคู่มือ
3. บทนำ มีเนื้อหาเกี่ยวกับความรู้ในการประเมินทักษะการพูดภาษาอังกฤษ

ส่วนที่ 2 วิธีการพัฒนาแฟ้มสะสมงานฯ โดยในส่วนนี้จะได้มาจากผลการวิจัยในระยะที่ 1

1. ส่วนประกอบของแฟ้มสะสมงานฯ
2. ขั้นตอนของแฟ้มสะสมงานฯ

ส่วนที่ 3 วิธีการดำเนินการประเมินทักษะการพูดภาษาอังกฤษ ที่ได้จากการประเมิน
ในระยะที่ 3

1. การประเมินระหว่างเรียน ประกอบด้วยเนื้อหาต่างๆ ดังนี้

1.1 ภาระงานการพูด และเกณฑ์การให้คะแนนแบบแยกส่วนที่ใช้ในการประเมินนักเรียน ที่ครูสามารถเลือกรูปแบบที่มีระดับความเที่ยงที่เหมาะสม และสามารถปฏิบัติได้

1.2 จำนวนผู้ประเมินที่ใช้ในการประเมินภาระงานการพูด ที่ครูสามารถเลือกรูปแบบที่มีระดับความเที่ยงที่เหมาะสม และสามารถปฏิบัติได้

1.3 การประเมินตนเองของนักเรียน พร้อมแบบประเมิน

1.4 การประเมินเพื่อนของนักเรียน พร้อมแบบประเมิน

1.5 การประเมินนักเรียนจากครูผู้สอน พร้อมแบบประเมิน

2. การประเมินสรุปรวม เป็นการประเมินแฟ้มสะสมงานฯ ของนักเรียนเพื่อสรุประดับความสามารถของทักษะการพูดภาษาอังกฤษของนักเรียน ประกอบด้วยเนื้อหาต่างๆ ดังนี้

2.1 ภาระงานการพูด และจำนวนภาระงานการพูดที่เก็บใส่แฟ้มสะสมงานที่มีความเหมาะสมในการประเมินแฟ้มสะสมงานฯ ทำให้การประเมินมีระดับความเที่ยงที่เหมาะสม และสามารถปฏิบัติได้

2.2 จำนวนผู้ประเมินที่ใช้ในการประเมินแฟ้มสะสมงานฯ และรูปแบบการให้คะแนนที่มีระดับความเที่ยงที่เหมาะสม และสามารถปฏิบัติได้

2.3 เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษแบบแยกส่วนที่ใช้ในการประเมินแฟ้มสะสมงานฯ ของนักเรียน ที่มีระดับความเที่ยงที่เหมาะสม

4.2 นำคู่มือการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานที่สร้างขึ้นให้อาจารย์ที่ปรึกษาตรวจสอบความเหมาะสม และปรับแก้ตามข้อเสนอแนะ

4.3 นำคู่มือการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานให้ผู้ทรงคุณวุฒิ จำนวน 9 ท่าน ตรวจสอบความเหมาะสม และปรับแก้ตามข้อเสนอแนะ โดยใช้เกณฑ์ในการคัดเลือกผู้ทรงคุณวุฒิ ดังนี้

1) ผู้ทรงคุณวุฒิด้านการสอนภาษาอังกฤษ ที่มีประสบการณ์ในการสอนวิชาภาษาอังกฤษไม่ต่ำกว่า 10 ปี จบสาขาการสอนภาษาอังกฤษหรือสาขาที่เกี่ยวข้อง หรือเป็นผู้ที่จบการศึกษาด้านการสอนภาษาอังกฤษในระดับปริญญาโทขึ้นไป และมีประสบการณ์ในการสอนวิชาภาษาอังกฤษไม่ต่ำกว่า 5 ปีจำนวน 3 คน

2) ผู้ทรงคุณวุฒิด้านแฟ้มสะสมงาน เป็นผู้มีประสบการณ์ในการใช้แฟ้มสะสมงานในการจัดการเรียนการสอน จำนวน 3 คน

3) ผู้ทรงคุณวุฒิด้านการวัดและประเมินผลการศึกษา เป็นผู้มีประสบการณ์ด้านการวัดและประเมินผลการศึกษาอย่างน้อย 5 ปี และจบการศึกษาด้านการวัดและประเมินผลการศึกษาระดับปริญญาโทขึ้นไป จำนวน 3 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัยในระยะที่ 4

กลุ่มตัวอย่างที่ใช้ในการเก็บรวบรวมข้อมูลในระยะที่ 4 คือ ผู้ทรงคุณวุฒิด้านการสอนภาษาอังกฤษ การวัดประเมินผลทางภาษา และการวัดประเมินผลการศึกษา จำนวน 9 ท่าน

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลในระยะที่ 4

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล 1 ฉบับ ได้แก่ แบบประเมินคุณภาพและความเหมาะสมของคู่มือการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

การวิเคราะห์ข้อมูลในระยะที่ 4

วิเคราะห์ข้อมูลจากแบบประเมินจากผู้ทรงคุณวุฒิเกี่ยวกับความเหมาะสมของคู่มือการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ด้วยความตรงตามเนื้อหาโดยการวิเคราะห์ค่า IOC

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้เป็นการวิจัยเพื่อการพัฒนาแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน: การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร โดยมีวัตถุประสงค์เพื่อการพัฒนาแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ทดลองใช้รูปแบบการประเมินที่พัฒนาขึ้น และพัฒนาคู่มือการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน โดยวิธีการวิจัยมีการเก็บรวบรวมข้อมูลจากหลายแหล่งและหลายวิธีเพื่อความถูกต้องของข้อมูล การวิจัยครั้งนี้ผู้วิจัยได้เก็บรวบรวมข้อมูลทั้งข้อมูลเชิงปริมาณ (quantitative) และเชิงคุณภาพ (qualitative) ซึ่งได้จากการสัมภาษณ์ผู้ทรงคุณวุฒิ การศึกษาแนวคิดทฤษฎีและเอกสารที่เกี่ยวข้อง การใช้แบบสัมภาษณ์ และการใช้แบบประเมิน การนำเสนอผลการวิจัยนั้นแบ่งตามขั้นตอนการดำเนินการวิจัย โดยแบ่งออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 ผลการพัฒนาแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

ตอนที่ 2 ผลการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานสำหรับประเมินทักษะการพูดภาษาอังกฤษ

ตอนที่ 3 ผลการเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

ผู้วิจัยได้กำหนดสัญลักษณ์และตัวอักษรย่อภาษาอังกฤษที่ใช้สื่อความหมายแทนตัวและค่าสถิติ ดังนี้

สัญลักษณ์แทนค่าสถิติ

Gen coeff.	หมายถึง ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์
Phi coeff.	หมายถึง ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์
Rel. error	หมายถึง ความคลาดเคลื่อนเชิงสัมพัทธ์
Abs. error	หมายถึง ความคลาดเคลื่อนเชิงสัมบูรณ์

สัญลักษณ์แทนตัวแปร

voc.	หมายถึง คำศัพท์
syn.	หมายถึง การสร้างประโยค
coh.	หมายถึง การเชื่อมโยงเนื้อหา
pro.	หมายถึง การออกเสียง
fun.	หมายถึง การบรรลุวัตถุประสงค์
flu.	หมายถึง ความคล่องแคล่ว
interview	หมายถึง การสัมภาษณ์
oral	หมายถึง การนำเสนอปากเปล่า
story	หมายถึง การเล่านิทาน
picture	หมายถึง การบรรยายภาพ

ตอนที่ 1 ผลการพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

1.1 ขั้นตอนในการดำเนินการและองค์ประกอบของแฟ้มสะสมงานการประเมินภาษาอังกฤษสำหรับนักเรียนชั้นประถมศึกษาปีที่ 6

ผลการพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ในการศึกษาครั้งนี้ผู้วิจัยได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องรวมทั้งทำการสัมภาษณ์ผู้ทรงคุณวุฒิทั้งด้านภาษาอังกฤษ และด้านการวัดและประเมินผลการศึกษา แล้วนำมาสังเคราะห์ พบว่า ส่วนประกอบของแฟ้มสะสมงานฯ ประกอบด้วย 3 ส่วน ได้แก่ 1) ส่วนนำ เป็นส่วนที่เก็บข้อมูลส่วนตัวของนักเรียน ดังนี้ ปก คำนำ สารบัญ ข้อมูลส่วนตัวนักเรียน และเป้าหมายการเรียนรู้ ความคาดหวังในการเรียน 2) ส่วนเนื้อหา ประกอบด้วยข้อมูล ดังนี้ ผลการประเมินของนักเรียนข้อมูลย้อนกลับจากครู และผลการสะท้อนคิดต่อผลงานตนเอง และ 3) ส่วนภาคผนวก ประกอบด้วยข้อมูล ดังนี้ คำอธิบายภาระงาน และเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ ดังตารางที่ 4.1 ส่วนขั้นตอนของการพัฒนาแฟ้มสะสมงานฯ ปรากฏผลการพัฒนาจากผลการสังเคราะห์เอกสารงานวิจัยที่เกี่ยวข้องและการสัมภาษณ์ พบว่ามีขั้นตอนในการดำเนินการแฟ้มสะสมงานการประเมินภาษาอังกฤษสำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 7 ขั้นตอน ได้แก่ 1) ขั้นการวางแผน 2) ขั้นการเตรียมความพร้อมนักเรียน 3) ขั้นการเก็บรวบรวมหลักฐาน 4) ขั้นการติดตามความก้าวหน้า 5) ขั้นการปรับปรุงผลงาน 6) ขั้นการสะท้อนคิด และ 7) ขั้นการจัดแสดงผลงาน ซึ่งแต่ละส่วนมีรายละเอียด ดังตารางที่ 4.2

ตารางที่ 4.1 การสังเคราะห์ส่วนประกอบของแฟ้มสะสมงานทางภาษาอังกฤษ

ส่วนประกอบ	ผู้วิจัย/ผู้ทรงคุณวุฒิ											รวม
	1	2	3	4	5	6	7	8	9	10	11	
1. ส่วนนำ												
1.1 ปก	✓	✓				✓		✓	✓	✓	✓	7
1.2 ประวัติ/ข้อมูลส่วนตัว	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	10
1.3 จุดประสงค์/เป้าหมาย	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	10
2. ส่วนเนื้อหา												
2.1 คะแนนจากการประเมิน	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11
2.2 ข้อมูลย้อนกลับจากครู								✓	✓	✓	✓	4
2.3 ข้อมูลการสะท้อนคิดต่อ ผลงานตนเอง	✓		✓	✓	✓	✓		✓	✓	✓	✓	9
3. ภาคผนวก												
3.1 คำอธิบายภาระงาน		✓						✓	✓	✓	✓	5
3.2 เกณฑ์การให้คะแนน		✓	✓	✓		✓		✓	✓	✓	✓	8
3.3 ผลงาน/หลักฐาน	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	11

หมายเหตุ: 1. Burke, fogart and Begrad (1994) 2. EPL (2004) 3. Cole et.al. (2000) 4. ชัยพฤกษ์ เสรีรักษ์ และคณะ (ม.ป.ป.) 5. อุทุมพร จามรมาน (2540) 6. กรมวิชาการ (2542) 7. ชัยฤทธิ์ ศิลาเดช (2540) 8. วิษณุ ทรัพย์สมบัติ (2556) 9. ไพรวลัย พิทักษ์สาธิต (2556) 10. ยุวดี อยู่สบาย (2556) 11. ภาวิณี ศรีสุขวัฒนานันท์ (2556)

ตารางที่ 4.2 การสังเคราะห์ขั้นตอนของแฟ้มสะสมงานทางภาษาอังกฤษ

ขั้นตอน	ผู้วิจัย/ผู้ทรงคุณวุฒิ														รวม
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1. ให้ความสำคัญในการตระหนัก (raising awareness)	✓												✓	✓	3
2. การเกาะติดกับเป้าหมาย (goal tracking)	✓										✓	✓	✓	✓	3
3. การสร้างทางเลือก (making choices)	✓														1
4. การสะท้อนคิด (reflection)	✓	✓	✓		✓					✓	✓	✓	✓	✓	9
5. การประเมินตนเอง (self-assessment)	✓		✓		✓	✓				✓	✓	✓	✓	✓	9
6. วางแผน และการกำหนด วัตถุประสงค์ (purpose)		✓	✓			✓		✓	✓	✓	✓		✓	✓	9
7. การเก็บรวบรวมและจัดระบบ ผลงาน (collection and organize)		✓	✓	✓				✓		✓	✓	✓	✓	✓	9
8. คัดเลือกผลงาน (select)			✓	✓						✓	✓		✓		5
9. สร้างสรรค์ผลงานให้มี ลักษณะเฉพาะ (interject personality)			✓												1
10. ทำให้สมบูรณ์และประเมินค่า (perfect and evaluate)			✓												1
11. การสร้างความสัมพันธ์ (connect conference)			✓												1
12. ปรับเปลี่ยนผลงานให้เป็นปัจจุบัน (inject/eject to update)			✓												1
13. การจัดแสดงผลงาน (respect accomplishment)			✓							✓	✓	✓	✓		5
14. การประเมินระหว่างเรียน (formative assessment)					✓					✓	✓	✓	✓		5
15. การประเมินสรุปรวม (summative assessment)					✓					✓	✓	✓	✓		5
16. การให้ข้อมูลย้อนกลับ (feedback) เพื่อปรับปรุงผลงาน					✓					✓	✓	✓	✓		5
17. การประเมินผลงานเพื่อน					✓	✓				✓	✓	✓	✓		6
18. การเตรียมความพร้อมนักเรียน							✓		✓	✓	✓	✓			5
19. การกำกับติดตามความก้าวหน้า							✓		✓	✓	✓	✓			6
20. การประเมินแฟ้มสะสมงาน							✓		✓				✓		3

หมายเหตุ: 1. EPL (2004) 2. Barton and Collins (1997) 3. Burke, fogart and Begrad (1994) 4. Hamp-Lyons and Condon (2000) 5. Lam and Lee (2010) 6. Barabouti (2012) 7. Lo (2010) 8. Tasdemir, Tasdemid and Yildirim (2008) 9. Kouzouli (2012) 10. กนกพงศ์ จิตต์ปลื้ม (2549) 11. วิษณุ ทรัพย์สมบัติ (2556) 12. ไพรวลัย พิทักษ์สาถ์ (2556) 13. ยูวดี อยู่สบาย (2556) 14. ภาวิณี ศรีสุขวัฒนานันท์ (2556)

ขั้นตอนของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วย 7 ขั้นตอน ดังนี้

1) ขั้นตอนการวางแผน

ขั้นตอนการวางแผนมีการดำเนินการ ดังนี้

1.1 ครูผู้สอน และครูกลุ่มสาระการเรียนรู้ภาษาต่างประเทศร่วมกันศึกษาวิเคราะห์หลักสูตร กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ (ภาษาอังกฤษ) ตัวชี้วัดในระดับชั้นประถมศึกษาปีที่ 6 การวัดและประเมินผลทางภาษา การประเมินทักษะการพูด

1.2 ครูผู้สอน ครูกลุ่มสาระการเรียนรู้ภาษาต่างประเทศ และนักเรียนร่วมกันกำหนดวัตถุประสงค์ในการเรียนรู้ด้วยแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ

2) ขั้นตอนเตรียมความพร้อมนักเรียน

ขั้นตอนการเตรียมความพร้อมนักเรียนมีการดำเนินการ ดังนี้

2.1 ผู้วิจัยร่วมกับครูผู้สอนให้ความรู้เกี่ยวกับการประเมินทักษะการพูดภาษาอังกฤษ และเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษแก่นักเรียน

2.2 ผู้วิจัยร่วมกับครูผู้สอนฝึกทักษะการประเมินตนเอง และประเมินเพื่อน

3) ขั้นตอนการเก็บรวบรวมหลักฐาน

ขั้นตอนการเก็บรวบรวมหลักฐานมีการดำเนินการ ดังนี้

3.1 ให้นักเรียนวิเคราะห์ความสามารถในการพูดภาษาอังกฤษของตนเองและกำหนดเป้าหมายในการเรียนของตนเอง

3.2 การประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ดังนี้

3.2.1 นักเรียนประเมินตนเอง โดยนักเรียนประเมินตนเองทุกครั้งที่มีการประเมินการพูดภาษาอังกฤษ

3.2.2 นักเรียนได้รับการประเมินจากเพื่อนในชั้นเรียน จำนวน 3 คน ในทุกครั้งที่มีการประเมินการพูดภาษาอังกฤษ

3.2.2 นักเรียนได้รับการประเมินจากครูในทุกครั้งที่มีการประเมินพูด

3.3 นักเรียนทำการประเมินทักษะการพูดภาษาอังกฤษของเพื่อนในชั้นเรียนจำนวน 3 คน ในแต่ละภาระงาน

4) ขั้นตอนการติดตามความก้าวหน้า

ขั้นตอนการติดตามความก้าวหน้ามีการดำเนินการ ดังนี้

4.1 ครูผู้สอนทำการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนทั้งในระหว่างเรียน จากภาระงานการพูด จำนวน 3 ภาระงาน และการประเมินสรุปรวม จำนวน 1 ภาระงาน รวมทั้งสิ้น 4 ภาระงาน แต่ละภาระงานทำการประเมิน 3 ครั้ง

4.2 ครูให้ข้อมูลย้อนกลับ (feedback) แก่นักเรียน โดยให้ข้อมูลแบบระบุความสำเร็จ และระบุข้อผิดพลาดเพื่อการแก้ไข

5) ขั้นตอนการปรับปรุงผลงาน

ขั้นตอนการปรับปรุงผลงานมีการดำเนินการ ดังนี้

5.1 นักเรียนวางแผนการปรับปรุงการพูดของตนเองตามคำแนะนำของครู (teacher feedback)

5.2 นักเรียนแก้ไขการพูดภาษาอังกฤษของตนเองอีก 2 ครั้ง รวมประเมินการพูดของนักเรียนทั้งสิ้น 3 ครั้ง

6) ขั้นตอนการสะท้อนคิด

ขั้นตอนการสะท้อนคิดมีการดำเนินการ ดังนี้

6.1 นักเรียนสะท้อนคิดต่อผลงานของตนเอง และประสบการณ์ต่างๆ ที่ได้รับการจัดการเรียนรู้โดยใช้แฟ้มสะสมงานฯ

6.2 นักเรียนระบุจุดเด่น และจุดที่ต้องพัฒนาของตนเอง

7) ขั้นตอนการจัดแสดงผลงาน

ขั้นตอนการจัดแสดงผลงานมีการดำเนินการ ดังนี้

7.1 จัดแสดงแฟ้มสะสมงานฯ ต่อนักเรียน และครู

7.2 จัดให้นักเรียนได้แสดงความสามารถในการพูดภาษาอังกฤษบนเวทีของโรงเรียน โดยมีครู และนักเรียนชั้นอื่นๆ เข้าร่วม

การตรวจสอบความเหมาะสมของแฟ้มสะสมผลงานการประเมินทักษะการพูดภาษาอังกฤษฯ พบว่า องค์ประกอบของแฟ้มสะสมงานทั้ง 3 ส่วน ได้แก่ ส่วนหน้า ส่วนเนื้อหา และภาคผนวก และขั้นตอนของแฟ้มสะสมงานทั้ง 7 ขั้นตอน ได้แก่ ขั้นการวางแผน ขั้นการเตรียมความพร้อมนักเรียน ขั้นการเก็บรวบรวมหลักฐาน ขั้นการปรับปรุงผลงาน ขั้นการสะท้อนคิด และขั้นการจัดแสดงผลงาน ตรวจสอบด้วยการหาค่า IOC พบว่า มีค่า IOC เท่ากับ 1 ทุกข้อ ยกเว้น ขั้นตอนของแฟ้มสะสมงานขั้นการติดตามความก้าวหน้า ในส่วนของการให้ข้อมูลย้อนกลับ พบว่า มีค่า IOC เท่ากับ .788 (ดังปรากฏในตารางที่ 4.15)

1.2 เกณฑ์การให้คะแนน

ผลการพัฒนาเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ในการศึกษาค้นคว้าวิจัยได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องรวมทั้งทำการสัมภาษณ์ผู้ทรงคุณวุฒิทั้งด้านภาษาอังกฤษ และด้านการวัดและประเมินผลการศึกษา แล้วนำมาสังเคราะห์ ดังตารางที่ 4.3 ปรากฏผลการพัฒนาจากผลการสังเคราะห์เอกสารงานวิจัยที่เกี่ยวข้องและการสัมภาษณ์ พบว่า เกณฑ์การให้คะแนนภาระงานที่ใช้การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่ใช้ในการประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) และใช้ภาระงานการพูดในการประเมินจำนวนทั้งสิ้น 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และ 4) การบรรยายภาพ แต่ละภาระงานมีเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) และมีขอบเขตเนื้อหาในการประเมิน 6 ด้าน และให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี แต่ละองค์ประกอบมีรายละเอียด ดังนี้

1.2.1 คำศัพท์ (vocabulary) คือ การให้คะแนนด้านคำศัพท์ จะพิจารณาจากการพูดของนักเรียนที่สะท้อน 2 สิ่ง คือ ขอบเขต (range) ของคำศัพท์ที่ใช้ และ ความถูกต้อง (accuracy) กล่าวคือ ขอบเขต (range) ของการใช้คำศัพท์ที่มีความหลากหลาย ความสามารถในการใช้คำศัพท์ได้อย่างถูกต้อง (accuracy) และเหมาะสม (appropriation) ตรงตามความหมายที่ต้องการสื่อสาร

1.2.2 การสร้างประโยค (syntax) คือ การให้คะแนนด้านการสร้างประโยค พิจารณาจากความสามารถของนักเรียนในสร้างประโยคได้อย่างมีความถูกต้อง (accuracy) ตามที่ต้องการสื่อสาร เช่น มีประธาน กริยา กรรมที่เหมาะสม เป็นต้น

1.2.3 การเชื่อมโยงเนื้อหา (cohesion) การให้คะแนนด้านการเชื่อมโยงเนื้อหา พิจารณาจากการที่นักเรียนสามารถพูดเชื่อมโยงประโยคแต่ละประโยคให้เป็นเรื่องราวที่มีความสัมพันธ์กันได้อย่างถูกต้องเหมาะสม

1.2.4 การออกเสียง (pronunciation) การให้คะแนนด้านการออกเสียง พิจารณาจากการที่นักเรียนสามารถเน้นเสียงหนักเบา (stress) เว้นวรรคการหยุด (segmental sound) และการใช้เสียงสูงต่ำ (intonation) ได้อย่างถูกต้อง เหมาะสม และทำให้ผู้ฟังเข้าใจในสิ่งที่ต้องการสื่อสารได้

1.2.5 บรรลุวัตถุประสงค์ (ideational function) คือ การให้คะแนนด้านบรรลุวัตถุประสงค์ พิจารณาจากการที่นักเรียนสามารถพูดสื่อสารได้บรรลุตามวัตถุประสงค์ที่ต้องการจะสื่อสารโดยการพูดนั้น ทำให้ผู้ฟังเข้าใจว่าต้องการสื่อสารอะไร เช่น พูดได้ตรงกับหัวเรื่องที่ระบุ ตอบตรงคำถาม ตอบคำถามได้อย่างเหมาะสมกับลักษณะคำถาม เป็นต้น

1.2.6 ความคล่องแคล่ว (fluency) คือ การใช้ภาษาที่มีความคล่องแคล่ว ลื่นไหล ซึ่งพิจารณาจากการที่นักเรียนสามารถพูดภาษาอังกฤษได้อย่างคล่องแคล่ว ราบรื่น และมีความลื่นไหล ในระหว่างที่พูด และไม่สะดุดหรือกระตุก หยุดชะงัก ระหว่างที่พูด

ตารางที่ 4.3 การสังเคราะห์เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ

ผู้วิจัย/ผู้ทรงคุณวุฒิ	เกณฑ์การให้คะแนน (analytic rating scale)					
	Voc.	Syn.	Coh.	Pro.	Fun.	Flu.
1. Bachman และ Palmer (2010)	✓	✓	✓	✓	✓	✓
2. Sawaki (2007)	✓	✓	✓	✓		
3. Westerveld and Moran (2010)		✓				✓
4. Sato (2011)	✓	✓		✓		✓
5. CEFR (2001)			✓			✓
6. ยুক্তี อยู่สบาย (2556)	✓	✓	✓	✓	✓	✓
7. สุทธิรักษ์ ทรัพย์สิรินทร์ (2556)	✓	✓	✓	✓	✓	✓
8. ธัญญาพร อารียา (2556)	✓	✓	✓	✓	✓	✓
9. รติณธร วาดเขียน (2556)	✓	✓	✓	✓	✓	✓
11. สุนณฑา วงษ์สวัสดิ์ (2556)	✓	✓	✓	✓	✓	✓
รวม	8	9	8	8	6	9

สำหรับขั้นตอนในการพัฒนาเกณฑ์การให้คะแนน ผู้วิจัยได้เสนอภาระงานที่จะใช้ในการประเมิน และร่างเกณฑ์การให้คะแนนจากเอกสารทางการประเมินภาษา งานวิจัยเกี่ยวกับการประเมินทางภาษาอังกฤษ และการสัมภาษณ์ผู้ทรงคุณวุฒิทางการจัดการเรียนการสอนภาษาอังกฤษ และการวัดผลทางภาษาอังกฤษแล้วนำเกณฑ์การให้คะแนนที่ได้เสนอต่อผู้ทรงคุณวุฒิเพื่อตรวจสอบความสอดคล้องและเหมาะสมของเกณฑ์การให้คะแนนการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ซึ่งพบว่าภาระงานและเกณฑ์การให้คะแนนทุกด้านมีความเหมาะสม มีค่า IOC เท่ากับ 1 ทุกด้าน นั้นแสดงว่าผ่านเกณฑ์การให้คะแนนมีความสอดคล้องมีความเหมาะสมทุกด้าน และพบข้อเสนอแนะเพิ่มเติมของผู้ทรงคุณวุฒิ ซึ่งมีรายละเอียดดังตารางที่ 4.4-4.5

ภาระงานที่ใช้ในการประเมินการพูดของนักเรียนในการวิจัยครั้งนี้ ประกอบด้วย 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ (interview) 2) การนำเสนอปากเปล่า (oral presentation) 3) การเล่านิทาน (storytelling) และ 4) การบรรยายภาพ (picture description) ซึ่งผู้ทรงคุณวุฒิได้ตรวจสอบความเหมาะสม และให้ข้อคิดเห็น และข้อเสนอแนะ ดังตารางที่ 4.4

ตารางที่ 4.4 ผลการตรวจสอบความสอดคล้องของนิยามภาระงานการพูดภาษาอังกฤษ และความคิดเห็น
ข้อเสนอแนะของผู้ทรงคุณวุฒิเกี่ยวกับภาระงานการพูดภาษาอังกฤษ

รายการที่ประเมิน	IOC	ความคิดเห็น/ข้อเสนอแนะ
1. การสัมภาษณ์ (interview)	1	- จะถามกี่คำถาม - ควรถามเท่ากันทุกคน
2. การนำเสนอปากเปล่า (oral presentation)	1	- ให้นักเรียนเตรียมตัวใหม่ - จะให้นักเรียนเตรียมตัวนานเท่าไร - ควรมีการกำหนดจำนวนประโยคที่พูด เช่น 5-10 ประโยค ขึ้นไป - ควรกำหนดเวลาในการพูด เช่น 1-2 นาที - มีการดู eye contact ด้วย เพราะนักเรียนพูดกับคนไม่ใช่เครื่องคอมพิวเตอร์
3. การเล่านิทาน (storytelling)	1	- เนื้อเรื่องที่นำมาเล่า ถ้านักเรียนเลือกเองจะหลากหลายยากต่อการควบคุม - ถ้านักเรียนเลือกเรื่องเองหลากหลาย จะเป็นปัญหาในการให้คะแนนของผู้ประเมิน - จะดูท่าทางของนักเรียนด้วยไหม
4. การบรรยายภาพ (picture description)	1	- เนื้อหาของภาพที่นำมา ถ้านักเรียนเลือกเองจะหลากหลายยากต่อการควบคุม - ถ้านักเรียนเลือกมาเองหลากหลาย จะเป็นปัญหาในการให้คะแนนของผู้ประเมิน เพราะภาพต่างกัน รายละเอียดมากน้อยต่างกัน - คนที่ได้ภาพที่ไม่มีรายละเอียดแต่พยายามพูดพยายามบรรยาย กับคนที่มีรายละเอียดของภาพมากแต่ไม่พูด ครูจะอย่างไร

สำหรับเกณฑ์การให้คะแนนการพูดภาษาอังกฤษที่พัฒนาขึ้นสำหรับการให้คะแนนนักเรียนชั้นประถมศึกษาปีที่ 6 นี้ ใช้เกณฑ์การให้คะแนนแบบแยกส่วนประกอบด้วย 6 ด้าน ได้แก่ 1. คำศัพท์ (vocabulary) 2. การสร้างประโยค (syntax) 3. การเชื่อมโยงเนื้อหา (cohesion) 4. การออกเสียง (pronunciation) 5. บรรลุวัตถุประสงค์ (ideational function) และ 6. ความคล่องแคล่ว (fluency) แต่ละด้านให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี ผลการพัฒนาเกณฑ์การให้คะแนน ปรากฏดังตารางที่ 4.5 และผลการตรวจสอบความสอดคล้องของนิยามและความเหมาะสม

กับวัตถุประสงค์การประเมินการพูดของนักเรียนชั้นประถมศึกษาปีที่ 6 พบว่าผู้ทรงคุณวุฒิได้ตรวจสอบความเหมาะสมและให้ข้อคิดเห็น และข้อเสนอแนะ ดังตารางที่ 4.6

ตารางที่ 4.5 เกณฑ์การให้คะแนนการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

เกณฑ์	ระดับ	คำอธิบาย
1. คำศัพท์ (vocabulary)	อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> • ความสามารถในการใช้คำศัพท์ยังจำกัดอยู่มาก • มักใช้คำศัพท์คำเดิมซ้ำๆ ทุกประโยค หรือเกือบทุกประโยค • ใช้คำศัพท์ไม่ถูกต้อง ไม่เหมาะสม
	อ่อน (2 คะแนน)	<ul style="list-style-type: none"> • ความสามารถในการใช้คำศัพท์ยังจำกัด ใช้คำศัพท์ไม่หลากหลาย • ยังพบการใช้คำศัพท์คำเดิมซ้ำๆ อยู่บ่อยครั้ง • มีการใช้คำศัพท์ที่ไม่ถูกต้อง ไม่เหมาะสมอยู่บ่อยครั้ง
	พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> • สามารถใช้คำศัพท์ได้หลากหลาย • แต่ยังใช้คำศัพท์คำเดิมซ้ำบ้างเป็นบางครั้ง • สามารถใช้คำศัพท์ได้อย่างถูกต้องและเหมาะสม อาจยังใช้คำศัพท์ผิดพลาดบ้างเป็นบางครั้ง
	ดี (4 คะแนน)	<ul style="list-style-type: none"> • สามารถใช้คำศัพท์ได้อย่างหลากหลาย • การใช้คำศัพท์ส่วนใหญ่สามารถใช้ได้ถูกต้องเหมาะสม และผิดพลาดน้อยมาก หรือ ถูกต้องทั้งหมดไม่มีความผิดพลาดเลย
2. การสร้างประโยค (syntax)	อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> • การพูดไม่มีการเชื่อมโยงให้เป็นประโยค พูดเป็นคำๆ • มีการใช้โครงสร้างประโยคที่ไม่ถูกต้องทุกครั้ง
	อ่อน (2 คะแนน)	<ul style="list-style-type: none"> • การพูดมีการเชื่อมโยงให้เป็นประโยค • การใช้โครงสร้างประโยคที่ไม่ถูกต้องอยู่บ่อยครั้ง ทำให้ผู้ฟังไม่สามารถเข้าใจได้อย่างกระจ่าง
	พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> • การพูดมีการเชื่อมโยงให้เป็นประโยค • การใช้โครงสร้างประโยคที่ไม่ถูกต้องบ้างเป็นบางครั้ง แต่ผู้ฟังสามารถเข้าใจได้
	ดี (4 คะแนน)	<ul style="list-style-type: none"> • การพูดมีการเชื่อมโยงให้เป็นประโยค • ส่วนใหญ่มีการใช้โครงสร้างประโยคที่ถูกต้อง และผิดพลาดน้อยมาก หรือไม่มีความผิดพลาดเลย

ตารางที่ 4.5 (ต่อ)

เกณฑ์	ระดับ	คำอธิบาย
3. การเชื่อมโยงเนื้อหา (cohesion)	อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อยมาก หรือไม่ใช่เลย การใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคเกือบทั้งหมดไม่ถูกต้อง
	อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อย การใช้คำเชื่อมโยงทำให้ความสัมพันธ์ระหว่างประโยคมีความสับสนอยู่บ่อยครั้ง และทำให้เข้าใจคลาดเคลื่อน
	พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยค ความสัมพันธ์ระหว่างประโยคอาจมีความสับสนอยู่บ้าง แต่ไม่ทำให้ความเข้าใจคลาดเคลื่อน
	ดี (4 คะแนน)	<ul style="list-style-type: none"> ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยค ถูกต้อง ชัดเจน เป็นส่วนใหญ่ และไม่ทำให้เข้าใจคลาดเคลื่อนผิดพลาดน้อยมาก หรือไม่มีความผิดพลาดเลย
4. การออกเสียง (pronunciation)	อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> การออกเสียงเกือบทั้งหมด มีความผิดพลาด ทำให้ไม่สามารถเข้าใจได้ นักเรียนพูดน้อยมาก หรือไม่พูดเลย
	อ่อน (2 คะแนน)	<ul style="list-style-type: none"> การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดอยู่มากทำให้ไม่สามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้ หรือทำให้เข้าใจคลาดเคลื่อน
	พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> การออกเสียงยังมีความผิดอยู่บ้าง การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดอยู่บ้างแต่ผู้ฟังยังสามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้
	ดี (4 คะแนน)	<ul style="list-style-type: none"> การออกเสียงส่วนใหญ่ถูกต้อง หรือไม่มีความผิดพลาดเลย การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดน้อยมาก หรือไม่มีความผิดพลาดเลย ผู้ฟังสามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้โดยง่าย
5. บรรลู่วัตถุประสงค์ (ideational function)	อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> พูดไม่ตรงประเด็น หรือไม่พูดเลย การพูดเกือบทั้งหมดไม่สามารถเข้าใจได้ว่าต้องการสื่อสารอะไร
	อ่อน (2 คะแนน)	<ul style="list-style-type: none"> เนื้อหาในการพูดไม่ค่อยสอดคล้องเหมาะสมกับหัวเรื่องหรือสิ่งที่ต้องการสื่อสาร ไม่สามารถเข้าใจอย่างชัดเจนได้ว่าวัตถุประสงค์ของการพูดคืออะไร
	พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> เนื้อหาในการพูดสอดคล้องเหมาะสม ถึงแม้จะมีความผิดพลาดในการพูดในบางครั้ง สามารถทำให้เข้าใจวัตถุประสงค์ที่ต้องการสื่อสารได้

ตารางที่ 4.5 (ต่อ)

เกณฑ์	ระดับ	คำอธิบาย
	ดี (4 คะแนน)	<ul style="list-style-type: none"> เนื้อหาในการทดสอบคล่องเหมาะสม มีความผิดพลาดในการพูดน้อยมาก หรือไม่ผิดพลาดเลย สามารถเข้าใจวัตถุประสงค์ในสิ่งที่ต้องการสื่อสารได้โดยง่าย
6. ความคล่องแคล่ว (fluency)	อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> การพูดยังตะกุกตะกัก พูดเป็นคำๆ ไม่เป็นประโยค หรือไม่พูดเลย
	อ่อน (2 คะแนน)	<ul style="list-style-type: none"> การพูดบ่อยครั้งยังช้ามาก และไม่สั่นไหล หยุดเป็นระยะ หยุดคิดนาน พูดไม่จบประโยค ทำให้ใจความขาดหาย
	พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> การพูดมีการหยุดบ้าง ไม่สั่นไหลเป็นบางครั้ง หยุดชะงักบ้าง หยุดคิดบ้างเป็นบางครั้ง สามารถพูดได้จบประโยค ใจความครบถ้วน
	ดี (4 คะแนน)	<ul style="list-style-type: none"> การพูดมีความเร็วสม่ำเสมอ เป็นธรรมชาติ พูดได้สั่นไหล ไม่ตะกุกตะกัก ไม่หยุดชะงัก สามารถพูดได้จบประโยค ใจความครบถ้วน

ตารางที่ 4.6 ผลการตรวจสอบความสอดคล้องของนิยามของเกณฑ์การให้คะแนนการพูดภาษาอังกฤษ และความคิดเห็น ข้อเสนอแนะของผู้ทรงคุณวุฒิเกี่ยวกับภาระงานการพูดภาษาอังกฤษ

รายการที่ประเมิน	IOC	ความคิดเห็น/ข้อเสนอแนะ
1. คำศัพท์ (vocabulary)	1	<ul style="list-style-type: none"> - เพิ่ม bullet เพื่อเน้นให้ชัดเจนว่าผู้ประเมินต้องประเมิน 3 ประเด็น - ในระดับ ดี (4 คะแนน) มีความผิดพลาดน้อย ให้เพิ่ม ใช้คำศัพท์ได้ถูกต้องทั้งหมดและไม่มีความผิดพลาดเลย
2. การสร้างประโยค (syntax)	1	<ul style="list-style-type: none"> - ต้องระบุว่าผู้ฟังสามารถฟังแล้วเข้าใจหรือไม่เข้าใจ เน้นว่ามีผู้ฟัง - ปรับคำอธิบายเกณฑ์ให้กระชับขึ้น ยาวไปผู้ประเมินอาจจะงง - เพิ่ม bullet เพื่อเน้นให้ชัดเจนว่าผู้ประเมินต้องประเมิน 3 ประเด็น - ย้าย “ใช้คำศัพท์ไม่หลากหลาย” มาอยู่ระดับ อ่อนมาก - ควรแยก accuracy กับ appropriation เป็นคนละด้าน เนื่องจากนักเรียนจะใช้คำศัพท์ที่ถูกต้อง (ไม่ผิด) แต่ไม่มีความเหมาะสม - เพิ่มคำอธิบายเพื่อให้สามารถแยกได้ระหว่างระดับอ่อน และพอใช้ ดังนี้ ระดับอ่อน เพิ่ม ทำให้ผู้ฟังไม่สามารถเข้าใจได้อย่างกระจ่าง - เพิ่ม คำอธิบายของ “การสร้างประโยค” เป็น มีประธาน กริยา กรรม ที่เหมาะสม

ตารางที่ 4.6 (ต่อ)

รายการที่ประเมิน	IOC	ความคิดเห็น/ข้อเสนอแนะ
3. การเชื่อมโยงเนื้อหา (cohesion)	1	<ul style="list-style-type: none"> - บางครั้งไม่ใช้ก็ได้ (and or but this such) The cohesion is still there อย่าเน้นการใช้คำพวกนี้เป็นเกณฑ์ - แก่เป็น 1 คือ ไม่มีการเชื่อมโยงเนื้อหา ไม่สามารถสื่อความให้เข้าใจได้ - แก่เป็น 2 คือ มีความเชื่อมโยงบ้าง - แก่เป็น 3 คือ เชื่อมโยงได้ดี - แก่เป็น 4 คือ มีความเชื่อมโยงเนื้อหาดีมาก มีความต่อเนื่องของใจความ - เพิ่ม bullet เพื่อเน้นให้ชัดเจนว่าผู้ประเมินต้องประเมิน 3 ประเด็น - cohesion เป็นเรื่องยากสำหรับเด็กประถม เด็กจะใช้เป็นแค่ and next เท่านั้น
4. การออกเสียง (pronunciation)	1	<ul style="list-style-type: none"> - แนะนำให้ดู TOEFLiBT speaking มาปรับใช้ด้วย - เพิ่ม bullet เพื่อเน้นให้ชัดเจนว่าผู้ประเมินต้องประเมิน 3 ประเด็น
5. การบรรลุวัตถุประสงค์ (ideational function)	1	<ul style="list-style-type: none"> - อันนี้น่าจะเป็นจุดเน้นของ task ว่าบรรลุไหม และจะนำมาใช้อย่างไร - เพิ่ม bullet เพื่อเน้นให้ชัดเจนว่าผู้ประเมินต้องประเมิน 3 ประเด็น - เพิ่ม คำอธิบาย “การบรรลุวัตถุประสงค์” คือ ตอบตรงคำถาม - ควรเพิ่มเพื่อความยาวในเนื้อหา - ควรเพิ่มการพูดได้ตรงหัวเรื่อง มีการให้เหตุผลสนับสนุนหัวเรื่อง
6. ความคล่องแคล่ว (fluency)	1	<ul style="list-style-type: none"> - แก่จากคำว่า ความสับสนไหล เป็น ความคล่องแคล่ว - เพิ่ม bullet เพื่อเน้นให้ชัดเจนว่าผู้ประเมินต้องประเมิน 3 ประเด็น - ไม่สับสนไหลมาก แก่เป็นสับสนไหลเป็นบางครั้ง - ไม่หยุดชะงัก แก่เป็น ไม่หยุดชะงักกลางประโยค - แก่จาก พูดไม่จบประโยค เป็น พูดไม่จบประโยคหรือทำให้ความเข้าใจคลาดเคลื่อน

1.3 ผลการพัฒนาคู่่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ ของนักเรียนชั้นประถมศึกษาปีที่ 6

ผลการพัฒนาคู่่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 พบว่าคู่มือการประเมินทักษะการพูดภาษาอังกฤษฯ ประกอบด้วยองค์ประกอบที่สำคัญดังนี้

1.3.1 โครงสร้างของคู่มือการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ประกอบด้วย 5 องค์ประกอบ ดังนี้

1. บทนำ
2. วัตถุประสงค์ของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

วัตถุประสงค์ของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 คือ เพื่อประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้กระบวนการของแฟ้มสะสมงาน เพื่อให้นักเรียนได้ปฏิบัติงานการพูดภาษาอังกฤษที่เป็นจริง คือ ภาระงานการพูดภาษาอังกฤษ 4 ภาระงาน ได้แก่

- 1) การสัมภาษณ์ (interview)
- 2) การนำเสนอปากเปล่า (oral presentation)
- 3) การเล่านิทาน (storytelling)
- 4) การบรรยายภาพ (picture description)

3. คำนิยามศัพท์

1) แฟ้มสะสมงาน หมายถึง สิ่งที่ใช้ในการรวบรวมผลงานนักเรียนจากการเรียนรู้ในชั้นเรียนและนำไปสู่การประเมินการเรียนรู้ของนักเรียนตามสภาพจริงวิธีหนึ่ง ซึ่งเป็นวิธีการที่ทำให้ผู้เรียนได้ลงมือปฏิบัติงานที่เป็นจริง และผู้เรียนได้รับข้อมูลย้อนกลับจากครู การสะท้อนคิด ได้ประเมินตนเอง และประเมินเพื่อนในชั้นเรียนด้วยทำให้มีความรับผิดชอบต่อการเรียนของตน

2) การประเมินทักษะการพูด หมายถึง การทดสอบทักษะการพูดภาษาอังกฤษของนักเรียน โดยให้นักเรียนได้แสดงความสามารถในการพูดสื่อความในการสื่อสารออกมา สำหรับการวิจัยครั้งนี้ประเมินทักษะการพูดภาษาอังกฤษโดยใช้ภาระงานการพูด (tasks)

3) ภาระงานการพูด (task) หมายถึง กิจกรรมที่ให้นักเรียนใช้สำหรับการทดสอบทักษะการพูดของนักเรียนโดยการให้นักเรียนได้แสดงการพูดสื่อสารหรือพูดสื่อความออกมาตามความรู้ความสามารถทางภาษาอังกฤษ ซึ่งภาระงานที่ใช้ในการประเมินครั้งนี้ ได้แก่ 1) การสัมภาษณ์ (interview) 2) การนำเสนอเกี่ยวกับตนเอง (oral presentation) 3) การเล่านิทาน (storytelling) และ 4) การบรรยายภาพ (picture description) โดยมีเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale)

4) การประเมินระหว่างเรียน (formative assessment) หมายถึง การประเมินทักษะการพูดของนักเรียนในระหว่างการดำเนินการเรียนการสอนโดยใช้แฟ้มสะสมงานฯ โดยมีจุดประสงค์เพื่อศึกษาความก้าวหน้าของผู้เรียน และปรับปรุงและพัฒนาการพูดของนักเรียน

5) การประเมินสรุปรวม (summative assessment) หมายถึง การประเมินเมื่อสิ้นสุดการดำเนินการจัดการเรียนการสอนด้วยแฟ้มสะสมงานฯ เพื่อสรุปรวมและให้ระดับทักษะการพูดแก่นักเรียน

4. เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ

ภาระงานทั้ง 4 ภาระงานใช้เกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) โดยมีขอบเขตเนื้อหาในการประเมิน ดังนี้

- 1) คำศัพท์ (vocabulary)
- 2) การสร้างประโยค (syntax)
- 3) การเชื่อมโยงเนื้อหา (cohesion)
- 4) การออกเสียง (pronunciation)
- 5) บรรลู่วัตถุประสงค์ (ideational function)
- 6) ความคล่องแคล่ว (fluency)

ให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี ตามลำดับ รายละเอียดของเกณฑ์การให้คะแนน ดังแสดงในภาคผนวก

การแปลความหมายคะแนนความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 สามารถแปลความหมาย ดังนี้

คะแนนเฉลี่ย ต่ำกว่า 1.51 หมายถึง อ่อนมาก

คะแนนเฉลี่ย 1.51-2.50 หมายถึง อ่อน

คะแนนเฉลี่ย 2.51-3.50 หมายถึง พอใช้

คะแนนเฉลี่ย 3.51-4.00 หมายถึง ดี

1.3.2 ส่วนประกอบของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ส่วนประกอบของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 มีทั้งหมด 3 ส่วน ได้แก่ ส่วนนำ ส่วนเนื้อหา และภาคผนวก ซึ่งแต่ละส่วนมีรายละเอียดดังนี้

1. ส่วนนำ ประกอบด้วย

- 1) ปก
- 2) คำนำ
- 3) สารบัญ
- 4) ข้อมูลส่วนตัวนักเรียน
- 5) เป้าหมายการเรียนรู้ ความคาดหวังในการเรียน

2. ส่วนเนื้อหา ประกอบด้วย

- 1) ผลการประเมินระหว่างเรียนและการประเมินสรุปรวม ได้มาจาก 3 ส่วน ได้แก่ ผลการประเมินตนเอง ผลการประเมินจากครู และ ผลการประเมินจากเพื่อน
- 2) ข้อมูลย้อนกลับจากครู
- 3) ผลการสะท้อนคิดต่อผลงานตนเอง
- 4) เป้าหมายการเรียนรู้ ความคาดหวังในการเรียน

3. ส่วนภาคผนวก ประกอบด้วย

- 1) คำอธิบายภาระงาน
- 2) เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ
- 3) CD บันทึกเสียงการพูดภาษาอังกฤษของนักเรียน

1.3.3 ขั้นตอนของการพัฒนาเพิ่มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

1. ขั้นการวางแผน

1) ศึกษาวิเคราะห์หลักสูตรกลุ่มสาระการเรียนรู้ภาษาต่างประเทศ (ภาษาอังกฤษ) ตัวชี้วัดในระดับชั้นประถมศึกษาปีที่ 6 การวัดและประเมินผลทางภาษาการประเมินทักษะการพูด

2) กำหนดวัตถุประสงค์ในการจัดการเรียนรู้ด้วยเพิ่มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ

2. ขั้นการเตรียมความพร้อมนักเรียน

1) ให้ความรู้เกี่ยวกับกระบวนการของเพิ่มสะสมงาน เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษแก่นักเรียน

2) ฝึกทักษะการประเมินตนเอง และประเมินเพื่อน

3. ขั้นการเก็บรวบรวมหลักฐาน

1) ให้นักเรียนวิเคราะห์ความสามารถในการพูดภาษาอังกฤษของตนเองและกำหนดเป้าหมายในการเรียนของตนเอง

2) ประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ดังนี้ ประเมินตนเอง ให้เพื่อนประเมิน และครูประเมิน

3) นักเรียนประเมินทักษะการพูดภาษาอังกฤษของเพื่อน

4. ขั้นการติดตามความก้าวหน้า

1) ครูประเมินทักษะการพูดภาษาอังกฤษของนักเรียน

2) ครูให้ข้อมูลย้อนกลับแก่นักเรียน โดยให้ข้อมูลแบบระบุความสำเร็จและระบุข้อผิดพลาดเพื่อการแก้ไข

5. ขั้นการปรับปรุงผลงาน

1) นักเรียนวางแผนการปรับปรุงการพูดของตนเองตามข้อมูลย้อนกลับของครูผู้สอน (teacher feedback)

2) นักเรียนแก้ไขการพูดภาษาอังกฤษของตนเองอีก 2 ครั้ง รวมนักเรียนได้รับประเมินทักษะการพูดทั้งสิ้น 3 ครั้ง

6. ขั้นการสะท้อนคิด

1) นักเรียนสะท้อนคิดต่อผลงานของตนเอง และประสบการณ์ต่างๆ ที่ได้รับจากการจัดการเรียนรู้โดยใช้แฟ้มสะสมงานฯ

2) นักเรียนระบุจุดเด่น และจุดที่ต้องพัฒนาของตนเอง

7. ขั้นการจัดแสดงผลงาน

1) จัดแสดงผลมสะสมงานฯ ต่อครู นักเรียน และผู้ปกครอง

2) จัดให้นักเรียนได้แสดงการพูดภาษาอังกฤษบนเวทีของโรงเรียนต่อ ครู นักเรียน ห้องอื่นๆ

1.3.4 เครื่องมือที่ใช้ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

1. เครื่องมือสำหรับนักเรียน

2. เครื่องมือสำหรับครู

เครื่องมือที่ใช้ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วยเครื่องมือสำหรับครู นักเรียน ในแต่ละชั้นมีรายละเอียด ดังนี้

ตารางที่ 4.7 เครื่องมือที่ใช้ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ

ขั้นที่	เครื่องมือ	
	นักเรียน	ครู
1. การวางแผน	<ul style="list-style-type: none"> - แบบบันทึกข้อมูลส่วนตัว - แบบบันทึกการกำหนดวัตถุประสงค์ เป้าหมายในการเรียนวิชาภาษาอังกฤษ - แบบบันทึกการกำหนดวัตถุประสงค์ เป้าหมายในการเรียนด้วยแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ 	<ul style="list-style-type: none"> - ตารางวิเคราะห์หลักสูตร ตัวชี้วัด - ใบความรู้การประเมินทักษะการพูดภาษาอังกฤษ
2. การเตรียมความพร้อมนักเรียน	<ul style="list-style-type: none"> - แบบบันทึกคะแนนการพูด 	<ul style="list-style-type: none"> - ไฟล์เสียงตัวอย่างสำหรับฝึกประเมิน - เกณฑ์การพูดภาษาอังกฤษ

ตารางที่ 4.7 (ต่อ)

ชั้นที่	เครื่องมือ	
	นักเรียน	ครู
3. การเก็บรวบรวมหลักฐาน	<ul style="list-style-type: none"> - แบบบันทึกการวิเคราะห์ ความสามารถในการพูด ภาษาอังกฤษของตนเอง - แบบบันทึกการกำหนดเป้าหมาย ในการเรียน/การพูดภาษาอังกฤษ ของตนเอง - ประเมินตนเอง - แบบประเมินเพื่อน 	<ul style="list-style-type: none"> - แบบประเมินทักษะการพูด นักเรียน - แบบบันทึกคะแนนทักษะการพูด ภาษาอังกฤษของนักเรียน
4. การติดตามความก้าวหน้า		<ul style="list-style-type: none"> - แบบบันทึกการให้ข้อมูล ย้อนกลับแก่นักเรียนเป็น รายคน
5. การปรับปรุงผลงาน	<ul style="list-style-type: none"> - แบบบันทึกการวางแผนการ ปรับปรุงการพูด 	<ul style="list-style-type: none"> - แบบบันทึกการให้ข้อมูล ย้อนกลับแก่นักเรียนเป็น รายคน
6. การสะท้อนคิด	<ul style="list-style-type: none"> - แบบบันทึกการสะท้อนคิดต่อ ผลงานของตน - แบบบันทึกจุดเด่น จุดที่ควรพัฒนา ของนักเรียน 	<ul style="list-style-type: none"> - แบบบันทึกจุดเด่น จุดที่ควร พัฒนาของนักเรียนเป็นรายคน
7. การจัดแสดงผลงาน		<ul style="list-style-type: none"> - แบบบันทึกจุดเด่น จุดที่ควร พัฒนาของนักเรียนในภาพรวม

1.4 การเลือกใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษ

การประเมินทักษะการพูดภาษาอังกฤษโดยใช้แฟ้มสะสมงานที่มีความเที่ยงในระดับที่เหมาะสม สามารถดำเนินการได้ ดังนี้

1.4.1 จำนวนผู้ประเมินและภาระงานที่เหมาะสม

จำนวนผู้ประเมินและจำนวนภาระงานที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษ ผู้ออกแบบการประเมินสามารถเลือกดำเนินการได้หลายรูปแบบโดยพิจารณาความเหมาะสมและเป็นไปได้ในการดำเนินการและมีระดับความเที่ยงในระดับต่างๆ ทั้งค่าสัมประสิทธิ์การสรุปลำดับสำหรับการตัดสินใจเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปลำดับสำหรับการตัดสินใจเชิงสัมบูรณ์ ดังตารางที่ 4.8-4.12

1. การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนเป็นรายองค์ประกอบ

1) ความเที่ยงแบบอิงกลุ่ม

ตารางที่ 4.8 ความเที่ยงแบบอิงกลุ่มในการประเมินทักษะการพูดรายองค์ประกอบ

จำนวน		องค์ประกอบรายด้าน					
ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	คำศัพท์	การสร้าง ประโยค	การ เชื่อมโยง	การออก เสียง	การบรรลุ เป้าหมาย	ความ คล่องแคล่ว
4	2	0.617	0.579	0.585	0.806	0.610	0.647
4	3	0.651	0.611	0.629	0.828	0.667	0.691
4	4	0.669	0.628	0.654	0.8400	0.701	0.715
4	5	0.680	0.638	0.670	0.848	0.722	0.730
4	6	0.688	0.645	0.681	0.853	0.737	0.741
4	7	0.694	0.651	0.689	0.856	0.749	0.749
4	8	0.698	0.655	0.695	0.859	0.757	0.755
4	9	0.701	0.658	0.700	0.861	0.764	0.759
4	10	0.704	0.661	0.704	0.863	0.770	0.763
4	4	0.669	0.628	0.654	0.840	0.701	0.715
5	4	0.713	0.674	0.703	0.867	0.739	0.758
6	4	0.746	0.710	0.739	0.886	0.768	0.790
7	4	0.772	0.738	0.768	0.901	0.789	0.815
8	4	0.792	0.760	0.791	0.911	0.806	0.834
9	4	0.808	0.778	0.810	0.920	0.820	0.85
10	4	0.822	0.793	0.825	0.927	0.831	0.863

2) ค่าความเที่ยงแบบอิงเกณฑ์

ตารางที่ 4.9 ความเที่ยงแบบอิงเกณฑ์ในการประเมินทักษะการพูดรายด้าน

จำนวน		องค์ประกอบรายด้าน					
ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	คำศัพท์	การสร้าง ประโยค	การ เชื่อมโยง	การ ออกเสียง	การบรรลุ เป้าหมาย	ความ คล่องแคล่ว
4	4	0.518	0.508	0.535	0.803	0.644	0.659
4	5	0.526	0.517	0.557	0.813	0.664	0.680
4	6	0.530	0.524	0.572	0.82	0.678	0.694
4	7	0.534	0.528	0.584	0.824	0.689	0.705
4	8	0.537	0.532	0.593	0.828	0.697	0.713
4	9	0.539	0.534	0.600	0.831	0.703	0.719
4	10	0.541	0.536	0.606	0.833	0.709	0.725
4	4	0.518	0.508	0.535	0.803	0.644	0.659
5	4	0.572	0.56	0.582	0.833	0.686	0.699
6	4	0.614	0.60	0.619	0.853	0.718	0.729
7	4	0.648	0.633	0.648	0.869	0.742	0.752
8	4	0.676	0.660	0.672	0.880	0.762	0.770
9	4	0.699	0.683	0.691	0.890	0.778	0.785
10	4	0.719	0.702	0.708	0.898	0.791	0.797

3) การประเมินทักษะการพูดรายด้านในภาพรวมของการประเมิน

ตารางที่ 4.10 ระดับความคลาดเคลื่อนและความเที่ยงของคะแนนรวมในการประเมินการพูดฯ
รายองค์ประกอบเพื่อการตัดสินใจจำนวนผู้ประเมินและภาระงาน

จำนวน		ความคลาดเคลื่อน		ความเที่ยง	
ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	อิงกลุ่ม (Relative error)	อิงเกณฑ์ (Absolute error)	อิงกลุ่ม	อิงเกณฑ์
4	2	0.017	0.031	0.819	0.708
4	3	0.015	0.027	0.837	0.733
4	4	0.014	0.026	0.846	0.745
4	5	0.013	0.025	0.851	0.753
4	6	0.013	0.024	0.855	0.758
4	7	0.013	0.023	0.857	0.762
4	8	0.012	0.023	0.859	0.765
4	9	0.012	0.023	0.861	0.767
4	10	0.012	0.023	0.862	0.769
3	4	0.018	0.033	0.806	0.692
4	4	0.014	0.026	0.846	0.745
5	4	0.011	0.021	0.872	0.781
6	4	0.009	0.018	0.890	0.807
7	4	0.008	0.016	0.904	0.826
8	4	0.007	0.014	0.914	0.842
9	4	0.006	0.013	0.922	0.854
10	4	0.006	0.012	0.929	0.864

2. การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนเป็นในภาพรวมของภาระงาน

1) การประเมินนักเรียนแยกตามภาระงาน

ตารางที่ 4.11 ความเที่ยงในการประเมินเมื่อประเมินรายการภาระงาน

ความเที่ยง	ภาระงาน	จำนวนผู้ประเมิน		
		2 คน	3 คน	4 คน
อิงกลุ่ม	การสัมภาษณ์	0.872	0.905	0.923
	การนำเสนอปากเปล่า	0.929	0.95	0.960
	การเล่นิทาน	0.823	0.875	0.903
	บรรยายภาพ	0.840	0.887	0.912
อิงเกณฑ์	การสัมภาษณ์	0.820	0.862	0.885
	การนำเสนอปากเปล่า	0.914	0.938	0.950
	การเล่นิทาน	0.810	0.865	0.895
	บรรยายภาพ	0.828	0.878	0.905

2) การประเมินทักษะการพูดในภาพรวมของคะแนนรวมของการประเมิน

ตารางที่ 4.12 ระดับความคลาดเคลื่อนและความเที่ยงของคะแนนรวมในการประเมินการพูดฯ เพื่อการตัดสินใจจำนวนผู้ประเมินและภาระงาน

จำนวน		ความคลาดเคลื่อน		ความเที่ยง	
ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	อิงกลุ่ม	อิงเกณฑ์	อิงกลุ่ม	อิงเกณฑ์
		(Relative error)	(Absolute error)		
2	4	0.006	0.007	0.939	0.925
3	4	0.004	0.005	0.958	0.947
4	4	0.003	0.004	0.967	0.959

1.4.2 การเลือกรูปแบบการประเมินของผู้ประเมินที่เหมาะสม

การเลือกจำนวนผู้ประเมิน จำนวนภาระงานและวิธีการประเมินของผู้ประเมินที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษ ผู้ออกแบบการประเมินสามารถเลือกดำเนินการได้หลายรูปแบบโดยพิจารณาความเหมาะสมและเป็นไปได้ในการดำเนินการและมีระดับความเที่ยงในระดับต่างๆ ทั้งความเที่ยงสำหรับอิงเกณฑ์และอิงกลุ่ม ดังตารางที่ 4.13-4.14

ตารางที่ 4.13 ระดับความคลาดเคลื่อนและความเที่ยงของคะแนนรวมในการประเมินการพูดฯ เพื่อการตัดสินใจเลือกรูปแบบการผู้ประเมิน

รูปแบบการประเมินของผู้ประเมิน	จำนวน		ความคลาดเคลื่อน		ความเที่ยง	
	ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	อิงกลุ่ม (Relative error)	อิงเกณฑ์ (Absolute error)	อิงกลุ่ม	อิงเกณฑ์
ประเมินซ้ำกันทุกงาน	2	4	0.006	0.007	0.939	0.925
	3	4	0.004	0.005	0.958	0.947
	4	4	0.003	0.004	0.967	0.959
ประเมินงานไม่ซ้ำกัน	2	4	0.004	0.005	0.955	0.945
	3	4	0.003	0.004	0.968	0.961
	4	4	0.002	0.003	0.975	0.969

1.4.3 การเลือกรูปแบบการประเมินสำหรับการประเมินระหว่างเรียนและสรุปรวม

ตารางที่ 4.14 ระดับความเที่ยงของการประเมินการพูดฯ เพื่อการตัดสินใจรูปแบบการประเมิน

ความเที่ยง	จำนวนผู้ประเมิน (คน)	จำนวนภาระงาน (เรื่อง)			การประเมิน/ความเที่ยง	
		ระหว่างเรียน	สรุปรวม	รวม	ระหว่างเรียน	สรุปรวม
อิงกลุ่ม	2	3	1	4	0.937	0.71
	3	3	1	4	0.956	0.753
	4	3	1	4	0.966	0.777
	5	3	1	4	0.972	0.792
	6	3	1	4	0.976	0.802
	2	3	2	5	0.937	0.783
	3	3	2	5	0.956	0.824
	4	3	2	5	0.966	0.847
	5	3	2	5	0.972	0.861
	6	3	2	5	0.976	0.871

ตารางที่ 4.14 (ต่อ)

ความเที่ยง	จำนวนผู้ประเมิน (คน)					
		ระหว่างเรียน	สรุปรวม	รวม	ระหว่างเรียน	สรุปรวม
อิงเกณฑ์	2	3	1	4	0.915	0.697
	3	3	1	4	0.941	0.741
	4	3	1	4	0.954	0.765
	5	3	1	4	0.962	0.781
	6	3	1	4	0.968	0.791
	2	3	2	5	0.915	0.774
	3	3	2	5	0.941	0.817
	4	3	2	5	0.954	0.84
	5	3	2	5	0.962	0.855
	6	3	2	5	0.968	0.865

1.5 การตรวจสอบคุณภาพคู่มือแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ

1.5.1 ผลการตรวจสอบคุณภาพของคู่มือแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษโดยผู้ทรงคุณวุฒิ

ผลการตรวจสอบความสอดคล้องความเหมาะสมของคู่มือแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษกับวัตถุประสงค์ของคู่มือฯ พบว่า มีค่าดัชนี IOC อยู่ระหว่าง 0.778-1.000 ซึ่งผ่านเกณฑ์ทุกข้อ โดยพบว่าผู้ทรงคุณวุฒิ จำนวน 2 ท่าน ใน 9 ท่าน เห็นว่า การให้ข้อมูลย้อนกลับในส่วนที่เป็นข้อบกพร่อง และผิดพลาด เช่น การออกเสียงผิดในทันทีหลังนักเรียนพูดต่อหน้านักเรียนคนอื่น ๆ นักเรียนอาจจะอายุได้ แต่อย่างไรก็ตามในขั้นตอนนี้ ผู้วิจัยไม่ได้แก้ไขเนื่องจากผู้ทรงคุณวุฒิส่วนใหญ่เห็นว่าเหมาะสมแล้ว แต่ในขั้นตอนการนำไปสอนจริงได้ให้คำแนะนำครูผู้สอนเกี่ยวกับการเลือกใช้คำเพื่อไม่ให้นักเรียนเกิดความอับอายต่อไป รายละเอียดดังตารางที่ 4.15

ตารางที่ 4.15 ผลการตรวจสอบความสอดคล้องของคู่มือเพิ่มประสิทธิภาพการประเมินทักษะการพูด
ภาษาอังกฤษ

รายการ	IOC	ความหมาย	ข้อเสนอแนะ
1. โครงสร้างของคู่มือฯ			
1.1 บทนำ	1	สอดคล้อง	
1.2 วัตถุประสงค์ของเพิ่มประสิทธิภาพฯ	1	สอดคล้อง	
1.3 คำนียามศัพท์	1	สอดคล้อง	
1.4 เกณฑ์การให้คะแนนทักษะการพูด ภาษาอังกฤษ	1	สอดคล้อง	- อาจขีดเส้นจุดเน้นของเกณฑ์แต่ละด้าน เช่น ไม่ ทำให้เข้าใจผิด ไม่หยุดระหว่างพูด
2. ส่วนประกอบของเพิ่มประสิทธิภาพฯ			
2.1 ส่วนนำ			
2.1.1 ปก	1	สอดคล้อง	- ข้อมูลนักเรียนควรใส่ทั้งภาษาไทยและ ภาษาอังกฤษ อาจอยู่ในวงเล็บ
2.1.2 คำนำ	1	สอดคล้อง	
2.1.3 สารบัญ	1	สอดคล้อง	
2.1.4 ข้อมูลนักเรียน	1	สอดคล้อง	
2.1.5 เป้าหมายการเรียนรู้ ความ คาดหวังในการเรียน	1	สอดคล้อง	
2.2 ส่วนเนื้อหา			
2.2.1 ผลการประเมินระหว่างเรียน และการประเมินสรุปรวม			
1) ผลการประเมินตนเอง	1	สอดคล้อง	- ควรสั้นๆ ไม่ยาว
2) ผลการประเมินจากครู	1	สอดคล้อง	- ควรให้นักเรียนทันที ดังนั้นควรสั้นๆ
3) ผลการประเมินจากเพื่อน	1	สอดคล้อง	- ควรให้นักเรียนมีการสลับเปลี่ยนผู้ประเมินไปใน แต่ละวัน
2.2.2 ข้อมูลย้อนกลับจากครู	.778	สอดคล้อง	- ครูจะเขียนทันเวลาไหม - ควรคำนึงถึงการใช้เวลาของครูว่าจะทำได้ไหม - ควรให้ครูให้ข้อมูลทันทีด้วยและเขียนลงในแบบ บันทึกของนักเรียน - ควรพูดภาพรวม เพราะนักเรียนอาจจะอาย เพื่อน - ควรให้ข้อชี้แนะปรับแก้ทันทีเมื่อพูดจบ นักเรียน จะได้รู้ข้อบกพร่องของตนเอง

ตารางที่ 4.15 (ต่อ)

รายการ	IOC	ความหมาย	ข้อเสนอแนะ
2.2.3 ผลการสะท้อนคิดต่อผลงาน ตนเอง	1	สอดคล้อง	- นักเรียนเอากลับบ้านจะเขียนใหม่ - ควรฝึกนักเรียนในการเขียนก่อน
2.3 ส่วนภาคผนวก			
2.3.1 คำอธิบายภาระงาน	1	สอดคล้อง	- ควรจะสั้นๆ กระชับกว่าของครู
2.3.2 เกณฑ์การให้คะแนนทักษะการ พูดภาษาอังกฤษ	1	สอดคล้อง	- อาจให้เก็บในเครื่องคอมพิวเตอร์โรงเรียนหรือ เว็บไซต์ที่นักเรียนเปิดฟังได้สะดวก
2.3.3 CD บันทึกเสียงการพูด ภาษาอังกฤษของนักเรียน	1	สอดคล้อง	
3. ขั้นตอนของการพัฒนาเพิ่มสะสมงานฯ	1	สอดคล้อง	
3.1 ขั้นการวางแผน	1	สอดคล้อง	- ครูในกลุ่มสาระฯ ทุกคนควรมีส่วนร่วมด้วย - นักเรียนควรได้ร่วมวางแผนด้วย
3.2 ขั้นการเตรียมความพร้อมนักเรียน	1	สอดคล้อง	- นักเรียนควรฝึกทั้งจากการฟังเทป และฟังเพื่อน พูดจริงๆ - ควรให้นักเรียนฝึกเขียนแสดงความคิดเห็นต่อ ตนเอง ผลงานเป็นก่อน
3.3 ขั้นการเก็บรวบรวมหลักฐาน	1	สอดคล้อง	- เอกสารแบบประเมินต่างๆ ควรแจกก่อนเรียน
3.4 ขั้นการติดตามความก้าวหน้า	1	สอดคล้อง	- ครูควรให้นักเรียนได้มีโอกาสฝึกพูดนอก ห้องเรียน
3.5 ขั้นการปรับปรุงผลงาน	1	สอดคล้อง	- ต้องแจ้งนักเรียนว่าปรับปรุงได้กี่ครั้ง
3.6 ขั้นการสะท้อนคิด	1	สอดคล้อง	- ต้องมีการฝึกเขียนการสะท้อนคิดให้นักเรียน - ควรระบุว่าให้นักเรียนสะท้อนคิดต่ออะไร เช่น ผลการพูด การฟังเพื่อน การเป็นผู้ประเมิน
3.7 ขั้นการจัดแสดงผลงาน	1	สอดคล้อง	- จัดแสดงเฉพาะแท่งงานหรือไม่ - ควรให้นักเรียนได้พูดแสดงความสามารถบนเวที เช่น ของโรงเรียน - การให้แสดงควรมีการฝึกนักเรียนเพิ่มเติมก่อน - การแสดงควรให้นักเรียนเลือกเพียงเรื่องเดียวที่ นักเรียนมั่นใจที่สุด - การแสดงอาจให้แสดงให้นักเรียนชั้นอื่นๆ ได้ดู เพื่อเป็นการสร้างแรงบันดาลใจ

ตารางที่ 4.15 (ต่อ)

รายการ	IOC	ความหมาย	ข้อเสนอแนะ
4. เครื่องมือที่ใช้ในแฟ้มสะสมงาน			
4.1 สำหรับนักเรียน			
4.1.1 แบบบันทึกข้อมูลส่วนตัว	1	สอดคล้อง	
4.1.2 แบบบันทึกการกำหนด วัตถุประสงค์ เป้าหมายในการเรียนวิชา ภาษาอังกฤษ	1	สอดคล้อง	
4.1.3 แบบบันทึกการกำหนด วัตถุประสงค์ เป้าหมายในการเรียนด้วยแฟ้ม สะสมงานการประเมินทักษะการพูด ภาษาอังกฤษ	1	สอดคล้อง	- ควรเขียนพร้อมกันทั้งห้อง - อาจให้ผู้ปกครองมีส่วนร่วม
4.1.4 แบบบันทึกคะแนนการพูด	1	สอดคล้อง	
4.1.5 แบบบันทึกการวิเคราะห์ ความสามารถในการพูดภาษาอังกฤษของ ตนเอง	1	สอดคล้อง	- ควรให้นักเรียนทำทันทีหลังเรียน ไม่ต้องเขียน ยาว - นักเรียนควรทำที่โรงเรียน
4.1.6 แบบบันทึกการกำหนด เป้าหมายในการเรียน/การพูดภาษาอังกฤษ ของตนเอง	1	สอดคล้อง	- ควรเขียนพร้อมกันทั้งห้อง - อาจให้ผู้ปกครองมีส่วนร่วม
4.1.7 ประเมินตนเอง	1	สอดคล้อง	- ควรให้นักเรียนฟังเสียงตัวเองด้วย เช่น เปิดฟัง ซ้ำได้ แล้วค่อยประเมิน
4.1.8 แบบประเมินเพื่อน	1	สอดคล้อง	- ควรให้มีการหมุนเวียนการประเมินในแต่ละครั้ง
4.1.9 แบบบันทึกการวางแผนการ ปรับปรุงการพูด	1	สอดคล้อง	
4.1.10 แบบบันทึกการสะท้อนคิดต่อ ผลงานของตน	1	สอดคล้อง	- ควรระบุแต่ละภาระงาน
4.1.11 แบบบันทึกจุดเด่น จุดที่ควร พัฒนาของนักเรียน	1	สอดคล้อง	
4.2 สำหรับครู			
4.2.1 เกณฑ์การพูดภาษาอังกฤษ	1	สอดคล้อง	
4.2.3 แบบประเมินทักษะการพูด นักเรียน	1	สอดคล้อง	
4.2.4 แบบบันทึกคะแนนทักษะการ พูดภาษาอังกฤษของนักเรียน	1	สอดคล้อง	

ตารางที่ 4.15 (ต่อ)

รายการ	IOC	ความหมาย	ข้อเสนอแนะ
4.2.5 แบบบันทึกการให้ข้อมูล ย้อนกลับนักเรียนเป็นรายคน	1	สอดคล้อง	
4.2.6 แบบบันทึกจุดเด่น จุดที่ควร พัฒนาของนักเรียนเป็นรายคน	1	สอดคล้อง	- ครูเขียนตอนไหน ให้ทันทีหรือตอนหลังเรียน - ควรให้ครูบอกนักเรียนในระหว่างเรียนด้วย
4.2.7 แบบบันทึกจุดเด่น จุดที่ควร พัฒนาของนักเรียนในภาพรวม	1	สอดคล้อง	

4.6.2 ผลการใช้คู่มือเพิ่มประสิทธิภาพการประเมินทักษะการพูดภาษาอังกฤษโดยครูผู้สอน

เมื่อได้พัฒนาคู่มือเพิ่มประสิทธิภาพการประเมินทักษะภาษาอังกฤษเรียบร้อยแล้ว ผู้วิจัยได้นำคู่มือที่พัฒนาขึ้นให้ครูผู้สอนได้ทดลองใช้ในการจัดการเรียนการสอนโดยใช้เพิ่มประสิทธิภาพการประเมินทักษะการพูดภาษาอังกฤษ เมื่อครูผู้สอนได้ทดลองใช้คู่มือเพิ่มประสิทธิภาพการประเมินทักษะภาษาอังกฤษเรียบร้อยแล้ว ผู้วิจัยได้สอบถามความคิดเห็นของผลการใช้คู่มือดังกล่าว โดยปรับมาจากมาตรฐานการประเมินตามมาตรฐานของ stuffebeam (1981 อ้างถึงใน ศิริชัย กาญจนวาสี, 2552) และ กมลวรรณ ดังธกานนท์ (2547) ใน 4 ด้าน ได้แก่ ความเหมาะสม ความชัดเจน ความเป็นไปได้ ในทางปฏิบัติ และความเป็นประโยชน์ในการนำไปใช้ ผลการใช้คู่มือเพิ่มประสิทธิภาพการพูดภาษาอังกฤษ ดังนี้

ด้านความเหมาะสม ครูให้ความเห็นว่าคู่มือการใช้เพิ่มประสิทธิภาพเป็นคู่มือที่ประกอบไปด้วยเกณฑ์การให้คะแนนและคำอธิบายภาระงาน ทำให้ครูและนักเรียนสามารถทำความเข้าใจเกณฑ์การให้คะแนนได้ก่อนเรียน และนักเรียนทราบว่าจะถูกประเมินอย่างไร นักเรียนสามารถโต้แย้งผลการประเมินจากเพื่อนได้หากไม่มั่นใจ และนักเรียนยังสามารถให้ข้อเสนอแนะแก่เพื่อนที่ตนประเมินได้อีกด้วย

“ก็มีที่ช่วงแรกๆ นักเรียนเขาฟ้องว่าเพื่อนโกรธเขาเลยให้น้อย พี่ก็บอกว่าคงไม่ใช่หรอก ...เขาจะให้เพื่อนแบบนั้นบ้าง...แต่ไม่ทำหรอก ต้องควรสอน”

ครูผู้สอน 28 กรกฎาคม 2557

“นักเรียนอย่างเด็กชาย...เขาก็แนะนำเพื่อนได้ อ้าปากกว้างๆ ลี ออกเสียงนี้ลี ไม่ต้องอาย เด็กเขาสนุกนะ”

ครูผู้สอน 28 กรกฎาคม 2557

“พี่ว่ามีเกณฑ์ให้คะแนนแบบนี้ดี ทุกคนมารู้ร่วมกันว่าถ้าเธออยากได้คะแนนเยอะ
อยากเก่งจะต้องทำไง ไม่ใช่พูดไปเรื่อย”

ครูผู้สอน 28 กรกฎาคม 2557

“นักเรียนเขามีเกณฑ์ในแฟ้มทุกคน อาจยากไปแต่เขาก็พอรู้เรื่องนะ จะได้คะแนนต้อง
ฝึกพูด ไม่อาย หาคำศัพท์ รู้จักถาม เด็กๆ เขาก็ถามกัน ค้น google เป็น เก่งกว่าครูอีก”

ครูผู้สอน 28 กรกฎาคม 2557

ด้านความชัดเจน ครูให้ความเห็นว่า คู่มือการใช้แฟ้มสะสมงานเป็นคู่มือที่มีเครื่องมือที่ใช้
ในการประเมินที่เหมาะสม ครอบคลุม โดยเครื่องมือสำหรับนักเรียนทั้งแบบประเมินเพื่อน และ
ประเมินตนเอง เป็นแบบประเมินง่ายๆ มีการ์ตูนแทนระดับความสามารถ มีการให้นักเรียนกำหนด
เป้าหมาย และสะท้อนคิดต่องานตนเอง การให้คะแนนของครูมีทั้งที่เป็นตัวเลขเพื่อตัดลิ้น
ความสามารถในการพูด และมีทั้งแบบเชิงคุณภาพเพื่อให้เป็นข้อมูลย้อนกลับแก่นักเรียนนำไปปรับปรุง
ต่อไป

“แบบประเมินที่เป็นประเมินตน กับประเมินเพื่อนง่ายดี นักเรียนแค่ตัดสินว่าดี ไม่ดี
ตามที่ตัวเองคิด แต่ก็ต้องทบทวนเกณฑ์กันบ่อยๆ ไม่งั้นลืมทั้งครูทั้งเด็ก”

ครูผู้สอน 28 กรกฎาคม 2557

“เด็กก็เขียนไม่เก่งนะตอนแรก ต้องฝึก ดีที่เราเขียนพร้อมกัน เขียนที่โรงเรียนในช่วง
เวลา ต้องบอกว่าการสะท้อนต่อการพูดของตัวเองยังงัย เป้าหมายก็คิดได้นะ เรียนๆ ไปเขาบอก
ภาษาอังกฤษสำคัญ”

ครูผู้สอน 28 กรกฎาคม 2557

“การกำหนดเป้าหมายเขามาเขียนไปพร้อมกัน ครูช่วยแนะเขาก็ทำได้”

ครูผู้สอน 28 กรกฎาคม 2557

“การสะท้อนคิดต่อการพูดของตนเอง ในคู่มือมันละเอียด เขียนกันมือหงิก แรกๆ เด็ก
บ่น เขียนไม่เป็น พอนานไปเขียนได้เอง”

ครูผู้สอน 28 กรกฎาคม 2557

ด้านความเป็นไปได้ในทางปฏิบัติ ครูให้ความเห็นว่า คู่มือการใช้แฟ้มสะสมงานเป็นคู่มือที่มีขั้นตอนและกิจกรรมที่ชัดเจน มีการระบุบทบาทครูและนักเรียน รวมทั้งมีขั้นตอนในการให้นักเรียนได้ปรับปรุงตนเองอย่างเป็นระบบมีการประเมินซ้ำหลายครั้ง และครูมีการให้ข้อมูลย้อนกลับแก่นักเรียนในการปรับปรุงการพูดของตนเองทั้งในทันที และมีการบันทึกในแฟ้มสะสมงานให้นักเรียนเป็นรายคน มีเครื่องมือที่ละเอียด แต่อย่างไรก็ตามครูผู้สอนให้ความเห็นว่าเนื่องจากเครื่องมือมีจำนวนมากอาจต้องทำความเข้าใจให้ดีกว่า และนักเรียนมีจำนวนมากครูจึงต้องใช้เวลาในการเรียนนาน และต้องเขียนข้อมูลให้นักเรียนเป็นรายคนจึงต้องใช้เวลาหลังเลิกเรียนและช่วงโหวงว่างเขียน ข้อดีคือเห็นความก้าวหน้าของนักเรียนแต่ละคน แต่ครูต้องให้ใช้เวลาในการสอนในมากขึ้น และต้องมีเวลาพอสมควร

“ในคู่มือมีขั้นตอนการสอน การประเมินดี ละเอียด ครุ่นำไปใช้ได้จริง ยกให้ครุคนอื่นก็น่าจะสอนได้นะ”

ครูผู้สอน 28 กรกฎาคม 2557

“การเขียน feedback ให้นักเรียนในท้องถิ่นที่ทำไมมันมันเยอะนะ แต่พูดให้ feedback ครุพูดทันทีที่เด็กพูดจบ ข้อดีคือคนอื่นจะได้ฟังด้วย”

ครูผู้สอน 28 กรกฎาคม 2557

“การให้ feedback เป็นสิ่งที่ดีนะ นักเรียนเขาก็ได้ลองให้คำแนะนำเพื่อนนะ”

ครูผู้สอน 28 กรกฎาคม 2557

“การมีแบบประเมินให้นักเรียนเขาใช้ครั้งแรกอาจงงบ้าง แต่ครั้งสอง สามก็เป็นกันหมดแล้ว ต่อไปง่ายเลยรู้หน้าที่กัน”

ครูผู้สอน 28 กรกฎาคม 2557

ด้านความเป็นประโยชน์ในการนำไปใช้ ครูให้ความเห็นว่า คู่มือการใช้แฟ้มสะสมงานเป็นคู่มือที่มีเนื้อหาที่ครอบคลุมดี คือมีแบบประเมินครบทุกด้าน พร้อมทั้งในแต่ละขั้นตอนมีวิธีดำเนินการและกิจกรรมทำให้ครูสามารถทำตามได้ และครูสามารถนำไปใช้สอนได้จริง

“พี่ชอบที่มีแบบประเมินครู แบบประเมินนักเรียน มีครบเลย ทำให้รู้ว่าจะเก็บคะแนนอะไรบ้าง วางแผนการสอนได้เลย”

ครูผู้สอน 28 กรกฎาคม 2557

“ที่ว่าใครก็เอาไปใช้ได้ล่ะ ครู ป.1 ก็ใช้ได้ แบบประเมิน แบบบันทึกคะแนนมีให้หมด
ใช้ง่ายด้วย มีการดูน เหมาะกับเด็กดี”

ครูผู้สอน 28 กรกฎาคม 2557

“ถ้าครูห้องอื่นสนใจ เขาก็หยิบเล่มนี้ไปอ่าน ทำความเข้าใจในงาน เกณฑ์ พี่ก็อธิบาย
เขาได้ มีประโยชน์นะ”

ครูผู้สอน 28 กรกฎาคม 2557

“เพราะมีแบบประเมินต่างๆ ให้หมดแล้ว นักเรียนเขาก็เห็นใจว่าครูจะเก็บคะแนน
อะไรเขา เพื่อนจะให้คะแนนเขาอย่างไร เขาก็ประเมินได้จริงนะ”

ครูผู้สอน 28 กรกฎาคม 2557

ตอนที่ 2 ผลการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานสำหรับประเมินทักษะการพูดภาษาอังกฤษ

การศึกษาผลของการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน
ชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ผู้วิจัยทำการศึกษาผลที่เกิดจากการพัฒนาทักษะการพูด
ภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานใน 2 ประเด็น คือ 1) การวิเคราะห์
สถิติพื้นฐานของคะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 2) การศึกษา
ความก้าวหน้าด้านทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน
ซึ่งการศึกษาความก้าวหน้าด้านทักษะการพูดภาษาอังกฤษที่ใช้การประเมินโดยใช้แฟ้มสะสมงาน
เป็นการศึกษาผลของความก้าวหน้าในการพูดใน 4 ภาระงานที่ใช้ในการจัดการเรียนการสอนของครูผู้สอน
คะแนนที่นำมาวิเคราะห์เป็นคะแนนที่ได้จากครูผู้สอนที่เป็นผู้ดำเนินการสอนโดยใช้รูปแบบการประเมิน
การพูดโดยใช้แฟ้มสะสมงานที่ผู้วิจัยพัฒนาขึ้นจากนักเรียนกลุ่มทดลอง จำนวน 34 คน โดยในแต่ละ
ภาระงานทำการประเมินจำนวน 3 ครั้ง และทำการวิเคราะห์ความก้าวหน้าด้านทักษะการพูด
ภาษาอังกฤษของนักเรียนโดยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำชนิดทางเดียว (one-way
repeated ANOVA measurement) ผลการศึกษาความก้าวหน้าของความสามารถด้านทักษะการพูด
ภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 และ 3) การศึกษาการให้ข้อมูลย้อนกลับของครู
(teacher feedback) ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ปรากฏผล ดังนี้

2.1 การวิเคราะห์สถิติพื้นฐานของคะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

การวิเคราะห์ข้อมูลพื้นฐานเพื่อการบรรยายลักษณะของคะแนนความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนที่ได้จากการประเมิน 3 ครั้ง ได้แก่ ค่าเฉลี่ย (mean) ค่ามัธยฐาน (median) ค่าฐานนิยม (mode) ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ค่าสัมประสิทธิ์ความแปรผัน (CV) ความแปรปรวน (variance) ค่าความเบ้ (skewness) ค่าความโด่ (kurtosis) ค่าต่ำสุด (minimum) ค่าสูงสุด (maximum) โดยผู้วิจัยได้แยกการวิเคราะห์ข้อมูลเป็น 3 ชุด ซึ่งพบผลการวิเคราะห์ ดังนี้

ผลการวิเคราะห์สถิติพื้นฐานของคะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนโดยใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ พบว่า ในภาพรวมของทั้ง 4 ภาระงาน นักเรียนมีความสามารถอยู่ในระดับอ่อนในการประเมินครั้งที่ 1 และ 2 ส่วนการประเมินครั้งที่ 3 มีความสามารถอยู่ในระดับพอใช้ทุกภาระงาน โดยพบว่านักเรียนมีความสามารถเพิ่มขึ้นทุกครั้งในทุกภาระงาน โดยพบว่านักเรียนมีคะแนนเฉลี่ยอยู่ระหว่าง 1.52-3.04 จากคะแนนเต็ม 4 คะแนน เก็บรวบรวมข้อมูลจากกลุ่มตัวอย่าง จำนวน 34 คน เมื่อพิจารณาแยกในแต่ละภาระงานพบผลการวิเคราะห์ ดังนี้

2.1.1 ภาระงานการสัมภาษณ์ (interview) มีคะแนนเต็ม 4 คะแนน คะแนนเฉลี่ยครั้งที่ 3 สูงที่สุด ($M=2.61, SD=0.42$) รองลงมาคือ ครั้งที่ 2 ($M=2.00, SD=0.43$) และครั้งที่ 1 ($M=1.52, SD=0.32$) ตามลำดับ การประเมินครั้งที่ 1 2 และ 3 มีคะแนนมัธยฐานเท่ากับ 1.50 2.00 และ 2.50 ตามลำดับ คะแนนสูงสุดพบในการประเมินครั้งที่ 3 คือ 4 คะแนน ส่วนคะแนนต่ำสุดพบในครั้งที่ 1 คือ 1 คะแนน ส่วนคะแนนฐานนิยมในการประเมินเท่ากับ 1.50 2.17 และ 2.33 ตามลำดับ สำหรับความแปรปรวนของคะแนนจากการประเมิน 3 ครั้ง พบว่า การประเมินครั้งที่ 2 มีความแปรปรวนมากที่สุด ($CV (%) = 21.71$) รองลงมาคือครั้งที่ 3 และ 1 ($CV (%) = 21.21$ และ 15.96 ตามลำดับ)

เมื่อพิจารณารูปแบบการโค้งแจกแจงจากความเบ้และความโด่งพบว่า จากการประเมินทั้ง 3 ครั้ง มีลักษณะการแจกแจงคล้ายกันคือเบ้ขวา โดยการประเมินครั้งที่ 1 มีลักษณะการเบ้มากที่สุด รองลงมาคือ ครั้งที่ 3 และ 2 ตามลำดับ ($Sk = 1.97, 1.33$ และ 1.20) สำหรับการโด่งของลักษณะการแจกแจงของคะแนนจากการประเมินทั้ง 3 ครั้ง โดยการประเมินครั้งที่ 1 มีความโด่งมากที่สุด รองลงมาคือ ครั้งที่ 2 และ 3 ($Ku = 7.23, 3.25$ และ 2.56 ตามลำดับ) เมื่อพิจารณาจากค่าความคลาดเคลื่อนของความเบ้และความโด่งมีค่าเท่ากันทั้ง 3 ครั้ง คือเท่ากับ 0.40 และ 0.79 ตามลำดับ ซึ่งไม่เกิน 2.00 หรือไม่น้อยกว่า -2 ยังถือได้ว่ามีลักษณะการแจกแจงเป็นโค้งปกติ รายละเอียดดังตารางที่ 4.16

2.1.2 ภาระงานการนำเสนอปากเปล่า (oral presentation) มีคะแนนเต็ม 4 คะแนน คะแนนเฉลี่ยครั้งที่ 3 สูงที่สุด รองลงมาคือ ครั้งที่ 2 และครั้งที่ 1 มีคะแนนเฉลี่ย คือ 2.61 2.00 และ 1.60 ตามลำดับ (ส่วนเบี่ยงเบนมาตรฐาน คือ 0.41 0.48 และ 0.41 ตามลำดับ) การประเมินครั้งที่ 1, 2 และ 3 มีคะแนนมัธยฐานเท่ากับ 1.50 2.00 และ 2.50 ตามลำดับ คะแนนสูงสุดพบในการประเมินครั้งที่ 3 คือ 4 คะแนน ส่วนคะแนนต่ำสุดพบในครั้งที่ 1 คือ 1.17 คะแนน ส่วนคะแนนฐานนิยมในการประเมินเท่ากับ 1.33 1.67 และ 2.33 ตามลำดับ สำหรับความแปรปรวนของคะแนนจากการประเมิน 3 ครั้ง พบว่า การประเมินครั้งที่ 1 มีความแปรปรวนมากที่สุด ($CV (%) = 25.82$) รองลงมาคือครั้งที่ 2 และ 3 ($CV (%) = 23.85$ และ 15.52 ตามลำดับ)

เมื่อพิจารณารูปแบบการโค้งแจกแจงจากความเบ้และความโด่งพบว่า จากการประเมินทั้ง 3 ครั้ง มีลักษณะการแจกแจงคล้ายกันคือเบ้ขวา โดยการประเมินครั้งที่ 1 มีลักษณะการเบ้มากที่สุด รองลงมาคือ ครั้งที่ 3 และ 2 ตามลำดับ ($Sk = 2.40$ 1.37 และ 0.95) สำหรับการโด่งของลักษณะการแจกแจงของคะแนนจากการประเมินทั้ง 3 ครั้ง โดยการประเมินครั้งที่ 1 มีความโด่งมากที่สุด รองลงมาคือ ครั้งที่ 3 และ 2 ($Ku = 8.56$ 2.85 และ 1.59 ตามลำดับ) เมื่อพิจารณาจากค่าความคลาดเคลื่อนของความเบ้และความโด่งมีค่าเท่ากันทั้ง 3 ครั้ง คือเท่ากับ 0.40 และ 0.79 ตามลำดับ ซึ่งไม่เกิน 2.00 หรือไม่น้อยกว่า -2 ยังถือได้ว่ามีลักษณะการแจกแจงเป็นโค้งปกติ รายละเอียดดังตารางที่ 4.16

2.1.3 ภาระงานการเล่านิทาน (storytelling) มีคะแนนเต็ม 4 คะแนน คะแนนเฉลี่ยครั้งที่ 3 สูงที่สุด ($M=3.04$, $SD=0.22$) รองลงมาคือ ครั้งที่ 2 ($M=2.56$, $SD=0.23$) และครั้งที่ 1 ($M=1.82$, $SD=0.20$) ตามลำดับ การประเมินครั้งที่ 1 2 และ 3 มีคะแนนมัธยฐานเท่ากับ 1.80 2.67 และ 3.00 ตามลำดับ คะแนนสูงสุดพบในการประเมินครั้งที่ 3 คือ 4 คะแนน ส่วนคะแนนต่ำสุดพบในครั้งที่ 1 คือ 1.50 คะแนน ส่วนคะแนนฐานนิยมในการประเมินเท่ากับ 2.00 2.67 และ 3.00 ตามลำดับ สำหรับความแปรปรวนของคะแนนจากการประเมิน 3 ครั้ง พบว่า การประเมินครั้งที่ 1 มีความแปรปรวนมากที่สุด ($CV (%) = 11.01$) รองลงมาคือครั้งที่ 2 และ 3 ($CV (%) = 8.91$ และ 7.17 ตามลำดับ)

เมื่อพิจารณารูปแบบการโค้งแจกแจงจากความเบ้และความโด่ง พบว่า จากการประเมินทั้ง 3 ครั้ง มีลักษณะการแจกแจงคล้ายกันคือเบ้ขวา โดยการประเมินครั้งที่ 3 มีลักษณะการเบ้มากที่สุด รองลงมาคือ ครั้งที่ 1 และ 2 ตามลำดับ ($Sk=0.68$ 0.23 และ 0.07) สำหรับการโด่งของลักษณะการแจกแจงของคะแนนพบว่า ครั้งที่ 3 มีความโด่งของลักษณะการแจกแจงปกติ ($Ku= 1.26$) ส่วนการประเมินทั้ง 1 และ 2 มีลักษณะแบนกว่าโค้งปกติ โดยการประเมินครั้งที่ 2 มีความแบนมากกว่าครั้งที่ 1 ($Ku=-0.43$ และ -0.15 ตามลำดับ) เมื่อพิจารณาจากค่าความคลาดเคลื่อนของความเบ้และ

ความโด่งมีค่าเท่ากับทั้ง 3 ครั้ง คือเท่ากับ 0.40 และ 0.79 ตามลำดับ ซึ่งไม่เกิน 2.00 หรือไม่น้อยกว่า -2 ยังถือได้ว่ามีลักษณะการแจกแจงเป็นโค้งปกติ รายละเอียดดังตารางที่ 4.16

2.1.4 ภาระงานการบรรยายภาพ (picture description) มีคะแนนเต็ม 4 คะแนน คะแนนเฉลี่ยครั้งที่ 3 สูงที่สุด ($M=2.88, SD=0.29$) รองลงมาคือ ครั้งที่ 2 ($M=2.40, SD=0.29$) และครั้งที่ 1 ($M=1.76, SD=0.35$) ตามลำดับ การประเมินครั้งที่ 1 2 และ 3 มีคะแนนมัธยฐาน เท่ากับ 1.67 2.42 และ 3.00 ตามลำดับ คะแนนสูงสุดพบในการประเมินครั้งที่ 3 คือ 4 คะแนน ส่วนคะแนนต่ำสุดพบในครั้งที่ 1 คือ 1.17 คะแนน ส่วนคะแนนฐานนิยมในการประเมินเท่ากับ 1.50 2.50 และ 3.00 ตามลำดับ สำหรับความแปรปรวนของคะแนนจากการประเมิน 3 ครั้ง พบว่า การประเมินครั้งที่ 1 มีความแปรปรวนมากที่สุด ($CV (%) = 19.67$) รองลงมาคือครั้งที่ 3 และ 2 ($CV (%) = 10.11$ และ 9.98 ตามลำดับ)

เมื่อพิจารณารูปแบบการโค้งแจกแจงจากความเบ้และความโด่งพบว่า จากการประเมินครั้งที่ 2 และ 3 มีลักษณะการแจกแจงคล้ายกันคือเบ้ซ้าย ส่วนการประเมินครั้งที่ 1 มีลักษณะการเบ้ขวา โดยพบว่า การประเมินครั้งที่ 1 มีความเบ้มากที่สุด รองลงมาคือครั้งที่ 3 และ 2 ($Sk = 1.57 -0.21$ และ -1.07 ตามลำดับ) สำหรับการโด่งของลักษณะการแจกแจงของคะแนนพบว่า ครั้งที่ 1 มีความโด่งของลักษณะการแจกแจงปกติ ($Ku=0.24$) ส่วนการประเมินทั้ง 2 และ 3 มีลักษณะแบนกว่า โค้งปกติ โดยการประเมินครั้งที่ 3 มีความแบนมากกว่าครั้งที่ 2 ($Ku = -0.47$ และ -0.21 ตามลำดับ) เมื่อพิจารณาจากค่าความคลาดเคลื่อนของความเบ้และความโด่งมีค่าเท่ากับทั้ง 3 ครั้ง คือเท่ากับ 0.40 และ 0.79 ตามลำดับ ซึ่งไม่เกิน 2.00 หรือไม่น้อยกว่า -2 ยังถือได้ว่ามีลักษณะการแจกแจงเป็นโค้งปกติ รายละเอียดดังตารางที่ 4.16

ตารางที่ 4.16 การวิเคราะห์สถิติพื้นฐานและการแปลผลของคะแนน ความสามารถด้านทักษะ

การพูด 4 ภาระงาน

	การสัมภาษณ์			การนำเสนอปากเปล่า			การเล่านิทาน			การบรรยายภาพ		
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
N	34	34	34	34	34	34	34	34	34	34	34	34
Mean	1.52	2.00	2.61	1.60	2.00	2.61	1.82	2.56	3.04	1.76	2.40	2.88
median	1.50	2.00	2.50	1.50	2.00	2.50	1.83	2.67	3.00	1.67	2.42	3.00
Mode	1.50	2.17	2.33	1.33	1.67 ^a	2.33 ^a	2.00	2.67	3.00	1.50	2.50	3.00
SD	0.32	0.43	0.42	0.41	0.48	0.41	0.20	0.23	0.22	0.35	0.24	0.29
CV (%)	21.21	21.71	15.96	25.82	23.85	15.52	11.01	8.91	7.17	19.67	9.98	10.11
Var.	0.10	0.19	0.17	0.17	0.23	0.16	0.04	0.05	0.05	0.12	0.06	0.08
Sk	1.97	1.20	1.33	2.40	0.95	1.37	0.23	0.07	0.68	0.57	-0.07	-0.21
SE(Sk)	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
Ku	7.23	3.25	2.56	8.56	1.59	2.85	-0.15	-0.43	1.26	0.24	-0.21	-0.47
SE(Ku)	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79
Min.	1.00	1.33	2.00	1.17	1.33	2.00	1.50	2.17	2.50	1.17	1.83	2.33
Max.	2.83	3.50	4.00	3.33	3.50	4.00	2.33	3.00	3.50	2.67	2.83	3.50

2.2 การศึกษาความก้าวหน้าด้านทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

2.2.1 ผลการทดสอบเงื่อนไข sphericity ของค่าเฉลี่ยของคะแนนความสามารถในทักษะการพูดภาษาอังกฤษที่ได้จากการประเมินระหว่างเวลา

ก่อนทำการทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนด้วยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำชนิดทางเดียว ผู้วิจัยได้ทำการตรวจสอบข้อตกลงเบื้องต้นเรื่อง compound symmetry ของคะแนนทักษะการพูดใน 4 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ ที่มีการประเมินซ้ำ 3 ครั้งก่อน โดยวิธีการทดสอบ sphericity และหากมีการละเมิดข้อตกลงเบื้องต้นนี้จะมีการปรับค่าด้วยวิธีการ greenhouse-geisser ต่อไป

จากการทดสอบ sphericity ของคะแนนเฉลี่ยจากภาระงานทั้ง 4 ภาระงาน โดยการวิเคราะห์ค่าสถิติ Mauchly พบว่าคะแนนการวัดซ้ำของภาระงานการสัมภาษณ์ การนำเสนอปากเปล่า และการเล่านิทาน มีค่าสถิติไคสแควร์ คือ 3.022 1.781 และ 4.029 ตามลำดับ และค่าสถิติ Mauchly ไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าคะแนนเฉลี่ยของการวัดซ้ำของทั้ง 3 ภาระงาน

เป็นไปตามข้อตกลงเบื้องต้นคือมีลักษณะความแปรปรวนเป็น compound symmetry ส่วนภาระงานการบรรยายภาพพบว่า มีค่าสถิติไคสแควร์ คือ 23.896 และค่าสถิติ Mauchly มีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่าคะแนนเฉลี่ยของการวัดซ้ำของภาระงานการบรรยายภาพมีลักษณะความแปรปรวนไม่เป็น compound symmetry ซึ่งเป็นการละเมิดข้อตกลงเบื้องต้นของการวิเคราะห์ sphericity ดังนั้นในการวิเคราะห์จึงมีการปรับค่าด้วยวิธีการ Greenhouse-Geisser รายละเอียดดังตารางที่ 4.17

ตารางที่ 4.17 ผลการทดสอบ sphericity ของคะแนนความสามารถในทักษะการพูดภาษาอังกฤษ จำแนกตามภาระงาน

ภาระงาน	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
การสัมภาษณ์	.910	3.022	2	.221	.917	.969	.500
การนำเสนอปากเปล่า	.946	1.781	2	.410	.949	1.000	.500
การเล่านิทาน	.882	4.029	2	.133	.894	.942	.500
การบรรยายภาพ	.474	23.896	2	.000	.655	.671	.500

2.2.2 การเปรียบเทียบคะแนนเฉลี่ยความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียน

2.2.2.1 การเปรียบเทียบคะแนนเฉลี่ยของภาระงานการสัมภาษณ์

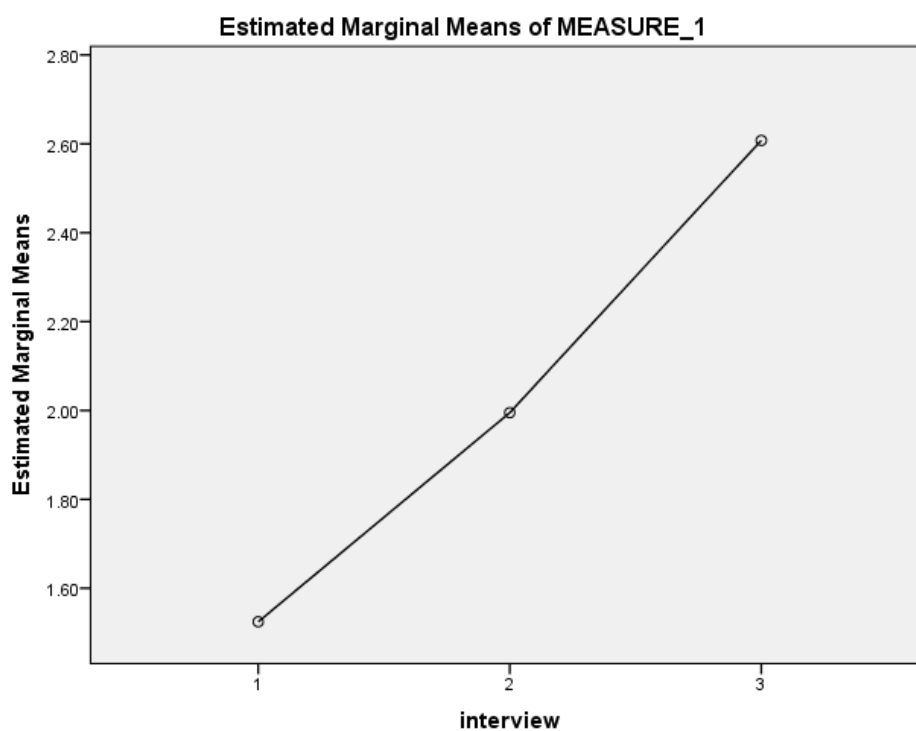
ผลการเปรียบเทียบคะแนนเฉลี่ยจากการประเมิน 3 ครั้ง ในภาระงานการสัมภาษณ์ โดยรวมของนักเรียนกลุ่มตัวอย่าง พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F=480.345$ $p=.000$) แสดงว่านักเรียนมีคะแนนความสามารถในทักษะการพูดภาษาอังกฤษในภาระงานการสัมภาษณ์แตกต่างกันอย่างน้อย 1 ช่วงเวลาในการประเมิน ดังตารางที่ 4.18 ดังนั้นผู้วิจัยจึงทำการวิเคราะห์คะแนนเฉลี่ยรายคู่ต่อไปด้วยวิธีของ Bonferroni ผลการเปรียบเทียบคะแนนเฉลี่ยภาระงานการสัมภาษณ์ของนักเรียนในการประเมิน 3 ครั้ง พบว่า คะแนนเฉลี่ยในการประเมินครั้งที่ 2 แตกต่างจากการประเมินครั้งที่ 1 และ ครั้งที่ 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยพบว่าคะแนนเฉลี่ยครั้งที่ 3 สูงกว่าคะแนนเฉลี่ยครั้งที่ 1 และคะแนนเฉลี่ยครั้งที่ 2 และคะแนนเฉลี่ยครั้งที่ 2 สูงกว่าคะแนนเฉลี่ยครั้งที่ 1 และพบว่า คะแนนเฉลี่ยครั้งที่ 3 เพิ่มขึ้นจากครั้งที่ 2 มากกว่า คะแนนเฉลี่ยครั้งที่ 2 เพิ่มขึ้นจากครั้งที่ 1 แสดงว่าเพิ่มสะสมงานการประเมินทักษะการพูดส่งผลต่อความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนให้สูงขึ้นในภาระงานการสัมภาษณ์ดังตารางที่ 4.19 และภาพที่ 4.1

ตารางที่ 4.18 การเปรียบเทียบคะแนนเฉลี่ยจากการสัมภาษณ์โดยรวม

แหล่งความแปรปรวน	SS	df	MS	F	p
การสัมภาษณ์	20.066	2	10.033	480.345	.000
ความคลาดเคลื่อน	1.379	66	.021		

ตารางที่ 4.19 การเปรียบเทียบคะแนนเฉลี่ยการสัมภาษณ์รายคู่ ด้วยวิธี Bonferroni

Time	การสัมภาษณ์		
	Mean difference	SE	p
ครั้งที่ 1 กับ ครั้งที่ 2	-0.471	0.033	0.000
ครั้งที่ 1 กับ ครั้งที่ 3	-1.083	0.04	0.000
ครั้งที่ 2 กับ ครั้งที่ 3	-0.613	0.032	0.000



ภาพที่ 4.1 คะแนนเฉลี่ยการสัมภาษณ์ในการประเมิน 3 ครั้ง

2.2.2.2 การเปรียบเทียบคะแนนเฉลี่ยของภาระงานการนำเสนอปากเปล่า

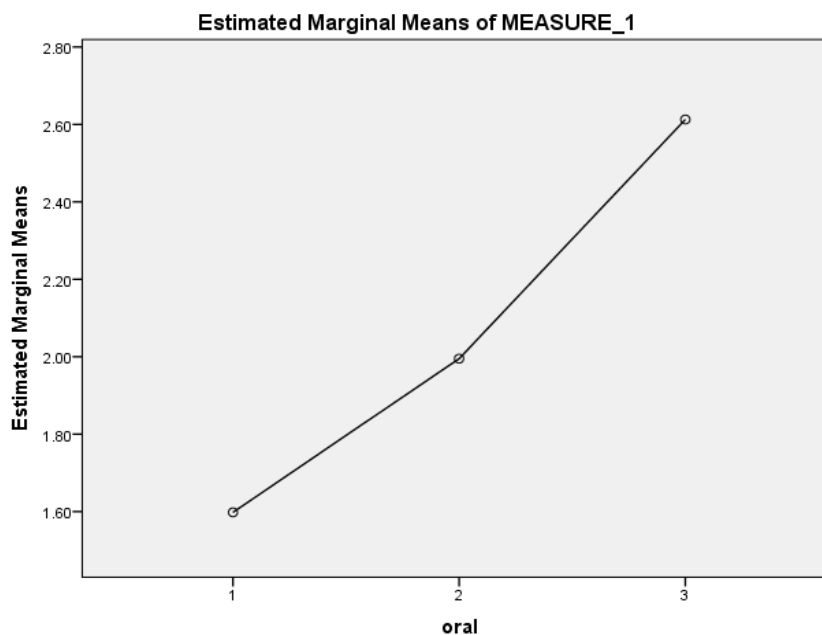
ผลการเปรียบเทียบคะแนนเฉลี่ยจากการประเมิน 3 ครั้ง ในภาระงานการนำเสนอปากเปล่าโดยรวมของนักเรียนกลุ่มตัวอย่าง พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F=278.136$ $p=.000$) แสดงว่านักเรียนมีคะแนนความสามารถในทักษะการพูดภาษาอังกฤษในภาระงานการสัมภาษณ์แตกต่างกันอย่างน้อย 1 ช่วงเวลาในการประเมิน ดังตารางที่ 4.20 ดังนั้นผู้วิจัยจึงทำการวิเคราะห์คะแนนเฉลี่ยรายคู่ต่อไปด้วยวิธีของ Bonferroni ผลการเปรียบเทียบคะแนนเฉลี่ยภาระงานการนำเสนอปากเปล่าของนักเรียนในการประเมิน 3 ครั้ง พบว่า คะแนนเฉลี่ยในการประเมินครั้งที่ 2 แตกต่างจากการประเมินครั้งที่ 1 และ ครั้งที่ 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยพบว่า คะแนนเฉลี่ยครั้งที่ 3 สูงกว่าคะแนนเฉลี่ยครั้งที่ 1 และคะแนนเฉลี่ยครั้งที่ 2 และคะแนนเฉลี่ยครั้งที่ 2 สูงกว่าคะแนนเฉลี่ยครั้งที่ 1 และพบว่า คะแนนเฉลี่ยครั้งที่ 3 เพิ่มขึ้นจากครั้งที่ 2 มากกว่า คะแนนเฉลี่ยครั้งที่ 2 เพิ่มขึ้นจากครั้งที่ 1 แสดงว่าแพ้มสละสมงานการประเมินทักษะการพูดส่งผลต่อความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนให้สูงขึ้นในภาระงานการนำเสนอปากเปล่าดังตารางที่ 4.21 และภาพที่ 4.2

ตารางที่ 4.20 การเปรียบเทียบคะแนนเฉลี่ยจากการนำเสนอปากเปล่าโดยรวม

แหล่งความแปรปรวน	Sum of Squares	df	Mean square	F	p
การนำเสนอปากเปล่า	17.779	2	8.890	278.136	.000
ความคลาดเคลื่อน	2.109	66	.032		

ตารางที่ 4.21 การเปรียบเทียบคะแนนเฉลี่ยการนำเสนอปากเปล่ารายคู่ วิธี Bonferroni

Time	กานำเสนอปากเปล่า		
	Mean difference	SE	p
ครั้งที่ 1 กับ ครั้งที่ 2	-0.397	0.047	0.000
ครั้งที่ 1 กับ ครั้งที่ 3	-1.015	0.038	0.000
ครั้งที่ 2 กับ ครั้งที่ 3	-0.618	0.045	0.000



ภาพที่ 4.2 คะแนนเฉลี่ยการนำเสนอปากเปล่าในการประเมิน 3 ครั้ง

2.2.2.3 การเปรียบเทียบคะแนนเฉลี่ยการเล่านิทาน

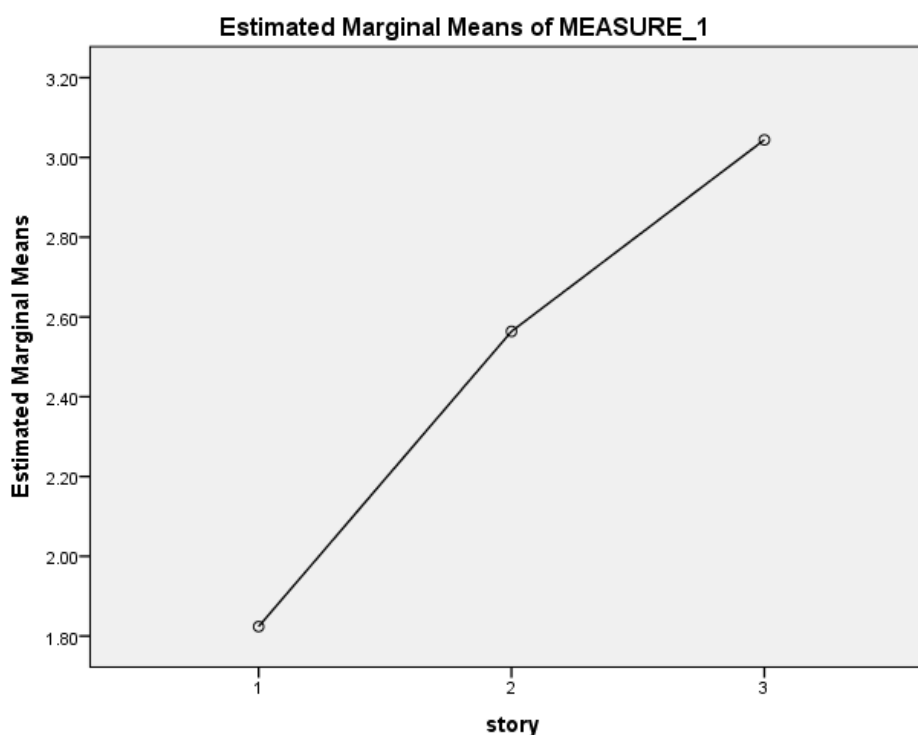
ผลการเปรียบเทียบคะแนนเฉลี่ยจากการประเมิน 3 ครั้ง ในภาระงานการเล่านิทาน โดยรวมของนักเรียนกลุ่มตัวอย่าง พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F=413.951$ $p=.000$) แสดงว่านักเรียนมีคะแนนความสามารถในทักษะการพูดภาษาอังกฤษในภาระงานการนำเสนอปากเปล่าแตกต่างกันอย่างน้อย 1 ช่วงเวลาในการประเมิน ดังตารางที่ 4.22 ดังนั้นผู้วิจัยจึงทำการวิเคราะห์คะแนนเฉลี่ยรายคู่ต่อไปด้วยวิธีของ Bonferroni ผลการเปรียบเทียบคะแนนเฉลี่ยภาระงานการเล่านิทานของนักเรียนในการประเมิน 3 ครั้ง พบว่า คะแนนเฉลี่ยในการประเมินครั้งที่ 2 แตกต่างจากการประเมินครั้งที่ 1 และ ครั้งที่ 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยพบว่า คะแนนเฉลี่ยครั้งที่ 3 สูงกว่าคะแนนเฉลี่ยครั้งที่ 1 และคะแนนเฉลี่ยครั้งที่ 2 และคะแนนเฉลี่ยครั้งที่ 2 สูงกว่าคะแนนเฉลี่ยครั้งที่ 1 และพบว่า คะแนนเฉลี่ยครั้งที่ 2 เพิ่มขึ้นจากครั้งที่ 1 มากกว่า คะแนนเฉลี่ยครั้งที่ 3 เพิ่มขึ้นจากครั้งที่ 2 แสดงว่าแพ้มะสมงานการประเมินทักษะการพูดส่งผลต่อความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนให้สูงขึ้นในภาระงานการเล่านิทาน ดังตารางที่ 4.23 และภาพที่ 4.3

ตารางที่ 4.22 การเปรียบเทียบคะแนนเฉลี่ยจากการเล่านิทานโดยรวม

แหล่งความแปรปรวน	Sum of Squares	df	Mean square	F	p
การเล่านิทาน	25.710	2	12.855	413.951	.000
ความคลาดเคลื่อน	2.050	66	.031		

ตารางที่ 4.23 การเปรียบเทียบคะแนนเฉลี่ยการเล่านิทานรายคู่ ด้วยวิธี Bonferroni

Time	การเล่านิทาน		
	Mean difference	SE	p
ครั้งที่ 1 กับ ครั้งที่ 2	-0.74	0.048	0.000
ครั้งที่ 1 กับ ครั้งที่ 3	-1.221	0.044	0.000
ครั้งที่ 2 กับ ครั้งที่ 3	-0.48	0.035	0.000



ภาพที่ 4.3 คะแนนเฉลี่ยการเล่านิทานในการประเมิน 3 ครั้ง

2.2.2.4 การเปรียบเทียบคะแนนเฉลี่ยการบรรยายภาพ

ผลการเปรียบเทียบคะแนนเฉลี่ยจากการประเมิน 3 ครั้ง ในภาระงานการบรรยายภาพ โดยรวมของนักเรียนกลุ่มตัวอย่าง พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($F=212.030$ $p=.000$) แสดงว่านักเรียนมีคะแนนความสามารถในทักษะการพูดภาษาอังกฤษในการงานการสัมภาษณ์แตกต่างกันอย่างน้อย 1 ช่วงเวลาในการประเมิน ดังตารางที่ 4.24 ดังนั้นผู้วิจัยจึงทำการวิเคราะห์คะแนนเฉลี่ยรายคู่ต่อไปด้วยวิธีของ Bonferroni ผลการเปรียบเทียบคะแนนเฉลี่ยภาระงานการบรรยายภาพของนักเรียนในการประเมิน 3 ครั้ง พบว่า คะแนนเฉลี่ยในการประเมินครั้งที่ 2

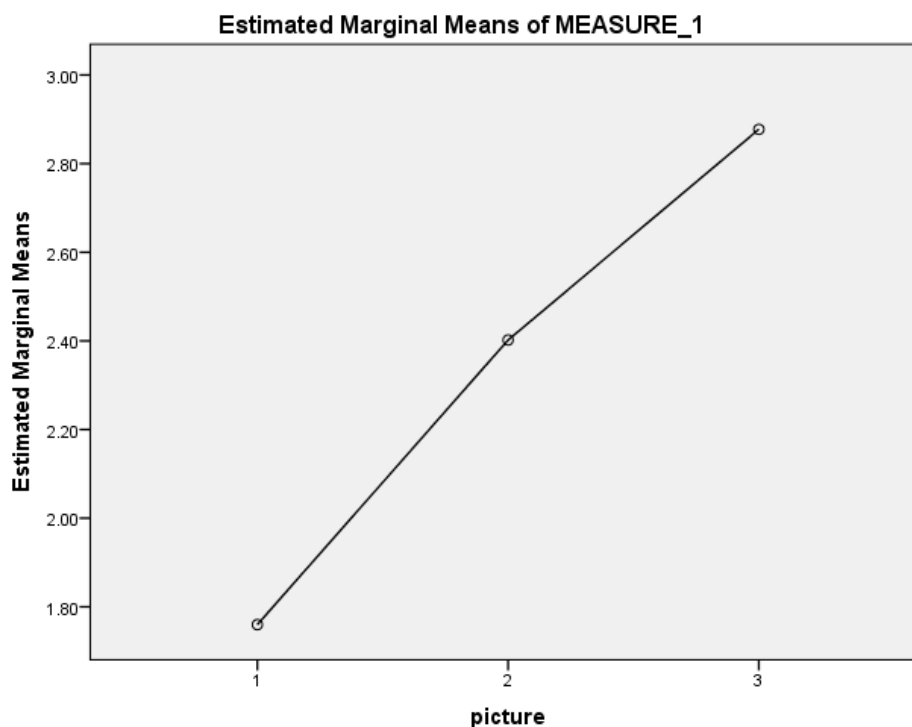
แตกต่างจากการประเมินครั้งที่ 1 และ ครั้งที่ 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยพบว่า คะแนนเฉลี่ยครั้งที่ 3 สูงกว่าคะแนนเฉลี่ยครั้งที่ 1 และคะแนนเฉลี่ยครั้งที่ 2 และคะแนนเฉลี่ยครั้งที่ 2 สูงกว่าคะแนนเฉลี่ยครั้งที่ 1 และพบว่า คะแนนเฉลี่ยครั้งที่ 2 เพิ่มขึ้นจากครั้งที่ 1 มากกว่า คะแนนเฉลี่ยครั้งที่ 3 เพิ่มขึ้นจากครั้งที่ 2 แสดงว่าเพิ่มสะสมงานการประเมินทักษะการพูดส่งผลต่อความสามารถ ในทักษะการพูดภาษาอังกฤษของนักเรียนให้สูงขึ้นในภาระงานการบรรยายภาพดังตารางที่ 4.25 และ ภาพที่ 4.4

ตารางที่ 4.24 การเปรียบเทียบคะแนนเฉลี่ยการบรรยายภาพในภาพรวม

แหล่งความแปรปรวน	Sum of Squares	df	Mean square	F	p
การบรรยายภาพ	21.393	1.311	16.324	212.030	.000
ความคลาดเคลื่อน	3.330	43.248	.077		

ตารางที่ 4.25 การเปรียบเทียบคะแนนเฉลี่ยการบรรยายภาพรายคู่ ด้วยวิธี Bonferroni

Time	การบรรยายภาพ		
	Mean difference	SE	p
ครั้งที่ 1 กับ ครั้งที่ 2	-0.642	0.048	0.000
ครั้งที่ 1 กับ ครั้งที่ 3	-1.118	0.071	0.000
ครั้งที่ 2 กับ ครั้งที่ 3	-0.475	0.039	0.000



ภาพที่ 4.4 คะแนนเฉลี่ยการบรรยายภาพในการประเมิน 3 ครั้ง

2.3 การให้ข้อมูลย้อนกลับของครูในการประเมินทักษะการพูดภาษาอังกฤษโดยใช้แฟ้มสะสมงาน การให้ข้อมูลย้อนกลับ (feedback) ในการวิจัยครั้งนี้เป็นให้ข้อมูลโดยครูผู้สอนในการประเมินทักษะการพูดภาษาอังกฤษโดยใช้แฟ้มสะสมงาน ซึ่งเป็นให้ข้อมูลย้อนกลับในระหว่างการเรียนรู้ การสอนที่มีการประเมินผู้เรียนด้วยแฟ้มสะสมงาน การให้ข้อมูลย้อนกลับสำหรับการวิจัยครั้งนี้ ผู้วิจัยได้นำแนวความคิดการให้ข้อมูลย้อนกลับแบบเชิงบรรยายที่ให้ข้อมูลย้อนกลับแบบการระบุความสำเร็จ (achievement feedback) และการระบุสิ่งที่ควรปรับปรุง (improvement feedback) และแก้ไขให้ถูกต้อง (corrective feedback) เพื่อเป็นการเสริมแรงในการเรียน รวมทั้งรู้เกี่ยวกับสิ่งที่จะนำตนเองไปสู่ความสำเร็จในการเรียน และรู้ถึงข้อผิดพลาดและแก้ไขให้ถูกต้อง (correction of errors) ผลการให้ข้อมูลย้อนกลับของครูพบว่า ครูมีการให้ข้อมูลย้อนกลับแก่นักเรียนทั้ง 2 แบบ โดยให้ข้อมูลย้อนกลับทันทีที่นักเรียนพูดจบ เพื่อให้นักเรียนทราบข้อดีและข้อที่ควรแก้ไขของตนเอง และนักเรียนคนอื่นๆ ในห้องเรียนได้เกิดการเรียนรู้ไปพร้อมกัน และมีการให้ข้อมูลย้อนกลับด้วยการเขียน เพื่อให้ นักเรียนได้เก็บไว้เป็นเอกสารหลักฐานและนำไปปรับปรุงแก้ไขในครั้งต่อไป ซึ่งผลของการให้ข้อมูลย้อนกลับของครู ทั้ง 6 องค์ประกอบ ได้แก่ ด้านคำศัพท์ ด้านการสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุเป้าหมาย และความคล่องแคล่ว ตามเกณฑ์การให้คะแนนแบบแยกส่วน พบผลการให้ข้อมูลย้อนกลับ ดังนี้

ผลการให้ข้อมูลย้อนกลับของครูด้านคำศัพท์ พบว่า ในการพูดครั้งที่ 1 นักเรียนส่วนใหญ่ มีปัญหาเกี่ยวกับการใช้คำศัพท์ไม่หลากหลาย ส่วนสิ่งที่ทำได้ดีแล้วคือสามารถใช้คำศัพท์ได้ถูกต้อง เหมาะสม ครูได้ให้ข้อเสนอแนะแก่นักเรียนเพื่อนำไปปรับแก้ในการพูดครั้งต่อไป

“พูดได้ดีค่ะ มีการเลือกใช้คำศัพท์ที่ถูกต้อง มีความเหมาะสมกับประโยค แต่ครูว่า ถ้าใช้คำใหม่ๆ แทนคำว่า like ก็จะดีขึ้นไปอีกเพราะนักเรียนใช้ I like..., I like ... ทุกประโยคมัน ไม่น่าสนใจ นักเรียนรู้ใหม่ว่ามีคำไหนอีก... คำว่า love มีความหมายเดียวกันค่ะ จะทำให้คำศัพท์ มีความหลากหลายขึ้นอีก ค่ะแนนก็จะเพิ่มขึ้นอีก”

ครูผู้สอน 5 มิถุนายน 2557

“นักเรียนรู้จักหาคำศัพท์จากพจนานุกรมเพิ่มขึ้น บางคนถามครู บางคน google เก่งมาก แต่ครูยังเห็นบางคนใช้ I ขึ้นต้นประโยคทุกประโยค ลองเปลี่ยนเป็น My ค่ะประโยคเรา ก็มีความหลากหลายขึ้นอีก ค่ะแนนก็จะเพิ่มขึ้นอีก”

ครูผู้สอน 5 มิถุนายน 2557

ผลการให้ข้อมูลย้อนกลับของครูด้านการสร้างประโยค พบว่า ครูให้ข้อมูลย้อนกลับ ในด้านที่ทำได้ดีอยู่แล้วคือ นักเรียนสามารถสร้างประโยคง่ายๆ ได้ คือใช้ present simple และ ในส่วนของข้อที่ยังต้องแก้ไขคือ นักเรียนพูดเป็นคำๆ นักเรียนยังไม่เข้าใจโครงสร้างประโยค เช่น กริยาไม่เหมาะสมกับประธานของประโยค ซึ่งเมื่อพูดจบครูจะแก้ไขประโยคให้ถูกต้องตามหลัก ไวยากรณ์ เพื่อนำไปปรับแก้ในการพูดครั้งต่อไป

“นักเรียนใช้โครงสร้างประโยคเป็นแล้ว แต่นักเรียนจะต้องเลือกกริยาให้เหมาะสมกับ ประโยค เมื่อไม่มั่นใจให้ถามครูก่อน เพราะคำว่า am นักเรียนจะต้องใช้กับ I เป็นต้น...You ใช้ กับ am ไม่ได้ ต้องเป็น ”

ครูผู้สอน 5 มิถุนายน 2557

“นักเรียนเก่งมากที่นำประโยคตัวอย่างที่ให้มาปรับและเปลี่ยนสถานการณ์ได้ แต่นักเรียนต้องดูด้วยว่าประธานเปลี่ยนไปก็เปลี่ยน verb ด้วย เช่น She จะ am ไม่ได้ค่ะ ต้อง is ”

ครูผู้สอน 23 มิถุนายน 2557

ผลการให้ข้อมูลย้อนกลับของครูด้านการเชื่อมโยงเนื้อหา พบว่า ครูให้ข้อมูลย้อนกลับในด้านที่ทำได้ดีอยู่แล้วคือ นักเรียนสามารถใช้คำเชื่อมง่ายๆ มาเชื่อมประโยคได้ โดยรู้จักใช้ and ได้ และในส่วนของข้อที่ยังต้องแก้ไขคือในบางครั้งนักเรียนไม่มีคำเชื่อมประโยคซึ่งครูได้ให้ข้อเสนอแนะเพื่อนำไปปรับแก้ในการพูดครั้งต่อไป

“นักเรียนรู้ว่าเมื่อจะพูดประโยคต่อไปจะใช้ and อันนี้ดีมาก แต่นอกจาก and ใช้คำอื่นได้อีก เช่น but หรือ or ลองใช้นะคะ ลองพูดมาเดี่ยวครูแก้ไขให้เอง”

ครูผู้สอน 10 มิถุนายน 2557

ผลการให้ข้อมูลย้อนกลับของครูด้านการออกเสียง พบว่า ครูให้ข้อมูลย้อนกลับในด้านที่ทำได้ดีอยู่แล้วคือ นักเรียนสามารถ stress ได้ถูกต้อง ในคำสั้นๆ 2 พยางค์ การแบ่งวรรคตอนมีความเหมาะสมดี เสียงสูงต่ำบางคนทำได้ดี เพราะไม่อาย มีความมั่นใจ และในส่วนของข้อที่ยังต้องแก้ไขคือเสียงทำยังไม่ดีเท่าที่ควรต้องได้รับการฝึก บางคนไม่มีโทนเสียงสูงต่ำ เพราะไม่มั่นใจยังเขินอาย ซึ่งครูได้ให้ข้อเสนอแนะเพื่อนำไปปรับแก้ในการพูดครั้งต่อไป

“เด็กชาย...ออกเสียงท้าย K ได้ชัดเจนมาก นักเรียนดูเป็นตัวอย่างนะคะ นอกจากนี้เวลานั้ฐพลถามถึงเกณฑ์เขาจะมีเสียงสูงท้ายๆ ประโยค เก่งมาก”

ครูผู้สอน 10 มิถุนายน 2557

จุฬาลงกรณ์มหาวิทยาลัย

“H กับ S ต่างกันนะคะ H ต้องลากยาวๆ...S ต้อง...นักเรียนฟังพร้อมกัน ออกเสียงตามคะ หนหน้าครูจะรอฟัง ค่ะแนนตรงนี้ถ้าออกเสียงถูกจะเพิ่มแนนอนคะ”

ครูผู้สอน 10 มิถุนายน 2557

“good morning กับ good สั้นๆ ไม่ใช่ลากยาว หนหน้าครูจะรอฟังคะ ไปฝึกนะคะ ในกลุ่ม ถ้าออก good สั้นลงคะแนนจะเพิ่มแนนอนคะ”

ครูผู้สอน 10 มิถุนายน 2557

ผลการให้ข้อมูลย้อนกลับของครูด้านการบรรล่วัตถุประสงค์ พบว่า ครูให้ข้อมูลย้อนกลับในด้านที่ทำได้ดีอยู่แล้วคือ นักเรียนสามารถตอบคำถามได้ถูกต้อง ซึ่งคำถามในการสัมภาษณ์เป็นคำถามสั้นๆ ที่นักเรียนเคยเรียน และคุ้นเคย พูดเรื่องเกี่ยวกับตนเองได้ การเล่านิทานยังพูดเร็วไปเพราะความตื่นเต้น และการบรรยายภาพพบว่านักเรียนบรรยายได้ตรงกับภาพ และในส่วนของข้อที่

ยังต้องแก้ไขคือ เมื่อตื่นตื่น ฟังคำถามไม่ออกนักเรียนไม่รู้จักถามครูหรือขอฟังคำถามใหม่ใช้การพูดครั้งแรก ตอบไม่ตรงคำถามเป็นต้น เพราะไม่มั่นใจ ซึ่งครูได้ให้ข้อเสนอแนะเพื่อนำไปปรับแก้ในการพูดครั้งต่อไป

“นักเรียนตอบถูกแสดงว่าฟังรู้เรื่องว่าครูถามอะไร เก่งค่ะ แต่ที่ครูถามเราเคยเรียนแล้วทั้งนั้น คุณเคยแล้ว ถ้าครูถามเราตอบไม่ได้ ต้องค้นนะค่ะ ถามครู ถามเพื่อน แล้วหน้าหน้าต้องตอบถูกแน่นอน”

ครูผู้สอน 10 มิถุนายน 2557

นักเรียนบอกว่าภาพมีอะไรบ้าง เช่น ลูกโป่งสีอะไร มีเค้ก แล้วเพื่อนๆ ในห้องตั้งใจฟังหรือเปล่า งานอะไรคะ พอรู้ใหม่เพื่อนเขาเล่าอะไร ถ้าเพื่อนเข้าใจแสดงว่านักเรียนบรรลุเป้าหมายในการบรรยายภาพแล้วนะค่ะ

ครูผู้สอน 23 มิถุนายน 2558

ผลการให้ข้อมูลย้อนกลับของครูด้านความคล่องแคล่ว พบว่า ครูให้ข้อมูลย้อนกลับในด้านที่ทำได้ดีอยู่แล้วคือ นักเรียนที่มั่นใจในตนเองสูงจะพูดได้อย่างต่อเนื่อง สั้นไหลลื่น แต่นักเรียนบางคนยังตื่นเต้น จำคำศัพท์ไม่ได้ ใช้คำไม่ถูกต้องตอบไม่ตรงคำถาม เพราะตื่นเต้น พูดติดขัด ซึ่งครูได้ให้กำลังใจและให้ข้อเสนอแนะเพื่อนำไปปรับแก้ในการพูดครั้งต่อไป

“มีความมั่นใจ พูดได้น่าฟัง ไม่หยุดนานเกินไป ถึงแม้จะยังไม่มีความหลากหลายของคำศัพท์ ประโยคได้น้อยแต่น่าฟัง ครูว่าถ้าเราลองแต่งประโยคเพิ่มอีกในครั้งหน้าจะดีมาก ๆ เอาสัก 3 ประโยคนะค่ะ”

ครูผู้สอน 10 มิถุนายน 2557

ในภาพรวมดีนะค่ะ พูดต่อเนื่องดี ถึงนิดหน่อยแต่ไม่นาน มีความมั่นใจในตัวเอง แสดงว่าเตรียมตัวมาดี เก่งมาก แต่ครูว่าถ้าฝึกพูดอีกพูดให้ต่อเนื่อง ไม่หยุดตอนยังไม่จบประโยคจะดีมาก ๆ ลืมหรืออะไรคะ เอาใหม่นะหน้าหน้า

ครูผู้สอน 10 มิถุนายน 2557

โดยรวมดีแล้วค่ะ พูดได้น่าฟัง สั้นไหลลื่นค่ะ แต่ครูว่ายังไม่ค่อยมั่นใจในตัวเอง แล้วถ้าไม่ห้องๆ จะดีมาก ลองพูดแบบธรรมชาติเอาที่เราเข้าใจแล้วมาพูดจะน่าสนใจกว่านี้

ครูผู้สอน 23 มิถุนายน 2557

ในส่วนการประเมินตนเองและประเมินเพื่อนของนักเรียนพบว่า เมื่อนักเรียนได้ฟังเพื่อนพูด และฟังครูให้ข้อมูลย้อนกลับแก่เพื่อน ทำให้นักเรียนได้ฝึกการฟังและเป็นแรงบันดาลใจในการพูดของตนเอง เมื่อฟังเพื่อนและต้องประเมินทำให้ต้องตั้งใจฟังมีความสนใจการเรียนมากขึ้น มีความรับผิดชอบต่อบทบาทผู้ประเมินของตนเอง และทำให้มีความพยายามในการพัฒนาตนเองขึ้นไปเนื่องจากได้ตัวอย่างการพูดของเพื่อนคนอื่นๆ ในห้องเรียน

“ตอนแรกๆ หนูก็อาย ไม่ค่อยกล้าพูด ไม่รู้จะพูดอะไรด้วย เจนเพื่อน จำไม่ได้ กลัวผิด ตอนนี้ก็อายนะแต่น้อยลง เพื่อนกัน ก็อยากพูดได้เหมือนเพื่อน”

นักเรียน 21 กรกฎาคม 2557

“หนูไม่ค่อยชอบบงกชพลนะเมื่อก่อน คือดูเขามั่นใจเกินไป ชอบเสียงดัง แต่เขาก็พูดดีค่ะ เขามั่นใจ เขาออกเสียงดีครูชมตลอด หนูก็อยากพูดได้เหมือนบงกชพลค่ะ”

นักเรียน 21 กรกฎาคม 2557

“การให้เราต้องประเมินเพื่อน คือ เราต้องฟัง เพื่อนผิดเราก็บอก บางทีเขาตอบผิด เช่น ครูถามอายุเท่าไร ตอบว่า ป. 6 เฉยเลย”

นักเรียน 21 กรกฎาคม 2557

“การให้เราต้องประเมินตัวเอง คือ ได้คิดว่าดีไหม ได้ให้คะแนนตัวเองทำให้เราอยากทำให้ดีขึ้น เราฟังเพื่อนว่าเราได้เท่าไร เราเองเราก็คิดของเรา ก็พูดครั้งหน้าแก้ไข”

นักเรียน 21 กรกฎาคม 2557

“ตอนวันนำเสนอผลงานมีห้องอื่นมาดู หนูก็คงเขิน แต่หนูอยากพูดได้ อยากเก่ง ขอเป็นพิธีกรคู่บงกชพล เดี่ยวหนูซ้อมกับบงกชพลค่ะ”

นักเรียน 21 กรกฎาคม 2557

ตอนที่ 3 ผลการเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

3.1 ข้อมูลทั่วไปของตัวแปรที่ศึกษารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ผลการวิเคราะห์ข้อมูลพื้นฐานของกลุ่มตัวอย่างที่ทำการศึกษาพบว่า ในภาพรวมของภาระงานทั้ง 4 ภาระงาน มีคะแนนเฉลี่ยความสามารถในการพูดภาษาอังกฤษอยู่ในระดับอ่อนและระดับพอใช้ ($M=2.441-2.878$, $SD=.288-.455$) จากคะแนนเต็ม 4 คะแนน โดยภาระงานที่มีความสามารถอยู่ในระดับพอใช้ ได้แก่ การสัมภาษณ์ (interview) ($M=2.521$, $SD=.350$) และการเล่านิทาน (storytelling) ($M=2.875$, $SD=.288$) ส่วนการนำเสนอปากเปล่า (oral presentation) ($M=2.441$, $SD=.455$) และการบรรยายภาพ (picture description) ($M=2.475$, $SD=.324$) พบว่าอยู่ในระดับอ่อน

เมื่อพิจารณาตามองค์ประกอบของเกณฑ์ในการประเมินทั้ง 6 ด้าน พบว่า สำหรับภาระงานการสัมภาษณ์ ด้านการบรรลุวัตถุประสงค์ มีคะแนนเฉลี่ยสูงสุด และอยู่ในระดับพอใช้ ($M=2.816$, $SD=.332$) รองลงมาคือ ความคล่องแคล่ว ($M=2.625$, $SD=.400$) การสร้างประโยค ($M=2.551$, $SD=.321$) อยู่ในระดับพอใช้เช่นกัน ส่วนการออกเสียง ($M=2.412$, $SD=.529$) การสร้างประโยค ($M=2.412$, $SD=.503$) และคำศัพท์ อยู่ในระดับอ่อน ($M=2.309$, $SD=.375$) ในทำนองเดียวกันพบว่าภาระงานการนำเสนอปากเปล่า พบว่า ด้านการบรรลุวัตถุประสงค์ มีคะแนนเฉลี่ยสูงสุด และอยู่ในระดับพอใช้ ($M=2.610$, $SD=.537$) รองลงมาคือ ความคล่องแคล่ว อยู่ในระดับพอใช้เช่นกัน ($M=2.507$, $SD=.433$) ส่วนด้านการเชื่อมโยงประโยค ($M=2.412$, $SD=.500$) คำศัพท์ ($M=2.390$, $SD=.432$) การสร้างประโยค ($M=2.375$, $SD=.622$) การออกเสียง ($M=2.353$, $SD=.500$) อยู่ในระดับอ่อน ภาระงานการเล่านิทาน พบว่า อยู่ในระดับพอใช้ทุกด้าน โดยด้านสร้างประโยค มีคะแนนเฉลี่ยสูงสุด ($M=2.963$, $SD=.176$) รองลงมาคือ คำศัพท์ ($M=2.949$, $SD=.312$) การบรรลุเป้าหมาย ($M=2.934$, $SD=.355$) การเชื่อมโยงประโยค ($M=2.904$, $SD=.289$) ความคล่องแคล่ว ($M=2.779$, $SD=.499$) และการออกเสียง ($M=2.721$, $SD=.698$) งานการบรรยายภาพ พบว่า มี 3 ด้าน ที่อยู่ในระดับพอใช้ โดยด้านการออกเสียงมีคะแนนเฉลี่ยสูงสุด ($M=2.581$, $SD=.514$) รองลงมาคือ การบรรลุเป้าหมาย ($M=2.574$, $SD=.329$) และความคล่องแคล่ว ($M=2.515$, $SD=.480$) ส่วนด้านการเชื่อมโยงประโยค ($M=2.456$, $SD=.428$) คำศัพท์ ($M=2.441$, $SD=.448$) และการสร้างประโยค ($M=2.287$, $SD=.409$) อยู่ในระดับอ่อน

เมื่อพิจารณาการให้คะแนนของผู้ประเมินทั้ง 4 คน พบว่า ผู้ประเมินส่วนใหญ่ประเมินความสามารถด้านการพูดของนักเรียนอยู่ในระดับพอใช้ ($M=2.670-2.507$, $SD=.273-.399$) ยกเว้น

ผู้ประเมินคนที่ 4 อยู่ระดับอ่อน ($M=2.473$, $SD=.329$) เมื่อพิจารณาเป็นรายคนแยกตามเกณฑ์ การให้คะแนนปรากฏผลดังนี้ ผู้ประเมินคนที่ 1 พบว่า ประเมินความสามารถด้านการบรรลุเป้าหมาย ($M=2.787$, $SD=.349$) และด้านความคล่องแคล่ว ($M=2.515$, $SD=.403$) อยู่ในระดับพอใช้ ส่วนด้านอื่นอยู่ในระดับอ่อน ได้แก่ ด้านคำศัพท์ ($M=2.500$, $SD=.326$) การสร้างประโยค ($M=2.246$, $SD=.362$) การออกเสียง ($M=2.412$, $SD=.536$) และการเชื่อมโยงเนื้อหา ($M=2.404$, $SD=.308$) ผู้ประเมินคนที่ 2 พบว่า ให้ประเมินความสามารถด้านการพูดอยู่ในระดับพอใช้ทุกด้าน ดังนี้ ด้านความคล่องแคล่ว มีคะแนนเฉลี่ยสูงสุด ($M=2.772$, $SD=.291$) รองลงมา คือ การเชื่อมโยงเนื้อหา ($M=2.728$, $SD=.350$) การออกเสียง ($M=2.647$, $SD=.399$) การบรรลุเป้าหมาย ($M=2.610$, $SD=.385$) การสร้างประโยค ($M=2.285$, $SD=.393$) และคำศัพท์ ($M=2.566$, $SD=.310$) ผู้ประเมินคนที่ 3 พบว่า ประเมินสอดคล้องกับผู้ประเมินคนที่ 1 กล่าวคือ ประเมินความสามารถด้านการบรรลุเป้าหมายสูงที่สุด และอยู่ในระดับพอใช้ ($M=2.801$, $SD=.363$) และพบว่าประเมินความสามารถด้านการพูดอยู่ในระดับพอใช้ทุกด้าน ดังนี้ ด้านการเชื่อมโยงเนื้อหา ($M=2.772$, $SD=.304$) ด้านความคล่องแคล่ว ($M=2.721$, $SD=.352$) ด้านการสร้างประโยค ($M=2.618$, $SD=.344$) ด้านการออกเสียง ($M=2.618$, $SD=.501$) ยกเว้นด้านคำศัพท์ ที่อยู่ในระดับอ่อน ($M=2.493$, $SD=.328$) และผู้ประเมินคนที่ 4 พบว่า ให้คะแนนสอดคล้องกันกับผู้ประเมินคนที่ 1 และ 3 กล่าวคือ ประเมินความสามารถด้านการบรรลุเป้าหมายสูงที่สุด และอยู่ในระดับพอใช้ ($M=2.735$, $SD=.325$) ส่วนด้านที่อยู่ในระดับอ่อน ดังนี้ ด้านคำศัพท์ ($M=2.463$, $SD=.290$) ความคล่องแคล่ว ($M=2.419$, $SD=.459$) การเชื่อมโยงเนื้อหา ($M=2.419$, $SD=.273$) การสร้างประโยค ($M=2.412$, $SD=.331$) และการออกเสียง ($M=2.390$, $SD=.5374$) รายละเอียดดังตารางที่ 4.26

ตารางที่ 4.26 คะแนนเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของภาระงาน และผู้ประเมิน จำแนกตามเกณฑ์การให้คะแนนแบบแยกส่วน

รายการ	เกณฑ์การให้คะแนน (analytic rating scale)												รวม	
	voc.		syn.		coh.		pro.		fun.		flu.			
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Tasks														
Inter.	2.309	0.375	2.412	0.503	2.551	0.312	2.412	0.529	2.816	0.322	2.625	0.400	2.521	0.350
oral	2.390	0.432	2.375	0.622	2.412	0.500	2.353	0.500	2.610	0.537	2.507	0.433	2.441	0.455
story	2.949	0.312	2.963	0.176	2.904	0.289	2.721	0.698	2.934	0.355	2.779	0.499	2.875	0.288
Pic..	2.441	0.448	2.287	0.409	2.456	0.428	2.581	0.514	2.574	0.329	2.515	0.480	2.475	0.324
raters														
คนที่ 1	2.500	0.326	2.426	0.362	2.404	0.308	2.412	0.536	2.787	0.349	2.515	0.403	2.507	0.339
คนที่ 2	2.566	0.310	2.581	0.393	2.728	0.350	2.647	0.399	2.610	0.385	2.772	0.291	2.651	0.273
คนที่ 3	2.493	0.328	2.618	0.344	2.772	0.304	2.618	0.501	2.801	0.363	2.721	0.352	2.670	0.302
คนที่ 4	2.463	0.290	2.412	0.331	2.419	0.273	2.390	0.537	2.735	0.325	2.419	0.459	2.473	0.329

3.2 การศึกษาสัมประสิทธิ์การสรุปร่างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

3.2.1 สัมประสิทธิ์การสรุปร่างอิงความน่าเชื่อถือของเกณฑ์การให้คะแนนการพูดภาษาอังกฤษโดยใช้ $p^* \times t^* \times r^*$ Design

ผลการตรวจสอบความเที่ยงของเกณฑ์การให้คะแนนทักษะการพูดทั้ง 6 ด้าน โดยใช้ $p^* \times t^* \times r^*$ Design โดยใช้ภาระงาน จำนวน 4 ภาระงาน และผู้ประเมิน จำนวน 4 คน ปรากฏดังตารางที่ 4.27 และ 4.28 โดยพบว่าผลการตรวจสอบความเที่ยงของเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุมิติประสงค์ และความคล่องแคล่ว โดยค่าที่แสดงในแนวทแยงคือ องค์ประกอบความแปรปรวน ส่วนค่าที่แสดงในตัวเอียงคือ ค่าสหสัมพันธ์ และค่าที่อยู่ใต้ทแยงคือ ค่าความแปรปรวนร่วม พบว่า ความแปรปรวนของนักเรียน (person, p) มีขนาดใหญ่ที่สุด โดยพบว่ามีค่าความแปรปรวนคิดเป็นร้อยละ 51.84 50.83 53.46 80.33 64.37 และ 65.90 (คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุมิติประสงค์ และความคล่องแคล่ว ตามลำดับ) รองลงมาคือ ปฏิสัมพันธ์ระหว่างนักเรียนกับผู้ประเมิน (pxr) โดยพบว่ามีค่าความแปรปรวนคิดเป็นร้อยละ 19.18 23.42 18.63 11.14 13.79 และ 16.62 (คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุมิติประสงค์ และความคล่องแคล่ว ตามลำดับ) รองลงมาคือ ปฏิสัมพันธ์ระหว่าง นักเรียน ภาระงาน และผู้ประเมิน โดยพบว่ามีค่าความแปรปรวนคิดเป็นร้อยละ 4.89 4.79 9.63 3.77 10.28 และ 9.64 (คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุมิติประสงค์ และความคล่องแคล่ว ตามลำดับ) ในขณะที่พบว่าความแปรปรวนของภาระงาน และปฏิสัมพันธ์ระหว่างนักเรียนและภาระงานมีขนาดเล็ก ไม่เกินร้อยละ 6

เมื่อพิจารณาแหล่งที่มีความคลาดเคลื่อนจากผู้ประเมินพบว่า องค์ประกอบของเกณฑ์การให้คะแนนทักษะการพูด ที่มีความคลาดเคลื่อนมากที่สุดคือ ด้านการสร้างประโยค รองลงมาคือ การเชื่อมโยงเนื้อหา การบรรลุมิติประสงค์ ส่วนด้านอื่นไม่เกินร้อยละ 5 แสดงว่าลักษณะการให้คะแนนของผู้ประเมินมีผลต่อคะแนนของนักเรียนในด้านการสร้างประโยค การเชื่อมโยงเนื้อหา การบรรลุมิติประสงค์ แต่อย่างไรก็ตามความคลาดเคลื่อนที่เกิดขึ้นนี้ยังมีขนาดเล็กกว่าความแปรปรวนของตัวนักเรียนทุกด้าน

เมื่อพิจารณาค่าความเที่ยงของเกณฑ์การให้คะแนนที่นำมาใช้ในการประเมินทักษะการพูดพบว่า เกณฑ์การให้คะแนนด้านที่มีผลต่อความแปรปรวนของคะแนนรวมของคะแนนการพูดมากที่สุดคือ การออกเสียง คิดเป็นร้อยละ 25.66 รองลงมาคือ ความคล่องแคล่ว การบรรลุมิติประสงค์ การสร้างประโยค คำศัพท์ และการเชื่อมโยงเนื้อหา (คิดเป็นร้อยละ 17.66 15.12 15.10 13.68 และ 12.77 ตามลำดับ) และเมื่อพิจารณาค่าความเที่ยงหรือค่าสัมประสิทธิ์การสรุปร่างอิงสำหรับการตัดสินใจ

เชิงสัมพัทธ์ (Generalizability coefficient) พบว่า มีค่าความเชื่อถือได้ในระดับปานกลางทุกด้าน (.628-.715) ยกเว้นด้านการออกเสียง ที่พบว่ามีค่าความเชื่อถือที่อยู่ในระดับสูง คือ มีค่าเท่ากับ .840 เมื่อพิจารณาค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (Phi coefficient) พบว่า ค่าความเชื่อถือที่อยู่ในระดับปานกลางทุกด้าน (.508-.659) ยกเว้นด้านการออกเสียง ที่พบว่ามีค่าความเชื่อถือที่อยู่ในระดับสูง คือ มีค่าเท่ากับ .803 ดังตารางที่ 4.27

เมื่อพิจารณาความเที่ยงหรือความเชื่อถือได้ของคะแนนรวม (composite score) พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (Generalizability coefficient) มีความเชื่อถือของการประเมินอยู่ในระดับสูงคือมีค่าเท่ากับ .846 ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (Phi coefficient) พบว่ามีความเชื่อถือได้ในระดับปานกลาง โดยมีค่าเท่ากับ .745 ดังตารางที่ 4.28

ตารางที่ 4.27 องค์ประกอบความแปรปรวนและความแปรปรวนร่วมสำหรับ $p^* \times t^* \times r^*$ Design

Source of variation	Estimated variance component											
	Voc.	%	Syn.	%	Coh.	%	Pro.	%	Fun.	%	Flu.	%
Person (p)	0.053	51.84	1.024		0.814		0.935		0.983		1.057	
	0.060		0.064	50.83	0.931		0.943		0.923		0.999	
	0.042		0.053		0.050	53.46	0.903		0.989		0.955	
	0.092		0.103		0.086		0.183	80.33	0.996		1.029	
	0.056		0.058		0.055		0.106		0.061	64.37	1.049	
	0.068		0.072		0.060		0.125		0.073		0.080	65.90
Task (t)	0.000	0.00										
	0.001		0.001	1.18								
	0.001		0.004		0.005	5.74						
	0.001		0.003		0.006		0.004	1.80				
	0.000		0.001		0.001		0.002		0.001	1.16		
	0.002		0.005		0.009		0.005		0.002		0.006	5.28
Rater (r)	0.023	4.34										
	0.021		0.022	17.00								
	0.014		0.015		0.008	8.19						
	0.011		0.009		0.007		0.006	2.54				
	0.008		0.011		0.009		0.003		0.006	6.55		
	0.007		0.009		0.008		0.003		0.005		0.003	2.34
pxt	0.002	1.59										
	0.000		0.002	1.97								
	0.002		0.001		0.000	0.00						
	0.001		0.001		0.000		0.001	0.35				
	0.003		0.001		0.001		0.001		0.003	3.44		
	0.000		0.000		0.002		0.003		0.001		0.000	0.00

ตารางที่ 4.27 (ต่อ)

Source of variation	Estimated variance component												
	Voc.	%	Syn.	%	Coh.	%	Pro.	%	Fun.	%	Flu.	%	
pxr	0.019	19.18											
	0.013		0.030	23.42									
	0.007		0.013		0.017	18.63							
	0.004		0.008		0.011		0.025	11.14					
	0.007		0.012		0.006		0.007		0.013	13.79			
	0.006		0.010		0.009		0.015		0.007		0.020	16.62	
txr	0.000	0.00											
	0.000		0.001	0.81									
	0.001		0.002		0.004	4.35							
	0.000		0.000		0.000		0.000	0.00					
	0.000		0.000		0.001		0.000		0.000	0.00			
	0.000		0.000		0.001		0.000		0.000		0.000	0.00	
ptr,e	0.005	4.89											
	0.001		0.006	4.79									
	0.002		0.001		0.009	9.63							
	0.001		0.001		0.000		0.009						
	0.001		0.001		0.002		0.000		0.010	10.28			
	0.001		0.001		0.002		0.001		0.002		0.012	9.64	
Total	0.102	100	0.127	100	0.093	100	0.228	100	0.095	100	0.121	100	

ตารางที่ 4.28 องค์ประกอบความแปรปรวนและความแปรปรวนร่วมของการประเมินสำหรับ

$$p^* \times t^* \times r^* \text{ Design}$$

Contributions to	Variance and covariance components					
	Voc.	Syn.	Coh.	Pro.	Fun.	Flu.
A priori weight	0.167	0.167	0.167	0.167	0.167	0.167
Universe score variance (%)	13.68	15.1	12.77	25.66	15.12	17.66
Relative error variance (%)	14.51	19.69	15.03	17.74	15.15	17.88
Absolute error variance (%)	17.23	21.46	17.57	15.68	12.16	15.89
Gen Coefficient	0.669	0.628	0.654	0.840	0.701	0.715
Phi Coefficient	0.518	0.508	0.535	0.803	0.644	0.659
Composite						
Relative Error Variance				0.014		
Absolute Error Variance				0.026		
Generalizability coefficient				0.846		
Phi coefficient				0.745		

3.2.2 สัมประสิทธิ์การสุร้อ้างอิงความน่าเชื่อถือของภาระงานการประเมินทักษะการพูดภาษาอังกฤษ โดยใช้ $p \times t \times r$ Design การตรวจสอบคุณภาพของภาระการประเมินทักษะการพูดภาษาอังกฤษแต่ละภาระงานด้วยการออกแบบ $p \times t \times r$ Design หมายถึง นักเรียนทุกคน (p) ทำภาระงานการพูดภาษาอังกฤษ (t) ทุกภาระงาน โดยใช้ภาระงานที่ต่างกัน จำนวน 4 ภาระงาน ผู้ประเมิน (r) จำนวน 4 คน ตรวจสอบให้คะแนนนักเรียนทุกคน และตรวจสอบทุกภาระงาน ผลการตรวจสอบความเที่ยงของภาระงานที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ทั้ง 4 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ โดยใช้ $p \times t \times r$ Design โดยค่าที่แสดงในแนวทแยงคือองค์ประกอบความแปรปรวน ส่วนค่าที่แสดงในตัวเอียง คือ ค่าสหสัมพันธ์ และค่าที่อยู่ใต้แนวทแยงคือค่าความแปรปรวนร่วม ผลการตรวจสอบพบว่าความแปรปรวนของนักเรียน (person, p) มีขนาดใหญ่ที่สุด โดยพบว่า มีค่าความแปรปรวนคิดเป็นร้อยละ 81.95 91.42 80.94 และ 85.22 (การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ ตามลำดับ) รองลงมาคือ ปฏิสัมพันธ์ระหว่างนักเรียนกับภาระงานกับผู้ประเมิน (pxtxr) โดยพบว่ามีค่าความแปรปรวนคิดเป็นร้อยละ 10.27 6.39 17.37 และ 13.15 (การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ ตามลำดับ) รองลงมาอันดับ 2 คือ ปฏิสัมพันธ์ระหว่างภาระงานกับผู้ประเมิน โดยพบว่าภาระงานการสัมภาษณ์มีความแปรปรวนมากที่สุด ส่วนภาระงานการเล่านิทาน การบรรยายภาพ และการนำเสนอปากเปล่านั้นมีความแปรปรวนใกล้เคียงกัน และมีความแปรปรวนไม่มากเพียงร้อยละ 1 เท่านั้น (ร้อยละ 4.54 1.58 1.15 และ 1.52 ตามลำดับ) ส่วนแหล่งความแปรปรวนหรือแหล่งความคลาดเคลื่อน (error source) รองลงมาอันดับที่ 3 คือ ปฏิสัมพันธ์ระหว่างนักเรียนกับภาระงาน พบว่า ภาระงานการสัมภาษณ์มีความแปรปรวนมากที่สุดแต่ก็พบว่ามีเพียงร้อยละ 1.75 ส่วนภาระงานอื่นๆ พบว่า มีความคลาดเคลื่อนน้อยมากคือไม่ถึงร้อยละ 1 ส่วนความคลาดเคลื่อนที่เกิดจาก ภาระงานพบว่ามีน้อยมาก โดยพบว่า ภาระงานการสัมภาษณ์มีความคลาดเคลื่อนหรือความแปรปรวนสูงที่สุด แต่ก็เพียงร้อยละ 1.5 เท่านั้น ส่วนภาระงานอื่นๆ มีความแปรปรวนไม่ถึงร้อยละ 1 และผู้ประเมินพบว่ามีค่าความแปรปรวนเกือบเป็นศูนย์ทุกภาระงาน ยกเว้นภาระงานการเล่านิทานที่พบว่ามีค่าความแปรปรวน .05 แต่ก็น้อยมากไม่ถึงร้อยละ 1 จากผลการวิเคราะห์ข้อมูลดังกล่าว อธิบายได้ว่า คะแนนการพูดของนักเรียนนั้นเกิดจากตัวนักเรียนมากที่สุด ไม่ใช่เกิดจากความคลาดเคลื่อนจากแหล่งอื่นๆ และพบว่า ภาระงานมีปฏิสัมพันธ์กับผู้ประเมิน และนักเรียนมีปฏิสัมพันธ์กับภาระงาน แต่เมื่อพิจารณาความคลาดเคลื่อนที่เกิดจากตัวภาระงาน (t) หรือ main effect พบว่ามีความคลาดเคลื่อนสูงที่สุดเพียงร้อยละ 1.50 คือ การสัมภาษณ์ นอกนั้นน้อยมากคือน้อยกว่าร้อยละ 1 เท่านั้น ปรากฏดังตารางที่ 4.29

เมื่อพิจารณาค่าความเที่ยงของภาระงานที่นำมาใช้ในการประเมินทักษะการพูด พบว่า ภาระงานที่มีผลต่อความแปรปรวนของคะแนนรวมของคะแนนการพูดมากที่สุดคือ การนำเสนอปากเปล่า

คิดเป็นร้อยละ 35.48 รองลงมาคือ การสัมภาษณ์ การบรรยายภาพ และการเล่านิทาน (คิดเป็นร้อยละ 24.47 21.55 และ 18.49 ตามลำดับ) และเมื่อพิจารณาค่าความเที่ยงหรือค่าสัมประสิทธิ์การสรุปล้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (Generalizability coefficient) พบว่า มีค่าความเชื่อถือได้ในระดับสูงทุกภาระงาน (มากกว่า .80) ส่วนค่าสัมประสิทธิ์การสรุปล้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (Phi coefficient) พบผลคล้ายคลึงกัน คือ ค่าความเชื่อถือได้ในระดับสูงทุกภาระงาน (มากกว่า .80)

เมื่อพิจารณาความเที่ยงหรือความเชื่อถือได้ของคะแนนรวม (composite score) พบว่า มีความเชื่อถือได้ของการประเมินอยู่ในระดับสูง ทั้งค่าสัมประสิทธิ์การสรุปล้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปล้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ คือ มีความเชื่อถือได้เท่ากับ .967 และ .959 ตามลำดับ ดังตารางที่ 4.30

ตารางที่ 4.29 องค์ประกอบความแปรปรวนและความแปรปรวนร่วมสำหรับ $p \times t \times r$ Design

Source of variation	variance and covariance component							
	สัมภาษณ์	นำเสนอ		เล่านิทาน		บรรยายภาพ		
		%	ปากเปล่า	%	%	%	%	
Person (p)	0.112	81.95	0.773		0.554		0.609	
	0.118		0.209	91.42	0.632		0.647	
	0.051		0.080		0.076	80.94	0.612	
	0.063		0.092		0.052		0.096	85.22
Task (t)	0.002	1.50						
			0.001	0.46				
					0.000	0.00		
						0.000	0.00	
Rater (r)	0.000	0.00						
	0.001		0.000	0.00				
	0.001		0.001		0.000	0.05		
	0.000		0.000		0.000		0.000	0.00
pxt	0.002	1.75						
			0.001	0.58				
					0.000	0.00		
						0.000	0.12	
pxr	0.000	0.00						
	0.001		0.000	0.00				
	0.001		0.001		0.000	0.05		
	0.000		0.000		0.000		0.000	0.00

ตารางที่ 4.29 (ต่อ)

Source of variation	variance and covariance component							
	สัมภษณ์	%	นำเสนอ ปากเปล่า	%	เล่า นิทาน	%	บรรยายภาพ	%
txr	0.006	4.54	0.003	1.15	0.001	1.58	0.002	1.52
ptr	0.014	10.27	0.015	6.39	0.016	17.37	0.015	13.15
total	0.136	100.00	0.229	100.00	0.094	100.00	0.113	100.00

ตารางที่ 4.30 ผลการวิเคราะห์คะแนนรวมของการประเมินสำหรับ $p^* \times t^* \times r^*$ Design

Contributions to	Variance components			
	สัมภษณ์	นำเสนอปากเปล่า	เล่านิทาน	บรรยายภาพ
A priori weight	0.250	0.250	0.250	0.250
Universe score variance (%)	24.47%	35.48%	18.49%	21.55%
Relative error variance (%)	24.00%	24.77%	26.59%	24.64%
Absolute error variance (%)	28.94%	25.43%	23.95%	21.69%
Gen Coefficient	0.923	0.960	0.903	0.912
Phi Coefficient	0.885	0.950	0.895	0.905
Composite				
Relative Error Variance			0.003	
Absolute Error Variance			0.004	
Generalizability coefficient			0.967	
Phi coefficient			0.959	

3.2.3 สัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินระหว่างเรียน และการประเมินสรุปรวม

ในการศึกษาสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินการพูดภาษาอังกฤษโดยใช้แฟ้มสะสมงานที่ให้มีการประเมินทั้งการประเมินระหว่างเรียนและการประเมินสรุปรวมในครั้งนี้ ใช้ $p \times t \times r$ Design โดยนักเรียน (p) ถูกประเมินทักษะการพูดด้วยการงานการพูด (t) จำนวน 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และ 4) การบรรยายภาพ และผู้ประเมิน (r) ซึ่งเป็นครูผู้สอนวิชาภาษาอังกฤษในโรงเรียน จำนวน 4 คน เป็นผู้ตรวจให้คะแนน โดยผู้ประเมินตรวจให้คะแนนนักเรียนทุกคน และตรวจทุกภาระงาน โดยในการประเมินโดยใช้แฟ้มสะสมงานในครั้งนี้แบ่งการประเมินเป็นสองระยะ คือ การประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) โดยการประเมินระหว่างเรียนใช้ 3 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และการประเมินสรุปรวมใช้ 1 ภาระงาน คือ การบรรยายภาพ ผลการตรวจสอบรูปแบบการประเมินพบว่า ความแปรปรวนของนักเรียน (p) มีขนาดใหญ่ที่สุด โดยพบว่า มีค่าความแปรปรวนคิดเป็นร้อยละ 95.41 และ 76.55 (การประเมินระหว่างเรียน และการประเมินสรุปรวม ตามลำดับ) รองลงมาคือ ปฏิสัมพันธ์ระหว่างนักเรียนกับภาระงานกับผู้ประเมิน (p_{txr}) โดยพบว่ามีค่าความแปรปรวนคิดเป็นร้อยละ 2.81 และ 3.67 (การประเมินระหว่างเรียน และการประเมินสรุปรวม ตามลำดับ) รองลงมาอันดับ 2 คือ ปฏิสัมพันธ์ระหว่างนักเรียนกับภาระงาน โดยพบว่า การประเมินสรุปรวมมีความแปรปรวนหรือความคลาดเคลื่อน ร้อยละ 12.69 ส่วนการประเมินระหว่างเรียนพบว่ามีค่าความแปรปรวนหรือความคลาดเคลื่อนน้อยมาก เพียงร้อยละไม่ถึง 1 เท่านั้น ส่วนความแปรปรวนหรือความคลาดเคลื่อนของปฏิสัมพันธ์ระหว่างนักเรียนกับผู้ประเมินพบว่าการประเมินสรุปรวมมีความแปรปรวนร้อยละ 5.6 ส่วนการประเมินระหว่างเรียนมีความแปรปรวนน้อยมากไม่ถึงร้อยละ 1 เท่านั้น ส่วนปฏิสัมพันธ์ระหว่างภาระงานกับผู้ประเมิน (txr) พบว่า ในการประเมินระหว่างเรียนมีความแปรปรวนหรือความคลาดเคลื่อนน้อยมาก คือ เพียงร้อยละ 1.17 และในการประเมินสรุปรวมพบว่ามีค่าความแปรปรวนเกือบเป็นศูนย์หรือไม่มีค่าความคลาดเคลื่อนเลย และเมื่อพิจารณา ความแปรปรวนของผู้ประเมิน (r) และภาระงาน (t) หรือ main effect พบว่ามีค่าน้อยมากไม่ถึงร้อยละ 1 ดังนั้น ความแปรปรวนของปฏิสัมพันธ์ระหว่างนักเรียนกับภาระงาน (p_{xt}) อธิบายได้ว่า คะแนนการพูดของนักเรียนแต่ละคนนั้นขึ้นอยู่กับภาระงานแต่ละเรื่องที่แตกต่างกัน หรือความยากของแต่ละภาระงาน ปรากฏดังตารางที่ 4.31

เมื่อพิจารณาค่าความเที่ยงของรูปแบบการประเมินระหว่างเรียนและการประเมินสรุปรวมที่นำมาใช้ในการประเมินทักษะการพูด พบว่า การประเมินในช่วงการประเมินระหว่างเรียนที่มีผลต่อความแปรปรวนของคะแนนรวมของคะแนนการพูดมากที่สุด คิดเป็นร้อยละ 78.98 ส่วนการประเมิน

สรุปรวม คิดเป็นร้อยละ 21.02 และเมื่อพิจารณาค่าความเที่ยงหรือค่าสัมประสิทธิ์การสรุปร่างอิง สำหรับการตัดสินใจเชิงสัมพัทธ์ พบว่า การประเมินระหว่างเรียนมีค่าความเชื่อถือได้ในระดับสูง คือ .966 ส่วนการประเมินสรุปรวม อยู่ในระดับปานกลาง คือมีค่าเท่ากับ .78 ส่วนค่าสัมประสิทธิ์การสรุปร่างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์พบผลคล้ายคลึงกัน คือ พบว่า การประเมินระหว่างเรียนมีค่าความเชื่อถือได้ในระดับสูง คือ .954 ส่วนการประเมินสรุปรวมอยู่ในระดับปานกลาง คือมีค่าเท่ากับ.77

เมื่อพิจารณาความเที่ยงหรือความเชื่อถือได้ของคะแนนรวม (composite score) พบว่า มีความเชื่อถือได้ของการประเมินอยู่ในระดับสูงทั้งค่าสัมประสิทธิ์การสรุปร่างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปร่างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์คือ มีความเชื่อถือได้เท่ากับ .961 และ .956 ตามลำดับ ดังตารางที่ 4.32

ตารางที่ 4.31 องค์ประกอบความแปรปรวนและความแปรปรวนร่วมของการประเมินระหว่างเรียน และการประเมินสรุปรวม สำหรับ $p \times t \times r$ Design

Source of variation	องค์ประกอบความแปรปรวนและความแปรปรวนร่วม			
	การประเมินระหว่างเรียน (formative assessment)		การประเมินสรุปรวม (summative assessment)	
		%		%
Person (p)	0.097	95.41	0.770	
	0.069		0.082	76.55
Task (t)	0.000	0.00		
			0.001	0.96
Rater (r)	0.0003	0.03		
	0.001		0.000	0.00
pxt	0.0004	0.35		
			0.014	12.69
pxr	0.0002	0.23		
	0.000		0.006	5.62
txr	0.001	1.17		
			0.000	0.00
ptr	0.003	2.81		
			0.004	3.67
total	0.102	100.00	0.107	100.00

ตารางที่ 4.32 ผลการวิเคราะห์คะแนนรวมของการประเมินของการประเมินระหว่างเรียนและ
การประเมินสรุปรวม สำหรับ $p^{\circ} \times t^{\circ} \times r^{\circ}$ Design

Contributions to:	Variance components	
	การประเมินระหว่างเรียน	การประเมินสรุปรวม
A priori weight	0.75	0.25
Universe score variance (%)	78.98%	21.02%
Relative error variance (%)	56.89%	43.11%
Absolute error variance (%)	63.60%	36.40%
Gen Coefficient	0.966	0.777
Phi Coefficient	0.954	0.765
Composite		
Relative Error Variance		0.003
Absolute Error Variance		0.004
Generalizability coefficient		0.962
Phi coefficient		0.956

3.3 การเปรียบเทียบรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ด้วยทฤษฎีการสุปร้องอิงความน่าเชื่อถือพหุตัวแปร

3.3.1 รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6
โดยใช้เกณฑ์การให้คะแนนการพูดภาษาอังกฤษแบบแยกส่วน (analytic rating scale)

ในการศึกษารูปแบบการประเมินที่มีประสิทธิภาพมากที่สุดและมีความเหมาะสมสำหรับการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน เพื่อศึกษาจำนวนผู้ประเมิน และจำนวนภาระงานที่เหมาะสม และทำให้การประเมินการพูดภาษาอังกฤษมีความน่าเชื่อถือ หรือมีความเที่ยงในระดับที่เหมาะสม ซึ่งการเปรียบเทียบรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ในครั้งนี้ มุ่งศึกษาเงื่อนไขที่เกี่ยวข้อง 2 แห่ง คือ 1) ผู้ประเมิน (raters) และ 2) ภาระงานการพูด (tasks) โดยใช้ผลของ D-study ของรูปแบบ $p^{\circ} \times t^{\circ} \times r^{\circ}$ Design โดยนักเรียน (p) ทำการฝึกทักษะการพูดภาษาอังกฤษและถูกประเมินทักษะการพูดด้วยภาระงานการพูด (t) จำนวน 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และ 4) การบรรยายภาพ และผู้ประเมิน (r) ซึ่งเป็นครูผู้สอนวิชาภาษาอังกฤษในโรงเรียน จำนวน 4 คน เป็นผู้ตรวจให้คะแนน โดยผู้ประเมินตรวจให้คะแนนนักเรียนทุกคน และตรวจทุกภาระงาน ในการศึกษาครั้งนี้ได้ทำการศึกษาผลของการเพิ่มจำนวนผู้ประเมินและจำนวนภาระงานที่มีต่อความน่าเชื่อถือและความคลาดเคลื่อนในการประเมิน

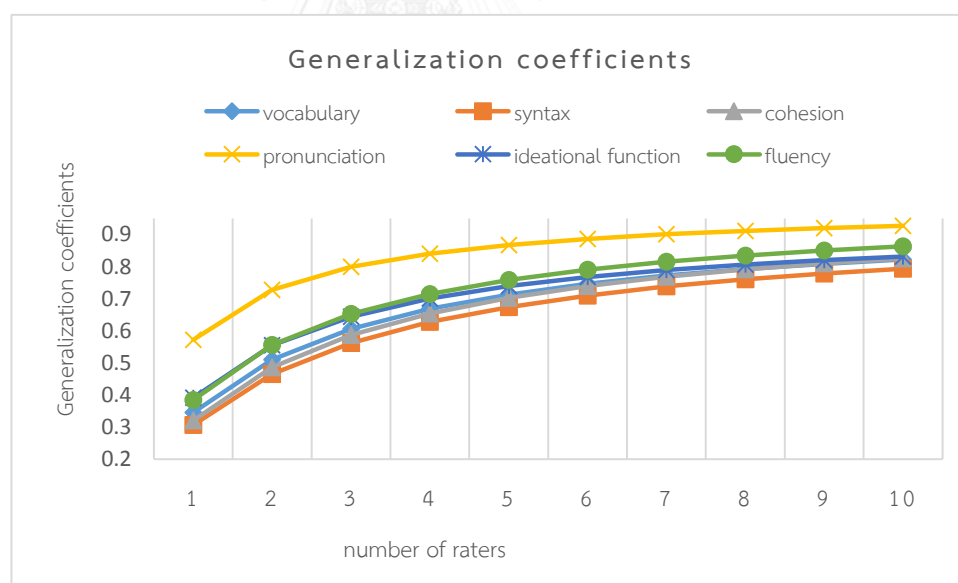
3.3.1.1 การเปรียบเทียบจำนวนผู้ประเมินที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษ

ผลการเปรียบเทียบพบว่า เมื่อเพิ่มจำนวนผู้ประเมินให้มากขึ้นทำให้ระดับความน่าเชื่อถือหรือความเที่ยงในการประเมินมีค่าสูงขึ้น โดยพบว่าค่าสัมประสิทธิ์การสรุปอ้างอิง ความน่าเชื่อถือเชิงสัมพัทธ์จะเพิ่มขึ้น หากใช้ผู้ประเมินเพิ่มขึ้น โดยพบว่าหากใช้ผู้ประเมินตั้งแต่ 9 คนขึ้นไปจะทำให้ค่าความเที่ยงอยู่ในระดับสูง (.80) ยกเว้นด้านการสร้างประโยค ที่อยู่ระดับปานกลาง คือมีค่าเท่ากับ .778 แต่หากพิจารณาเป็นรายองค์ประกอบพบว่า การออกเสียง (pronunciation) จะใช้ผู้ประเมินเพียง 4 คน เท่านั้น จะมีความเที่ยงอยู่ในระดับที่เชื่อถือได้หรือมีความเที่ยงสูง คือมีค่าเท่ากับ .840 รองลงมาคือ ความคล่องแคล่ว (fluency) การบรรลุเป้าหมาย (ideational function) (ผู้ประเมินจำนวน 7 และ 8 คน ตามลำดับ) ส่วนการเชื่อมโยง (cohesion) และ คำศัพท์ (vocabulary) ใช้จำนวน 9 คนเท่านั้น ในขณะที่การสร้างประโยค (syntax) ต้องใช้ผู้ประเมินมากกว่า 10 คน ขึ้นไป ดังตารางที่ 4.33 และภาพที่ 4.5 ในทำนองเดียวกัน พบว่าค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ (Phi coefficients) จะเพิ่มขึ้น หากใช้ผู้ประเมินเพิ่มขึ้น โดยพบว่า หากใช้ผู้ประเมินตั้งแต่ 9 คนขึ้นไปจะทำให้ค่าความเที่ยงระดับปานกลาง (.70) ขึ้นไป ยกเว้นด้านการออกเสียง ที่อยู่ระดับสูง แต่หากพิจารณาเป็นรายองค์ประกอบพบว่า การออกเสียงจะใช้ผู้ประเมินเพียง 4 คน เท่านั้น จะมีความเที่ยงอยู่ในระดับที่เชื่อถือได้ (.80) ส่วนด้านความคล่องแคล่ว การบรรลุเป้าหมายต้องใช้ผู้ประเมิน จำนวน 6 คน ขึ้นไป จึงมีความเที่ยงปานกลาง และด้านสร้างประโยค การเชื่อมโยง และคำศัพท์ต้องใช้ผู้ประเมิน จำนวน 10 คน ขึ้นไป จึงมีความเที่ยงปานกลาง ดังตารางที่ 4.34 และภาพที่ 4.6

เมื่อพิจารณาคะแนนรวม (composite score) ของการเพิ่มจำนวนผู้ประเมินพบว่า เมื่อเพิ่มจำนวนผู้ประเมินขึ้นจะทำให้ค่าความคลาดเคลื่อนทั้งความคลาดเคลื่อนเชิงสัมพัทธ์ (relative error) และความคลาดเคลื่อนเชิงสมบูรณ์ (absolute error) ลดลง ดังตารางที่ 4.35 และภาพที่ 4.7 ส่วนค่าความเที่ยงพบว่ามีค่าความน่าเชื่อถือของการประเมินในภาพรวมสูงขึ้นเมื่อเพิ่มจำนวนผู้ประเมินให้มากขึ้น โดยพบว่าค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์จะมีค่าสูงหรือมีความเชื่อถือได้เมื่อใช้ผู้ประเมินเพียง 3 คนเท่านั้น สอดคล้องกับค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์พบว่าสูงขึ้นเมื่อเพิ่มจำนวนผู้ประเมินให้มากขึ้น โดยพบว่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์จะมีค่าสูง (.80) หรือมีความเชื่อถือได้ เมื่อใช้ผู้ประเมินจำนวน 6 คน ดังตารางที่ 4.35 และภาพที่ 4.8

ตารางที่ 4.33 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพันธ์ของเกณฑ์การให้คะแนนการพูดแบบแยกส่วน
เมื่อเพิ่มจำนวนผู้ประเมิน

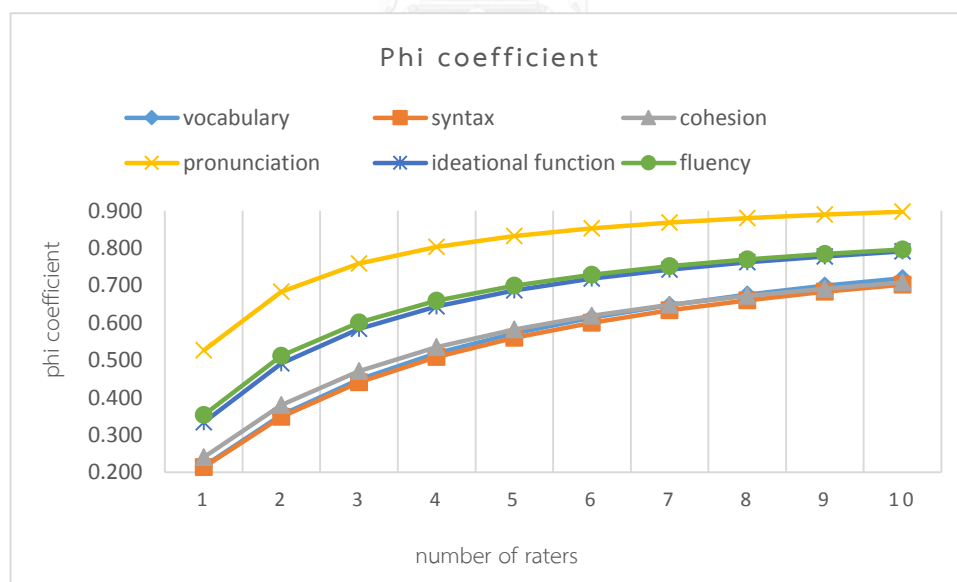
ผู้ประเมิน (คน)	Generalization coefficients (ภาระงาน=4)					
	Voc.	Syn.	Coh.	Pro.	Fun.	Flu.
1	0.346	0.307	0.321	0.572	0.392	0.385
2	0.51	0.465	0.486	0.727	0.555	0.556
3	0.606	0.562	0.587	0.799	0.644	0.653
4	0.669	0.628	0.654	0.840	0.701	0.715
5	0.713	0.674	0.703	0.867	0.739	0.758
6	0.746	0.71	0.739	0.886	0.768	0.790
7	0.772	0.738	0.768	0.901	0.789	0.815
8	0.792	0.760	0.791	0.911	0.806	0.834
9	0.808	0.778	0.810	0.920	0.820	0.850
10	0.822	0.793	0.825	0.927	0.831	0.863



ภาพที่ 4.5 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพันธ์ของเกณฑ์การให้คะแนนการพูดแบบแยกส่วน
เมื่อเพิ่มจำนวนผู้ประเมิน

ตารางที่ 4.34 สัมประสิทธิ์การสรุปอ้างอิงเชิงสมบูรณ์ของเกณฑ์การให้คะแนนการพูดแบบแยกส่วน
เมื่อเพิ่มจำนวนผู้ประเมิน

ผู้ประเมิน (คน)	phi coefficients (ภาระงาน = 4)					
	Voc.	Syn.	Coh.	Pro.	Fun.	Flu.
1	0.216	0.214	0.240	0.527	0.333	0.353
2	0.354	0.348	0.380	0.684	0.491	0.512
3	0.449	0.441	0.471	0.759	0.583	0.601
4	0.518	0.508	0.535	0.803	0.644	0.659
5	0.572	0.560	0.582	0.833	0.686	0.699
6	0.614	0.600	0.619	0.853	0.718	0.729
7	0.648	0.633	0.648	0.869	0.742	0.752
8	0.676	0.660	0.672	0.880	0.762	0.770
9	0.699	0.683	0.691	0.890	0.778	0.785
10	0.719	0.702	0.708	0.898	0.791	0.797



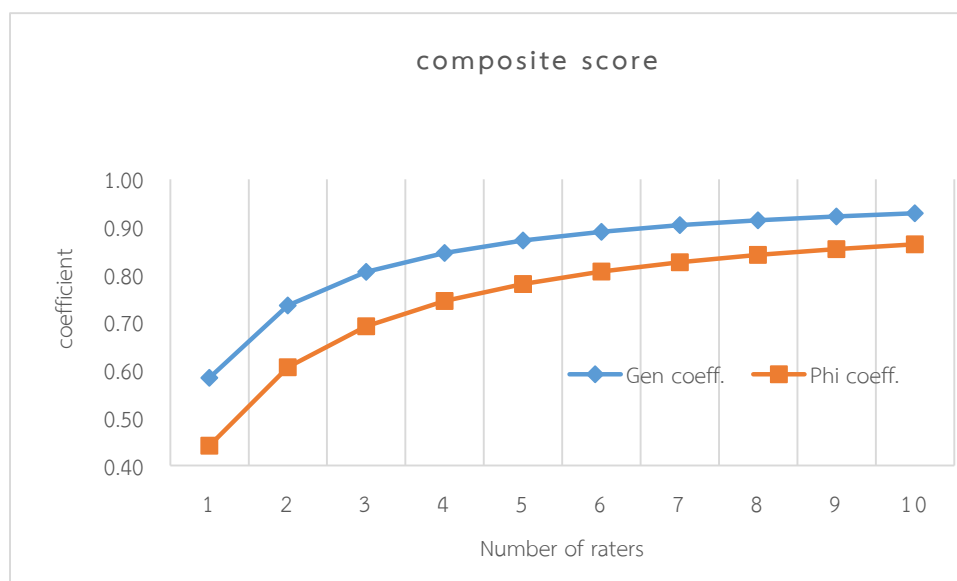
ภาพที่ 4.6 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสมบูรณ์ของเกณฑ์การประเมินการพูดแบบแยกส่วน
เมื่อเพิ่มจำนวนผู้ประเมิน

ตารางที่ 4.35 คะแนนรวมของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนผู้ประเมิน

จำนวนผู้ประเมิน (คน)	composite score (ภาระงาน= 4)			
	Relative error	Absolute error	Gen coeff.	Phi coeff.
1	0.054	0.095	0.584	0.442
2	0.027	0.049	0.736	0.606
3	0.018	0.033	0.806	0.692
4	0.014	0.026	0.846	0.745
5	0.011	0.021	0.872	0.781
6	0.009	0.018	0.890	0.807
7	0.008	0.016	0.904	0.826
8	0.007	0.014	0.914	0.842
9	0.006	0.013	0.922	0.854
10	0.006	0.012	0.929	0.864



ภาพที่ 4.7 ความคลาดเคลื่อนของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนผู้ประเมิน



ภาพที่ 4.8 คะแนนรวมของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนผู้ประเมิน

2.3.1.2 การเปรียบเทียบจำนวนภาระงานที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษ

ผลการเปรียบเทียบพบว่า เมื่อเพิ่มจำนวนภาระงานให้มากขึ้นทำให้ระดับความน่าเชื่อถือหรือความเที่ยงในการประเมินมีค่าสูงขึ้น โดยพบว่าค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ จะเพิ่มขึ้นหากใช้จำนวนภาระงานเพิ่มขึ้น ซึ่งพบว่า ด้านการออกเสียงมีค่าอยู่ในระดับที่เชื่อถือได้ (.80) ถึงแม้จะใช้ภาระงานเพียง 2 ภาระงานเท่านั้น ส่วน ด้านความคล่องแคล่ว การบรรลุวัตถุประสงค์ ต้องใช้ภาระงานตั้งแต่ 4 ภาระงานขึ้นไป ด้านการเชื่อมโยงเนื้อหา และคำศัพท์ ต้องใช้ภาระงานจำนวน 9 ภาระงาน และด้านการสร้างประโยค ต้องใช้ภาระงานมากกว่า 10 ภาระงาน ขึ้นไป จึงจะทำให้ค่าความเที่ยงอยู่ในระดับปานกลาง (.70) ดังตารางที่ 4.36 และภาพที่ 4.9 ในทำนองเดียวกัน พบว่าค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสมบูรณ์จะเพิ่มขึ้น หากใช้ภาระงานเพิ่มขึ้น โดยพบว่า ด้านการออกเสียงจะมีความเที่ยงอยู่ในระดับสูง (.80) เมื่อใช้ภาระงานเพียง 4 ภาระงานเท่านั้น ส่วน ด้านความคล่องแคล่ว ต้องใช้ภาระงานตั้งแต่ 7 ภาระงานขึ้นไป ด้านการบรรลุวัตถุประสงค์ ต้องใช้ภาระงานจำนวน 8 ภาระงาน และด้านการสร้างประโยค การเชื่อมโยงเนื้อหา และ คำศัพท์ ต้องใช้ภาระงานมากกว่า 10 ภาระงานขึ้นไป จึงจะทำให้ค่าความเที่ยงของอยู่ในระดับปานกลาง (.70) ดังตารางที่ 4.37 และภาพที่ 4.10

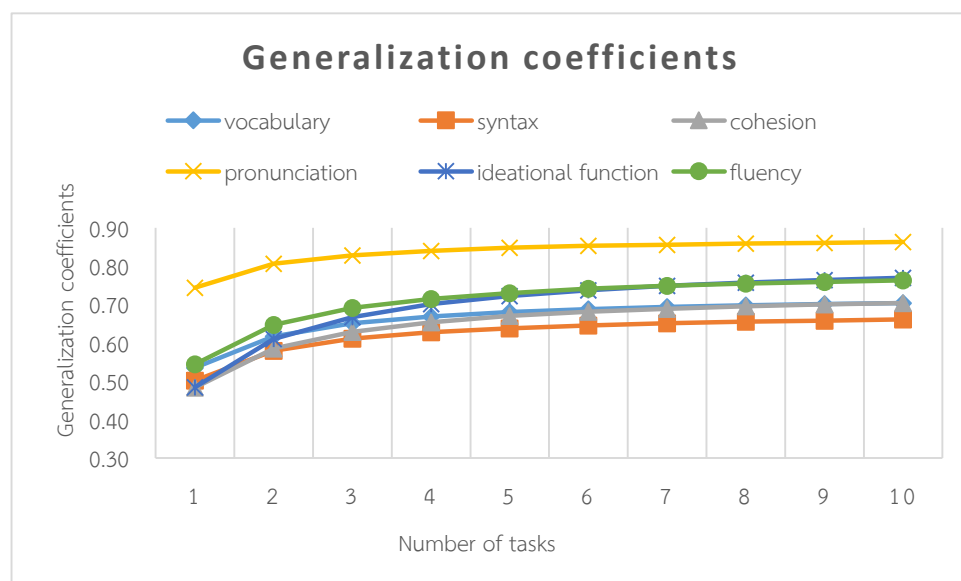
เมื่อพิจารณาคะแนนรวม (composite score) พบว่า เมื่อเพิ่มจำนวนภาระงานขึ้นจะทำให้ค่าความคลาดเคลื่อนทั้งความคลาดเคลื่อนเชิงสัมพัทธ์ (relative error) และความคลาดเคลื่อนเชิงสัมบูรณ์ (absolute error) ลดลง ดังตารางที่ 4.38 และภาพที่ 4.11 ส่วนค่าความเที่ยงพบว่ามี

ความน่าเชื่อถือของการประเมินในภาพรวมสูงขึ้นเมื่อเพิ่มจำนวนภาระงานให้มากขึ้น โดยพบว่าค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพันธ์จะมีความเที่ยงสูง (.80) หรือมีความเชื่อถือได้เมื่อใช้ผู้ประเมินเพียง 2 คนเท่านั้น สอดคล้องกับค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพันธ์ พบว่า สูงขึ้นเมื่อเพิ่มจำนวนภาระงานให้มากขึ้น โดยพบว่าค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพันธ์จะมีความเที่ยงสูงปานกลาง (.70) หรือมีความเชื่อถือได้ในระดับปานกลางเมื่อใช้ผู้ประเมินจำนวน 2 คน แต่หากต้องการให้ความเที่ยงอยู่ในระดับสูงถึงระดับ .80 จะต้องใช้ภาระงานมากกว่า 10 ภาระงาน ดังตารางที่ 4.38 และภาพที่ 4.12

จากผลการศึกษากล่าวได้ว่าหากผู้ประเมินต้องการให้การประเมินการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้เกณฑ์การให้คะแนนแบบแยกส่วนและต้องการให้การประเมินในแต่ละด้าน ทั้ง 6 ด้าน ได้แก่ คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุมิติคุณประสงค์ และความคล่องแคล่ว มีความน่าเชื่อถือผู้ประเมินอาจเลือกใช้วิธีการเพิ่มภาระงานหรือจำนวนครั้งในการประเมินแต่ละภาระงานให้มากขึ้น และใช้จำนวนผู้ประเมินเพียง 3 คน ก็จะได้ความเที่ยงในการประเมินในระดับที่เชื่อถือได้ เพราะสามารถทำได้ง่ายกว่าการเพิ่มจำนวนผู้ประเมินซึ่งจะเป็นการเพิ่มค่าใช้จ่ายในการประเมินมากกว่า

ตารางที่ 4.36 สัมประสิทธิ์การสรุปร่างเชิงสัมพันธ์ของเกณฑ์การให้คะแนนการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนภาระงาน

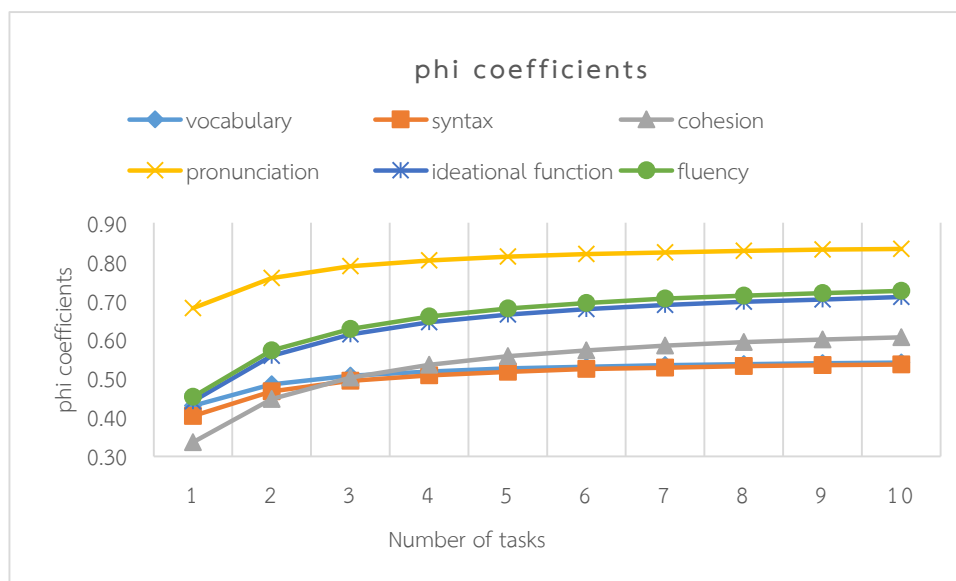
ภาระงาน (เรื่อง)	Generalization coefficients (ผู้ประเมิน=4 คน)					
	Voc.	Syn.	Coh.	Pro.	Fun.	Flu.
1	0.535	0.502	0.483	0.744	0.484	0.544
2	0.617	0.579	0.585	0.806	0.610	0.647
3	0.651	0.611	0.629	0.828	0.667	0.691
4	0.669	0.628	0.654	0.840	0.701	0.715
5	0.680	0.638	0.670	0.848	0.722	0.730
6	0.688	0.645	0.681	0.853	0.737	0.741
7	0.694	0.651	0.689	0.856	0.749	0.749
8	0.698	0.655	0.695	0.859	0.757	0.755
9	0.701	0.658	0.700	0.861	0.764	0.759
10	0.704	0.661	0.704	0.863	0.770	0.763



ภาพที่ 4.9 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพันธ์ ของเกณฑ์การประเมินการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนภาระงาน

ตารางที่ 4.37 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพันธ์ของเกณฑ์การให้คะแนนการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนภาระงาน

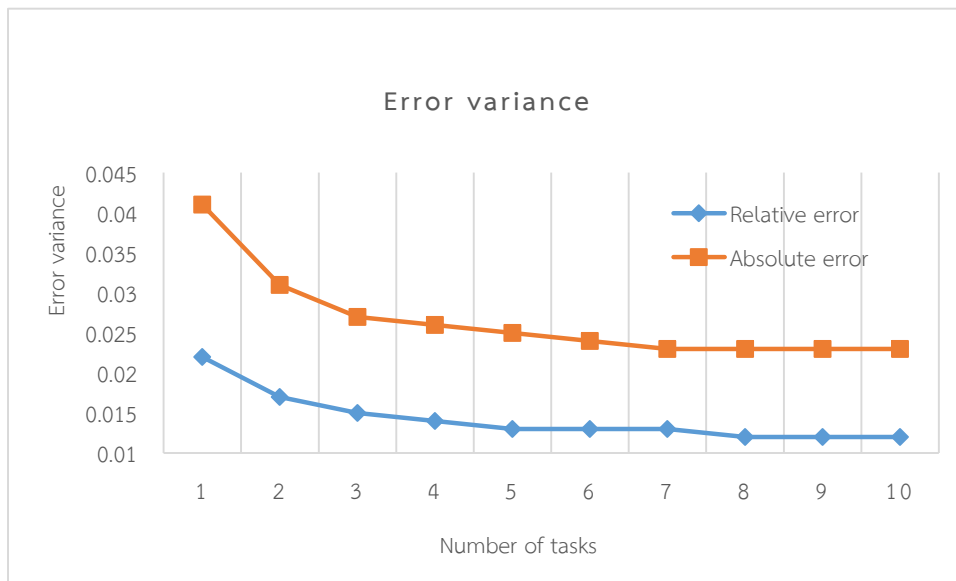
ภาระงาน (เรื่อง)	phi coefficients (ผู้ประเมิน=4 คน)					
	Voc.	Syn.	Coh.	Pro.	Fun.	Flu.
1	0.430	0.403	0.336	0.681	0.441	0.453
2	0.485	0.467	0.447	0.758	0.558	0.572
3	0.507	0.494	0.502	0.788	0.613	0.627
4	0.518	0.508	0.535	0.803	0.644	0.659
5	0.526	0.517	0.557	0.813	0.664	0.680
6	0.530	0.524	0.572	0.820	0.678	0.694
7	0.534	0.528	0.584	0.824	0.689	0.705
8	0.537	0.532	0.593	0.828	0.697	0.713
9	0.539	0.534	0.600	0.831	0.703	0.719
10	0.541	0.536	0.606	0.833	0.709	0.725



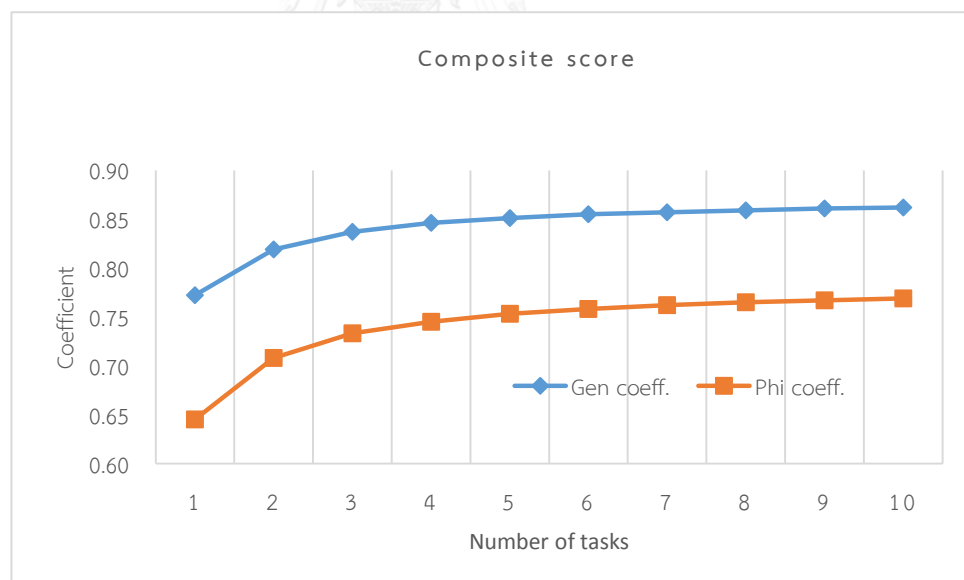
ภาพที่ 4.10 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสมบูรณของเกณฑ์การให้คะแนนการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนภาระงาน

ตารางที่ 4.38 คะแนนรวมของเกณฑ์การให้คะแนนการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนภาระงาน

ภาระงาน (เรื่อง)	composite score (ผู้ประเมิน=4 คน)			
	Relative error	Absolute error	Gen coeff.	Phi coeff.
1	0.022	0.041	0.772	0.645
2	0.017	0.031	0.819	0.708
3	0.015	0.027	0.837	0.733
4	0.014	0.026	0.846	0.745
5	0.013	0.025	0.851	0.753
6	0.013	0.024	0.855	0.758
7	0.013	0.023	0.857	0.762
8	0.012	0.023	0.859	0.765
9	0.012	0.023	0.861	0.767
10	0.012	0.023	0.862	0.769



ภาพที่ 4.11 ความคลาดเคลื่อนของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนภาระงาน



ภาพที่ 4.12 คะแนนรวมของเกณฑ์การให้คะแนนทักษะการพูดแบบแยกส่วนเมื่อเพิ่มจำนวนภาระงาน

3.3.2 รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้ภาระงานการพูดภาษาอังกฤษ

การประเมินการพูดในภาพรวมเป็นการประเมินผู้เรียนตลอดทั้งการจัดการเรียนการสอน โดยเก็บรวบรวมคะแนนจากภาระงานทั้ง 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และ 4) การบรรยายภาพ เพื่อศึกษาความเหมาะสมของจำนวนผู้ประเมินที่จะใช้ในการประเมินนักเรียนเพื่อให้ได้ความน่าเชื่อถือในการประเมินการพูด ซึ่งในการศึกษาค้นคว้าครั้งนี้ใช้ผลของ D-study ของรูปแบบ $p^* \times t^* \times r^*$ Design โดยในแต่ละภาระงานใช้เกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) ประกอบด้วย 6 ด้าน ได้แก่ คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุเป้าหมาย และความคล่องแคล่ว ใช้ผู้ประเมินจำนวน 4 คน ซึ่งการศึกษาค้นคว้าครั้งนี้มุ่งศึกษาจำนวนผู้ประเมินที่เหมาะสมว่าจะต้องมีจำนวนเท่าใดจึงจะเพียงพอ และจะใช้รูปแบบการให้คะแนนแบบใดจึงมีความเป็นไปได้ ประหยัดค่าใช้จ่าย และยังคงมีความน่าเชื่อถือในการประเมินหรือมีความเที่ยงที่สูงและเป็นที่ยอมรับได้ ซึ่งการเปรียบเทียบได้เปรียบเทียบใน 2 ประเด็น คือ 1) การศึกษาจำนวนผู้ประเมินที่เหมาะสม และ 2) การเปรียบเทียบรูปแบบการให้คะแนนของผู้ประเมิน ซึ่งผลการศึกษาพบดังนี้

3.3.2.1 การเปรียบเทียบจำนวนผู้ประเมินที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษ

ผลการเปรียบเทียบจำนวนผู้ประเมินที่จะใช้ในการประเมินการพูดภาษาอังกฤษโดยใช้ภาระงานเป็นตัวแปรในการศึกษา จำนวน 4 ภาระงาน พบว่า ค่าสัมประสิทธิ์การสุร่ออ้างอิงเชิงสัมพันธ์ของแต่ละภาระงานมีค่าสูงขึ้นเมื่อจำนวนผู้ประเมินเพิ่มขึ้น นอกจากนี้ยังพบว่าในการประเมินการพูดภาษาอังกฤษโดยใช้ภาระงาน 4 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ มีค่าความเที่ยงหรือความน่าเชื่อถือในการประเมินในระดับที่สูง (.80) โดยสามารถใช้ผู้ประเมินจำนวน 2 คนเท่านั้น ดังตารางที่ 4.39 และเมื่อพิจารณาค่าสัมประสิทธิ์การสุร่ออ้างอิงเชิงสัมพันธ์ พบผลการศึกษาที่สอดคล้องกัน กล่าวคือ ผลการประเมินในภาพรวมมีค่าความเชื่อถือได้หรือค่าความเที่ยงที่สูงทุกภาระงาน และในการประเมินสามารถใช้ผู้ประเมินเพียง 2 คน เท่านั้น แต่มีค่าความเที่ยงอยู่ในระดับที่เชื่อถือได้ (มากกว่า .80) ทุกภาระงาน ดังตารางที่ 4.40 และเมื่อพิจารณาคะแนนรวมของการประเมิน พบว่าในภาพรวมของการประเมินเมื่อจำนวนผู้ประเมินเพิ่มขึ้นทำให้ความคลาดเคลื่อนลดลงทั้งความคลาดเคลื่อนเชิงสัมพันธ์และความคลาดเคลื่อนเชิงสัมบูรณ์ ส่วนค่าความเที่ยงของการประเมินพบว่าเมื่อเพิ่มจำนวนผู้ประเมินมากขึ้นทำให้ค่าความเที่ยงสูงขึ้น และพบผลที่สอดคล้องกันทั้งค่าสัมประสิทธิ์การสุร่ออ้างอิงเชิงสัมพันธ์ และค่าสัมประสิทธิ์การสุร่ออ้างอิงเชิงสัมบูรณ์ ที่พบว่าหากต้องการผลการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษา

ปีที่ 6 โดยใช้ภาระงานการพูดฯ 4 ภาระงาน และใช้เกณฑ์การให้คะแนนแบบแยกส่วนมีความเที่ยงในระดับที่สูงและมีความเชื่อถือได้ ผู้บริหารการประเมินสามารถใช้ผู้ประเมินอย่างน้อยที่สุดเพียง 2 คน เท่านั้น และยังคงทำให้การประเมินมีความเที่ยงในระดับที่เชื่อถือได้ ดังตารางที่ 4.41

ตารางที่ 4.39 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์เมื่อจำนวนผู้ประเมินต่างกัน

ผู้ประเมิน (คน)	Generalizability coefficient			
	สัมภาษณ์	นำเสนอปากเปล่า	เล่านิทาน	บรรยายภาพ
2	0.872	0.929	0.823	0.840
3	0.905	0.950	0.875	0.887
4	0.923	0.960	0.903	0.912

ตารางที่ 4.40 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์เมื่อจำนวนผู้ประเมินต่างกัน

ผู้ประเมิน (คน)	Phi coefficient			
	สัมภาษณ์	นำเสนอปากเปล่า	เล่านิทาน	บรรยายภาพ
2	0.820	0.914	0.810	0.828
3	0.862	0.938	0.865	0.878
4	0.885	0.950	0.895	0.905

ตารางที่ 4.41 คะแนนรวมของการประเมินเมื่อจำนวนผู้ประเมินต่างกัน

ผู้ประเมิน (คน)	Composite score			
	Relative error	Absolute error	Gen coeff.	Phi coeff.
2	0.006	0.007	0.939	0.925
3	0.004	0.005	0.958	0.947
4	0.003	0.004	0.967	0.959

3.3.2.2 การเปรียบเทียบรูปแบบการให้คะแนนของผู้ประเมิน

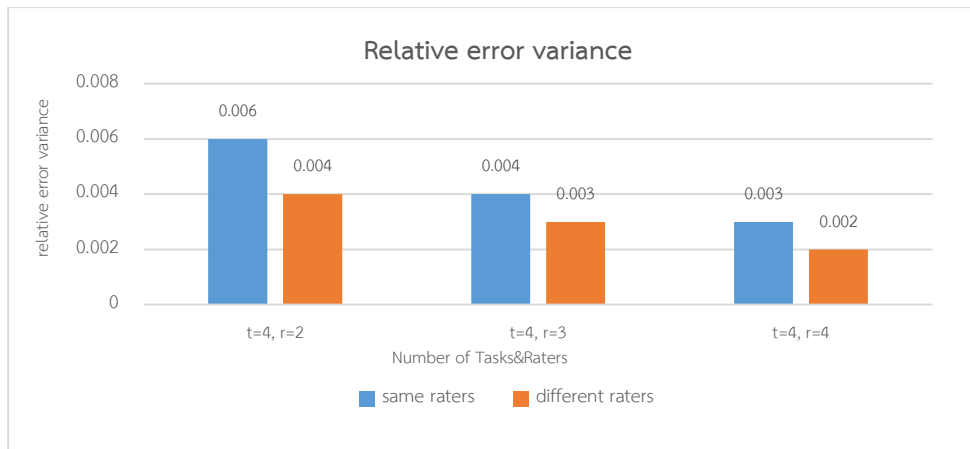
การเปรียบเทียบรูปแบบการให้คะแนนของผู้ประเมิน เป็นการศึกษาถึงวิธีการให้คะแนนของผู้ประเมินเพื่อทราบว่ามียุทธวิธีรูปแบบการให้คะแนนอย่างไรที่จะสามารถประหยัดเวลาในการประเมินและเป็นการประหยัดค่าใช้จ่ายในการจ่ายค่าตอบแทนผู้เชี่ยวชาญ แต่ยังคงมีระดับความน่าเชื่อถือในการวัดที่เชื่อถือได้และมีความเที่ยงในการวัดในระดับที่สูง (มากกว่า .80 ขึ้น) หรือในระดับที่ผู้บริหารการประเมินต้องการได้ โดยในการศึกษาครั้งนี้ใช้ผลของ D-study ของรูปแบบ $p^{\circ} \times t^{\circ} \times r^{\circ}$ Design ซึ่งในการศึกษาครั้งนี้ทำการเปรียบเทียบรูปแบบการให้คะแนนของผู้ประเมิน 2 รูปแบบ คือ 1) ผู้ประเมินทำการประเมินทุกภาระงาน โดยใช้รูปแบบ $p^{\circ} \times t^{\circ} \times r^{\circ}$ Design 2) ผู้ประเมิน

ไม่ต้องทำการประเมินทุกภาระงาน หรือแต่ละภาระงานจะถูกประเมินโดยผู้ประเมินต่างกัน โดยใช้รูปแบบ $p \times t \times r$ Design

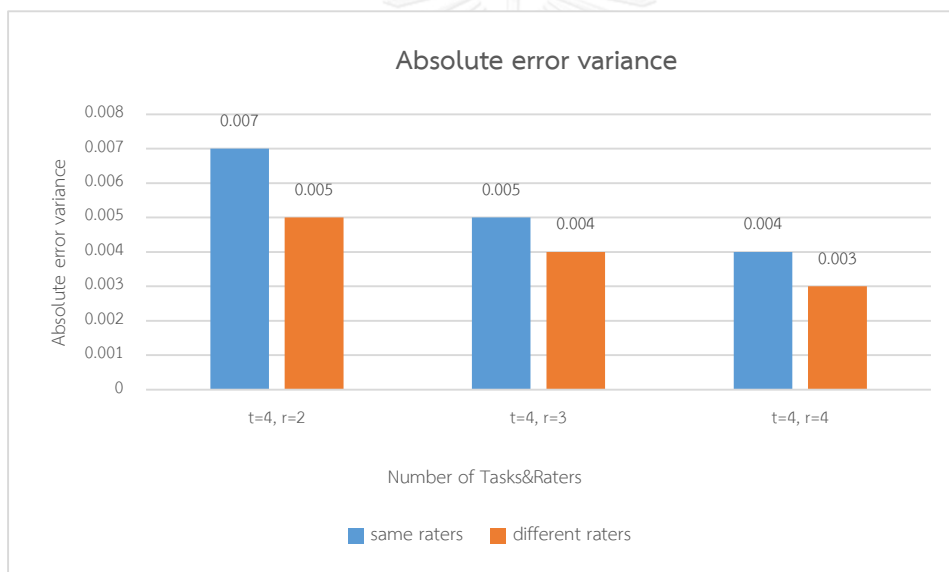
ผลการเปรียบเทียบรูปแบบการประเมินของผู้ประเมินในแต่ละภาระงานที่ต่างกัน พบว่าเมื่อเพิ่มจำนวนผู้ประเมินมากขึ้นความคลาดเคลื่อนทั้งเชิงสัมพัทธ์ และความคลาดเคลื่อนเชิงสัมบูรณ์ลดลงทั้งสองรูปแบบ แต่อย่างไรก็ตามพบว่าในรูปแบบการประเมินที่ใช้ผู้ประเมินต่างกันมีความคลาดเคลื่อนต่ำกว่าใช้ผู้ประเมินเหมือนกัน ดังนั้นเมื่อพิจารณาค่าความเที่ยงหรือความน่าเชื่อถือของการประเมินจึงพบว่ารูปแบบการประเมินที่ใช้ผู้ประเมินต่างกันมีค่าความเที่ยงสูงกว่า และพบว่าสามารถใช้ผู้ประเมินเพียง 2 คนเท่านั้น ความเที่ยงของประเมินยังคงมีค่าสูง (มากกว่า .80) ดังตารางที่ 4.42 และภาพที่ 4.13-4.16 จากผลการศึกษาดังกล่าวข้างต้นสามารถสรุปได้ว่าการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้ 4 ภาระงาน และเกณฑ์การให้คะแนนแบบแยกส่วนแบบ 6 องค์ประกอบสามารถใช้ผู้ประเมินเพียง 2 คน และสามารถใช้รูปแบบการประเมินที่ในแต่ละภาระงานใช้ผู้ประเมินที่ต่างกัน เพื่อเป็นการประหยัดเวลาและประหยัดค่าใช้จ่ายในการบริหารจัดการ และยังคงมีความน่าเชื่อถือในการประเมินหรือมีความเที่ยงที่สูงได้

ตารางที่ 4.42 คะแนนรวมสำหรับรูปแบบการให้คะแนนที่ต่างกันของผู้ประเมิน

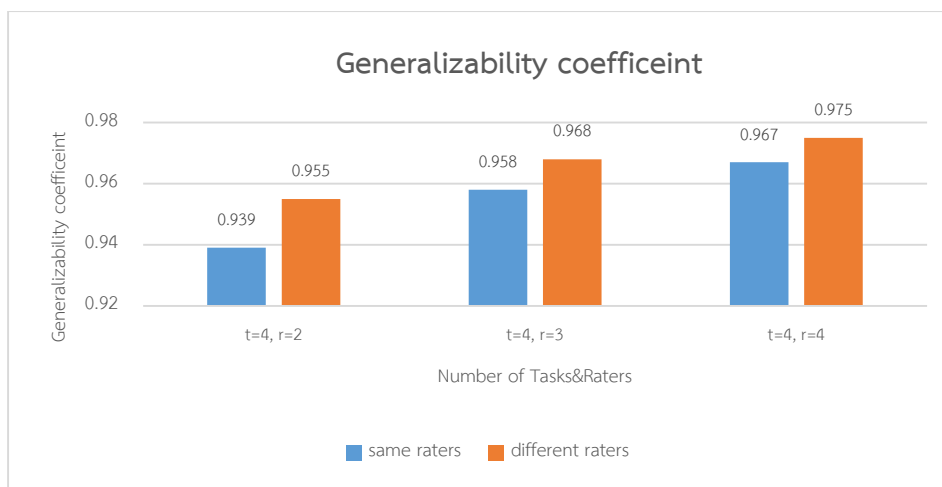
จำนวน ภาระงาน/ผู้ ประเมิน	$p \times t \times r$ design (same rater)				$p \times t \times r$ design (different rater)			
	Rel. error	Abs. error	G coeff.	Phi coeff.	Rel. error	Abs. error	G coeff.	Phi coeff.
t=4, r=2	0.006	0.007	0.939	0.925	0.004	0.005	0.955	0.945
t=4, r=3	0.004	0.005	0.958	0.947	0.003	0.004	0.968	0.961
t=4, r=4	0.003	0.004	0.967	0.959	0.002	0.003	0.975	0.969



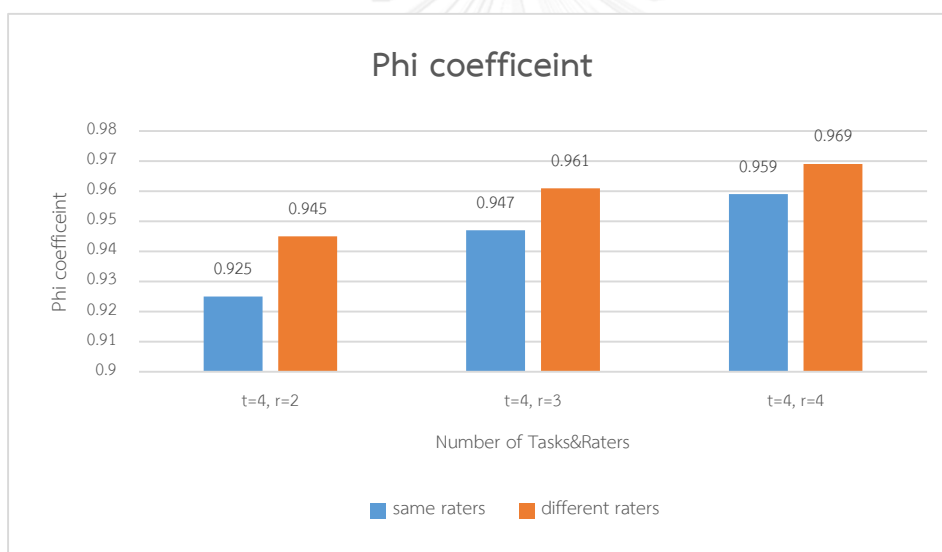
ภาพที่ 4.13 ความคลาดเคลื่อนเชิงสัมพัทธ์ของการประเมินที่ใช้ผู้ประเมินต่างกัน



ภาพที่ 4.14 ความคลาดเคลื่อนเชิงสมบูรณ์ของการประเมินที่ใช้ผู้ประเมินต่างกัน



ภาพที่ 4.15 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ของการประเมินที่ใช้ผู้ประเมินต่างกัน



ภาพที่ 4.16 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ของการประเมินที่ใช้ผู้ประเมินต่างกัน

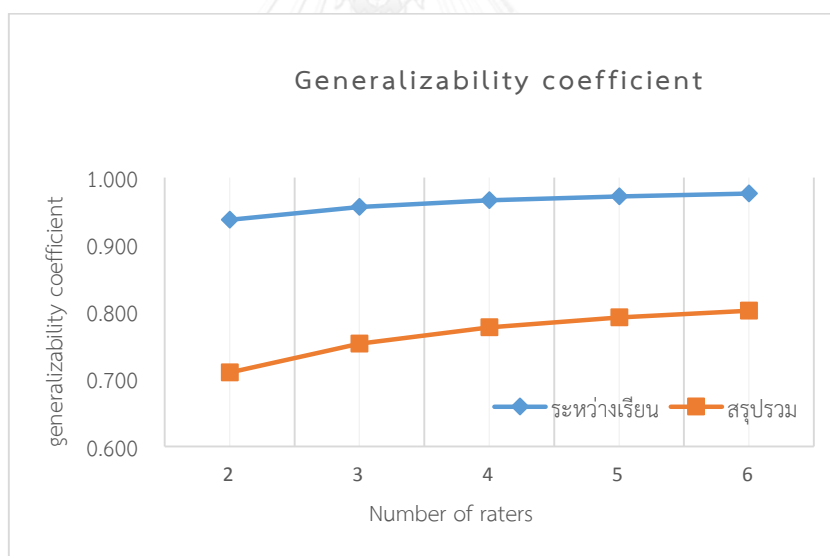
3.3.3 รูปแบบการประเมินทักษะการพูดภาษาอังกฤษโดยการประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment)

การประเมินการพูดในภาพรวมเป็นประเมินภาษาอังกฤษของนักเรียนโดยใช้แฟ้มสะสมงาน แบ่งระยะการประเมินออกเป็นประเมินระหว่างเรียน และการประเมินสรุปรวม โดยการประเมินในระหว่างเรียนใช้ภาระงาน จำนวน 3 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า และ 3) การเล่านิทาน และการประเมินสรุปรวมใช้ภาระงาน จำนวน 1 ภาระงาน คือ 4) การบรรยายภาพ

เพื่อทำการศึกษาวาดต้องใช้จำนวนผู้ประเมินจำนวนเท่าใดจึงจะเหมาะสม และใช้ภาระงานในการประเมินจำนวนเท่าใดเมื่อทำการรวมภาระงานแต่ละภาระงานในการประเมินระหว่างเรียนเข้าด้วยกัน และในส่วนของ การประเมินสรุปต้องใช้ภาระงานเท่าใดและผู้ประเมินกี่คน โดยใช้รูปแบบ $p \times t \times r$ Design โดยการศึกษาเปรียบเทียบรูปแบบการประเมินในภาพรวมนี้ทำการเปรียบเทียบเมื่อเพิ่มจำนวนผู้ประเมิน และเพื่อจำนวนภาระงาน ผลการเปรียบเทียบรูปแบบการประเมินโดยการเพิ่มจำนวนผู้ประเมินโดยใช้ ภาระงานจำนวนทั้งสิ้นตลอดการประเมินจำนวน 4 ภาระงาน แบ่งเป็นภาระงานที่ใช้ระหว่างเรียน 3 ภาระงาน และภาระงานสรุปรวม จำนวน 1 ภาระงาน พบว่า เมื่อเพิ่มจำนวนผู้ประเมินมากขึ้น ทำให้ค่าความเที่ยงเพิ่มขึ้นทั้งสองระยะของการประเมิน กล่าวคือ ค่าสัมประสิทธิ์ความน่าเชื่อถือเชิงสัมพัทธ์ของการประเมินระหว่างเรียน (formative assessment) มีค่าความเที่ยงอยู่ในระดับสูง คือ .937 โดยสามารถใช้ผู้ประเมินจำนวนเพียง 2 คนเท่านั้น ในขณะที่การประเมินสรุปรวม (summative assessment) ต้องใช้ผู้ประเมินถึง 6 คน จึงจะทำให้การประเมินมีความเที่ยงในระดับที่ สูง คือ .802 ดังตารางที่ 4.43 และภาพที่ 4.17 แต่หากทำการเพิ่มจำนวนภาระงานในการประเมิน โดยในการศึกษาคั้งนี้ได้ทำเพิ่มภาระงานเฉพาะในการประเมินสรุปรวมเนื่องจากการหาคุณภาพ ของการประเมินพบว่า การประเมินสรุปรวมมีความคลาดเคลื่อนสูงกว่าการประเมินระหว่างเรียน (ดังตารางที่ 4.43) โดยใช้ภาระงานในการประเมินในภาพรวมทั้งสิ้นจำนวน 5 ภาระงาน เมื่อเพิ่มภาระงาน ในการประเมินเป็น 2 ภาระงาน และการประเมินระหว่างเรียนยังคงเป็น 3 ภาระงานดังเดิม และ ทำการเพิ่มจำนวนผู้ประเมิน พบว่า ในการประเมินสรุปรวมสามารถใช้ผู้ประเมินจำนวนเพียง 3 คน เท่านั้นที่จะทำให้ความเที่ยงในการประเมินอยู่ในระดับสูง (.824) ดังตารางที่ 4.43 สอดคล้องกับ ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ที่พบว่าเมื่อเพิ่มจำนวนผู้ประเมินมากขึ้นจะมีค่าความเที่ยง สูงขึ้น โดยในการประเมินที่ใช้ภาระงานจำนวนทั้งสิ้น 4 ภาระงาน ผลการศึกษาพบว่าในการประเมิน ระหว่างเรียน สามารถใช้ผู้ประเมินเพียง 2 คน เท่านั้นก็เพียงพอที่จะทำให้ค่าความเที่ยงในระดับที่ เชื่อถือได้ (.915) ในขณะที่การประเมินสรุปรวมต้องใช้ภาระงานมากกว่า 6 เรื่อง ดังตารางที่ 4.44 และภาพที่ 4.18 จึงจะได้ความเที่ยงที่ระดับสูง (มากกว่า .80) แต่หากเป็นการเพิ่มจำนวนภาระงาน ในการประเมินสรุปรวมอีก 1 ภาระงาน เป็น 2 ภาระงาน ซึ่งจะมีจำนวนภาระงานทั้งการประเมิน จำนวนทั้งสิ้น 5 เรื่อง พบว่า ในการประเมินสรุปรวมสามารถใช้ผู้ประเมินเพียง 3 คนเท่านั้น ทำให้การ ประเมินยังคงมีความเที่ยงในระดับสูง (.817) ดังตารางที่ 4.44

ตารางที่ 4.43 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์รูปแบบการประเมินทักษะการพูด
ภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น

จำนวนภาระงาน (เรื่อง)			จำนวนผู้ประเมิน (คน)	Generalizability coefficient	
รวม	ระหว่างเรียน	สรุปรวม		ระหว่างเรียน	สรุปรวม
4	3	1	2	0.937	0.710
4	3	1	3	0.956	0.753
4	3	1	4	0.966	0.777
4	3	1	5	0.972	0.792
4	3	1	6	0.976	0.802
5	3	2	2	0.937	0.783
5	3	2	3	0.956	0.824
5	3	2	4	0.966	0.847
5	3	2	5	0.972	0.861
5	3	2	6	0.976	0.871



ภาพที่ 4.17 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ของรูปแบบการประเมินทักษะการพูด
ภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น

ตารางที่ 4.44 ค่าสัมประสิทธิ์การสรุปร่างอิงเชิงสัมบูรณ์ของรูปแบบการประเมินทักษะการพูด
ภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น

รวม	จำนวนภาระงาน		จำนวนผู้ประเมิน	Phi coefficient	
	ระหว่างเรียน	สรุปรวม		ระหว่างเรียน	สรุปรวม
4	3	1	2	0.915	0.697
4	3	1	3	0.941	0.741
4	3	1	4	0.954	0.765
4	3	1	5	0.962	0.781
4	3	1	6	0.968	0.791
5	3	2	2	0.915	0.774
5	3	2	3	0.941	0.817
5	3	2	4	0.954	0.840
5	3	2	5	0.962	0.855
5	3	2	6	0.968	0.865



ภาพที่ 4.18 ค่าสัมประสิทธิ์การสรุปร่างอิงเชิงสัมบูรณ์ของรูปแบบการประเมินทักษะการพูด
ภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น

เมื่อพิจารณาผลของคะแนนรวม (composite score) พบว่า เมื่อเพิ่มจำนวนผู้ประเมิน
มากขึ้นค่าความคลาดเคลื่อนทั้งความคลาดเคลื่อนเชิงสัมพัทธ์ และความคลาดเคลื่อนเชิงสัมบูรณ์

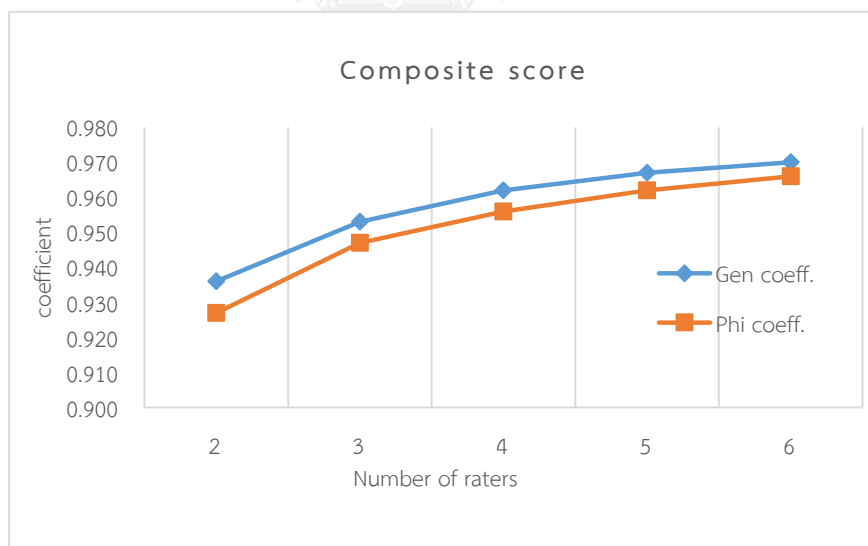
จะลดลง และจะเริ่มคงที่เมื่อใช้ผู้ประเมิน 4 คน ใช้ภาระงานจำนวนทั้งสิ้น 4 ภาระงาน และเมื่อพิจารณาผลการประเมินในภาพรวมพบว่า ค่าความเที่ยงอยู่ในระดับสูง (มากกว่า .80) เมื่อใช้ผู้ประเมินตั้งแต่ 2 คนขึ้นไป ทั้งค่าสัมประสิทธิ์การสรุปร่างอิงเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปร่างอิงเชิงสัมบูรณ์ ดังตารางที่ 4.45 และภาพที่ 4.19-4.20 ตามลำดับ จากผลการศึกษารูปแบบการประเมินในภาพรวมดังกล่าวข้างต้น สามารถสรุปได้ว่าการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ที่มีรูปแบบการประเมินที่แบ่งเป็นการประเมินระหว่างเรียนและการประเมินสรุปรวม เป็นการประเมินที่สามารถใช้ภาระงานจำนวน 4 ภาระงาน โดยการประเมินระหว่างเรียนใช้ 3 ภาระงาน และการประเมินสรุปรวมใช้ 1 ภาระงาน และสามารถให้ผู้ประเมินจำนวนเพียง 2 คน ผลการประเมินยังคงมีความเที่ยงในระดับสูง (มากกว่า .80) ซึ่งเป็นรูปแบบการประเมินที่ความเป็นไปได้ในทางปฏิบัติ โรงเรียนสามารถนำมาใช้ได้จริง

ตารางที่ 4.45 คะแนนรวมของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น

จำนวนภาระงาน (เรื่อง)			จำนวนผู้ประเมิน (คน)	Rel Error	Abs. Error	Gen coeff.	Phi coeff.
รวม	ระหว่างเรียน	สรุปรวม					
4	3	1	2	0.006	0.007	0.936	0.927
4	3	1	3	0.004	0.005	0.953	0.947
4	3	1	4	0.003	0.004	0.962	0.956
4	3	1	5	0.003	0.003	0.967	0.962
4	3	1	6	0.003	0.003	0.970	0.966
5	3	2	2	0.006	0.006	0.931	0.928
5	3	2	3	0.004	0.005	0.948	0.946
5	3	2	4	0.004	0.004	0.957	0.955
5	3	2	5	0.003	0.003	0.963	0.961
5	3	2	6	0.003	0.003	0.966	0.965



ภาพที่ 4.19 ความคลาดเคลื่อนของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น



ภาพที่ 4.20 คะแนนรวมของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษเมื่อจำนวนผู้ประเมินและภาระงานเพิ่มขึ้น

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิจัยเพื่อพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน โดยมุ่งพัฒนาขั้นตอนการดำเนินการและองค์ประกอบของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 เพื่อพัฒนาเกณฑ์การให้คะแนนในการประเมินทักษะการพูดภาษาอังกฤษทั้งที่ใช้ในการประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) และเพื่อเปรียบเทียบรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือพหุตัวแปร โดยมุ่งศึกษาเงื่อนไขที่เกี่ยวข้อง 3 แห่ง ดังนี้ ผู้ประเมิน (raters) ภาระงานการพูด (tasks) และเกณฑ์การให้คะแนนภาระงานการพูดแบบแยกส่วน (analytic rating scale)

การวิจัยครั้งนี้มีการดำเนินการ 4 ระยะ คือ ระยะที่ 1 การพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ระยะที่ 2 การทดลองใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ระยะที่ 3 การวิเคราะห์ผลการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนระดับประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ระยะที่ 4 การพัฒนาคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

กลุ่มตัวอย่างที่ใช้การวิจัยครั้งนี้ ได้แก่ นักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 1 ห้อง ครูผู้สอนวิชาภาษาอังกฤษ ชั้นประถมศึกษาปีที่ 6 และผู้ประเมิน ที่ทำหน้าที่เป็นผู้ประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 4 คน

เครื่องมือที่ใช้ในการพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานในส่วนของพัฒนาเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ ภาระงานที่ใช้ในการประเมิน ขั้นตอนของแฟ้มสะสมงาน องค์ประกอบของแฟ้มสะสมงาน ใช้การสัมภาษณ์โดยใช้แบบสัมภาษณ์แบบกึ่งมีโครงสร้าง และการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้อง ส่วนการหาคุณภาพของเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ และคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ใช้แบบสอบถาม การวิเคราะห์ข้อมูลในการศึกษาคุณภาพของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษ และการเปรียบเทียบรูปแบบการประเมินที่มีความเหมาะสมใช้การวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิง

ความน่าเชื่อถือแบบพหุตัวแปรโดยใช้โปรแกรม mGENOVA version 2.1 และวิเคราะห์ความก้าวหน้าของทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำชนิดทางเดียว (one-way repeated ANOVA measurement)

สรุปผลวิจัย

1. ผลการพัฒนารูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

1.1 ผลการพัฒนาขั้นตอนและส่วนประกอบของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ผลการพัฒนาแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ในการศึกษาครั้งนี้ผู้วิจัยได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง รวมทั้งทำการสัมภาษณ์ผู้ทรงคุณวุฒิทั้งทางภาษาอังกฤษ และทางการวัดและประเมินผลการศึกษา แล้วนำมาสังเคราะห์ ปรากฏผลการพัฒนาจากผลการสังเคราะห์เอกสารงานวิจัยที่เกี่ยวข้องและการสัมภาษณ์ พบว่ามีขั้นตอนในการดำเนินการแฟ้มสะสมงานการประเมินภาษาอังกฤษสำหรับนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 7 ขั้นตอน ได้แก่ 1) ขั้นการวางแผน 2) ขั้นการเตรียมความพร้อมนักเรียน 3) ขั้นการเก็บรวบรวมหลักฐาน 4) ขั้นการติดตามความก้าวหน้า 5) ขั้นการปรับปรุงผลงาน 6) ขั้นการสะท้อนคิด และ 7) ขั้นการจัดแสดงผลงาน ส่วนประกอบของแฟ้มสะสมงานประกอบด้วย 3 ส่วน ได้แก่ 1) ส่วนนำ 2) ส่วนเนื้อหา และ 3) ส่วนภาคผนวก ทำนองเดียวกันกับการพัฒนาเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานในการศึกษาครั้งนี้ผู้วิจัยได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง รวมทั้งทำการสัมภาษณ์ผู้ทรงคุณวุฒิทั้งทางภาษาอังกฤษ และทางการวัดและประเมินผลการศึกษา แล้วนำมาสังเคราะห์ ปรากฏผลการพัฒนาจากผลการสังเคราะห์เอกสารงานวิจัยที่เกี่ยวข้องและการสัมภาษณ์ พบว่า เกณฑ์การให้คะแนนภาระงานที่ใช้การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่ใช้ในการประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment)

1.2 ผลของการพัฒนาเกณฑ์การให้คะแนน

ภาระงานการพูดในการประเมินจำนวนทั้งสิ้น 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และ 4) การบรรยายภาพ แต่ละภาระงานมีเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) และมีขอบเขตเนื้อหาในการประเมิน 6 ด้าน ได้แก่ คำศัพท์ (vocabulary) การสร้างประโยค (syntax) การเชื่อมโยงเนื้อหา (cohesion) การออกเสียง

(pronunciation) บรรลุวัตถุประสงค์ (ideational function) และความคล่องแคล่ว (fluency) ให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี

2. การศึกษาสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปร

2.1 ผลการตรวจสอบคุณภาพของเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ

ผลการตรวจสอบความเที่ยงของเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุวัตถุประสงค์ และความคล่องแคล่ว พบว่า ความแปรปรวนของนักเรียน (person, p) มีขนาดใหญ่ที่สุด โดยพบว่า มีค่าความแปรปรวนคิดเป็นร้อยละ 51.84 50.83 53.46 80.33 64.37 และ 65.90. (คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุวัตถุประสงค์ และความคล่องแคล่ว ตามลำดับ) เกณฑ์การให้คะแนนด้านที่มีผลต่อความแปรปรวนของคะแนนรวมของคะแนนการพูดมากที่สุดคือ การออกเสียง คิดเป็นร้อยละ 25.66 รองลงมาคือ ความคล่องแคล่ว การบรรลุวัตถุประสงค์ การสร้างประโยค คำศัพท์ และการเชื่อมโยงเนื้อหา (คิดเป็นร้อยละ 17.66 15.12 15.10 13.68 และ 12.77 ตามลำดับ) ความเที่ยงหรือความเชื่อถือได้ของคะแนนรวม (composite score) พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์มีค่าอยู่ในระดับสูง (มากกว่า .80) โดยมีค่าเท่ากับ .846 ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ (Phi coefficient) พบว่ามีความเชื่อถือได้ในระดับค่อนข้างสูง คือ .745

2.2 การตรวจสอบคุณภาพของภาระงานการประเมินทักษะการพูดภาษาอังกฤษ

ผลการตรวจสอบพบว่า ความแปรปรวนของนักเรียน (person, p) มีขนาดใหญ่ที่สุด โดยพบว่า มีค่าความแปรปรวนคิดเป็นร้อยละ 81.95 91.42 80.94 และ 85.22 (การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ ตามลำดับ) ภาระงานที่มีผลต่อความแปรปรวนของคะแนนรวมของคะแนนการพูดมากที่สุดคือ ภาระงานการนำเสนอปากเปล่า คิดเป็นร้อยละ 35.48 รองลงมาคือ การสัมภาษณ์ การบรรยายภาพ และการเล่านิทาน (คิดเป็นร้อยละ 24.47 21.55 และ 18.49 ตามลำดับ) ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์พบว่า มีค่าอยู่ในระดับสูงทุกภาระงาน (มากกว่า .80) ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์พบผลคล้ายคลึงกัน คือ ค่าอยู่ในระดับสูงทุกภาระงาน (มากกว่า .80) ความเที่ยงของคะแนนรวม (composite score) พบว่า มีค่าอยู่ในระดับสูง (มากกว่า .80) ทั้งค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์คือ มีความเชื่อถือได้เท่ากับ .967 และ .959 ตามลำดับ

2.3 การตรวจสอบคุณภาพของรูปแบบการประเมินระหว่างเรียนและสรุปรวม

ในการศึกษาสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินการพูดภาษาอังกฤษโดยใช้แฟ้มสะสมงานที่ให้มีการประเมินทั้งการประเมินระหว่างเรียนและการประเมินสรุปรวม โดยในการประเมินโดยใช้แฟ้มสะสมงานในครั้งนี้แบ่งการประเมินเป็นสองระยะ คือ การประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) โดยการประเมินระหว่างเรียนใช้ 3 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และการประเมินสรุปรวมใช้ 1 ภาระงาน คือ การบรรยายภาพ ผลการตรวจสอบรูปแบบการประเมิน พบว่า ความแปรปรวนของนักเรียน (p) มีขนาดใหญ่ที่สุด โดยพบว่า มีค่าความแปรปรวนคิดเป็นร้อยละ 95.41 และ 76.55 (การประเมินระหว่างเรียน และการประเมินสรุปรวม ตามลำดับ) ค่าความเที่ยงของรูปแบบการประเมินระหว่างเรียนและการประเมินสรุปรวมที่นำมาใช้ในการประเมินทักษะการพูด พบว่า การประเมินในช่วงการประเมินระหว่างเรียนที่มีผลต่อความแปรปรวนของคะแนนรวมของคะแนนการพูดมากที่สุด คิดเป็นร้อยละ 78.98 ส่วนการประเมินสรุปรวม คิดเป็นร้อยละ 21.02 ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ พบว่า การประเมินระหว่างเรียนมีค่าอยู่ในระดับสูง คือ .966 (มากกว่า .80) ส่วนการประเมินสรุปรวมอยู่ในระดับปานกลาง (.78) ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ พบว่า การประเมินระหว่างเรียนมีค่าอยู่ในระดับสูง คือ .954 (มากกว่า .80) ส่วนการประเมินสรุปรวม อยู่ในระดับปานกลาง (.77) ความเที่ยงของคะแนนรวม (composite score) พบว่า มีความเชื่อถือได้ของการประเมินอยู่ในระดับสูง (มากกว่า .80) ทั้งค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ (Generalizability coefficient) และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ (Phi coefficient) คือ มีความเชื่อถือได้เท่ากับ .961 และ .956 ตามลำดับ

2.4 ผลการเปรียบเทียบรูปแบบการให้คะแนนของผู้ประเมิน

รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษแบบแยกส่วน (analytic rating scale) ผลการเปรียบเทียบจำนวนผู้ประเมินที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษ พบว่า เมื่อเพิ่มจำนวนผู้ประเมินให้มากขึ้นทำให้ระดับความน่าเชื่อถือ หรือความเที่ยงในการประเมินมีค่าสูงขึ้น โดยพบว่าค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์จะเพิ่มขึ้น หากใช้ผู้ประเมินเพิ่มขึ้น โดยพบว่าหากใช้ผู้ประเมินตั้งแต่ 9 คนขึ้นไปจะทำให้ค่าความเที่ยงอยู่ในระดับสูง ยกเว้นด้านคำศัพท์ที่อยู่ระดับปานกลาง (.70) แต่หากพิจารณาเป็นรายองค์ประกอบพบว่า การออกเสียงจะใช้ผู้ประเมินเพียง 3 คน เท่านั้น จะมีความเที่ยงอยู่ในระดับสูง (.80) รองลงมาคือ ความคล่องแคล่ว การบรรลุมิติวัตถุประสงค์ การเชื่อมโยงเนื้อหา และคำศัพท์ (ผู้ประเมิน จำนวน 7 8 9 และ 10 คน ตามลำดับ) ในขณะที่การสร้างประโยคต้องใช้ผู้ประเมินมากกว่า 10 คน ขึ้นไป ส่วนคะแนนรวม (composite score) ของการเพิ่มจำนวนผู้ประเมินพบว่า เมื่อเพิ่มจำนวนผู้ประเมินขึ้นจะทำให้ค่า

ความคลาดเคลื่อนเชิงสัมพัทธ์ (relative error) และความคลาดเคลื่อนเชิงสมบูรณ์ (absolute error) ลดลง ค่าความเที่ยงของการประเมินในภาพรวมพบว่าสูงขึ้นเมื่อเพิ่มจำนวนผู้ประเมินให้มากขึ้น โดยพบว่าค่าสัมประสิทธิ์การสุรุปอ้างอิงเชิงสัมพัทธ์จะมีความเที่ยงสูง (.80) หรือมีความเชื่อถือได้เมื่อใช้ผู้ประเมินเพียง 3 คนเท่านั้น สอดคล้องกับค่าสัมประสิทธิ์การสุรุปอ้างอิงเชิงสมบูรณ์ พบว่าสูงขึ้นเมื่อเพิ่มจำนวนผู้ประเมินให้มากขึ้น โดยพบว่าค่าสัมประสิทธิ์การสุรุปอ้างอิงเชิงสมบูรณ์ จะมีความเที่ยงสูง (.80) เมื่อใช้ผู้ประเมินจำนวน 6 คน การเปรียบเทียบจำนวนภาระงานที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษพบว่าเมื่อเพิ่มจำนวนภาระงานให้มากขึ้นทำให้ระดับความเที่ยง หรือความเที่ยงในการประเมินมีค่าสูงขึ้น โดยพบว่าค่าสัมประสิทธิ์การสุรุปอ้างอิงเชิงสัมพัทธ์จะเพิ่มขึ้น หากใช้จำนวนภาระงานเพิ่มขึ้น ซึ่งพบว่า ด้านการออกเสียง มีความเที่ยงอยู่ในระดับที่เชื่อถือได้หรือมีความเที่ยงสูง (.80) ถึงแม้จะใช้ภาระงานเพียง 2 ภาระงานเท่านั้น ส่วนด้านความคล่องแคล่ว การบรรลุวัตถุประสงค์ ต้องใช้ภาระงานตั้งแต่ 4 ภาระงานขึ้นไป ด้านการเชื่อมโยงเนื้อหาและคำศัพท์ ต้องใช้ภาระงานจำนวน 9 ภาระงาน และด้านการสร้างประโยคต้องใช้ภาระงานมากกว่า 10 ภาระงานขึ้นไป จึงจะทำให้ค่าความเที่ยงอยู่ในระดับปานกลาง (.70) สัมประสิทธิ์การสุรุปอ้างอิงเชิงสมบูรณ์จะเพิ่มขึ้นหากใช้ภาระงานเพิ่มขึ้น โดยพบว่า ด้านการออกเสียง จะมีความเที่ยงอยู่ในระดับสูง (.80) เมื่อใช้ภาระงานเพียง 4 ภาระงานเท่านั้น ส่วนด้านความคล่องแคล่วต้องใช้ภาระงานตั้งแต่ 7 ภาระงานขึ้นไป ด้านการบรรลุวัตถุประสงค์ ต้องใช้ภาระงานจำนวน 8 ภาระงาน และด้านการสร้างประโยค การเชื่อมโยงเนื้อหา และคำศัพท์ ต้องใช้ภาระงานมากกว่า 10 ภาระงานขึ้นไป จึงจะทำให้ค่าความเที่ยงอยู่ในระดับปานกลาง (.70) ส่วนคะแนนรวมของการประเมิน (composite score) ของการเพิ่มจำนวนผู้ประเมินพบว่า เมื่อเพิ่มจำนวนภาระงานขึ้นจะทำให้ค่าความคลาดเคลื่อนทั้งความคลาดเคลื่อนเชิงสัมพัทธ์ (relative error) และความคลาดเคลื่อนเชิงสมบูรณ์ (absolute error) ลดลง ค่าความเที่ยงพบว่ามีความน่าเชื่อถือของการประเมินในภาพรวมสูงขึ้นเมื่อเพิ่มจำนวนภาระงานให้มากขึ้น โดยพบว่าค่าสัมประสิทธิ์การสุรุปอ้างอิงเชิงสัมพัทธ์จะมีความเที่ยงสูง (.80) เมื่อใช้ผู้ประเมินเพียง 2 คนเท่านั้น สอดคล้องกับค่าสัมประสิทธิ์การสุรุปอ้างอิงเชิงสมบูรณ์พบว่าสูงขึ้นเมื่อเพิ่มจำนวนภาระงานให้มากขึ้น โดยพบว่าค่าสัมประสิทธิ์การสุรุปอ้างอิงเชิงสมบูรณ์จะมีความเที่ยงสูงปานกลาง (.70) เมื่อใช้ผู้ประเมินจำนวน 2 คน แต่หากต้องการให้ความเที่ยงอยู่ในระดับสูงต้องใช้ภาระมากกว่า 10 ภาระงาน

รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้ภาระงานการพูดภาษาอังกฤษ โดยเก็บรวบรวมคะแนนจากภาระงานทั้ง 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ 2) การนำเสนอปากเปล่า 3) การเล่านิทาน และ 4) การบรรยายภาพ ผลการเปรียบเทียบจำนวนผู้ประเมินที่จะใช้ในการประเมินการพูดภาษาอังกฤษโดยใช้ภาระงานเป็นตัวแปรในการศึกษาจำนวน 4 ภาระงาน พบว่า ค่าสัมประสิทธิ์การสุรุปอ้างอิงเชิงสัมพัทธ์ของแต่ละภาระงานมีค่าเพิ่มขึ้นเมื่อเพิ่มจำนวนผู้ประเมินขึ้น และพบว่าในการประเมินการพูดภาษาอังกฤษโดยใช้ภาระงาน 4 ภาระงาน

ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่นนิทาน และการบรรยายภาพ มีค่าความเที่ยงในระดับที่สูง (.80) โดยสามารถใช้ผู้ประเมินจำนวน 2 คนเท่านั้น ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมบูรณ์พบว่า ผลการประเมินในภาพรวมมีค่าความเที่ยงที่สูงทุกภาระงาน และในการประเมินสามารถใช้ผู้ประเมินเพียง 2 คน เท่านั้น แต่มีค่าความเที่ยงอยู่ในระดับที่สูง (มากกว่า .80) ทุกภาระงานเช่นกัน คะแนนรวมพบว่าภาพรวมของการประเมินเมื่อเพิ่มจำนวนผู้ประเมินขึ้นทำให้ความคลาดเคลื่อนลดลงทั้งความคลาดเคลื่อนเชิงสัมพัทธ์และความคลาดเคลื่อนเชิงสัมบูรณ์ ส่วนค่าความเที่ยงพบว่าเมื่อเพิ่มจำนวนผู้ประเมินมากขึ้นทำให้ค่าความเที่ยงสูงขึ้น และพบผลที่สอดคล้องกันทั้งค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปร่างเชิงสัมบูรณ์ พบว่าในการประเมินที่มีความเที่ยงในระดับที่สูงสามารถใช้ผู้ประเมินอย่างน้อยที่สุดเพียง 2 คนเท่านั้น การเปรียบเทียบรูปแบบการให้คะแนนของผู้ประเมินโดยในการศึกษาครั้งนี้ใช้ผลของ D-study ของรูปแบบ $p^{\circ} \times t^{\circ} \times r^{\circ}$ และ $p^{\circ} \times t^{\circ} \times r^{\circ}$ Design คือ การเปรียบเทียบระหว่างการให้ผู้ประเมินทำการประเมินทุกภาระงานกับการประเมินที่ใช้ผู้ประเมินต่างกันในแต่ละภาระงาน ผลการเปรียบเทียบรูปแบบการประเมินของผู้ประเมินในแต่ละภาระงานที่ต่างกัน พบว่า เมื่อเพิ่มจำนวนผู้ประเมินมากขึ้นความคลาดเคลื่อนทั้งเชิงสัมพัทธ์ และความคลาดเคลื่อนเชิงสัมบูรณ์ ลดลงทั้งสองรูปแบบ พบว่าในรูปแบบการประเมินที่ใช้ผู้ประเมินต่างกันในแต่ละภาระงานมีความคลาดเคลื่อนต่ำกว่า ดังนั้นเมื่อพิจารณาค่าความเที่ยงพบว่า รูปแบบการประเมินที่ใช้ผู้ประเมินต่างกันมีค่าความเที่ยงสูงกว่าและพบว่าสามารถใช้ผู้ประเมินเพียง 2 คนเท่านั้น แต่ความเที่ยงยังคงอยู่ในระดับสูง (มากกว่า .80)

รูปแบบการประเมินทักษะการพูดภาษาอังกฤษโดยการประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปร่างรวม (summative assessment) ผลการเปรียบเทียบรูปแบบการประเมินโดยการเพิ่มจำนวนผู้ประเมินโดยใช้ภาระงานจำนวนทั้งสิ้นตลอดการประเมินจำนวน 4 ภาระงาน แบ่งเป็นภาระงานที่ใช้ระหว่างเรียน 3 ภาระงาน และภาระงานสรุปร่างรวม จำนวน 1 ภาระงาน พบว่า เมื่อเพิ่มจำนวนผู้ประเมินมากขึ้นทำให้ค่าความเที่ยงเพิ่มขึ้นทั้งสองระยะของการประเมิน กล่าวคือ ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมพัทธ์ของการประเมินระหว่างเรียน (formative assessment) มีค่าความเที่ยงอยู่ในระดับสูง (.937) โดยสามารถใช้ผู้ประเมินจำนวนเพียง 2 คนเท่านั้น ในขณะที่การประเมินสรุปร่างรวม (summative assessment) ต้องใช้ผู้ประเมินถึง 6 คน จึงจะทำให้การประเมินมีความเที่ยงในระดับที่สูง (.802) แต่หากทำการเพิ่มจำนวนภาระงานในการประเมิน โดยเพิ่มภาระงานเฉพาะในการประเมินสรุปร่างรวม โดยใช้ภาระงานในการประเมินในภาพรวมทั้งสิ้นจำนวน 5 ภาระงาน โดยเพิ่มภาระงานในการประเมินเป็น 2 ภาระงาน และการประเมินระหว่างเรียนยังคงเป็น 3 ภาระงานดั้งเดิม และทำการเพิ่มจำนวนผู้ประเมิน พบว่า ในการประเมินสรุปร่างรวมสามารถใช้ผู้ประเมินจำนวนเพียง 2 คนเท่านั้น ที่จะทำให้ความเที่ยงในการประเมินอยู่ในระดับสูง (.824) ค่าสัมประสิทธิ์การสรุปร่างเชิงสัมบูรณ์พบว่าเมื่อเพิ่มจำนวนผู้ประเมินขึ้นจะมีค่าความเที่ยง

สูงขึ้น และพบว่าในการประเมินระหว่างเรียน สามารถใช้ผู้ประเมินเพียง 2 คน ค่าความเที่ยงในระดับที่เชื่อถือได้ (.915) ในขณะที่การประเมินสรุปรวมต้องใช้ภาระงานมากกว่า 6 ภาระงาน จึงจะได้ความเที่ยงที่ระดับสูง (มากกว่า .80) แต่หากเป็นการเพิ่มจำนวนภาระงานในการประเมินสรุปรวมอีก 1 ภาระงาน เป็น 2 ภาระงาน และจะมีจำนวนภาระงานทั้งการประเมินจำนวนทั้งสิ้น 5 ภาระงาน พบว่าในการประเมินสรุปรวมสามารถใช้ผู้ประเมินเพียง 2 คน เท่านั้น ทำให้การประเมินยังคงมีค่าความเที่ยงในระดับสูง (.817) คะแนนรวม พบว่า เมื่อเพิ่มจำนวนผู้ประเมินมากขึ้นค่าความคลาดเคลื่อนทั้งความคลาดเคลื่อนเชิงสัมพัทธ์และความคลาดเคลื่อนเชิงสัมบูรณ์จะลดลงและจะเริ่มคงที่เมื่อใช้ผู้ประเมิน 4 คน ผลการประเมินในภาพรวมพบว่า ค่าความเที่ยงอยู่ในระดับสูง (มากกว่า .80) เมื่อใช้ผู้ประเมินตั้งแต่ 2 คนขึ้นไป ทั้งค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์

3. ผลของการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

ผลการวิเคราะห์สถิติพื้นฐานของคะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนโดยใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษฯ พบว่า ในภาพรวมของทั้ง 4 ภาระงาน นักเรียนมีความสามารถอยู่ในระดับอ่อนในการประเมินครั้งที่ 1 และ 2 ส่วนการประเมินครั้งที่ 3 มีความสามารถอยู่ในระดับพอใช้ทุกภาระงาน โดยพบว่านักเรียนมีความสามารถเพิ่มขึ้นทุกครั้งในทุกภาระงาน ผลการเปรียบเทียบคะแนนเฉลี่ยจากการประเมิน 3 ครั้ง ในภาระงานการสัมภาษณ์ โดยรวมของนักเรียนกลุ่มตัวอย่าง พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และเมื่อเปรียบเทียบคะแนนเฉลี่ยภาระงานการสัมภาษณ์ของนักเรียนในการประเมิน 3 ครั้ง พบว่า คะแนนเฉลี่ยในการประเมินครั้งที่ 2 สูงกว่าการประเมินครั้งที่ 1 และคะแนนเฉลี่ยครั้งที่ 3 สูงกว่าคะแนนเฉลี่ยครั้งที่ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนผลการเปรียบเทียบคะแนนเฉลี่ยจากการประเมิน 3 ครั้ง ในภาระงานการสัมภาษณ์ โดยรวมของนักเรียนกลุ่มตัวอย่าง พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และเมื่อเปรียบเทียบคะแนนเฉลี่ยภาระงานการนำเสนอปากเปล่าของนักเรียนในการประเมิน 3 ครั้ง พบว่า คะแนนเฉลี่ยในการประเมินครั้งที่ 2 สูงกว่าจากการประเมินครั้งที่ 1 และ คะแนนเฉลี่ยครั้งที่ 3 เพิ่มขึ้นจากครั้งที่ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลการเปรียบเทียบคะแนนเฉลี่ยจากการประเมิน 3 ครั้ง ในภาระงานการนำเสนอปากเปล่าโดยรวมของนักเรียนกลุ่มตัวอย่าง พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และเมื่อเปรียบเทียบคะแนนเฉลี่ยภาระงานการเล่านิทานของนักเรียนในการประเมิน 3 ครั้ง พบว่า คะแนนเฉลี่ยในการประเมินครั้งที่ 2 แตกต่างจากการประเมินครั้งที่ 1 และ ครั้งที่ 3 สูงกว่าคะแนนเฉลี่ยครั้งที่ 2 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และผลการเปรียบเทียบคะแนนเฉลี่ยจากการประเมิน 3 ครั้ง ในภาระงานการสัมภาษณ์ โดยรวมของนักเรียนกลุ่มตัวอย่าง พบว่ามีความแตกต่าง

กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และเปรียบเทียบคะแนนเฉลี่ยภาระงานการบรรยายภาพของนักเรียนในการประเมิน 3 ครั้ง พบว่า คะแนนเฉลี่ยในการประเมินครั้งที่ 2 สูงกว่าการประเมินครั้งที่ 1 และ ครั้งที่ 3 สูงกว่าการประเมินครั้งที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

4. ผลการพัฒนาคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ผลการพัฒนาคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน พบว่า คู่มือการประเมินทักษะการพูดภาษาอังกฤษประกอบด้วยองค์ประกอบที่สำคัญดังนี้ 1) โครงสร้างของคู่มือการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ประกอบด้วย 5 องค์ประกอบ ได้แก่ บทนำ วัตถุประสงค์ของแฟ้มสะสมงาน ฯ คำนิยามศัพท์ เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ 2) ส่วนประกอบของแฟ้มสะสมงาน ฯ ประกอบด้วย 3 ส่วน ดังนี้ ส่วนนำ ส่วนเนื้อหา ส่วนภาคผนวก 3) ขั้นตอนของการพัฒนาแฟ้มสะสมงาน ฯ ประกอบด้วย 7 ขั้นตอน ดังนี้ ขั้นการวางแผน ขั้นการเตรียมความพร้อมนักเรียน ขั้นการเก็บรวบรวมหลักฐาน ขั้นการติดตามความก้าวหน้า ขั้นการปรับปรุงผลงาน ขั้นการสะท้อนคิด และขั้นการจัดแสดงผลงาน 4) เครื่องมือที่ใช้ในแฟ้มสะสมงาน ฯ แบ่งออกเป็นเครื่องมือสำหรับนักเรียน จำนวน 11 ฉบับ และเครื่องมือสำหรับครู จำนวน 8 ฉบับ และ 5) การเลือกรูปแบบการประเมินทักษะการพูดภาษาอังกฤษ ประกอบด้วย 3 รูปแบบ ได้แก่ 1) จำนวนผู้ประเมินและภาระงานที่เหมาะสม 2) การเลือกรูปแบบการประเมินของผู้ประเมินที่เหมาะสม 3) การเลือกรูปแบบการประเมินสำหรับการประเมินระหว่างเรียนและสรุปรวมผลการตรวจสอบความสอดคล้องของวัตถุประสงค์ของคู่มือแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ พบว่า มีค่าดัชนี IOC อยู่ระหว่าง 0.778-1.000 ซึ่งผ่านเกณฑ์ทุกข้อ

อภิปรายผลวิจัย

1. ผลการพัฒนาเกณฑ์การให้คะแนนและภาระงานที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ผลการตรวจสอบความเที่ยงของเกณฑ์การให้คะแนนทักษะการพูดทั้ง 6 ด้าน โดยใช้รูปแบบ $p \times t \times r$ Design โดยพบว่าผลการตรวจสอบความเที่ยงของเกณฑ์ที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุวัตถุประสงค์ และความคล่องแคล่ว พบว่า ความแปรปรวนของนักเรียน (person, p) มีขนาดใหญ่ที่สุด เมื่อพิจารณาแยกตามองค์ประกอบของเกณฑ์การให้คะแนนพบว่าการออกเสียง (pronunciation) มีความแปรปรวนมากที่สุด คิดเป็นร้อยละ 25.66 รองลงมาคือความคล่องแคล่ว (fluency) การบรรลุวัตถุประสงค์ (ideational function) การสร้างประโยค

(syntax) คำศัพท์ (vocabulary) และการเชื่อมโยงเนื้อหา (cohesion) (คิดเป็นร้อยละ 17.66 15.12 15.10 13.68 และ 12.77 ตามลำดับ) จากผลการวิจัยชี้ให้เห็นว่า ความสามารถในการพูดภาษาอังกฤษของนักเรียนนั้นการออกเสียง (pronunciation) เป็นองค์ประกอบที่มีผลต่อคะแนนรวมของนักเรียนมากที่สุด ซึ่งผลการวิจัยนี้มีความขัดแย้งกับงานวิจัยในต่างประเทศ (Sato, 2011; Iwashita & Griove, 2003) ที่พบว่า ความคล่องแคล่ว (fluency) เป็นองค์ประกอบที่มีความแปรปรวนมากที่สุด ทั้งนี้อาจเนื่องมาจากการประเมินทักษะการพูดภาษาอังกฤษในประเทศไทย โดยเฉพาะในระดับประถมศึกษา ผู้ประเมินในครั้งนี้ซึ่งเป็นครูผู้สอนวิชาภาษาอังกฤษในโรงเรียนที่ทำการทดลองให้ความสำคัญกับองค์ประกอบด้านการออกเสียงเป็นสำคัญ ทั้งนี้อาจเนื่องมาจากการใช้ภาษาอังกฤษในฐานะภาษาต่างประเทศ ในการจัดการเรียนการสอนรวมทั้งการประเมินผู้ประเมิน ซึ่งเป็นครูผู้สอนจึงให้ความสำคัญกับการออกเสียงให้ถูกต้องเป็นหลัก ซึ่งผลการวิจัยนี้ต่างจากการทดลองใช้เกณฑ์การประเมินเช่นกัน ซึ่งผู้วิจัยทดลองใช้กับนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 119 คน โดยผู้ประเมินเป็นผู้เชี่ยวชาญด้านการสอนภาษาอังกฤษ กล่าวคือ เป็นนิสิตปริญญาเอก ชั้นปีที่ 2 และ 3 สาขาภาษาอังกฤษเป็นภาษานานาชาติ (English as an International Language) จากสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย และทุกคนเป็นอาจารย์ในระดับอุดมศึกษา ซึ่งผลในการทดลองใช้เกณฑ์การประเมินพบว่า ความคล่องแคล่ว (fluency) เป็นองค์ประกอบที่มีความแปรปรวนมากที่สุด ดังตารางที่ 3.3 แต่อย่างไรก็ตามในการวิจัยครั้งนี้ พบว่า ความคล่องแคล่ว (fluency) การบรรลุวัตถุประสงค์ (ideational function) เป็นองค์ประกอบที่มีความแปรปรวนรองลงมาซึ่งสอดคล้องกับงานวิจัยของ Sato (2011) และ Iwashita and Grove (Iwashita & Grove, 2003) และการศึกษา นำร่องในตารางที่ 3.3 เช่นกัน และสอดคล้องกับโดย McKay (2006) ได้กล่าวถึงลักษณะของ oral language ability ที่อ้างอิงจากแนวคิดการประเมินทางภาษาของ Bachman and Palmer's (1966) ที่กล่าวว่า function knowledge ของเด็กเป็นสิ่งสำคัญ คือต้องสื่อสารเพื่อให้ผู้อื่นเข้าใจว่าตนต้องการอะไร คิดอะไร และเป้าหมายที่แท้จริงของการสื่อสารในแต่ละครั้งคืออะไร โดยเรียนรู้จากประสบการณ์ที่ผ่านมา และเด็กจะต้องเชื่อมโยงเรื่องราวที่จะพูดเพื่อให้ผู้อื่นเข้าใจโดยเชื่อมโยงและจัดลำดับสิ่งที่จะพูด

ความเที่ยงของคะแนนรวมของเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษแบบแยกส่วน (speaking analytic rating scale) พบว่าค่าอยู่ในระดับสูง คือมากกว่า 0.80 โดยค่าความเที่ยงถือว่าเป็นค่าที่สูงใน CCT (classical testing) (Fulcher, 2003) ซึ่งแสดงว่าเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษที่ใช้สำหรับการประเมินการพูดของนักเรียนในระดับชั้นประถมศึกษาปีที่ 6 มีความเหมาะสมและเชื่อถือได้ ทั้งนี้เป็นเพราะในการพัฒนาเกณฑ์ในครั้งนี้ เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษในแต่ละองค์ประกอบมีการให้คะแนนในการวัด 4 ระดับ ตั้งแต่ 1-4 คะแนน จาก อ่อนมาก อ่อนพอใช้ และ ดี ตามลำดับ ซึ่งถือว่ามีความเหมาะสมตามที่ Luoma (2004) กล่าวไว้ว่าเกณฑ์ในการวัด

การพูดควรมีระดับการวัด 4-6 ระดับ จึงจะสะดวกต่อการให้คะแนนของผู้ประเมินสอดคล้องกับ Weigle (2002) ที่กล่าวว่าเกณฑ์การให้คะแนนแบบแยกส่วนส่วนใหญ่จะไม่เกิน 4 ระดับ ซึ่งจะง่ายต่อการประเมินค่าของผู้ประเมิน และแต่ละระดับมีการให้นิยามที่ชัดเจนทำให้ผู้ประเมิน สามารถให้คะแนนได้สอดคล้องกัน และมีความน่าเชื่อถือในการประเมิน ซึ่งสอดคล้องกับงานวิจัยก่อนหน้านี้ที่พบว่า เกณฑ์การให้คะแนนแบบแยกส่วนจะเป็นเกณฑ์ที่มี 4-5 ระดับ ประกอบด้วยองค์ประกอบ 4-6 องค์ประกอบ (Sato, 2011; Sawaki, 2007 and Lee, 2006)

สัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของภาระงานการประเมินทักษะการพูดภาษาอังกฤษ โดยใช้ $p \times t \times r$ Design ผลการตรวจสอบพบว่า ความแปรปรวนของนักเรียน (person, p) มีขนาดใหญ่ที่สุด ความคลาดเคลื่อนที่เกิดจากภาระงานพบว่ามีน้อยมาก โดยพบว่า ภาระงานการสัมภาษณ์มีความคลาดเคลื่อนหรือความแปรปรวนสูงที่สุด แต่ก็เพียงร้อยละ 1.5 เท่านั้น ส่วนภาระงานอื่นๆ มีความแปรปรวนไม่ถึงร้อยละ 1 จากผลการวิเคราะห์ข้อมูลดังกล่าว อธิบายได้ว่า คะแนนการพูดของนักเรียนนั้นเกิดจากตัวนักเรียนมากที่สุด ไม่ใช่เกิดจากภาระงาน หรือลำดับการให้คะแนนของผู้ประเมินหรือลำดับของภาระงาน ความคลาดเคลื่อนของผู้ประเมินพบที่มีความแปรปรวนเกือบเป็นศูนย์ทุกภาระงาน ยกเว้นภาระงานการเล่านิทานที่พบที่มีความแปรปรวน .05 แต่ก็น้อยมากไม่ถึงร้อยละ 1 ขนาดของความคลาดเคลื่อนของภาระงาน (tasks) มีขนาดใหญ่กว่าผู้ประเมิน (raters) สำหรับการศึกษาในครั้งนี้ ทั้งนี้อาจเนื่องมาจากความยากของภาระงานที่ต่างกันทำให้ผลต่อความสามารถในการพูดของนักเรียน กล่าวคือ ในภาระงานที่ยากขึ้น หรือเป็นเรื่องที่นักเรียนไม่คุ้นเคย ทำให้นักเรียนไม่สามารถเชื่อมโยงเรื่องราวให้สอดคล้องกันได้ และสร้างประโยคที่มีความถูกต้องตามหลักไวยากรณ์ได้ไม่ตึก ซึ่งสอดคล้องกับงานวิจัยก่อนนี้ที่พบว่าภาระงานมีความคลาดเคลื่อนมากกว่าผู้ประเมิน (raters หรือ rating) (Lee, 2006 and Sawaki, 2007)

ผลการวิจัยพบว่า ความคลาดเคลื่อนของผู้ประเมินพบที่มีความแปรปรวนเกือบเป็นศูนย์ทุกภาระงาน ยกเว้นภาระงานการเล่านิทาน ที่พบที่มีความแปรปรวน .05 แต่ก็น้อยมากไม่ถึงร้อยละ 1 จากผลการวิเคราะห์ข้อมูลดังกล่าว อธิบายได้ว่า คะแนนการพูดของนักเรียนนั้นเกิดจากตัวนักเรียนมากที่สุด ไม่ใช่เกิดจากความคลาดเคลื่อนจากแหล่งอื่นๆ และพบว่า ภาระงานมีปฏิสัมพันธ์กับผู้ประเมิน และนักเรียนมีปฏิสัมพันธ์กับภาระงาน แต่เมื่อพิจารณาความคลาดเคลื่อนที่เกิดจากภาระงาน (t) หรือ main effect พบว่ามีความคลาดเคลื่อนสูงที่สุดเพียงร้อยละ 1.50 คือ การสัมภาษณ์ นอกนั้นน้อยมากคือเพียงน้อยกว่าร้อยละ 1 เท่านั้น ผลการวิจัยนี้ชี้ให้เห็นว่าผู้ประเมินมีความเห็นในการประเมินความสามารถของนักเรียนที่สอดคล้องกัน ทั้งนี้อาจเนื่องมาจากในการวิจัยครั้งนี้ ผู้วิจัยได้มีการฝึกการประเมินให้แก่ผู้ประเมินทุกคน เปิดโอกาสให้ผู้ประเมินซึ่งคือครูผู้สอนวิชาภาษาอังกฤษในโรงเรียน ได้มีโอกาสอภิปรายผลการให้คะแนนระหว่างที่ฝึก มีการทดลองให้คะแนนจนเข้าใจในตัวเกณฑ์การให้คะแนนเป็นอย่างดี ตลอดจนมีความเห็นที่สอดคล้องกัน เห็นได้จากค่าสัมประสิทธิ์สหสัมพันธ์แบบ

เพียร์สัน (Pearson' Product Moment Coefficient of Correlation ดังตารางที่ 3.3 ที่พบว่า ผลการหาค่าความสอดคล้องของการประเมินของผู้ประเมิน (inter-raters) ทั้ง 4 คน พบว่า การประเมิน โดยใช้เกณฑ์การให้คะแนนแบบแยกส่วน ทั้ง 6 ด้าน มีค่าสัมประสิทธิ์ความสัมพันธ์อยู่ในระดับกลาง ถึงระดับสูง คือ มีค่าสัมประสิทธิ์ความสัมพันธ์อยู่ระหว่าง .625-1.000 สอดคล้องกับงานวิจัยของ Xi (2007) ที่พบว่า การฝึกผู้ประเมินก่อนทำการประเมินส่งผลให้เกิดความสอดคล้องในผลการประเมิน และเกิดความคลาดเคลื่อนในการประเมินที่ต่ำ

2. รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

2.1 ผลการวิจัยพบว่าเมื่อเพิ่มจำนวนผู้ประเมินและจำนวนภาระงานให้มากขึ้นทำให้ระดับความน่าเชื่อถือหรือความเที่ยงในการประเมินมีค่าสูงขึ้น จะทำให้ความเที่ยงของทั้งเกณฑ์การให้คะแนน การพูดภาษาอังกฤษแบบแยกส่วน (analytic rating scale) และภาระงานการประเมินสูงขึ้น โดยพบว่าการเพิ่มจำนวนภาระงานทำให้ความเที่ยงเพิ่มมากกว่าการเพิ่มจำนวนผู้ประเมินโดยใช้จำนวนภาระงานเพียง 6 งาน และใช้จำนวนผู้ประเมิน เพียง 2-4 คน ก็จะทำให้ได้ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ ในระดับที่สูงที่เชื่อถือได้ (มากกว่า 0.80) ซึ่งการใช้ Multivariate g-theory ทำให้ทราบว่าการเพิ่มจำนวนภาระงาน และผู้ประเมินที่ทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์เพิ่มขึ้น และพบว่าหากให้จำนวนผู้ประเมินคงที่ คือเพียง 4 คน และเพิ่มจำนวนภาระงานให้เป็น 6 งาน จะทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์อยู่ในระดับสูง (Lee, 2006, Bolus et al, 1981 and Fulcher, 2003) แต่ถ้าจะเพิ่มจำนวนผู้ประเมินให้มีความน่าเชื่อถือ ต้องใช้จำนวนผู้ประเมินถึง 10 คน ซึ่งในทางปฏิบัติการเพิ่มจำนวนผู้ประเมินจะเป็นการเพิ่มค่าใช้จ่าย และภาระหน้าที่ให้แก่ผู้ประเมิน ดังนั้นจากผลการศึกษาค้นคว้าทำให้ได้ทราบว่าผู้ดำเนินการประเมิน สามารถเพิ่มจำนวนภาระงานก็สามารถเพิ่มค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ให้อยู่ในระดับที่เชื่อถือได้ ซึ่งสอดคล้องกับงานวิจัยที่ผ่านมาที่พบว่าเมื่อเพิ่มจำนวนภาระงานและผู้ประเมินทำให้ความเที่ยงเพิ่มขึ้นด้วยและผลการศึกษาในครั้งนี้ สอดคล้องกับการศึกษาของ Gebriel (2010) และ Xi (2007) ที่ผลการวิจัยพบว่าเมื่อจำนวนผู้ประเมิน และจำนวนภาระงานเพิ่มขึ้นค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์และเชิงสัมบูรณ์จะเพิ่มขึ้นด้วย แต่การเพิ่มจำนวนภาระงานจะทำให้มีค่าใช้จ่ายในการดำเนินการน้อยกว่า (Bolus et al, 1981 and Fulcher, 2003) นอกจากนี้การเพิ่มจำนวนภาระงานหรือการเพิ่มจำนวนครั้งในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนยังเป็นการทำให้นักเรียนได้ฝึกทักษะการพูดของตนเองเพิ่มขึ้นด้วย และได้ฝึกพูดมากขึ้น ซึ่งสอดคล้องกับแนวคิดในการจัดการเรียนการสอนภาษาอังกฤษตามแนวการสอน ภาษาอังกฤษแบบภาษาเพื่อการสื่อสาร (communicative language teaching) ทั้งนี้เนื่องจากวิชา ภาษาอังกฤษเป็นวิชาทักษะที่นักเรียนจะต้องได้รับได้ฝึกปฏิบัติซ้ำๆ จนเกิดทักษะนั้นๆ และทำให้นักเรียนมีความสามารถในทักษะการพูดเพิ่มขึ้นอีกด้วย ซึ่งในปัจจุบันการสอนภาษาต่างประเทศ

(ภาษาอังกฤษ) ได้หันมายึดแนวทางการสอนเพื่อการสื่อสารมากขึ้น (สถาบันภาษาอังกฤษ, 2557) ซึ่งการสอนตามแนวการสอนภาษาเพื่อการสื่อสารเน้นความสำคัญของผู้เรียน ทำความเข้าใจ จดจำ แล้วนำสิ่งที่เรียนรู้ไปใช้ ซึ่งเป็นการเรียนรู้ที่มีความหมาย ได้ฝึกใช้ภาษาในสถานการณ์ต่างๆ ที่มีโอกาสพบในชีวิตประจำวัน (Littlewood, 1991) (สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ, 2557)

2.2 การเปรียบเทียบรูปแบบการประเมินทักษะการพูดภาษาอังกฤษ โดยการเปรียบเทียบรูปแบบการให้คะแนนของผู้ประเมินที่ต่างกัน พบว่า ผลการเปรียบเทียบรูปแบบการประเมินของผู้ประเมินในแต่ละภาระงานที่ต่างกัน พบว่า เมื่อเพิ่มจำนวนผู้ประเมินมากขึ้นความคลาดเคลื่อนทั้งเชิงสัมพัทธ์และความคลาดเคลื่อนเชิงสัมบูรณ์ ลดลงทั้งสองรูปแบบ และพบว่า ในรูปแบบการประเมินที่ใช้ผู้ประเมินต่างกันในแต่ละภาระงานมีความคลาดเคลื่อนต่ำกว่า ดังนั้นเมื่อพิจารณาค่าความเที่ยงของการประเมินจึงพบว่ารูปแบบการประเมินที่ใช้ผู้ประเมินต่างกันมีค่าความเที่ยงสูงกว่ารูปแบบการประเมินของผู้ประเมินที่มีการประเมินที่เหมือนกัน ผลการวิจัยชี้ให้เห็นว่าการให้ผู้ประเมินตรวจให้คะแนนนักเรียนไม่เข้าภาระงานเป็นการลดความคลาดเคลื่อนในการประเมินลง ทั้งนี้ เมื่อความคลาดเคลื่อนลดลงย่อมส่งผลให้ความเที่ยงสูงขึ้นด้วย (Brennan, 2001; Shavelson & Webb, 1991) ซึ่งผลการวิจัยนี้สอดคล้องกับงานวิจัยของ Gebriel (2010) ที่พบว่าเมื่อเปรียบเทียบรูปแบบการให้คะแนนของผู้ประเมินที่ให้คะแนนต่างภาระงานกันกับให้คะแนนทุกภาระงาน พบว่ารูปแบบการให้คะแนนที่ให้ผู้ประเมินให้คะแนนภาระงานต่างกันมีความเที่ยงสูงกว่าการให้ผู้ประเมินตรวจให้คะแนนเข้าภาระงานกัน

3. ผลของการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน

ผลของการใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน พบว่า ความก้าวหน้าด้านทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน เป็นการศึกษาผลของความก้าวหน้าในการพูดใน 4 หัวเรื่อง หรือ 4 ภาระงานที่ใช้ในการจัดการเรียนการสอนของครูผู้สอนด้วยการวิเคราะห์ความก้าวหน้าในทักษะการพูดภาษาอังกฤษของนักเรียนโดยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำชนิดทางเดียว (one-way repeated ANOVA measurement) ผลการวิจัยพบว่า นักเรียนมีคะแนนเฉลี่ยในการประเมินการพูดในครั้งที่ 3 สูงกว่าการประเมินครั้งที่ 2 และการประเมินครั้งที่ 2 สูงกว่า การประเมินครั้งที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกภาระงาน จากผลการวิจัยกล่าวได้ว่าการประเมินโดยใช้แฟ้มสะสมงาน การประเมินทักษะการพูดภาษาอังกฤษทำให้นักเรียนมีความสามารถในการพูดภาษาอังกฤษสูงขึ้น ทั้งนี้เพราะแฟ้มสะสมงานที่ผู้วิจัยพัฒนาขึ้นมีตั้งแต่ขั้นตอนการวางแผน การกำหนดเป้าหมาย ซึ่งนักเรียนมีการกำหนดเป้าหมายในการพูดในแต่ละภาระงาน ทำให้ทราบเป้าหมายในการพูดของตน นอกจากนี้ยังมีโอกาสได้รับการประเมินจากทั้งเพื่อนและครูผู้สอน และได้เป็นผู้ประเมินทำให้

ต้องตั้งใจฟังเพื่อนซึ่งทำให้มีโอกาสได้พัฒนาตนเองไปด้วย นอกจากนี้ในขั้นตอนการติดตามความก้าวหน้า ครูผู้สอนจะมีการให้ข้อมูลย้อนกลับแก่นักเรียนในทันทีที่ทำการประเมินแต่ละครั้ง โดยทำการให้ข้อมูลย้อนกลับแบบระบุสิ่งที่ทำดีแล้ว และบอกข้อที่ต้องปรับปรุง ซึ่ง Tuntall and Gipps (1996) กล่าวว่า การให้ข้อมูลย้อนกลับของครูต่อนักเรียนเป็นกระบวนการที่มีความสำคัญต่อการประเมินระหว่างเรียน (formative assessment) ที่ทำให้เกิดความก้าวหน้าในการเรียนรู้ของนักเรียน การให้ข้อมูลย้อนกลับของครูแก่นักเรียนในการเรียนภาษาที่สองเป็นสิ่งจำเป็นเพื่อให้นักเรียนปรับปรุงและแก้ไขข้อผิดพลาดให้ถูกต้อง (corrective feedback) ต่อไป ซึ่งการให้ข้อมูลย้อนกลับเพื่อให้นักเรียนนำไปปรับปรุงนี้ TESOL (Teachers of English to Speakers of Other Languages) ได้ให้ความสำคัญมาอย่างยาวนานแล้ว (Mori, 2011) นอกจากนี้นักเรียนยังมีโอกาสในการปรับปรุงการพูดหลายครั้งทำให้นักเรียนมีโอกาสในการพัฒนามากยิ่งขึ้น สอดคล้องกับ Lam และ Lee (2010) ที่พบว่า การพัฒนาแฟ้มสะสมงานการเขียนภาษาอังกฤษที่มีการรวบรวมผลงาน การสะท้อนคิด การประเมินระหว่างเรียน (formative assessment) และการประเมินสรุปรวม (summative assessment) โดยในขั้นตอนของการประเมินระหว่างเรียนได้เพิ่มขั้นตอนบางขั้นตอนเข้าไป คือ การให้ข้อมูลย้อนกลับของครู (ongoing teacher feedback) การปรึกษาระหว่างครูและผู้เรียน (conferencing) และการประเมินจากเพื่อนในชั้นเรียน (peer review) ทำให้นักเรียนมีความสามารถในการเขียนภาษาอังกฤษสูงขึ้น สอดคล้องกับ Kouzouli (2012) ที่พบว่าการใช้แฟ้มสะสมงานกับนักเรียนเกรด 3 ในโรงเรียนประถมศึกษาเป็นวิธีการที่เหมาะสมกับนักเรียนในวัยเด็ก (young learner) นักเรียนมีเจตคติที่ดีต่อการเรียน สามารถประเมินตนเองและประเมินเพื่อนในชั้นได้ มีการสะท้อนคิดต่อผลงานของตนเองได้ ทำให้เกิดการเรียนรู้ที่เป็นอิสระ มีความเข้าใจในปัญหาที่เกิดขึ้นในการเรียนได้อย่างถ่องแท้ และสามารถจัดการกับปัญหาได้ และสอดคล้องกับ โชติภา ภาชีผล และคณะ (2558) ที่พัฒนารูปแบบแฟ้มสะสมงานอิเล็กทรอนิกส์ที่สะท้อนข้อมูลย้อนกลับในการประเมินพัฒนาการเรียนรู้ของนักเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน โดยใช้การสะท้อนข้อมูลย้อนกลับแบบทั่วไป (feedback) และการสะท้อนข้อมูลชี้แนะเพื่อการปรับปรุง (feedforward) ผลวิจัยพบว่าผลด้านพัฒนาการเรียนรู้ของนักเรียนพบว่านักเรียนระดับประถมศึกษาและมัธยมศึกษาที่เป็นกลุ่มตัวอย่างมีพัฒนาการการเรียนรู้ในวิชาภาษาไทยและวิทยาศาสตร์สูงขึ้น โดยสำหรับนักเรียนประถมศึกษาพบว่าการใช้แฟ้มสะสมงานอิเล็กทรอนิกส์ที่สะท้อนข้อมูลย้อนกลับแบบทั่วไปทำให้นักเรียนส่วนใหญ่ของนักเรียนมีพัฒนาการต่ำกว่าไม่มีพัฒนาการ ในขณะที่การใช้แฟ้มสะสมงานอิเล็กทรอนิกส์ที่สะท้อนข้อมูลชี้แนะเพื่อการปรับปรุงร้อยละของนักเรียนที่มีพัฒนาสูงกว่าไม่มีพัฒนาการ

4. ผลการพัฒนาคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ชั้นประถมศึกษาปีที่ 6

ผลการพัฒนาคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 พบว่า คู่มือการประเมินทักษะการพูดภาษาอังกฤษ ประกอบด้วยองค์ประกอบที่สำคัญดังนี้ 1) โครงสร้างของคู่มือการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน ประกอบด้วย 5 องค์ประกอบ 2) ส่วนประกอบของแฟ้มสะสมงาน ประกอบด้วย 3 ส่วน 3) ขั้นตอนของการพัฒนาแฟ้มสะสมงาน ประกอบด้วย 7 ขั้นตอน 4) เครื่องมือที่ใช้ในแฟ้มสะสมงานฯ แบ่งออกเป็นเครื่องมือสำหรับนักเรียน จำนวน 11 ฉบับ และเครื่องมือสำหรับครู จำนวน 8 ฉบับ และ 5) การเลือกใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษ ประกอบด้วย 3 รูปแบบ ได้แก่ 1) จำนวนผู้ประเมินและภาระงานที่เหมาะสม 2) การเลือกรูปแบบการประเมินของผู้ประเมินที่เหมาะสม 3) การเลือกรูปแบบการประเมินสำหรับการประเมินระหว่างเรียนและสรุปรวม ผลการตรวจสอบความสอดคล้องของวัตถุประสงค์ของคู่มือแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ พบว่า มีค่าดัชนี IOC อยู่ระหว่าง 0.778-1.000 ซึ่งผ่านเกณฑ์ทุกข้อ และมีความเหมาะสมสำหรับการนำไปใช้ได้ ทั้งนี้เป็นเพราะในการพัฒนาคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษในครั้งนี้ ผู้วิจัยได้ศึกษาค้นคว้าแนวคิดทฤษฎีเกี่ยวกับการประเมินทางภาษาสำหรับเด็ก (young learner) ขั้นตอนและองค์ประกอบของแฟ้มสะสมงาน รวมทั้งได้มีการสัมภาษณ์ผู้ทรงคุณวุฒิทั้งทางด้านการจัดการเรียนการสอนภาษาอังกฤษ ครูผู้สอนวิชาภาษาอังกฤษ การวัดและประเมินผลทางภาษา และการวัดและประเมินผลทางการศึกษา จึงทำให้ได้เครื่องมือที่มีคุณภาพ และสามารถนำมาใช้ได้จริง นอกจากนี้แล้วคู่มือการใช้แฟ้มสะสมงานฉบับนี้ผู้วิจัยได้ให้ครูผู้สอนวิชาภาษาอังกฤษได้ทดลองใช้และให้ข้อคิดเห็นจากการใช้เพื่อนำมาปรับปรุง และในส่วนของทางเลือกใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษที่จะอำนวยความสะดวกให้กับครูในการเลือกรูปแบบการประเมินได้ตามความเหมาะสมของตนเองและสอดคล้องกับทรัพยากรที่มีอยู่ เช่น จำนวนภาระงานและผู้ประเมินที่เหมาะสม เป็นต้น ซึ่งในส่วนนี้ผู้วิจัยนำรูปแบบมาจากผลการวิจัย ซึ่งน่าจะทำให้การประเมินครูสามารถใช้รูปแบบการประเมินได้อย่างมีประสิทธิภาพมากที่สุด และเมื่อนำร่างคู่มือให้ผู้ทรงคุณวุฒิตรวจสอบคุณภาพจึงทำให้ผ่านเกณฑ์ที่กำหนดไว้ทุกข้อ

ข้อเสนอแนะ

ข้อเสนอแนะในการนำไปใช้

1. จากผลการศึกษาพบว่า หากผู้บริหารการประเมินต้องการเพิ่มความเที่ยงในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 เช่น การออกแบบการประเมินระดับโรงเรียน หรือเพื่อการแข่งขัน หรือสำหรับการประเมินในระดับชาติ ผู้บริหารการประเมินสามารถใช้ภาระงานทั้ง 4 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ โดยใช้เกณฑ์การให้คะแนนแบบแยกส่วน ที่ประกอบด้วยเกณฑ์การให้คะแนน 6 ด้าน ได้แก่ คำศัพท์

(vocabulary) การสร้างประโยค (syntax) การเชื่อมโยงเนื้อหา (cohesion) การออกเสียง (pronunciation) การบรรลุลักษณะประสงค์ (ideational function) และความคล่องแคล่ว (fluency) เพื่อให้ได้ระดับความเที่ยงที่เชื่อถือได้ ผู้บริหารการประเมินควรใช้วิธีการเพิ่มจำนวนภาระงานมากกว่า การเพิ่มผู้ประเมิน เนื่องจากการเพิ่มภาระงานทำให้มีค่าใช้จ่ายน้อยกว่าการเพิ่มจำนวนผู้ประเมิน แต่ยังคงได้ระดับความเที่ยงที่สูงและเชื่อถือได้เช่นกัน

2. สำหรับการประเมินทักษะการพูดภาษาอังกฤษในระดับชาติ และระดับเขตพื้นที่การศึกษา ของนักเรียนชั้นประถมศึกษาปีที่ 6 จากผลการวิจัยทำให้ได้สารสนเทศเกี่ยวกับรูปแบบการประเมิน ทักษะการพูดภาษาอังกฤษที่มีความเที่ยงหรือความน่าเชื่อถือในการประเมิน โดยใช้ภาระงานการพูด 4 เรื่อง ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพ ใช้เกณฑ์การให้ คะแนนแบบแยกส่วนที่ประกอบด้วยขอบเขตการให้คะแนน 6 ด้าน ได้แก่ คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุลักษณะประสงค์ และความคล่องแคล่ว และควรใช้ผู้ประเมิน อย่างน้อย 2 คน ขึ้นไป

3. ในการบริหารการประเมินทักษะการพูดภาษาอังกฤษนั้นเพื่อให้เกิดความสะดวกและ เป็นไปได้ในทางปฏิบัติรวมถึงประหยัดค่าใช้จ่าย และแรงงาน ผู้บริหารการประเมินควรใช้รูปแบบ การประเมินที่ผู้ประเมินมีรูปแบบการให้คะแนนที่ต่างกัน กล่าวคือ ในภาระงานที่ต่างกันเมื่อใช้ ผู้ประเมินที่ต่างกัน หรือผู้ประเมินไม่จำเป็นต้องตรวจให้คะแนนนักเรียนในทุกภาระงาน ซึ่งพบว่า รูปแบบการประเมินนี้ทำให้เกิดความคลาดเคลื่อนน้อยลงและทำให้มีความเที่ยงในการประเมินสูงขึ้น จึงเป็นรูปแบบการประเมินที่มีความเหมาะสมและสามารถนำมาใช้ได้อย่างมีประสิทธิภาพ

4. จากผลการวิจัยพบว่าในการประเมินตนเองและประเมินเพื่อนของนักเรียน และการประเมิน นักเรียนของครู พบว่า นักเรียนบางคนมีการประเมินที่ยังไม่สอดคล้องกับเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) ทั้ง 6 องค์ประกอบ และประเมินตนเองและเพื่อนไม่ตรงประเด็นและ ขอบเขตเนื้อหาของเกณฑ์การประเมิน จึงทำให้ครูจะต้องย้ำและทบทวนเกณฑ์การให้คะแนนทุกครั้ง เพื่อป้องกันนักเรียนประเมินโดยไม่มีทิศทาง และไม่คำนึงถึงเกณฑ์การให้คะแนน ดังนั้นในการนำ แพ้ผสมงานการประเมินทักษะการพูดภาษาอังกฤษไปใช้ ควรใส่เกณฑ์การประเมินกำกับไว้ใน ทุกภาระงาน แทนที่จะไว้ที่ภาคผนวกเท่านั้น

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ในการศึกษาครั้งนี้เป็นการพัฒนาแพ้ผสมงานเพื่อการประเมินทักษะการพูด ภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 และเพิ่มความสามารถในทักษะการพูดภาษาอังกฤษ เท่านั้น ในการศึกษาครั้งต่อไปผู้วิจัยอาจนำรูปแบบแพ้ผสมงานไปพัฒนาเป็นแพ้ผสมงาน

การประเมินความสามารถด้านภาษาอังกฤษในทักษะการฟัง ทักษะการอ่าน และทักษะการเขียน เพื่อให้ครอบคลุมทุกทักษะ

2. ในการศึกษาครั้งนี้เป็นการพัฒนาเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนในระดับชั้นประถมศึกษาปีที่ 6 และเป็นเกณฑ์แบบแยกส่วน ซึ่งจะช่วยให้ครูผู้สอนสามารถวินิจฉัยความสามารถของนักเรียนตามองค์ประกอบความสามารถทางภาษาอังกฤษ ซึ่งในการศึกษาครั้งต่อไป ผู้วิจัยอาจนำแนวคิดในการพัฒนาเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษในนักเรียนระดับประถมศึกษาในระดับเด็กเล็ก ได้แก่ ระดับปฐมวัย และระดับชั้นประถมศึกษาปีที่ 1-3 เพื่อให้มีเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษที่ครอบคลุมทุกชั้นในระดับประถมศึกษา

3. ในการศึกษาครั้งนี้เป็นการพัฒนาแฟ้มสะสมงานเพื่อการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่มีลักษณะเป็นแฟ้มปกตีทั่วไป และใช้อุปกรณ์คอมพิวเตอร์ช่วยในการบันทึกเสียงการพูดภาษาอังกฤษของนักเรียนเท่านั้น ในการศึกษาครั้งต่อไป ผู้วิจัยอาจนำรูปแบบแฟ้มสะสมงานไปพัฒนาเป็นแฟ้มสะสมงานอิเล็กทรอนิกส์เพื่อประเมินความสามารถด้านภาษาอังกฤษในทักษะการพูด ทักษะการฟัง ทักษะการอ่าน และทักษะการเขียน เพื่อให้นักเรียนสนใจการเรียนมากขึ้น รวมทั้งมีความสะดวกในการที่นักเรียนจะเข้าไปเปิดฟังผลงานการพูดของตนเอง และเปิดดูผลงานในส่วนอื่นๆ ของตนเองยิ่งขึ้น

รายการอ้างอิง

ภาษาไทย

- โชติกา ภาชีผล และคณะ. (2558). การพัฒนารูปแบบแฟ้มสะสมงานอิเล็กทรอนิกส์ที่สะท้อนข้อมูลย้อนกลับในการประเมินพัฒนาการเรียนรู้ของนักเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. วารสารวิธีวิทยาการวิจัย, 28(1), 1-10.
- กนกพงศ์ จิตต์ปลื้ม. (2549). ผลของการใช้แฟ้มสะสมงานอิเล็กทรอนิกส์ประกอบการเรียนการสอนภาษาอังกฤษเพื่อการสื่อสารที่มีต่อผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์มหาบัณฑิต สาขาวิชาโสตทัศนศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- กมลวรรณ ตั้งชนกานนท์. (2557). การวัดและประเมินทักษะการปฏิบัติ. กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- กมลวรรณ ตั้งชนกานนท์. (2547). การพัฒนากระบวนการจัดทำแฟ้มสะสมงานเพื่อประเมินผลการเรียนรู้ของผู้เรียนที่ได้รับการจัดการเรียนการสอนตามแนวการเรียนรู้เพื่อสร้างสรรค์ด้วยปัญญาแบบเต็มรูปแบบในโรงเรียนดรณสิกขาลัย. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- จรรยา กสิวิทย์อำนวย. (2539). ผลของการสอนโดยใช้แนวการสอนแบบกลุ่มสัมพันธ์ต่อผลสัมฤทธิ์ทางการเรียนภาษาอังกฤษด้านทักษะการฟังและการพูดของนักเรียนชั้นประถมศึกษาปีที่ 6. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาประถมศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ชมรมศึกษานิเทศก์, กรมสามัญ. (2541). แนวทางการประเมินตามสภาพจริงโดยใช้แฟ้มสะสมงาน portfolio สำหรับครู อาจารย์ นักเรียน. กรุงเทพมหานคร: เสมาธรรม.
- ชัยพฤกษ์ เสรีรักษ์ และคณะ. (ม.ป.ป.). คู่มือการประเมินผลโดยใช้แฟ้มสะสมงานดีเด่น. กรุงเทพมหานคร: สถาบันพัฒนาคุณภาพวิชาการ.
- ชัยฤทธิ์ ศิลาเดช. (2540). การพัฒนาแฟ้มสะสมงานในการประเมินผลการเรียนวิชาภาษาอังกฤษระดับชั้นมัธยมศึกษาปีที่สาม. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาการทดสอบและวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
- นริศรา ดานวงศ์. (2537). ผลของการใช้เทคนิคการสื่อความหมายให้เข้าใจตรงกัน ที่มีต่อผลสัมฤทธิ์ทางการเรียนภาษาอังกฤษทักษะฟัง-พูด ของนักเรียนชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์

ปริญญามหาบัณฑิต สาขาวิชาการสอนภาษาอังกฤษ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ประไพ เทียนเงิน. (2534). *การพัฒนาทักษะการฟังและการพูดภาษาอังกฤษ ของนักเรียนชั้นประถมศึกษาปีที่ 5 โดยใช้วีดิทัศน์*. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาประถมศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่* (พิมพ์ครั้งที่ 3). กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2552). *ทฤษฎีการประเมิน* (พิมพ์ครั้งที่ 7). กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, ก. (2553). *แนวการวัดและประเมินผลวิชาภาษาอังกฤษ ระดับมัธยมศึกษา ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพมหานคร: ม.ป.ท.

สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, กระทรวงศึกษาธิการ. (2553). *แนวการวัดและประเมินผลวิชาภาษาอังกฤษ ระดับมัธยมศึกษา ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพมหานคร: ม.ป.ท.

สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, กระทรวงศึกษาธิการ. (2557). *การจัดการเรียนการสอนภาษาอังกฤษแนวใหม่ตามกรอบอ้างอิงความสามารถทางภาษาของสหภาพยุโรป ระดับมัธยมต้น*. กรุงเทพมหานคร: โรงพิมพ์องค์การสงเคราะห์ทหารผ่านศึก.

สมศักดิ์ ภู่วิภาดาบรรณ. (2544). *การยึดผู้เรียนเป็นศูนย์กลางและการประเมินตามสภาพจริง* (พิมพ์ครั้งที่ 2). เชียงใหม่: The knowledge center.

สำนักวิชาการและมาตรฐานการศึกษา กระทรวงศึกษาธิการ. (2553). *ตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพมหานคร: ม.ป.ท..

สุวิทย์ มูลคำ. (2540). *การยึดผู้เรียนเป็นศูนย์กลางและการประเมินตามสภาพจริง* (พิมพ์ครั้งที่ 6). กรุงเทพมหานคร: ภาพพิมพ์.

อัจฉรา วงศ์โสธร. (2543). *การพัฒนาข้อทดสอบทางภาษาอังกฤษให้เป็นมาตรฐาน*. กรุงเทพมหานคร: สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย.

อัจฉรา วงศ์โสธร. (2544). *การพัฒนาข้อทดสอบอิงวัตถุประสงค์แบบอ้อม: วัดความสามารถสูงสุดและการกระทำปกติทางภาษาอังกฤษ*. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

- อุทุมพร (ทองอุไทย) จามรमान. (2540ก). *การตีค่าความสามารถที่แท้จริงของผู้เรียนเพื่อการปฏิรูปการศึกษา (Authentic performance and portfolio assessment of learners for educational reform)*. กรุงเทพมหานคร: ฟีนนี่พับลิชชิ่ง.
- อุทุมพร (ทองอุไทย) จามรमान. (2540ข). *แฟ้มสะสมงาน (Portfolio) ของนักเรียน ครู ศึกษานิเทศก์ และผู้บริหารโรงเรียน*. กรุงเทพมหานคร: ฟีนนี่พับลิชชิ่ง.

ภาษาอังกฤษ

- American Council on the Teaching of Foreign Languages. (2012). The ACTFL Proficiency Guidelines 2012 – Speaking.
- Atilla, Y., & Ezel, T. (2010). The Examination of Reliability According to Classical Test and Generalizability on a Job Performance scale. *Educational Sciences: Theory & Practice*, 10(3), 1847-1854.
- Aurelio, J. F., & Sally, O. (2010). Commentary for special issue of journal of memory and language: Generalizability theory analysis for psycholinguistic applications. *Journal of Memory and Language*(59), 426–433.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: OUP.
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford: OUP.
- Barabouti, M. (2012). The Writing Portfolio: an alternative assessment tool with young learners of English. *Research Papers in Language Teaching and Learning*, 3(1), 123-145.
- Barton, J., & Collins, A. (1997). *Portfolio Assessment: A handbook for Education*. California: Addison-Wesley Publishing Company.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.
- Brennan, R. L. (2003). Coefficients and Indices in Generalizability Theory. *the Annual Meeting of the American Statistical Association*.
- Brennan, R. L. (2011a). Generalizability Theory and Classical Test Theory. *Applied Measurement in Education*, 24(1), 1-21.
- Brennan, R. L. (2011). Utility Indexes for Decisions about Subscores *CASMA Research Report: Center for Advanced Studies in Measurement and Assessment*.

- Brennan, R. L. (2011b). *Utility Indexes for Decisions about Subscores CASMA Research Report: Center for Advanced Studies in Measurement and Assessment*.
- Brurch, V. C., Norman, G. R., Schmidt, H. G., & Van der vleuten, C. P. M. (2008). Are specialist certification examinations a reliable measure of physician competence? *Adv in Health Sci Educ*, 13, 521–533.
- Burke, K., Forgarty, R., & Belgrad, S. (1994). *Rth mindful school: The portfolio connection*. Los Angeles, CA: Skylight Training & Publishing Inc.
- Cambridge ESOL. (2011). *Using the CEFR: Principles of Good Practice*. Cambridge ESOL.
- Cambridge ESOL. (2007). *Cambridge Young Learners English Tests*. Cambridge: University of Cambridge ESOL Examinations.
- Clauser, B. E., Harik, P., & Margolis, M. J. (2006). A Multivariate Generalizability Analysis of Data from a Performance Assessment of Physicians' Clinical Skills. *Journal of Educational Measurement*, 43(3), 173–191.
- Cole, D. J., Ryan, C. W., Kick, F., & Mathies, B. K. (2000). *Portfolio across the curriculum and beyond*. Thousand Oaks, CA: Sage publication.
- Efthymiou, G. (2012). Portfolio Assessment of Speaking Skills in English as a Foreign Language in Primary Education. *Research Papers in Language Teaching and Learning*, 3(1), 200-224.
- Fulcher, G. (2003). *Testing second language speaking*. London: Pearson Longman.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. New York: Routledge.
- Fulcher, G., Davidson, F., & Kemp, J. (2010). Effective rating scale development for speaking tests: Performance decision trees. *Language Testing*, 28(1), 5-29. doi: 10.1177/0265532209359514
- Gebriel, A. (2010). Bringing reading-to-write and writing-only assessment tasks together: A generalizability analysis. *Assessing Writing*, 15(2), 100-117. doi: 10.1016/j.asw.2010.05.002
- Hamp-Lyons, L., & Condon, W. (2000). New Jersey: Hampton Press.

- Huang, C. (2009). Magnitude of Task-Sampling Variability in Performance Assessment: A Meta-Analysis. *Educational and Psychological Measurement, 69*(6), 887-912. doi: 10.1177/0013164409344550
- Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments?—A generalizability theory approach. *Assessing Writing, 13*(3), 201-218. doi: 10.1016/j.asw.2008.10.002
- Iwashita, N., & Grove, E. (2003). A comparison of analytic and holistic scales in the context of a specific-purpose speaking test. *Prospect, 18*(3), 25–35.
- Janudom, R. (2009). *The development of an English instructional model using the integration of drama and questioning techniques to enhance students' speaking achievement and critical thinking skill*. Doctoral dissertation, English as an International Language Graduate School, Chulalongkorn University.
- Jinyan, H. (2008). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. USA: Addison-Wesley Publishing Company.
- Kouzouli, S. (2012). Investigating Portfolio Assessment with Learners of the 3rd Grade in a Greek State Primary School. *Research Papers in Language Teaching and Learning, 3*(1), 146-164.
- Lam, R., & Lee, I. (2009). Balancing the dual functions of portfolio assessment. *ELT Journal, 64*(1), 54-64. doi: 10.1093/elt/ccp024
- Lazaraton, A. (2006). Process and outcome in paired oral assessment. *ELT Journal, 60*(3), 287-289. doi: 10.1093/elt/ccl009
- Lee, Y.-W. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing, 23*(2), 131-166. doi: 10.1191/0265532206lt325oa
- Lisa, A. K., Brian, E. C., & S., D. B. (2010). Using multivariate generalizability theory to assess the effect of content stratification on the reliability of a performance assessment. *Adv in Health Sci Educ, 15*, 717–733.
- Littlewood, W. T. (1991). *Techniques and Principles in Language Teaching*. Oxford: Oxford University Press.

- Lo, Y. F. (2010). Implementing reflective portfolios for promoting autonomous learning among EFL college students in Taiwan. *Language Teaching Research*, 14(1), 77-95. doi: 10.1177/1362168809346509
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge university press.
- May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, 26(3), 397-421. doi: 10.1177/0265532209104668
- Mckay, P. (2006). *Assessing young language learners*. Cambridge: Cambridge university press.
- Mori, R. (2011). Teacher cognition in corrective feedback in Japan. *Sciverse Sciencedirect*, 39(2011), 451-467.
- O'Sullivan, B. (2008). *Notes on Assessing Speaking*. Paper presented at the RLC Conference. <http://lrc.cornell.edu/events/past/2008-2009>
- Office of Qualifications and Examinations Regulation. (2010). Estimating the Reliability of Composite Scores. [Online] Available from: <http://webarchive.nationalarchives.gov.uk/+http://www.ofqual.gov.uk/2794.aspx>.
- Payne, D. A. (2003). *Applied educational assessment* (2nd ed.). Belmont: Thomson Learning.
- Plakht, Y., Shiyovich, A., Nusbaum, L., & Raizer, H. (2012). The association of positive and negative feedback with clinical performance, self-evaluation and practice contribution of nursing students. *Nurse Education Today*, 2012(July).
- Popham, W. J. (2012). *Portfolio Assessment and Performance Testing*. Boston: Pearson Education, Inc.
- Sander, K. E. B.-, Pardee Lowe, J., Miles, J., & Swender, E. (2000). ACTFL Proficiency Guidelines — Speaking Revised 1999. *Foreign Language Annals*, 33(1), 13-18.
- Sapsirin, S. (2006). *A study of trait factors of oral language ability in a computer-based speaking test for Thai university students*. Doctoral dissertation, English as an International Language Graduate School, Chulalongkorn University.

- Sato, T. (2011). The contribution of test-takers' speech content to scores on an English oral proficiency test. *Language Testing*, 29(2), 223-241. doi: 10.1177/0265532211421162
- Sawaki, Y. (2007). Construct validation of analytic rating scales in a speaking assessment: Reporting a score profile and a composite. *Language Testing*, 24(3), 355-390. doi: 10.1177/0265532207077205
- Seitz, H., & Bartholomew, C. (2008). Powerful Portfolios for Young Children. *Early Childhood Education Journal*, 36(1), 63-68. doi: 10.1007/s10643-008-0242-7
- Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning. *Assessment in Education: Principles, Policy & Practice*, 18(4), 433-449. doi: 10.1080/0969594x.2011.582838
- Tavil, Z. M. (2010). Integrating listening and speaking skills to facilitate English language learners' communicative competence. *Procedia - Social and Behavioral Sciences*, 9, 765-770. doi: 10.1016/j.sbspro.2010.12.231
- Tunstall, P., & Gipps, C. (1996). Teacher Feedback to Young Children in Formative Assessment: A Typology. *British Educational Research Journal*, 22(4), 389-404.
- Weigle, S. C. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- Westerveld, M. F., & Moran, C. A. (2011). Expository Language Skills of Young School-Age Children. *LANGUAGE, SPEECH, AND HEARING SERVICES IN SCHOOLS*, 42, 182-193. doi: 10.1044/0161-1461(2010/10-0044)
- Woehr, D. J., Putka, D. J., & Bowler, M. C. (2011). An Examination of G-Theory Methods for Modeling Multitrait-Multimethod Data: Clarifying Links to Construct Validity and Confirmatory Factor Analysis. *Organizational Research Methods*, 15(1), 134-161. doi: 10.1177/1094428111408616
- Xiaoming, X. (2007). Evaluating analytic scoring for the TOEFL(R) Academic Speaking Test (TAST) for operational use. *Language Testing*, 24(2), 251-286. doi: 10.1177/0265532207076365
- Yilmaz, S., & Akcan, S. (2011). Implementing the European Language Portfolio in a Turkish context. *ELT Journal*, 66(2), 166-174. doi: 10.1093/elt/ccr042

Yin, P., & Sconing, J. (2008). Estimating Standard Errors of Cut Scores for Item Rating and Mapmark Procedures : A Generalizability Theory Approach. *Educational and Psychological Measurement*, 68(1), 25-41.



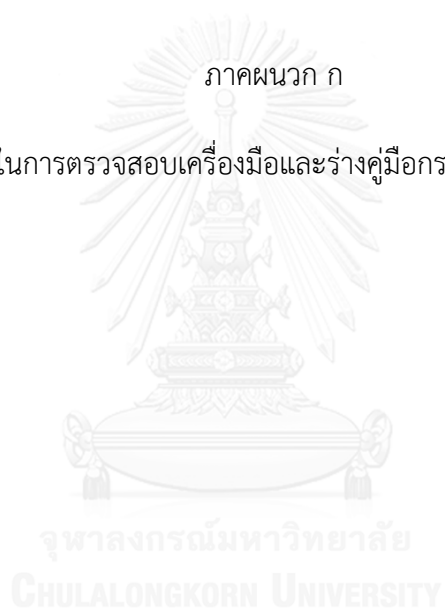


ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ก

รายนามผู้ทรงคุณวุฒิในการตรวจสอบเครื่องมือและร่างคู่มือกระบวนการจัดทำแฟ้มสะสมงาน



รายนามผู้ทรงคุณวุฒิในการตรวจสอบเครื่องมือและร่างคู่มือกระบวนการจัดทำแฟ้มสะสมงาน

1. รายนามผู้ทรงคุณวุฒิในการตรวจสอบเครื่องมือ
 - 1.1 ผู้ทรงคุณวุฒิด้านการจัดการเรียนการสอนภาษาอังกฤษ
 - 1.1.1 ผศ.ดร. จิระดา วุฑฒิยากร
อาจารย์ประจำ สถาบันภาษาอังกฤษ จุฬาลงกรณ์มหาวิทยาลัย
 - 1.1.2 ดร. สุทธิรักษ์ ทรัพย์สิรินทร์
อาจารย์ประจำ สถาบันภาษาอังกฤษ จุฬาลงกรณ์มหาวิทยาลัย
 - 1.1.3 ดร. ธิญญาพร อารียา
อาจารย์ประจำ สถาบันภาษาอังกฤษ จุฬาลงกรณ์มหาวิทยาลัย
 - 1.1.4 ดร.ยวดี อยู่สบาย
นักวิชาการศึกษานานาชาติพิเศษ สถาบันภาษาอังกฤษ สำนักงานคณะกรรมการ
การศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ
 - 1.1.5 อาจารย์ รติธร วาดเขียน
ครูชำนาญการพิเศษ โรงเรียนสายน้ำทิพย์
 - 1.1.6 อาจารย์ สุนันทา วงษ์สวัสดิ์
ครูชำนาญการพิเศษ โรงเรียนประถมนนทรีศึกษา
 - 1.2 ผู้ทรงคุณวุฒิด้านการวัดและประเมินผลการศึกษาและจัดการเรียนการสอนโดยใช้แฟ้มสะสมงาน
 - 1.2.1 รศ.ดร.ภาวิณี ศรีสุขวัฒนานนท์
อดีตอาจารย์ประจำ คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
 - 1.2.2 ดร.ไพวัลย์ พิทักษ์สาลี
ผู้อำนวยการสำนักทดสอบทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้น
พื้นฐาน กระทรวงศึกษาธิการ
 - 1.2.3 ดร.วิษณุ ทรัพย์สมบัติ
นักวิชาการศึกษานานาชาติพิเศษ สำนักทดสอบทางการศึกษา สำนักงาน
คณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ

2. ผู้ทรงคุณวุฒิที่ทำหน้าที่เป็นผู้ประเมินความสามารถด้านทักษะการพูดภาษาอังกฤษของนักเรียน
ในชั้นทดลองใช้เกณฑ์การให้คะแนน
 - 2.1 นางสาวสกุลกานต์ อินทร์ไทร
นิสิตปริญญาเอก สาขาภาษาอังกฤษเป็นภาษานานาชาติ
(English as an international language) สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย
 - 2.2 นางพรชนก สุขพันธ์
นิสิตปริญญาเอก สาขาภาษาอังกฤษเป็นภาษานานาชาติ
(English as an international language) สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย
 - 2.3 นางสาวรัชต์สุดา ประดิษฐ์เพชร
นิสิตปริญญาเอก สาขาภาษาอังกฤษเป็นภาษานานาชาติ
(English as an international language) สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย



ภาคผนวก ข

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล



**แบบสัมภาษณ์แบบมีโครงสร้างเกี่ยวกับการประเมินทักษะการพูดภาษาอังกฤษ
ของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงาน**

คำชี้แจง

1. แบบสัมภาษณ์ฉบับนี้ จัดทำเพื่อศึกษาความคิดเห็นของผู้ทรงคุณวุฒิเกี่ยวกับการพัฒนาแฟ้มสะสมงานสำหรับการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

2. แบบสัมภาษณ์แบ่งเป็น 2 ส่วน คือ

ส่วนที่ 1 ข้อมูลผู้ให้สัมภาษณ์

ส่วนที่ 2 ความคิดเห็นเกี่ยวกับการพัฒนาแฟ้มสะสมงานและการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ส่วนที่ 1 ข้อมูลผู้ให้สัมภาษณ์

ชื่อ.....ตำแหน่ง.....

สังกัด/หน่วยงาน.....

วัน เวลา ที่สัมภาษณ์.....

ส่วนที่ 2 ความคิดเห็นเกี่ยวกับการพัฒนาแฟ้มสะสมงานและการประเมินทักษะการพูด

ภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

1. องค์ประกอบของแฟ้มสะสมงานสำหรับการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ควรมีองค์ประกอบอะไรบ้าง เพราะเหตุใด และผลงานของนักเรียนควรถูกเก็บรวบรวม ในลักษณะใดในการประเมินระหว่างเรียน (formative assessment) และสรุปรวม (summative assessment) เพื่อให้เห็นถึงพัฒนาการของนักเรียน เช่น ควรใช้ในรูปแบบ e-portfolio (เพราะเหตุใด)/ การบันทึกเสียง (เพราะเหตุใด)/ การบันทึก VDO (เพราะเหตุใด) เป็นต้น

2. ขั้นตอนการพัฒนาแฟ้มสะสมงานสำหรับการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ควรมีขั้นตอนใดบ้าง เพราะเหตุใด ขั้นตอนใดเป็นขั้นที่สำคัญที่สุดต้องมี (ขาดไม่ได้) เพราะเหตุใด

3. เครื่องมือที่ใช้ในกระบวนการเพิ่มสะสมงานสำหรับการประเมินทักษะการพูดภาษาอังกฤษในแต่ละขั้นตอน (ตามทีกล่าวใน ข้อ 2) ควรมีอะไรบ้าง เช่น นักเรียนประเมินตนเอง ประเมินเพื่อน ครู ประเมินนักเรียน การสะท้อนคิดต่อผลงานของนักเรียน การให้ข้อมูลป้อนกลับจากครู เป็นต้น และควรมีลักษณะการให้คะแนนแบบใดจึงจะเหมาะสม เช่น แบบตรวจสอบรายการ แบบมาตราประเมินค่า เป็นต้น

4. การให้ข้อมูลย้อนกลับของครู (teacher feedback) ในทันที (immediate feedback) โดยใช้การให้ข้อมูลย้อนกลับแบบบรรยายทั้งระบุความสำเร็จที่นักเรียนทำได้ดีแล้ว และระบุข้อผิดพลาดและบอกแนวทางแก้ไข (correction of errors) มีข้อควรคำนึงถึงอะไรบ้าง ควรมีการให้ข้อมูลย้อนกลับกี่ครั้ง และมีแนวทางในการดำเนินการจัดการเรียนการสอนอย่างไร และนอกจากครูแล้ว สำหรับนักเรียนชั้น ป. 6 นักเรียนในชั้นจะสามารถให้ข้อมูลย้อนกลับแก่เพื่อนในชั้น (peer feedback) ได้หรือไม่ เพราะเหตุใด

5. ภาระงาน (tasks) ที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ระหว่างเรียน (formative assessment) ควรมีภาระงานอะไรบ้าง และใช้จำนวนกี่ภาระงาน เพราะเหตุใด

6. ภาระงาน (tasks) ที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 เพื่อสรุปรวม (summative assessment) ควรเป็นภาระงานใด และมีลักษณะอย่างไร เพราะเหตุใด

7. เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ แบบแยกส่วน (analytic rating scale) ควรมีองค์ประกอบอะไรบ้าง และแต่ละองค์ประกอบมีการให้คะแนนกี่ระดับ เพราะเหตุใด

8. จำนวนผู้ประเมินภาระงานการพูดในระหว่างเรียน (formative assessment) ควรมีจำนวนกี่คน และเป็นใครบ้าง เพราะเหตุใด

.....

.....

9. จำนวนผู้ประเมินแฟ้มสะสมงานๆ ในการสรุปรวม (summative assessment) ควรมีจำนวนกี่คน และเป็นใครบ้าง เพราะเหตุใด

.....

.....

ขอบคุณค่ะ



แบบสอบถามเพื่อพิจารณาความเหมาะสมของเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ
ของนักเรียนชั้นประถมศึกษาปีที่ 6

คำชี้แจง

1. แบบสอบถามฉบับนี้ จัดทำขึ้นเพื่อพิจารณาความเหมาะสมของเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6
2. แบบสอบถามแบ่งเป็น 4 ส่วน คือ
 1. ส่วนที่ 1 ข้อมูลส่วนตัวของผู้ทรงคุณวุฒิ
 2. ส่วนที่ 2 คำอธิบายภาระงาน (tasks) ที่ใช้ในการทดสอบทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6
 3. ตอนที่ 3 เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6
 4. ส่วนที่ 4 ความคิดเห็นต่อเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6
3. ขอความกรุณาท่านผู้ทรงคุณวุฒิตอบตามความคิดเห็นเพื่อเกิดประโยชน์สูงสุดทางวิชาการ
4. แบบสอบถามฉบับนี้นำไปวิเคราะห์เพื่อประโยชน์ทางวิชาการเท่านั้น จะไม่ก่อให้เกิดผลกระทบใดๆ ต่อท่าน ขอขอบพระคุณในความร่วมมือในการตอบแบบสอบถามนี้เป็นอย่างดี

การพัฒนาเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ส่วนที่ 1 ข้อมูลส่วนตัวของผู้ทรงคุณวุฒิ

ชื่อ.....ตำแหน่ง.....

สถานที่ทำงาน.....

วุฒิการศึกษา.....

ส่วนที่ 2 คำอธิบายภาระงาน (tasks) ที่ใช้ในการทดสอบทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ภาระงาน (tasks) ที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วย 4 ภาระงาน โดยเป็นภาระงานที่ใช้ในการประเมินระหว่างเรียน จำนวน 3 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า และการเล่านิทาน และภาระงานที่ใช้ประเมินสรุปรวม จำนวน 1 ภาระงาน คือ การบรรยายภาพ ซึ่งแต่ละภาระงานมีรายละเอียด ดังนี้

1. การสัมภาษณ์ (interview) เป็นการสัมภาษณ์โดยครูเป็นผู้สัมภาษณ์เกี่ยวกับเรื่องใกล้ตัว เช่น กิจกรรมที่ทำในตอนเช้าก่อนมาโรงเรียน กิจกรรมที่ทำในห้องเรียน กีฬาที่ชอบ และเหตุผลที่ชอบ เป็นต้น

2. การนำเสนอปากเปล่า (oral presentation) เป็นการนำเสนอโดยครูอาจให้นักเรียนเตรียมตัวมาก่อนล่วงหน้า และพูดนำเสนอเกี่ยวกับเรื่องราวของตนเอง เช่น ประวัติโดยย่อ สิ่งที่ทำในเวลาว่าง กิจกรรมที่สนใจ เป็นต้น

3. การเล่านิทาน (storytelling) เป็นกิจกรรมที่ครูให้นักเรียนเตรียมตัวก่อน โดยเตรียมนิทานมาเล่าคนละ 1 เรื่อง หรือครูอาจเตรียมมาให้ และอาจนำหนังสือนิทานหรืออุปกรณ์มาประกอบการเล่า

4. การบรรยายภาพ (picture description) เป็นกิจกรรมที่ครูให้นักเรียนเตรียมตัวก่อน โดยการเตรียมภาพประกอบที่นักเรียนสนใจ หรือที่ครูเตรียมไว้ให้ แล้วให้นักเรียนบรรยายเรื่องราวในภาพของตนเอง

ส่วนที่ 3 เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

ภาระงานทั้ง 4 ภาระงานใช้เกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) โดยมีขอบเขตเนื้อหาในการประเมิน ดังนี้

1. คำศัพท์ (vocabulary)
2. การสร้างประโยค (syntax)
3. การเชื่อมโยงเนื้อหา (cohesion)
4. การออกเสียง (pronunciation)
5. บรรลู่วัตถุประสงค์ (ideational function)
6. ความคล่องแคล่ว (fluency)

ให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี ตามลำดับ

ส่วนที่ 4 ความคิดเห็นต่อเกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

คำชี้แจง โปรดทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับความคิดเห็นของท่านมากที่สุด และหากท่านเห็นว่าเกณฑ์ที่ใช้ในการประเมิน ไม่ควรมี โปรดแสดงเหตุผลประกอบว่าเกณฑ์ควรมีการเปลี่ยนแปลงอย่างไร

4.1 ความเหมาะสมของเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale)

เกณฑ์	ความเหมาะสม		เหตุผลประกอบและข้อเสนอแนะ
	ควรมี	ไม่ควรมี	
1. คำศัพท์ (vocabulary)			
2. การสร้างประโยค (syntax)			
3. การเชื่อมโยงเนื้อหา (cohesion)			
4. การออกเสียง (pronunciation)			
5. บรรลู่วัตถุประสงค์ (ideational function)			
6. ความคล่องแคล่ว (fluency)			

4.2 ความเหมาะสมของเกณฑ์การให้คะแนนแบบรูบริก (scoring rubric)

คำชี้แจง โปรดทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับความคิดเห็นของท่านมากที่สุด โดยพิจารณาความเหมาะสมของเกณฑ์การให้คะแนน ดังนี้

+1 หมายถึง มีความเหมาะสม

0 หมายถึง ไม่แน่ใจ

-1 หมายถึง ควรปรับปรุง

1. คำศัพท์ (vocabulary) การให้คะแนนด้านคำศัพท์ จะพิจารณาจากการพูดของนักเรียนที่สะท้อน 2 สิ่ง คือ ขอบเขต (range) ของคำศัพท์ที่ใช้ และ ความถูกต้อง (accuracy) กล่าวคือ ขอบเขต (range) ของการใช้คำศัพท์ที่มีความหลากหลาย และความสามารถในการใช้คำศัพท์ได้อย่างถูกต้อง (accuracy) เหมาะสม ตรงตามความหมายที่ต้องการสื่อสาร

ระดับ ความสามารถ	คำอธิบาย	ความคิดเห็น			ข้อเสนอแนะ
		+1	0	-1	
อ่อนมาก (1 คะแนน)	ความสามารถในการใช้คำศัพท์ยังมีความจำกัด ใช้คำศัพท์ไม่หลากหลาย ใช้คำศัพท์คำเดิมซ้ำๆ ทุกประโยค หรือเกือบทุกประโยค และใช้ไม่ถูกต้อง ไม่มีความเหมาะสม				
อ่อน (2 คะแนน)	ความสามารถในการใช้คำศัพท์ยังมีความจำกัด ยังมีการใช้คำศัพท์คำเดิมซ้ำๆ อยู่บ่อยครั้ง และมีการใช้คำศัพท์ที่ไม่ถูกต้อง ไม่มีความเหมาะสมอยู่บ่อยครั้ง				
พอใช้ (3 คะแนน)	สามารถใช้คำศัพท์ได้หลากหลาย แต่ยังมีการใช้คำศัพท์คำเดิมซ้ำบ้างบางครั้ง และสามารถใช้คำศัพท์ได้อย่างถูกต้องและเหมาะสม แต่ยังมีผิดพลาดบ้างบางครั้ง				
ดี (4 คะแนน)	สามารถใช้คำศัพท์ได้อย่างหลากหลาย และการใช้คำศัพท์ส่วนใหญ่สามารถใช้ได้ถูกต้องเหมาะสม และผิดพลาดน้อยมาก				

ข้อเสนอแนะอื่นๆ.....
.....

2. การสร้างประโยค (syntax) การให้คะแนนด้านการสร้างประโยค พิจารณาจากความสามารถของนักเรียนในเชื่อมโยงคำต่างๆ เพื่อสร้างประโยคได้อย่างมีความถูกต้อง (accuracy) ตามที่ต้องการ

ระดับ ความสามารถ	คำอธิบาย	ความคิดเห็น			ข้อเสนอแนะ
		+1	0	-1	
อ่อนมาก (1 คะแนน)	ความสามารถในการใช้คำศัพท์ยังมีความจำกัด ใช้คำศัพท์ไม่หลากหลาย ใช้คำศัพท์คำเดิมซ้ำๆ ทุกประโยค หรือเกือบทุกประโยค และใช้ไม่ถูกต้อง ไม่มีความเหมาะสม				
อ่อน (2 คะแนน)	ความสามารถในการใช้คำศัพท์ยังมีความจำกัด ยังมีการใช้คำศัพท์คำเดิมซ้ำๆ อยู่บ่อยครั้ง และมีการใช้คำศัพท์ที่ไม่ถูกต้อง ไม่มีความเหมาะสมอยู่บ่อยครั้ง				
พอใช้ (3 คะแนน)	สามารถใช้คำศัพท์ได้หลากหลาย แต่ยังมีการใช้คำศัพท์คำเดิมซ้ำบ้างบางครั้ง และสามารถ ใช้คำศัพท์ได้อย่างถูกต้องและเหมาะสม แต่ยังมี ความผิดพลาดบ้างบางครั้ง				
ดี (4 คะแนน)	สามารถใช้คำศัพท์ได้อย่างหลากหลาย และการใช้คำศัพท์ส่วนใหญ่สามารถใช้ได้ ถูกต้องเหมาะสม และผิดพลาดน้อยมาก				

ข้อเสนอแนะอื่นๆ

.....

.....

3. การเชื่อมโยงเนื้อหา (cohesion) การให้คะแนนด้านการเชื่อมโยงเนื้อหา พิจารณาจากการที่นักเรียนสามารถพูดเชื่อมโยงประโยคแต่ละประโยคให้เป็นเรื่องราวที่มีความสัมพันธ์กันได้อย่างถูกต้องเหมาะสม เช่น การใช้ and or but also this such เป็นต้น

ระดับ ความสามารถ	คำอธิบาย	ความคิดเห็น			ข้อเสนอแนะ
		+1	0	-1	
อ่อนมาก (1 คะแนน)	มีการใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อยมาก หรือไม่ใช้เลย และการใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคเกือบทั้งหมดไม่ถูกต้อง				
อ่อน (2 คะแนน)	มีการใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อย การใช้คำเชื่อมโยงทำให้ความสัมพันธ์ระหว่างประโยคมีความสับสนอยู่บ่อยครั้ง และทำให้ความเข้าใจคลาดเคลื่อน				
พอใช้ (3 คะแนน)	มีการใช้คำเชื่อมโยงเนื้อหาระหว่างประโยค แต่ความสัมพันธ์ระหว่างประโยคมีความสับสนอยู่บ้าง แต่ไม่ทำให้ความเข้าใจคลาดเคลื่อน				
ดี (4 คะแนน)	การใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคส่วนมากมีความถูกต้อง ชัดเจน และไม่ทำให้ความเข้าใจคลาดเคลื่อน มีความผิดพลาดน้อยมาก				

ข้อเสนอแนะอื่นๆ

.....

.....

.....

.....

4. การออกเสียง (pronunciation) การให้คะแนนด้านการออกเสียง พิจารณาจากการที่นักเรียนสามารถเน้นเสียงหนักเบา (stress) เว้นวรรคการหยุด (segmental sound) และการใช้เสียงสูงต่ำ (intonation) ได้อย่างถูกต้อง เหมาะสม และทำให้ผู้ฟังเข้าใจในสิ่งที่ต้องการสื่อสารได้

ระดับ ความสามารถ	คำอธิบาย	ความคิดเห็น			ข้อเสนอแนะ
		+1	0	-1	
อ่อนมาก (1 คะแนน)	การออกเสียงในการพูดเกือบทั้งหมดทำให้ไม่สามารถเข้าใจได้ หรือนักเรียนพูดน้อยมากจนไม่เข้าใจสิ่งที่ต้องการสื่อสาร				
อ่อน (2 คะแนน)	การออกเสียงในการพูดยังมีความผิดพลาดบ่อยครั้ง มีการออกเสียงเน้นหนักเบา เว้นวรรคการหยุด และการใช้เสียงสูงต่ำมีความผิดพลาดอยู่มาก และทำให้ไม่สามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้ หรือทำให้เข้าใจคลาดเคลื่อน				
พอใช้ (3 คะแนน)	การออกเสียงในการพูดยังมีความผิดพลาดอยู่บ้าง มีการออกเสียงเน้นหนักเบา เว้นวรรคการหยุด และการใช้เสียงสูงต่ำยังมีความผิดพลาดบ้าง แต่ผู้ฟังยังคงสามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้				
ดี (4 คะแนน)	การออกเสียงในการพูดส่วนใหญ่สามารถเข้าใจได้ง่าย มีการออกเสียงเน้นหนักเบา เว้นวรรคการหยุด และการใช้เสียงสูงต่ำมีความถูกต้องเหมาะสม มีความผิดพลาดน้อยมาก				

ข้อเสนอแนะอื่นๆ

.....

.....

.....

.....

5. **บรรล่วัตถุประสงค์ (ideational function)** การให้คะแนนด้านบรรล่วัตถุประสงค์ พิจารณาจากการที่นักเรียนสามารถพูดสื่อสารได้บรรลุตามวัตถุประสงค์ที่ต้องการจะสื่อสาร โดยการพูดนั้นทำให้ผู้ฟังเข้าใจว่าต้องการสื่อสารอะไร เช่น พูดได้ตรงกับหัวเรื่องที่ระบุ ตอบคำถามได้อย่างเหมาะสมกับลักษณะคำถาม เป็นต้น

ระดับ ความสามารถ	คำอธิบาย	ความคิดเห็น			ข้อเสนอแนะ
		+1	0	-1	
อ่อนมาก (1 คะแนน)	ในการพูดเกือบทั้งหมดไม่สามารถเข้าใจได้ว่าต้องการสื่อสารอะไร พูดไม่ตรงประเด็น หรือไม่พูดเลย				
อ่อน (2 คะแนน)	เนื้อหาในการพูดมีความสอดคล้องเหมาะสมน้อยกับหัวเรื่องหรือสิ่งที่ต้องการสื่อสาร ทำให้ไม่สามารถเข้าใจได้ว่าวัตถุประสงค์ของการพูดคืออะไร				
พอใช้ (3 คะแนน)	เนื้อหาในการพูดมีความสอดคล้องเหมาะสม ถึงแม้จะมีความผิดพลาดในการพูดในบางครั้ง แต่ทำให้เข้าใจวัตถุประสงค์ที่ต้องการสื่อสารได้				
ดี (4 คะแนน)	เนื้อหาในการพูดมีความสอดคล้องเหมาะสม ผู้ฟังสามารถเข้าใจถึงวัตถุประสงค์ในสิ่งที่ต้องการสื่อสารได้โดยง่าย มีความผิดพลาดในการพูดน้อยมาก				

ข้อเสนอแนะอื่นๆ

.....

.....

.....

.....

6. ความคล่องแคล่ว (fluency) หมายถึง การใช้ภาษาที่มีความคล่องแคล่ว ลื่นไหล ซึ่งพิจารณาจากการที่นักเรียนสามารถพูดภาษาอังกฤษได้อย่างคล่องแคล่ว ราบรื่น และมีความลื่นไหลในระหว่างที่พูด และไม่สะดุดหรือกระตุก หยุดชะงัก ระหว่างที่พูด

ระดับ ความสามารถ	คำอธิบาย	ความคิดเห็น			ข้อเสนอแนะ
		+1	0	-1	
อ่อนมาก (1 คะแนน)	การพูดยังตะกุกตะกัก พูดเป็นคำๆ ไม่เป็นประโยค หรือไม่พูดเลย				
อ่อน (2 คะแนน)	การพูดบ่อยครั้งยังซ้ำมาก และไม่ลื่นไหล มีการหยุดเป็นระยะ พูดไม่จบประโยค				
พอใช้ (3 คะแนน)	การพูดมีการหยุดบ้าง ไม่ลื่นไหล มีการหยุดคิดบ้างบางครั้ง แต่สามารถพูดได้จบประโยค				
ดี (4 คะแนน)	การพูดมีความเร็วสม่ำเสมอ พูดได้ลื่นไหล ไม่ ตะกุกตะกัก ไม่หยุดชะงัก และสามารถพูดได้จบประโยค				

ข้อเสนอแนะอื่นๆ

.....

.....

.....

.....

**การตรวจสอบความเหมาะสมของคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูด
ภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6**

ส่วนที่ 1 ข้อมูลส่วนตัวของผู้ทรงคุณวุฒิ

ชื่อ.....ตำแหน่ง.....

สถานที่ทำงาน.....

วุฒิการศึกษา.....

**ส่วนที่ 2 ความเหมาะสมของคู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ
ของนักเรียนชั้นประถมศึกษาปีที่ 6**

คำชี้แจง โปรดทำเครื่องหมาย ✓ ลงในช่องที่ตรงกับความคิดเห็นของท่านมากที่สุด โดยพิจารณา
ความเหมาะสมของเกณฑ์การให้คะแนน ดังนี้

+1 หมายถึง มีความเหมาะสม

0 หมายถึง ไม่แน่ใจ

-1 หมายถึง ควรปรับปรุง

รายการ	ความคิดเห็น			ข้อเสนอแนะ
	+1	0	-1	
1. โครงสร้างของคู่มือฯ				
1.1 บทนำ				
1.2 วัตถุประสงค์ของแฟ้มสะสมงาน ฯ				
1.3 คำนิยามศัพท์				
1.4 เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ				
2. ส่วนประกอบของแฟ้มสะสมงาน ฯ				
2.1 ส่วนนำ				
2.1.1 ปก				
2.1.2 คำนำ				
2.1.3 สารบัญ				
2.1.4 ข้อมูลนักเรียน				

รายการ	ความคิดเห็น			ข้อเสนอแนะ
	+1	0	-1	
2.1.5 เป้าหมายการเรียนรู้ ความคาดหวังในการเรียน 2.2 ส่วนเนื้อหา 2.2.1 ผลการประเมินระหว่างเรียนและการประเมินสรุป รวม 1) ผลการประเมินตนเอง 2) ผลการประเมินจากครู 3 ผลการประเมินจากเพื่อน 2.2.2 ข้อมูลย้อนกลับจากครู 2.2.3 ผลการสะท้อนคิดต่อผลงานตนเอง 2.3 ส่วนภาคผนวก 2.3.1 คำอธิบายภาระงาน 2.3.2 เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ 2.3.3 CD บันทึกเสียงการพูดภาษาอังกฤษของนักเรียน				
3. ขั้นตอนของการพัฒนาแฟ้มสะสมงานฯ 3.1 ขั้นการวางแผน 3.2 ขั้นการเตรียมความพร้อมนักเรียน 3.3 ขั้นการเก็บรวบรวมหลักฐาน 3.4 ขั้นการติดตามความก้าวหน้า 3.5 ขั้นการปรับปรุงผลงาน 3.6 ขั้นการสะท้อนคิด 3.7 ขั้นการจัดแสดงผลงาน				
4. เครื่องมือที่ใช้ในแฟ้มสะสมงาน 4.1 สำหรับนักเรียน 4.1.1 แบบบันทึกข้อมูลส่วนตัว 4.1.2 แบบบันทึกการกำหนดวัตถุประสงค์ เป้าหมาย ในการเรียนวิชาภาษาอังกฤษ 4.1.3 แบบบันทึกการกำหนดวัตถุประสงค์ เป้าหมาย ในการเรียนด้วยแฟ้มสะสมงานการประเมินทักษะการพูด ภาษาอังกฤษ				

รายการ	ความคิดเห็น			ข้อเสนอแนะ
	+1	0	-1	
4.1.4 แบบบันทึกคะแนนการพูด 4.1.5 แบบบันทึกการวิเคราะห์ความสามารถในการพูดภาษาอังกฤษของตนเอง 4.1.6 แบบบันทึกการกำหนดเป้าหมายในการเรียน/การพูดภาษาอังกฤษของตนเอง 4.1.7 ประเมินตนเอง 4.1.8 แบบประเมินเพื่อน 4.1.9 แบบบันทึกการวางแผนการปรับปรุงการพูด 4.1.10 แบบบันทึกการสะท้อนคิดต่อผลงานของตน 4.1.11 แบบบันทึกจุดเด่น จุดที่ควรพัฒนาของนักเรียน				
4.2 สำหรับครู 4.2.1 เกณฑ์การพูดภาษาอังกฤษ 4.2.3 แบบประเมินทักษะการพูดของนักเรียน 4.2.4 แบบบันทึกคะแนนทักษะการพูดภาษาอังกฤษของนักเรียน 4.2.5 แบบบันทึกการให้ข้อมูลย้อนกลับแก่นักเรียนเป็นรายคน 4.2.6 แบบบันทึกการให้ข้อมูลย้อนกลับแก่นักเรียนเป็นรายคน 4.2.7 แบบบันทึกจุดเด่น จุดที่ควรพัฒนาของนักเรียนเป็นรายคน 4.2.8 แบบบันทึกจุดเด่น จุดที่ควรพัฒนาของนักเรียนในภาพรวม				

ข้อเสนอแนะ

.....

.....

.....

ขอบคุณค่ะ

คู่มือการใช้

แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของ
นักเรียนชั้นประถมศึกษาปีที่ 6



โดย

นางสาวดารุวรรณ ศรีแก้ว

ผู้ช่วยศาสตราจารย์ ดร.กมลวรรณ ตั้งธนกานนท์ (อาจารย์ที่ปรึกษา)
ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสิ (อาจารย์ที่ปรึกษาร่วม)

เอกสารนี้เป็นส่วนหนึ่งของวิทยานิพนธ์ปริญญาโท
สาขาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยา
การศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556

สารบัญ

	หน้า
บทนำ.....	x
วัตถุประสงค์ของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ ของนักเรียนชั้นประถมศึกษาปีที่ 6.....	x
คำนิยามศัพท์.....	x
ผู้ประเมิน.....	x
ภาระงานการพูดภาษาอังกฤษ.....	x
เกณฑ์การพูดภาษาอังกฤษ.....	x
โครงสร้างของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของ นักเรียนชั้นประถมศึกษาปีที่ 6.....	x
ขั้นตอนของการพัฒนาแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ชั้นประถมศึกษาปีที่ 6.....	x
เครื่องมือที่ใช้ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของ นักเรียนชั้นประถมศึกษาปีที่ 6.....	x

บทนำ

คู่มือการใช้แฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 จัดทำขึ้นเพื่อให้ครูผู้สอนใช้ประกอบการจัดการเรียนรู้เกี่ยวกับการสอนทักษะการพูดวิชาภาษาอังกฤษ โดยใช้แฟ้มสะสมงาน และนักเรียนใช้เป็นแนวทางในการเก็บสะสมผลงานของตนเองในรูปแบบแฟ้มสะสมผลงานเพื่อนำไปพัฒนาทักษะการเรียนรู้ในวิชาต่างๆ ซึ่งคู่มือฉบับนี้ประกอบไปด้วยขั้นตอนของแฟ้มสะสมงาน เครื่องมือสำหรับครูและนักเรียน ซึ่งแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษนี้เกิดจากการสังเคราะห์เอกสาร และงานวิจัยต่างๆ เพื่อนำมาเป็นการพัฒนาแฟ้มสะสมงานเพื่อการประเมินทักษะการพูดภาษาอังกฤษที่มีการให้คะแนนทักษะการพูดของนักเรียน ทั้งระหว่างเรียน (formative assessment) และสรุปรวม (summative assessment) ตามกรอบแนวคิดของ Burke, Fogart and Begrad (1994), Lam and Lee (2010) และจากการสัมภาษณ์ผู้ทรงคุณวุฒิด้านการจัดการเรียนการสอนภาษาอังกฤษ การวัดและการประเมินผลการศึกษา และการประเมินโดยใช้แฟ้มสะสมงาน ซึ่งการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานในครั้งนี้เน้นที่การให้ข้อมูลย้อนกลับของครู (teacher feedback) ที่จะทำให้นักเรียนในการประเมินแต่ละครั้ง ซึ่งเป็นการให้ข้อมูลย้อนกลับแบบเชิงบรรยายที่มีการระบุความสำเร็จและการระบุสิ่งที่ควรปรับปรุง (corrective feedback) เพื่อเป็นการเสริมแรงในการเรียนแก่นักเรียน ให้นักเรียนได้รู้เกี่ยวกับสิ่งที่จะนำตนเองไปสู่ความสำเร็จในการเรียนได้ และรู้ถึงข้อผิดพลาดและการแก้ไขให้ถูกต้อง (correction of errors) (Tuntall & Gipps, 1996) ซึ่งขั้นตอนของแฟ้มสะสมงาน ได้แก่ การวางแผน การเตรียมความพร้อมนักเรียน การเก็บรวบรวมหลักฐาน การติดตามความก้าวหน้า การปรับปรุงผลงาน การสะท้อนคิด และการจัดแสดงผลงาน และแฟ้มสะสมงานประกอบด้วยมีส่วนประกอบ 3 ส่วน คือ ส่วนนำ ส่วนเนื้อหา และภาคผนวก ส่วนเครื่องมือที่ใช้ในแฟ้มสะสมงานในแต่ละขั้นตอนนั้น ได้แนบไว้ในส่วนของภาคผนวกของคู่มือเล่มนี้แล้ว

วัตถุประสงค์ของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

เพื่อประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้กระบวนการของแฟ้มสะสมงาน เพื่อให้นักเรียนได้ปฏิบัติงานการพูดภาษาอังกฤษที่เป็นจริง คือ ภาระงานการพูดภาษาอังกฤษ 4 ภาระงาน ได้แก่ 1) การสัมภาษณ์ (interview) 2) การนำเสนอปากเปล่า (oral presentation) 3) การเล่านิทาน (storytelling) และ 4) การบรรยายภาพ (picture description)

คำนิยามศัพท์

1. แฟ้มสะสมงาน หมายถึง สิ่งที่ใช้ในการรวบรวมผลงานนักเรียนจากการเรียนรู้ในชั้นเรียน และนำไปสู่การประเมินการเรียนของนักเรียนตามสภาพจริงวิธีหนึ่ง ซึ่งเป็นวิธีการที่ทำให้ผู้เรียนได้ลงมือปฏิบัติงานที่เป็นจริง และผู้เรียนได้รับข้อมูลย้อนกลับจากครู การสะท้อนคิด ได้ประเมินตนเอง และประเมินเพื่อนในชั้นเรียนด้วยทำให้มีความรับผิดชอบต่อการเรียนของตน

2. การประเมินทักษะการพูด หมายถึง การทดสอบทักษะการพูดภาษาอังกฤษของนักเรียน โดยให้นักเรียนได้แสดงความสามารถในการพูดสื่อความในการสื่อสารออกมา ซึ่งสำหรับการวิจัยครั้งนี้ ประเมินทักษะการพูดภาษาอังกฤษโดยใช้ภาระงานการพูด (tasks)

3. ภาระงานการพูด (task) หมายถึง กิจกรรมที่ให้นักเรียนใช้สำหรับการทดสอบทักษะการพูดของนักเรียนโดยการให้นักเรียนได้แสดงการพูดสื่อสารหรือพูดสื่อความออกมาตามความรู้ความสามารถทางภาษาอังกฤษ ซึ่งภาระงานที่ใช้ในการประเมินครั้งนี้ ได้แก่ 1) การสัมภาษณ์ (interview) 2) การนำเสนอปากเปล่า (oral presentation) 3) การเล่านิทาน (storytelling) และ 4) การบรรยายภาพ (picture description) โดยมีเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale)

4. การประเมินระหว่างเรียน (formative assessment) หมายถึง การประเมินทักษะการพูดของนักเรียนในระหว่างการดำเนินการเรียนการสอนโดยใช้แฟ้มสะสมงานฯ โดยมีจุดประสงค์เพื่อศึกษาความก้าวหน้าของผู้เรียน และปรับปรุงและพัฒนาการพูดของนักเรียน

5. การประเมินสรุปรวม (summative assessment) หมายถึง การประเมินเมื่อสิ้นสุดการดำเนินการจัดการเรียนการสอนด้วยแฟ้มสะสมงานฯ เพื่อสรุปรวมและให้ระดับทักษะการพูดแก่นักเรียน

ผู้ประเมิน

ผู้ประเมินผลงานหรือผู้ที่ทำการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนโดยใช้แฟ้มสะสมงานในครั้งนี้ นักเรียนจะได้รับการประเมินจาก 3 แห่ง ได้แก่ นักเรียนประเมินตนเอง เพื่อนนักเรียนที่เรียนในชั้นเรียนเดียวกัน และครู ซึ่งครูประกอบด้วยครูผู้สอนในชั้นเรียน และครูในกลุ่มสาระการเรียนรู้ภาษาต่างประเทศในโรงเรียน โดยครูผู้สอนในชั้นเรียนจะทำหน้าที่ประเมินการพูดทุกครั้งและทุกภาระงาน ส่วนครูในกลุ่มสาระการเรียนรู้ภาษาต่างประเทศในโรงเรียนจะทำการประเมินเฉพาะการพูดครั้งที่ 3 หรือครั้งสุดท้ายของแต่ละภาระงาน

ภาระงานการพูดภาษาอังกฤษ

ภาระงาน (tasks) ที่ใช้ในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วย 4 ภาระงาน โดยเป็นภาระงานที่ใช้ในการประเมินระหว่างเรียน จำนวน 3 ภาระงาน

ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า และการเล่านิทาน และเป็นภาระงานที่ใช้ประเมินสรุปรวมจำนวน 1 ภาระงาน คือ การบรรยายประกอบภาพ ซึ่งแต่ละภาระงานมีรายละเอียด ดังนี้

1. การสัมภาษณ์ (interview) เป็นการสัมภาษณ์โดยครูเป็นผู้สัมภาษณ์เกี่ยวกับเรื่องใกล้ตัว เช่น กิจกรรมที่ทำในตอนเช้าก่อนมาโรงเรียน กิจกรรมที่ทำในห้องเรียน กีฬาที่ชอบ และเหตุผลที่ชอบ เป็นต้น

2. การนำเสนอปากเปล่า (oral presentation) เป็นกิจกรรมการพูดที่ให้นักเรียนได้พูดนำเสนอเรื่องราวเกี่ยวกับตนเอง โดยครูให้นักเรียนเตรียมตัวมาก่อนล่วงหน้า เช่น ให้ค้นหาคำศัพท์ที่เกี่ยวข้อง ลองแต่งประโยค ปรึกษากับครูในชั้นเรียน หรือผู้ปกครอง เนื้อหาที่พูด เช่น ประวัติโดยย่อ งานอดิเรก กิจกรรมที่สนใจ เป็นต้น

3. การเล่านิทาน (storytelling) เป็นกิจกรรมที่ครูให้นักเรียนเตรียมตัวก่อนล่วงหน้า โดยเตรียมนิทานมาเล่าคนละ 1 เรื่อง หรือเลือกเรื่องจากที่ครูเตรียมให้ ซึ่งอาจนำหนังสือนิทานหรืออุปกรณ์มาประกอบการเล่า โดยนิทานที่จะใช้ในการเล่าครูจะเป็นผู้พิจารณาความเหมาะสม

4. การบรรยายภาพ (picture description) เป็นกิจกรรมที่ครูให้นักเรียนพูดบรรยายภาพ โดยนักเรียนจะได้เตรียมตัวก่อนล่วงหน้า คือ ศึกษารายละเอียดต่างๆ ของภาพ ชักถามรายละเอียดของภาพจากครูผู้สอน ค้นหาคำศัพท์ที่เกี่ยวข้อง ลองแต่งประโยค ปรึกษาการพูดกับครูในชั้นเรียน หรือผู้ปกครอง เป็นต้น โดยภาพประกอบที่ใช้เป็นภาพที่นักเรียนสนใจนำมาเอง หรือที่ครูเตรียมไว้ให้ แล้วให้นักเรียนบรรยายเรื่องราวในภาพของตนเอง โดยภาพที่จะนำมาบรรยายครูจะเป็นผู้พิจารณาความเหมาะสม

เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ

ภาระงานทั้ง 4 ภาระงานใช้เกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) โดยมีขอบเขตเนื้อหาในการประเมิน ดังนี้

1. คำศัพท์ (vocabulary)
2. การสร้างประโยค (syntax)
3. การเชื่อมโยงเนื้อหา (cohesion)
4. การออกเสียง (pronunciation)
5. บรรลุวัตถุประสงค์ (ideational function)
6. ความคล่องแคล่ว (fluency)

ให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี ตามลำดับ ซึ่งมีรายละเอียด ดังนี้

การแปลความหมายคะแนนความสามารถในทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 สามารถแปลความหมาย ดังนี้

คะแนนเฉลี่ย ต่ำกว่า 1.51 หมายถึง อ่อนมาก

คะแนนเฉลี่ย 1.51-2.50 หมายถึง อ่อน

คะแนนเฉลี่ย 2.51-3.50 หมายถึง พอใช้

คะแนนเฉลี่ย 3.51-4.00 หมายถึง ดี

1. คำศัพท์ (vocabulary) คือ การให้คะแนนด้านคำศัพท์ จะพิจารณาจากการพูดของนักเรียนที่สะท้อน 2 สิ่ง คือ ขอบเขต (range) ของคำศัพท์ที่ใช้ และ ความถูกต้อง (accuracy) กล่าวคือ ขอบเขต (range) ของการใช้คำศัพท์ที่มีความหลากหลาย ความสามารถในการใช้คำศัพท์ได้อย่างถูกต้อง (accuracy) และเหมาะสม (appropriation) ตรงตามความหมายที่ต้องการสื่อสาร

ตารางที่ 1 เกณฑ์การให้คะแนนด้านคำศัพท์

ระดับความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● ความสามารถในการใช้คำศัพท์ยังจำกัดอยู่มาก ● มักใช้คำศัพท์คำเดิมซ้ำๆ ทุกประโยค หรือเกือบทุกประโยค ● ใช้คำศัพท์ไม่ถูกต้อง ไม่เหมาะสม
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● ความสามารถในการใช้คำศัพท์ยังจำกัด ใช้คำศัพท์ไม่หลากหลาย ● ยังพบการใช้คำศัพท์คำเดิมซ้ำๆ อยู่บ่อยครั้ง ● มีการใช้คำศัพท์ที่ไม่ถูกต้อง ไม่เหมาะสมอยู่บ่อยครั้ง
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● สามารถใช้คำศัพท์ได้หลากหลาย ● แต่ยังใช้คำศัพท์คำเดิมซ้ำบ้างเป็นบางครั้ง ● สามารถใช้คำศัพท์ได้อย่างถูกต้องและเหมาะสม อาจยังใช้คำศัพท์ผิดพลาดบ้างเป็นบางครั้ง
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● สามารถใช้คำศัพท์ได้อย่างหลากหลาย ● การใช้คำศัพท์ส่วนใหญ่สามารถใช้ได้ถูกต้องเหมาะสม และผิดพลาดน้อยมาก หรือถูกต้องทั้งหมด ไม่มีความผิดพลาดเลย

2. การสร้างประโยค (syntax) คือ การให้คะแนนด้านการสร้างประโยค พิจารณาจากความสามารถของนักเรียนในสร้างประโยคได้อย่างมีความถูกต้อง (accuracy) ตามที่ต้องการสื่อสาร เช่น มีประธาน กริยา กรรม ที่เหมาะสม เป็นต้น

ตารางที่ 2 เกณฑ์การให้คะแนนด้านการสร้างประโยค

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● การพูดไม่มีการเชื่อมโยงให้เป็นประโยค พูดเป็นคำๆ ● มีการใช้โครงสร้างประโยคที่ไม่ถูกต้องทุกครั้ง
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการเชื่อมโยงให้เป็นประโยค ● การใช้โครงสร้างประโยคที่ไม่ถูกต้องอยู่บ่อยครั้ง ทำให้ผู้ฟังไม่สามารถเข้าใจได้อย่างกระจ่าง
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการเชื่อมโยงให้เป็นประโยค ● การใช้โครงสร้างประโยคที่ไม่ถูกต้องบ้างเป็นบางครั้ง แต่ผู้ฟังสามารถเข้าใจได้
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการเชื่อมโยงให้เป็นประโยค ● ส่วนใหญ่มีการใช้โครงสร้างประโยคที่ถูกต้อง และผิดพลาดน้อยมาก หรือไม่มีความผิดพลาดเลย

3. การเชื่อมโยงเนื้อหา (cohesion) คือ การให้คะแนนด้านการเชื่อมโยงเนื้อหา พิจารณาจากการที่นักเรียนสามารถผูกเชื่อมโยงประโยคแต่ละประโยคให้เป็นเรื่องราวที่มีความสัมพันธ์กันได้อย่างถูกต้องเหมาะสม

ตารางที่ 3 เกณฑ์การให้คะแนนด้านการเชื่อมโยงเนื้อหา

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อยมาก หรือไม่ใช้เลย ● การใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคเกือบทั้งหมดไม่ถูกต้อง
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อย ● การใช้คำเชื่อมโยงทำให้ความสัมพันธ์ระหว่างประโยคมีความสับสนอยู่บ่อยครั้ง และทำให้เข้าใจคลาดเคลื่อน
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยค ● ความสัมพันธ์ระหว่างประโยคอาจมีความสับสนอยู่บ้าง แต่ไม่ทำให้ความเข้าใจคลาดเคลื่อน
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยค ● ถูกต้อง ชัดเจน เป็นส่วนใหญ่ และไม่ทำให้เข้าใจคลาดเคลื่อน ผิดพลาดน้อยมาก หรือ ไม่มีความผิดพลาดเลย

4. การออกเสียง (pronunciation) คือ การให้คะแนนด้านการออกเสียง พิจารณาจากการที่นักเรียนสามารถเน้นเสียงหนักเบา (stress) เว้นวรรคการหยุด (segmental sound) และการใช้เสียงสูงต่ำ (intonation) ได้อย่างถูกต้อง เหมาะสม และทำให้ผู้ฟังเข้าใจในสิ่งที่ต้องการสื่อสารได้

ตารางที่ 4 เกณฑ์การให้คะแนนด้านการออกเสียง

ระดับความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงเกือบทั้งหมด มีความผิดพลาด ทำให้ไม่สามารถเข้าใจได้ ● นักเรียนพูดน้อยมาก หรือไม่พูดเลย
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงยังผิดพลาดอยู่บ่อยครั้ง ● การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดอยู่มาก ● ทำให้ไม่สามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้ หรือทำให้เข้าใจคลาดเคลื่อน
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงยังมีความผิดอยู่บ้าง ● การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดอยู่บ้าง ● แต่ผู้ฟังยังคงสามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงส่วนใหญ่ถูกต้อง หรือไม่มีความผิดพลาดเลย ● การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดน้อยมาก หรือไม่มีความผิดพลาดเลย ● ผู้ฟังสามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้โดยง่าย

5. บรรลู่วัตถุประสงค์ (ideational function) คือ การให้คะแนนด้านบรรลู่วัตถุประสงค์ พิจารณาจากการที่นักเรียนสามารถพูดสื่อสารได้บรรลุตามวัตถุประสงค์ที่ต้องการจะสื่อสาร โดยการพูดนั้นทำให้ผู้ฟังเข้าใจว่าต้องการสื่อสารอะไร เช่น พูดได้ตรงกับหัวเรื่องที่ระบุ ตอบตรงคำถาม ตอบคำถามได้อย่างเหมาะสมกับลักษณะคำถาม เป็นต้น

ตารางที่ 5 เกณฑ์การให้คะแนนด้านการบรรลู่วัตถุประสงค์

ระดับความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● พูดไม่ตรงประเด็น หรือไม่พูดเลย ● การพูดเกือบทั้งหมดไม่สามารถเข้าใจได้ว่าต้องการสื่อสารอะไร
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● เนื้อหาในการพูดไม่ค่อยสอดคล้องเหมาะสมกับหัวเรื่องหรือสิ่งที่ต้องการสื่อสาร ● ไม่สามารถเข้าใจอย่างชัดเจนได้ว่าวัตถุประสงค์ของการพูดคืออะไร
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● เนื้อหาในการพูดสอดคล้องเหมาะสม ถึงแม้จะมีความผิดพลาดในการพูดในบางครั้ง ● สามารถทำให้เข้าใจวัตถุประสงค์ที่ต้องการสื่อสารได้
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● เนื้อหาในการพูดสอดคล้องเหมาะสม มีความผิดพลาดในการพูดน้อยมาก หรือไม่ผิดพลาดเลย ● สามารถเข้าใจวัตถุประสงค์ในสิ่งที่ต้องการสื่อสารได้โดยง่าย



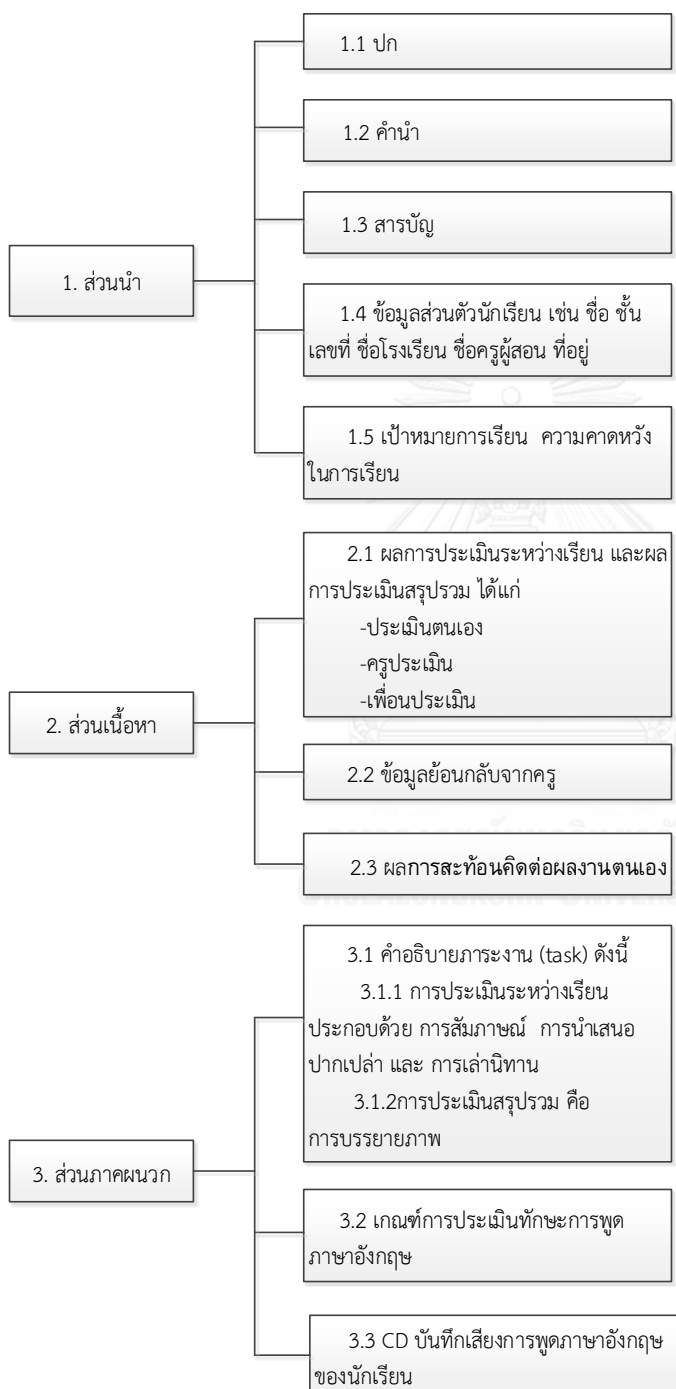
6. ความคล่องแคล่ว (fluency) คือ การใช้ภาษาที่มีความคล่องแคล่ว ลื่นไหล ซึ่งพิจารณาจากการที่นักเรียนสามารถพูดภาษาอังกฤษได้อย่างคล่องแคล่ว ราบรื่น และมีความลื่นไหลในระหว่างที่พูด และไม่สะดุดหรือกระตุก หยุดชะงัก ระหว่างที่พูด

ตารางที่ 6 เกณฑ์การให้คะแนนด้านความคล่องแคล่ว

ระดับ	คำอธิบาย
ความสามารถ	
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● การพูดยังตะกุกตะกัก พูดเป็นคำๆ ● ไม่เป็นประโยค หรือไม่พูดเลย
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● การพูดบ่อยครั้งยังซ้ำมาก และไม่ลื่นไหล ● หยุดเป็นระยะ หยุดคิดนาน พูดไม่จบประโยค ทำให้ใจความขาดหาย
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการหยุดบ้าง ไม่ลื่นไหลเป็นบางครั้ง ● หยุดชะงักบ้าง หยุดคิดบ้างเป็นบางครั้ง ● สามารถพูดได้จบประโยค ใจความครบถ้วน
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีความเร็วสม่ำเสมอ เป็นธรรมชาติ พูดได้ลื่นไหล ไม่ตะกุกตะกัก ● ไม่หยุดชะงัก ● สามารถพูดได้จบประโยค ใจความครบถ้วน

โครงสร้างของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

โครงสร้างของแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 มีทั้งหมด 3 ได้แก่ ส่วนนำ ส่วนเนื้อหา และภาคผนวก ซึ่งแต่ละส่วนมีรายละเอียดดังนี้



ขั้นตอนของการพัฒนาเพิ่มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ชั้นประถมศึกษาปีที่ 6

ขั้นตอนของการพัฒนาเพิ่มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน
ชั้นประถมศึกษาปีที่ 6 ประกอบด้วย 7 ขั้นตอน ดังนี้

1. ขั้นวางแผน	<p>1.1 ศึกษาวิเคราะห์หลักสูตรกลุ่มสาระการเรียนรู้ภาษาต่างประเทศ (ภาษาอังกฤษ) ตัวชี้วัดในระดับชั้น ป.6 การวัดและประเมินผลทางการประเมินทักษะการพูด</p> <p>1.2 กำหนดวัตถุประสงค์ในการเรียนรู้ด้วยเพิ่มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ</p>
2. ขั้นเตรียมความพร้อมนักเรียน	<p>2.1 ให้ความรู้เกี่ยวกับเกณฑ์การประเมินทักษะการพูดภาษาอังกฤษแก่นักเรียน</p> <p>2.2 ฝึกทักษะการประเมินตนเอง และประเมินเพื่อน</p>
3. ขั้นการเก็บรวบรวมหลักฐาน	<p>3.1 ให้นักเรียนวิเคราะห์ความสามารถในการพูดภาษาอังกฤษของตนเอง และกำหนดเป้าหมายในการเรียนของตนเอง</p> <p>3.2 ประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ดังนี้ ประเมินตนเอง ให้เพื่อนประเมิน และครูประเมิน</p> <p>3.3 นักเรียนประเมินทักษะการพูดภาษาอังกฤษของเพื่อน</p>
4. ขั้นการติดตามความก้าวหน้า	<p>4.1 ครูประเมินทักษะการพูดภาษาอังกฤษของนักเรียน</p> <p>4.2 ครูให้ข้อมูลย้อนกลับแก่นักเรียน โดยให้ข้อมูลแบบประสบความสำเร็จ และระบุข้อผิดพลาดเพื่อการแก้ไข</p>
5. ขั้นการปรับปรุงผลงาน	<p>5.1 นักเรียนวางแผนการปรับปรุงการพูดของตนเองตามคำแนะนำของครู (Teacher feedback)</p> <p>5.2 นักเรียนแก้ไขการพูดภาษาอังกฤษของตนเองอีก 2 ครั้ง รวมประเมินการพูดทั้งสิ้น 3 ครั้ง</p>
6. ขั้นการสะท้อนคิด	<p>6.1 นักเรียนสะท้อนคิดต่อผลงานของตนเอง และประสบการณ์ต่างๆ ที่ได้รับจากการจัดการเรียนรู้โดยใช้เพิ่มสะสมงานฯ</p> <p>6.2 นักเรียนระบุจุดเด่น และจุดที่ต้องพัฒนาของตนเอง</p>
7. ขั้นการจัดแสดงผลงาน	<p>7.1 จัดแสดงผลงานเพิ่มสะสมงานฯ ต่อครู นักเรียน และผู้ปกครอง</p> <p>7.2 จัดให้นักเรียนได้แสดงการพูดภาษาอังกฤษบนเวทีของโรงเรียน ต่อผู้ปกครอง และครู นักเรียน ชั้นอื่นๆ</p>

ขั้นตอนที่ 1 การวางแผน

ขั้นตอนการวางแผนมีการดำเนินการ ดังนี้

1.1 ครูผู้สอน และครูกลุ่มสาระการเรียนรู้ภาษาต่างประเทศร่วมกันศึกษาวิเคราะห์หลักสูตร กลุ่มสาระการเรียนรู้ภาษาต่างประเทศ (ภาษาอังกฤษ) ตัวชี้วัดในระดับชั้นประถมศึกษาปีที่ 6 การวัด และประเมินผลทางภาษา การประเมินทักษะการพูด

1.2 ครูผู้สอน ครูกลุ่มสาระการเรียนรู้ภาษาต่างประเทศ และนักเรียน ร่วมกันกำหนด วัตถุประสงค์ในการจัดการเรียนรู้ด้วยแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ

ตารางที่ 7 กิจกรรมของครูและนักเรียนในขั้นตอนการวางแผนการจัดทำแฟ้มสะสมงาน

กิจกรรมของนักเรียน	กิจกรรมของครู
- ร่วมกับครูกำหนดวัตถุประสงค์ เป้าหมายในการ เรียนวิชาภาษาอังกฤษ	- ศึกษาวิเคราะห์หลักสูตรภาษาอังกฤษ ตัวชี้วัดในระดับชั้น ป.6 การวัดและ ประเมินผล
- ร่วมกับครูกำหนดวัตถุประสงค์ เป้าหมายในการ เรียนด้วยแฟ้มสะสมงานการประเมินทักษะการ พูดภาษาอังกฤษ	- ร่วมกับนักเรียนกำหนดวัตถุประสงค์ เป้าหมายในการเรียนด้วยแฟ้มสะสมงาน การประเมินทักษะการพูดภาษาอังกฤษ

ขั้นตอนที่ 2 การเตรียมความพร้อมนักเรียน

ขั้นตอนการเตรียมความพร้อมนักเรียนมีการดำเนินการ ดังนี้

2.1 ผู้วิจัยร่วมกับครูผู้สอนให้ความรู้เกี่ยวกับการประเมินทักษะการพูดภาษาอังกฤษ และ เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษแก่นักเรียน

2.2 ผู้วิจัยร่วมกับครูผู้สอนฝึกทักษะการประเมินตนเอง และประเมินเพื่อน

ตารางที่ 8 กิจกรรมของครูและนักเรียนในขั้นตอนการวางแผนการจัดทำแฟ้มสะสมงาน

กิจกรรมของนักเรียน	กิจกรรมของครู
- เรียนรู้เกี่ยวกับการประเมินทักษะการพูด ภาษาอังกฤษ เกณฑ์การให้คะแนนทักษะการ พูดภาษาอังกฤษ	- ให้ความรู้เกี่ยวกับการประเมินทักษะการพูด ภาษาอังกฤษ เกณฑ์การให้คะแนนทักษะ การพูดภาษาอังกฤษแก่นักเรียน
- ฝึกทักษะการประเมินตนเอง และประเมินเพื่อน	- ฝึกทักษะการประเมินตนเอง และประเมิน เพื่อน

ขั้นตอนที่ 3 การเก็บรวบรวมหลักฐาน

ขั้นตอนการเก็บรวบรวมหลักฐานมีการดำเนินการ ดังนี้

3.1 ให้นักเรียนวิเคราะห์ความสามารถในการพูดภาษาอังกฤษของตนเองและกำหนดเป้าหมายในการเรียนของตนเอง

3.2 การประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ดังนี้

3.2.1 นักเรียนประเมินตนเอง โดยนักเรียนประเมินตนเองทุกครั้งที่มีการทดสอบการพูด

3.2.2 นักเรียนได้รับการประเมินจากเพื่อนในชั้นเรียน จำนวน 3 คน ในทุกครั้งที่มีการทดสอบการพูด

3.2.3 นักเรียนได้รับการประเมินจากครูในทุกครั้งที่มีการทดสอบพูด

3.3 นักเรียนทำการประเมินทักษะการพูดภาษาอังกฤษของเพื่อนในชั้นเรียน จำนวน 3 คน ในแต่ละภาระงาน

ตารางที่ 9 กิจกรรมของครูและนักเรียนในขั้นตอนการวางแผนการจัดทำแฟ้มสะสมงาน

กิจกรรมของนักเรียน	กิจกรรมของครู
- วิเคราะห์ความสามารถในการพูดภาษาอังกฤษของตนเอง	- ให้คำชี้แนะเกี่ยวกับการวิเคราะห์ความสามารถในการพูดภาษาอังกฤษของนักเรียน
- กำหนดเป้าหมายในการเรียน/การพูดภาษาอังกฤษของตนเอง	- ให้คำชี้แนะเกี่ยวกับกำหนดเป้าหมายในการเรียนและการพูดภาษาอังกฤษของนักเรียน
- ประเมินตนเอง	- ประเมินทักษะการพูดภาษาอังกฤษของนักเรียนทุกคน
- ประเมินเพื่อนในชั้น จำนวน 3 คน	

ขั้นตอนที่ 4 การติดตามความก้าวหน้า

ขั้นตอนการติดตามความก้าวหน้ามีการดำเนินการ ดังนี้

4.1 ครูผู้สอนทำการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนทั้งในระหว่างเรียนจากภาระงานการพูด จำนวน 3 ภาระงาน และการประเมินสรุปรวม จำนวน 1 ภาระงาน รวมทั้งสิ้น 4 ภาระงานแต่ละภาระงานทำการประเมิน 3 ครั้ง

4.2 ครูให้ข้อมูลย้อนกลับ (feedback) แก่นักเรียน โดยให้ข้อมูลแบบระบุความสำเร็จและระบุข้อผิดพลาดเพื่อการแก้ไข

ตารางที่ 10 กิจกรรมของครูและนักเรียนในขั้นตอนการวางแผนการจัดทำแฟ้มสะสมงาน

กิจกรรมของนักเรียน	กิจกรรมของครู
- ได้รับผลประเมินทักษะการพูดภาษาอังกฤษจากครูทั้งในระหว่างเรียนจากภาระงานการพูดจำนวน 3 ภาระงาน และการประเมินสรุปรวมจำนวน 1 ภาระงาน	- ประเมินทักษะการพูดภาษาอังกฤษของนักเรียนทั้งในระหว่างเรียนจากภาระงานการพูด จำนวน 3 ภาระงาน และการประเมินสรุปรวม จำนวน 1 ภาระงาน
- ศึกษาข้อมูลที่เป็นการให้ข้อมูลย้อนกลับ (feedback) จากครู และปรึกษากับครูเพื่อความเข้าใจยิ่งขึ้น	- ให้ข้อมูลย้อนกลับแก่นักเรียน โดยให้ข้อมูลแบบระบุความสำเร็จและระบุข้อผิดพลาดเพื่อการแก้ไข

ขั้นตอนที่ 5 การปรับปรุงผลงาน

ขั้นตอนการปรับปรุงผลงานมีการดำเนินการ ดังนี้

5.1 นักเรียนวางแผนการปรับปรุงการพูดของตนเองตามคำแนะนำของครู (Teacher feedback)

5.2 นักเรียนแก้ไขการพูดภาษาอังกฤษของตนเองอีก 2 ครั้ง รวมประเมินการพูดทั้งสิ้น 3 ครั้ง

ตารางที่ 11 กิจกรรมของครูและนักเรียนในขั้นตอนการวางแผนการจัดทำแฟ้มสะสมงาน

กิจกรรมของนักเรียน	กิจกรรมของครู
- วางแผนการปรับปรุงการพูดของตนเองตามคำแนะนำของครู (Teacher feedback)	- ให้คำแนะนำเพื่อการวางแผนในการปรับปรุงการพูดให้ดีขึ้นแก่นักเรียนเป็นรายคน
- แก้ไขการพูดภาษาอังกฤษของตนเองอีก 2 ครั้ง	- ให้ข้อมูลย้อนกลับและคำชี้แนะในการพูดอีกครั้ง และเพื่อเป็นการปรับปรุงการพูดครั้งต่อไป

ขั้นตอนที่ 6 การสะท้อนคิด

ขั้นตอนการสะท้อนคิดมีการดำเนินการ ดังนี้

6.1 นักเรียนสะท้อนคิดต่อผลงานของตนเอง และประสบการณ์ต่างๆ ที่ได้รับการจัดการเรียนรู้โดยใช้แฟ้มสะสมงานฯ

6.2 นักเรียนระบุจุดเด่น และจุดที่ต้องพัฒนาของตนเอง

ตารางที่ 12 กิจกรรมของครูและนักเรียนในขั้นตอนการวางแผนการจัดทำแฟ้มสะสมงาน

กิจกรรมของนักเรียน	กิจกรรมของครู
- สะท้อนคิดต่อผลงานของตนเอง และ ประสบการณ์ต่างๆ ที่ได้รับ	- ให้คำแนะนำในการสะท้อนคิดต่อผลงานการ พูดของนักเรียน
- ระบุจุดเด่น และจุดที่ต้องพัฒนาของตนเอง	- ให้คำชี้แนะเกี่ยวกับการระบุจุดเด่น และจุด ที่ควรพัฒนาของนักเรียนแต่ละคน

ขั้นตอนที่ 7 การจัดแสดงผลงาน

ขั้นตอนการจัดแสดงผลงานมีการดำเนินการ ดังนี้

7.1 จัดแสดงผลงานแฟ้มสะสมงานฯ ต่อครู นักเรียน และผู้ปกครอง

7.2 จัดให้นักเรียนได้แสดงความสามารถในการพูดภาษาอังกฤษบนเวทีของโรงเรียนต่อผู้ปกครอง และครู นักเรียน ชั้นอื่นๆ

ตารางที่ 13 กิจกรรมของครูและนักเรียนในขั้นตอนการวางแผนการจัดทำแฟ้มสะสมงาน

กิจกรรมของนักเรียน	กิจกรรมของครู
- นำแฟ้มสะสมงานมาจัดแสดงร่วมกันทั้งชั้นเรียน	- จัดแสดงผลงานสะสมงานของนักเรียนต่อ ผู้เกี่ยวข้องทั้งในโรงเรียน และต่อผู้ปกครอง
- แสดงความสามารถในการพูดภาษาอังกฤษ บนเวทีของโรงเรียน ต่อผู้ปกครอง และครู นักเรียนชั้นอื่นๆ	- จัดให้นักเรียนได้แสดงความสามารถในการ พูดภาษาอังกฤษบนเวทีของโรงเรียน ต่อผู้ปกครอง และครู นักเรียนชั้นอื่นๆ - ครูให้ข้อคิดเห็นเกี่ยวกับจุดเด่น และจุดที่ยัง ต้องพัฒนาของนักเรียนในภาพรวม

เครื่องมือที่ใช้ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6

เครื่องมือที่ใช้ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วยเครื่องมือสำหรับครู นักเรียน ในแต่ละชั้นมีรายละเอียด ดังนี้

ตารางที่ 14 เครื่องมือที่ใช้ในแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ

ชั้นที่	เครื่องมือ	
	นักเรียน	ครู
1. การวางแผน	<ul style="list-style-type: none"> - แบบบันทึกข้อมูลส่วนตัว - แบบบันทึกการกำหนดวัตถุประสงค์ เป้าหมายในการเรียนวิชาภาษาอังกฤษ - แบบบันทึกการกำหนดวัตถุประสงค์ เป้าหมายในการเรียนด้วยแฟ้มสะสมงานการประเมินทักษะการพูดภาษาอังกฤษ 	<ul style="list-style-type: none"> - ตารางวิเคราะห์หลักสูตร - ตัวชี้วัด ใบความรู้การประเมินทักษะการพูดภาษาอังกฤษ
2. การเตรียมความพร้อมนักเรียน	<ul style="list-style-type: none"> - แบบบันทึกคะแนนการพูด 	<ul style="list-style-type: none"> - ไฟล์เสียงตัวอย่างสำหรับฝึกประเมิน - เกณฑ์การพูดภาษาอังกฤษ
3. การเก็บรวบรวมหลักฐาน	<ul style="list-style-type: none"> - แบบบันทึกการวิเคราะห์ความสามารถในการพูดภาษาอังกฤษของตนเอง - แบบบันทึกการกำหนดเป้าหมายในการเรียน/การพูดภาษาอังกฤษของตนเอง - ประเมินตนเอง - แบบประเมินเพื่อน 	<ul style="list-style-type: none"> - แบบประเมินทักษะการพูดนักเรียน - แบบบันทึกคะแนนทักษะการพูดภาษาอังกฤษของนักเรียน
4. การติดตามความก้าวหน้า		<ul style="list-style-type: none"> - แบบบันทึกการให้ข้อมูลย้อนกลับแก่นักเรียนเป็นรายคน

ชั้นที่	เครื่องมือ	
	นักเรียน	ครู
5. การปรับปรุงผลงาน	- แบบบันทึกการวางแผนการปรับปรุงการพูด	- แบบบันทึกการให้ข้อมูลย้อนกลับแก่นักเรียนเป็นรายคน
6. การสะท้อนคิด	- แบบบันทึกการสะท้อนคิดต่อผลงานของตน - แบบบันทึกจุดเด่น จุดที่ควรพัฒนาของนักเรียน	- แบบบันทึกจุดเด่น จุดที่ควรพัฒนาของนักเรียนเป็นรายคน
7. การจัดแสดงผลงาน		- แบบบันทึกจุดเด่น จุดที่ควรพัฒนาของนักเรียนในภาพรวม

การเลือกใช้รูปแบบการประเมินทักษะการพูดภาษาอังกฤษ

การประเมินทักษะการพูดภาษาอังกฤษโดยใช้แฟ้มสะสมงานที่มีความเที่ยงในระดับที่เหมาะสมสามารถดำเนินการได้ ดังนี้

1. จำนวนผู้ประเมินและจำนวนภาระงานที่เหมาะสม

จำนวนผู้ประเมินและจำนวนภาระงานที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษ ผู้ออกแบบการประเมินสามารถเลือกดำเนินการได้หลายรูปแบบโดยพิจารณาความเหมาะสมและเป็นไปได้ในการดำเนินการและมีระดับความเที่ยงในระดับต่างๆ แบบอิงกลุ่มและอิงเกณฑ์ดังตารางที่ 15-19

1.1 การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนเป็นรายองค์ประกอบ

1) ความเที่ยงเมื่อใช้ในการประเมินแบบอิงกลุ่ม

ตารางที่ 15 ความเที่ยงอิงกลุ่มในการประเมินทักษะการพูดรายองค์ประกอบ

จำนวน		องค์ประกอบรายด้าน/ความเที่ยงอิงกลุ่ม					
ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	คำศัพท์	การสร้าง ประโยค	การเชื่อมโยง	การออก เสียง	การบรรลุ เป้าหมาย	ความ คล่องแคล่ว
4	2	0.617	0.579	0.585	0.806	0.610	0.647
4	3	0.651	0.611	0.629	0.828	0.667	0.691
4	4	0.669	0.628	0.654	0.84	0.701	0.715
4	5	0.680	0.638	0.67	0.848	0.722	0.730
4	6	0.688	0.645	0.681	0.853	0.737	0.741
4	7	0.694	0.651	0.689	0.856	0.749	0.749
4	8	0.698	0.655	0.695	0.859	0.757	0.755
4	9	0.701	0.658	0.700	0.861	0.764	0.759
4	10	0.704	0.661	0.704	0.863	0.770	0.763
4	4	0.669	0.628	0.654	0.840	0.701	0.715
5	4	0.713	0.674	0.703	0.867	0.739	0.758
6	4	0.746	0.71	0.739	0.886	0.768	0.790
7	4	0.772	0.738	0.768	0.901	0.789	0.815
8	4	0.792	0.760	0.791	0.911	0.806	0.834
9	4	0.808	0.778	0.810	0.920	0.820	0.85
10	4	0.822	0.793	0.825	0.927	0.831	0.863

2) ความเที่ยงเมื่อใช้ในการประเมินแบบอิงเกณฑ์

ตารางที่ 16 ความเที่ยงอิงเกณฑ์ในการประเมินทักษะการพูดรายด้าน

จำนวน		องค์ประกอบรายด้าน/ความเที่ยงอิงเกณฑ์					
ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	คำศัพท์	การสร้าง ประโยค	การเชื่อมโยง	การ ออกเสียง	การบรรลุ เป้าหมาย	ความ คล่องแคล่ว
4	4	0.518	0.508	0.535	0.803	0.644	0.659
4	5	0.526	0.517	0.557	0.813	0.664	0.680
4	6	0.530	0.524	0.572	0.82	0.678	0.694
4	7	0.534	0.528	0.584	0.824	0.689	0.705
4	8	0.537	0.532	0.593	0.828	0.697	0.713
4	9	0.539	0.534	0.600	0.831	0.703	0.719
4	10	0.541	0.536	0.606	0.833	0.709	0.725
4	4	0.518	0.508	0.535	0.803	0.644	0.659
5	4	0.572	0.560	0.582	0.833	0.686	0.699
6	4	0.614	0.6	0.619	0.853	0.718	0.729
7	4	0.648	0.633	0.648	0.869	0.742	0.752
8	4	0.676	0.66	0.672	0.88	0.762	0.770
9	4	0.699	0.683	0.691	0.89	0.778	0.785
10	4	0.719	0.702	0.708	0.898	0.791	0.797

3) การประเมินทักษะการพูดรายด้านในภาพรวมของการประเมิน

ตารางที่ 17 ระดับความคลาดเคลื่อนและความเที่ยงของคะแนนรวมในการประเมินการพูดฯ
รายองค์ประกอบเพื่อการตัดสินใจจำนวนผู้ประเมินและภาระงาน

จำนวน		ความคลาดเคลื่อน		ความเที่ยง	
ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	อิงกลุ่ม (Relative error)	อิงเกณฑ์ (Absolute error)	อิงกลุ่ม	อิงเกณฑ์
4	2	0.017	0.031	0.819	0.708
4	3	0.015	0.027	0.837	0.733
4	4	0.014	0.026	0.846	0.745
4	5	0.013	0.025	0.851	0.753
4	6	0.013	0.024	0.855	0.758
4	7	0.013	0.023	0.857	0.762
4	8	0.012	0.023	0.859	0.765
4	9	0.012	0.023	0.861	0.767
4	10	0.012	0.023	0.862	0.769
3	4	0.018	0.033	0.806	0.692
4	4	0.014	0.026	0.846	0.745
5	4	0.011	0.021	0.872	0.781
6	4	0.009	0.018	0.890	0.807
7	4	0.008	0.016	0.904	0.826
8	4	0.007	0.014	0.914	0.842
9	4	0.006	0.013	0.922	0.854
10	4	0.006	0.012	0.929	0.864

1.2 การประเมินทักษะการพูดภาษาอังกฤษของนักเรียนเป็นในภาพรวมของภาระงาน

1) การประเมินนักเรียนแยกตามภาระงาน

ตารางที่ 18 ความเที่ยงในการประเมินเมื่อประเมินรายการภาระงาน

ความเที่ยง	ภาระงาน	จำนวนผู้ประเมิน		
		2 คน	3 คน	4 คน
อิงกลุ่ม	การสัมภาษณ์	0.872	0.905	0.923
	การนำเสนอปากเปล่า	0.929	0.95	0.960
	การเล่านิทาน	0.823	0.875	0.903
	บรรยายภาพ	0.840	0.887	0.912
อิงเกณฑ์	การสัมภาษณ์	0.820	0.862	0.885
	การนำเสนอปากเปล่า	0.914	0.938	0.950
	การเล่านิทาน	0.810	0.865	0.895
	บรรยายภาพ	0.828	0.878	0.905

2) การประเมินทักษะการพูดในภาพรวมของคะแนนรวมของการประเมิน

ตารางที่ 19 ระดับความคลาดเคลื่อนและความเที่ยงของคะแนนรวมในการประเมินการพูดฯ
เพื่อการตัดสินใจจำนวนผู้ประเมินและภาระงาน

จำนวน		ความคลาดเคลื่อน		ความเที่ยง	
ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	อิงกลุ่ม (Relative error)	อิงเกณฑ์ (Absolute error)	อิงกลุ่ม	อิงเกณฑ์
2	4	0.006	0.007	0.939	0.925
3	4	0.004	0.005	0.958	0.947
4	4	0.003	0.004	0.967	0.959

2. การเลือกรูปแบบการประเมินของผู้ประเมินที่เหมาะสม

การเลือกจำนวนผู้ประเมิน จำนวนภาระงานและวิธีการประเมินของผู้ประเมินที่เหมาะสมในการประเมินทักษะการพูดภาษาอังกฤษ ผู้ออกแบบการประเมินสามารถเลือกดำเนินการได้หลายรูปแบบ โดยพิจารณาความเหมาะสมและเป็นไปได้ในการดำเนินการและมีระดับความเที่ยงในระดับต่างๆ ทั้งความเที่ยงในการประเมินแบบอิงกลุ่มและแบบอิงเกณฑ์ดังตารางที่ 20-21

2.1 การเลือกรูปแบบเพื่อการตัดสินใจในภาพรวมของการประเมิน

ตารางที่ 20 ระดับความคลาดเคลื่อนและความเที่ยงของคะแนนรวมในการประเมินการพูดฯ เพื่อการตัดสินใจเลือกรูปแบบการผู้ประเมิน

รูปแบบการประเมินของผู้ประเมิน	จำนวน		ความคลาดเคลื่อน		ความเที่ยง	
	ผู้ประเมิน (คน)	ภาระงาน (เรื่อง)	อิงกลุ่ม (Relative error)	อิงเกณฑ์ (Absolute error)	อิงกลุ่ม	อิงเกณฑ์
ประเมินงานซ้ำกัน	2	4	0.006	0.007	0.939	0.925
	3	4	0.004	0.005	0.958	0.947
	4	4	0.003	0.004	0.967	0.959
ประเมินงานไม่ซ้ำกัน	2	4	0.004	0.005	0.955	0.945
	3	4	0.003	0.004	0.968	0.961
	4	4	0.002	0.003	0.975	0.969

2.2 การเลือกรูปแบบการประเมินสำหรับการประเมินระหว่างเรียนและสรุปรวม

ตารางที่ 21 ระดับความเที่ยงของการประเมินการพูดฯ เพื่อการตัดสินใจรูปแบบการประเมิน

ความเที่ยง	จำนวนผู้ประเมิน (คน)	จำนวนภาระงาน (เรื่อง)			การประเมิน/ความเที่ยง		
		ระหว่างเรียน	สรุปรวม	รวม	ระหว่างเรียน	สรุปรวม	
อิงกลุ่ม	2	3	1	4	0.937	0.71	
	3	3	1	4	0.956	0.753	
	4	3	1	4	0.966	0.777	
	5	3	1	4	0.972	0.792	
	6	3	1	4	0.976	0.802	
	2	3	2	5	0.937	0.783	
	3	3	2	5	0.956	0.824	
	4	3	2	5	0.966	0.847	
	5	3	2	5	0.972	0.861	
	6	3	2	5	0.976	0.871	
	อิงเกณฑ์	2	3	1	4	0.915	0.697
		3	3	1	4	0.941	0.741
4		3	1	4	0.954	0.765	
5		3	1	4	0.962	0.781	
6		3	1	4	0.968	0.791	
2		3	2	5	0.915	0.774	
3		3	2	5	0.941	0.817	
4		3	2	5	0.954	0.84	
5		3	2	5	0.962	0.855	
6		3	2	5	0.968	0.865	

เครื่องมือสำหรับนักเรียน



ประวัติส่วนตัว

ชื่อ-สกุล.....

Name-Surname.....

ชื่อเล่น.....Nickname.....

ชั้น.....Grade.....

ปัจจุบันเรียนที่โรงเรียน.....

School name.....

วิชาที่ชอบ.....I likesubjects

สถานที่ที่ชอบ.....Favorite place.....

เพื่อนที่สนิทที่สุด.....My close friend.....

กีฬาที่ชอบ.....My favorite sport.....

ความใฝ่ฝัน.....My dream.....

วัน/เดือน/ปี เกิด.....Date of birth.....

อายุ.....Age.....

ที่อยู่ปัจจุบัน

.....

.....

Address.....

.....

.....

1. ข้อมูลส่วนตัวนักเรียน

ฉันชื่อ.....

My name is.....

ชั้น (Grade).....

โรงเรียน.....

School's name.....

Teacher's name.....

2. ฉันพูดภาษาอังกฤษเมื่อใด/กับใคร

ที่บ้าน/at home.....

.....

.....

ที่โรงเรียน/at school.....

.....

.....

.....

.....

.....

.....

3. ฉันทเรียนภาษาอังกฤษในโรงเรียนหรือในห้องเรียน:

วัน	ตั้งแต่เวลา ถึง เวลา	จำนวน ชม./นาที



การพูดเรื่องที่ 1

“.....”

ชื่อ.....ชั้น.....เลขที่.....

1. สิ่งที่คุณต้องการเรียนรู้ในการพูดเรื่องนี้เป็นคือ

.....

.....

.....

2. เป้าหมายในการพูดครั้งนี้คือ

2.1 ฉันต้องเรียนภาษาอังกฤษเพราะว่า

.....

.....

.....

2.2 ในการเรียนภาษาอังกฤษฉันจะต้องทำ/พูดสิ่งต่อไปนี้ให้ได้

.....

.....

.....

3. ตอนนี้ฉันสามารถที่จะพูดเกี่ยวกับ “.....”

ฉันรู้สึก (พอใจต่อผลการพูดมากน้อยเพียงใด / เพราะอะไร).....

.....

.....

.....

4. คำศัพท์ใหม่ๆ ที่ฉันได้เรียนรู้ได้แก่ (สามารถวาดภาพประกอบได้)

.....

.....

.....

.....

5. การสะท้อนคิดจากการพูดภาษาอังกฤษครั้งนี้

5.1 สิ่งที่คุณชอบทำในการเรียนภาษาอังกฤษในครั้งนี้คือ

.....

.....

.....

.....

.....

.....

.....

5.2 สิ่งที่คุณทำได้ดีในการเรียนภาษาอังกฤษในครั้งนี้คือ

.....

.....

.....

.....

.....

.....

.....

5.3 สิ่งที่ยากสำหรับฉันในการเรียนภาษาอังกฤษในครั้งนี้คือ

.....

.....

.....

.....

.....

.....

.....

5.4 สิ่งที่เป็นปัญหาสำหรับฉันในการเรียนภาษาอังกฤษในครั้งนี้คือ

.....

.....

.....

.....

.....

5.5 สิ่งที่ทำให้ฉันสามารถพูดภาษาอังกฤษได้ตามที่คาดหวังคือ

.....

.....

.....

5.6 สิ่งที่ฉันควรแก้ไขในการพูดครั้งนี้คือ

.....

.....

.....

5.7 สิ่งที่ฉันได้เรียนรู้ในครั้งนี้คือ

.....

.....

.....

.....

5.8 สิ่งที่จะฉันจะทำการพูดครั้งต่อไปคือ

.....

.....

.....

5.9 ตอนนี้ฉันรู้สึก

.....

.....

.....

การประเมินตนเอง

การพูดเรื่อง “.....” ครั้งที่.....

ชื่อ.....ชั้น.....เลขที่.....

คำชี้แจง โปรดใส่เครื่องหมาย ✓ ให้ตรงกับความเป็นจริงที่สุด

ความหมาย

☹☹ อ่อนมาก

☹ อ่อน

☺ พอใช้

☺ ดี

ประเด็น	ระดับความสามารถ			
	☺	☺	☹	☹☹
คำศัพท์ (vocabulary)				
การออกเสียง (pronunciation)				
การสร้างประโยค (syntax)				
การเชื่อมโยงประโยค (cohesion)				
บรรลุลักษณะประสม (ideational function)				
ความคล่องแคล่ว (fluency)				

ข้อเสนอแนะอื่นๆ

.....

.....

.....

.....

.....

.....

การประเมินเพื่อนคนที่

ชื่อผู้ถูกประเมิน.....

ชื่อผู้ประเมิน.....

การพูดเรื่อง “.....” ครั้งที.....

คำชี้แจง โปรดใส่เครื่องหมาย ✓ ให้ตรงกับความเป็นจริงที่สุด

ความหมาย

☹☹ อ่อนมาก

☹ อ่อน

☺ พอใช้

☺ ดี

ประเด็น	ระดับความสามารถ			
	☺	☺	☹	☹☹
คำศัพท์ (vocabulary)				
การออกเสียง (pronunciation)				
การสร้างประโยค (syntax)				
การเชื่อมโยงประโยค (cohesion)				
บรรลุมิติประสงค์ (ideational function)				
ความคล่องแคล่ว (fluency)				

ข้อเสนอแนะอื่นๆ

.....

.....

.....

.....

.....

.....

แบบสรุปความคิดเห็นของนักเรียน

ความคิดเห็นต่อการพูดแต่ละเรื่อง

1. การพูดเรื่อง “

.....
.....
.....

2. การพูดเรื่อง “

.....
.....
.....

3. การพูดเรื่อง “

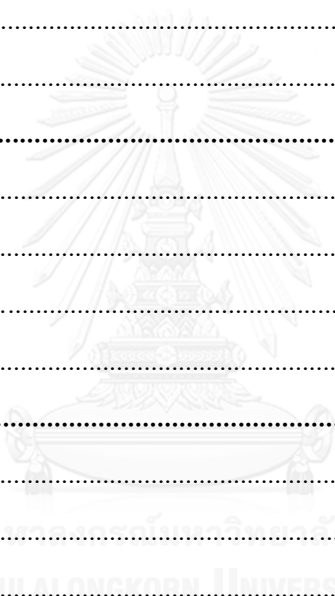
.....
.....
.....

4. การพูดเรื่อง “

.....
.....
.....

ความคิดเห็นต่อการพูดโดยรวม

.....
.....
.....
.....
.....
.....



แบบสรุปความคิดเห็นของครูผู้สอน

ความคิดเห็นต่อการพูดแต่ละเรื่อง

1. การพูดเรื่อง “

.....
.....
.....

2. การพูดเรื่อง “

.....
.....
.....

3. การพูดเรื่อง “

.....
.....
.....

4.การพูดเรื่อง “

.....
.....
.....

ความคิดเห็นต่อการพูดโดยรวม

.....
.....
.....
.....
.....



ภาคผนวก

ภาระงานการพูดภาษาอังกฤษ

ภาระงาน (tasks) ที่ใช้ในการประเมินการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วย 4 ภาระงาน โดยเป็นภาระงานที่ใช้ในการประเมินระหว่างเรียน จำนวน 3 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า และการเล่านิทาน และเป็นภาระงานที่ใช้ประเมินสรุปรวม จำนวน 1 ภาระงาน คือ การบรรยายภาพ ซึ่งแต่ละภาระงานมีรายละเอียด ดังนี้

การประเมินระหว่างเรียน ประกอบด้วย

1. การสัมภาษณ์ (interview) เป็นการสัมภาษณ์โดยครูเป็นผู้สัมภาษณ์เกี่ยวกับเรื่องใกล้ตัว เช่น กิจกรรมที่ทำในตอนเช้าก่อนมาโรงเรียน กิจกรรมที่ทำในห้องเรียน กีฬาที่ชอบ และเหตุผลที่ชอบ เป็นต้น

2. การนำเสนอปากเปล่า (oral presentation) เป็นกิจกรรมการพูดที่ให้นักเรียนได้พูดนำเสนอเรื่องราวเกี่ยวกับตนเอง โดยครูให้นักเรียนเตรียมตัวมาก่อนล่วงหน้า เช่น ให้ค้นหาคำศัพท์ที่เกี่ยวข้อง ลองแต่งประโยค ปรึกษากับครูในชั้นเรียน หรือผู้ปกครอง เนื้อหาที่พูด เช่น ประวัติโดยย่อ งานอดิเรก กิจกรรมที่สนใจ เป็นต้น

3. การเล่านิทาน (storytelling) เป็นกิจกรรมที่ครูให้นักเรียนเตรียมตัวก่อนล่วงหน้า โดยเตรียมนิทานมาเล่าคนละ 1 เรื่อง หรือเลือกเรื่องจากที่ครูเตรียมให้ และนิทานต้องไม่ให้ซ้ำกัน ซึ่งอาจนำหนังสือนิทานหรืออุปกรณ์มาประกอบการเล่า โดยนิทานที่จะใช้ในการเล่าครูจะเป็นผู้พิจารณาความเหมาะสม

การประเมินสรุปรวม คือ การบรรยายประกอบภาพ ซึ่งมีรายละเอียด ดังนี้

การบรรยายภาพ (picture description) เป็นกิจกรรมที่ครูให้นักเรียนพูดบรรยายภาพ โดยนักเรียนจะได้เตรียมตัวก่อนล่วงหน้า คือ ศึกษารายละเอียดต่างๆ ของภาพ ชักถามรายละเอียดของภาพจากครูผู้สอน ค้นหาคำศัพท์ที่เกี่ยวข้อง ลองแต่งประโยค ปรึกษาการพูดกับครูในชั้นเรียน หรือผู้ปกครอง เป็นต้น โดยภาพประกอบที่ใช้เป็นภาพที่นักเรียนสนใจนำมาเอง หรือที่ครูเตรียมไว้ให้โดยไม่ซ้ำกัน แล้วให้นักเรียนบรรยายเรื่องราวในภาพของตนเอง โดยภาพที่จะนำมาบรรยายครูจะเป็นผู้พิจารณาความเหมาะสม

เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ

ภาระงานการพูดทั้ง 4 ภาระงานมีเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) มีขอบเขตเนื้อหาในการประเมิน ดังนี้

1. คำศัพท์ (vocabulary) คือ การให้คะแนนด้านคำศัพท์ จะพิจารณาจากการพูดของนักเรียนที่สะท้อน 2 สิ่ง คือ ขอบเขต (range) ของคำศัพท์ที่ใช้ และ ความถูกต้อง (accuracy) กล่าวคือ ขอบเขต (range) ของการใช้คำศัพท์ที่มีความหลากหลาย ความสามารถในการใช้คำศัพท์ได้อย่างถูกต้อง (accuracy) และเหมาะสม (appropriation) ตรงตามความหมายที่ต้องการสื่อสาร

2. การสร้างประโยค (syntax) คือ การให้คะแนนด้านการสร้างประโยค พิจารณาจากความสามารถของนักเรียนในสร้างประโยคได้อย่างมีความถูกต้อง (accuracy) ตามที่ต้องการสื่อสาร เช่น มีประธาน กริยา กรรมที่เหมาะสม เป็นต้น

3. การเชื่อมโยงเนื้อหา (cohesion) การให้คะแนนด้านการเชื่อมโยงเนื้อหา พิจารณาจากการที่นักเรียนสามารถพูดเชื่อมโยงประโยคแต่ละประโยคให้เป็นเรื่องราวที่มีความสัมพันธ์กันได้อย่างถูกต้องเหมาะสม

4. การออกเสียง (pronunciation) การให้คะแนนด้านการออกเสียง พิจารณาจากการที่นักเรียนสามารถเน้นเสียงหนักเบา (stress) เว้นวรรคการหยุด (segmental sound) และการใช้เสียงสูงต่ำ (intonation) ได้อย่างถูกต้อง เหมาะสม และทำให้ผู้ฟังเข้าใจในสิ่งที่ต้องการสื่อสารได้

5. บรรลุวัตถุประสงค์ (ideational function) คือ การให้คะแนนด้านบรรลุวัตถุประสงค์ พิจารณาจากการที่นักเรียนสามารถพูดสื่อสารได้บรรลุตามวัตถุประสงค์ที่ต้องการจะสื่อสาร โดยการพูดนั้นทำให้ผู้ฟังเข้าใจว่าต้องการสื่อสารอะไร เช่น พูดได้ตรงกับหัวเรื่องที่ระบุ ตอบตรงคำถาม ตอบคำถามได้อย่างเหมาะสมกับลักษณะคำถาม เป็นต้น

6. ความคล่องแคล่ว (fluency) คือ การใช้ภาษาที่มีความคล่องแคล่ว ลื่นไหล ซึ่งพิจารณาจากการที่นักเรียนสามารถพูดภาษาอังกฤษได้อย่างคล่องแคล่ว ราบรื่น และมีความลื่นไหลในระหว่างที่พูด และไม่สะดุดหรือกระตุก หยุดชะงัก ระยะเวลาที่พูด

ให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี ตามลำดับ) ซึ่งมีรายละเอียด ดังนี้

1. คำศัพท์ (vocabulary)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● ความสามารถในการใช้คำศัพท์ยังจำกัดอยู่มาก ● มักใช้คำศัพท์คำเดิมซ้ำๆ ทุกประโยค หรือเกือบทุกประโยค ● ใช้คำศัพท์ไม่ถูกต้อง ไม่เหมาะสม
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● ความสามารถในการใช้คำศัพท์ยังจำกัด ใช้คำศัพท์ไม่หลากหลาย ● ยังพบการใช้คำศัพท์คำเดิมซ้ำๆ อยู่บ่อยครั้ง ● มีการใช้คำศัพท์ที่ไม่ถูกต้อง ไม่เหมาะสมอยู่บ่อยครั้ง
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● สามารถใช้คำศัพท์ได้หลากหลาย ● แต่ยังใช้คำศัพท์คำเดิมซ้ำบ้างเป็นบางครั้ง ● สามารถใช้คำศัพท์ได้อย่างถูกต้องและเหมาะสม อาจยังใช้คำศัพท์ผิดพลาดบ้างเป็นบางครั้ง
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● สามารถใช้คำศัพท์ได้อย่างหลากหลาย ● การใช้คำศัพท์ส่วนใหญ่สามารถใช้ได้ถูกต้องเหมาะสม และผิดพลาดน้อยมาก หรือถูกต้องทั้งหมด ไม่มีความผิดพลาดเลย

2. การสร้างประโยค (syntax)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● การพูดไม่มีการเชื่อมโยงให้เป็นประโยค พูดเป็นคำๆ ● มีการใช้โครงสร้างประโยคที่ไม่ถูกต้องทุกครั้ง
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการเชื่อมโยงให้เป็นประโยค ● การใช้โครงสร้างประโยคที่ไม่ถูกต้องอยู่บ่อยครั้ง ทำให้ผู้ฟังไม่สามารถเข้าใจได้อย่างกระจ่าง
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการเชื่อมโยงให้เป็นประโยค ● การใช้โครงสร้างประโยคที่ไม่ถูกต้องบ้างเป็นบางครั้ง แต่ผู้ฟังสามารถเข้าใจได้

ระดับ ความสามารถ	คำอธิบาย
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการเชื่อมโยงให้เป็นประโยค ● ส่วนใหญ่มีการใช้โครงสร้างประโยคที่ถูกต้อง และผิดพลาดน้อยมาก หรือไม่มีความผิดพลาดเลย

3. การเชื่อมโยงเนื้อหา (cohesion)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อยมาก หรือไม่ใช้เลย ● การใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคเกือบทั้งหมดไม่ถูกต้อง
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อย ● การใช้คำเชื่อมโยงทำให้ความสัมพันธ์ระหว่างประโยคมีความสับสนอยู่บ่อยครั้ง และทำให้เข้าใจคลาดเคลื่อน
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยค ● ความสัมพันธ์ระหว่างประโยคอาจมีความสับสนอยู่บ้าง แต่ไม่ทำให้ความเข้าใจคลาดเคลื่อน
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยค ● ถูกต้อง ชัดเจน เป็นส่วนใหญ่ และไม่ทำให้เข้าใจคลาดเคลื่อน ผิดพลาดน้อยมาก หรือไม่มีความผิดพลาดเลย

4. การออกเสียง (pronunciation)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงเกือบทั้งหมด มีความผิดพลาด ทำให้ไม่สามารถเข้าใจได้ ● นักเรียนพูดน้อยมาก หรือไม่พูดเลย
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงยังผิดพลาดอยู่บ่อยครั้ง ● การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดอยู่มาก

ระดับ ความสามารถ	คำอธิบาย
	<ul style="list-style-type: none"> ● ทำให้ไม่สามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้ หรือทำให้เข้าใจคลาดเคลื่อน
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงยังมีความผิดอยู่บ้าง ● การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดอยู่บ้าง ● แต่ผู้ฟังยังคงสามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงส่วนใหญ่ถูกต้อง หรือไม่มีความผิดพลาดเลย ● การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดน้อยมาก หรือไม่มีความผิดพลาดเลย ● ผู้ฟังสามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้โดยง่าย

5. บรรลุวัตถุประสงค์ (ideational function)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● พูดไม่ตรงประเด็น หรือไม่พูดเลย ● การพูดเกือบทั้งหมดไม่สามารถเข้าใจได้ว่าต้องการสื่อสารอะไร
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● เนื้อหาในการพูดไม่ค่อยสอดคล้องเหมาะสมกับหัวเรื่องหรือสิ่งที่ต้องการสื่อสาร ● ไม่สามารถเข้าใจอย่างชัดเจนได้ว่าวัตถุประสงค์ของการพูดคืออะไร
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● เนื้อหาในการพูดสอดคล้องเหมาะสม ถึงแม้จะมีความผิดพลาดในการพูดในบางครั้ง ● สามารถทำให้เข้าใจวัตถุประสงค์ที่ต้องการสื่อสารได้
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● เนื้อหาในการพูดสอดคล้องเหมาะสม มีความผิดพลาดในการพูดน้อยมาก หรือไม่ผิดพลาดเลย ● สามารถเข้าใจวัตถุประสงค์ในสิ่งที่ต้องการสื่อสารได้โดยง่าย

6. ความคล่องแคล่ว (fluency)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● การพูดยังตะกุกตะกัก พูดเป็นคำๆ ● ไม่เป็นประโยค หรือไม่พูดเลย
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● การพูดบ่อยครั้งยังซ้ำมาก และไม่สลับไหล ● หยุดเป็นระยะ หยุดคิดนาน พูดไม่จบประโยค ทำให้ใจความขาดหาย
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการหยุดบ้าง ไม่สลับไหลเป็นบางครั้ง ● หยุดชะงักบ้าง หยุดคิดบ้างเป็นบางครั้ง ● สามารถพูดได้จบประโยค ใจความครบถ้วน
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีความเร็วสม่ำเสมอ เป็นธรรมชาติ พูดได้สลับไหล ไม่ตะกุกตะกัก ● ไม่หยุดชะงัก ● สามารถพูดได้จบประโยค ใจความครบถ้วน



ผลการประเมินจากครู

การพูดเรื่อง “.....” ครั้งที่.....

ชื่อผู้ถูกประเมิน.....

คำชี้แจง โปรดใส่เครื่องหมาย ✓ ให้ตรงกับความเป็นจริงที่สุด

ความหมาย

☹☹ อ่อนมาก

☹ อ่อน

☺ พอใช้

☺ ดี

ประเด็น	ระดับความสามารถ			
	☺	☺	☹	☹☹
คำศัพท์ (vocabulary)				
การออกเสียง (pronunciation)				
การสร้างประโยค (syntax)				
การเชื่อมโยงประโยค (cohesion)				
บรรลุลักษณะประสงค์ (ideational function)				
ความคล่องแคล่ว (fluency)				

การให้ข้อมูลป้อนกลับ (feedback)

1. สิ่งทีนักเรียนทำได้ดีแล้ว: การให้ข้อมูลย้อนกลับแบบการระบุความสำเร็จ (achievement feedback)

ประเด็น	คำอธิบาย
คำศัพท์ (vocabulary)	
การสร้างประโยค (syntax)	
การเชื่อมโยงประโยค (cohesion)	
การออกเสียง (pronunciation)	
บรรล่วัตถุประสงค์ (ideational function)	
ความคล่องแคล่ว (fluency)	

ประเด็นอื่นๆ ที่นักเรียนทำได้ดี

.....

.....

.....

.....

.....

.....

2. สิ่งทีนักเรียนยังต้องปรับปรุง: การให้ข้อมูลย้อนกลับเพื่อปรับปรุงแก้ไขให้ถูกต้อง (corrective feedback)

ประเด็น	คำอธิบาย
คำศัพท์ (vocabulary)	
การสร้างประโยค (syntax)	
การเชื่อมโยงประโยค (cohesion)	
การออกเสียง (pronunciation)	
บรรล่วัตถุประสงค์ (ideational function)	
ความคล่องแคล่ว (fluency)	

ประเด็นอื่นๆ ถ้านักเรียนควรปรับเพื่อให้พูดดียิ่งขึ้น

.....

.....

.....

.....

.....

แบบบันทึกคะแนนทักษะการพูดภาษาอังกฤษ
การประเมินระหว่างเรียนและสรุปรวม

ชั้นประถมศึกษาปีที่.....โรงเรียน.....
ครูผู้สอน.....ตำแหน่ง..... วันที่..... เดือน.....พ.ศ.

คำชี้แจง โปรดใส่คะแนน ในช่อง ที่ตรงกับระดับความสามารถของนักเรียน

เกณฑ์การให้คะแนน ประกอบด้วย 5 ด้าน ดังนี้

1. คำศัพท์ (vocabulary)
2. การสร้างประโยค (syntax)
3. การเชื่อมโยงประโยค (cohesion)
4. การออกเสียง (pronunciation)
5. บรรลุมิติวัตถุประสงค์ (ideational function)
6. ความคล่องแคล่ว (fluency)

ความหมายของคะแนน

- | | | |
|---|---------|---------|
| 1 | หมายถึง | อ่อนมาก |
| 2 | หมายถึง | อ่อน |
| 3 | หมายถึง | พอใช้ |
| 4 | หมายถึง | ดี |

ความคิดเห็นต่อความสามารถในทักษะการพูดของนักเรียน

สิ่งที่นักเรียนทำได้ดี

.....

.....

.....

.....

.....

สิ่งที่นักเรียนต้องปรับปรุง

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

ปัญหาอุปสรรค

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

วิธีแก้ปัญหา

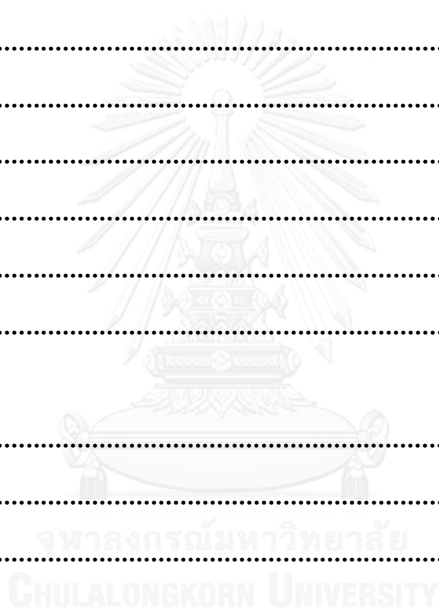
.....

.....

.....

.....

.....



สรุปผลการประเมินจากครู (ทั้ง 3 ครั้ง)

การพูดเรื่อง “.....”

ชื่อผู้ถูกประเมิน.....

ชื่อ.....ชั้น.....เลขที่.....

คำชี้แจง โปรดใส่เครื่องหมาย ✓ ให้ตรงกับความเป็นจริงที่สุด

ความหมาย

☹☹ อ่อนมาก

☹ อ่อน

☺ พอใช้

☺ ดี

ประเด็น	ระดับความสามารถ		
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
คำศัพท์ (vocabulary)	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺
การสร้างประโยค (syntax)	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺
การเชื่อมโยงประโยค (cohesion)	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺
การออกเสียง (pronunciation)	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺
บรรลุนิเทศประสงค์ (ideational function)	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺
ความคล่องแคล่ว (fluency)	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺	☹☹ ☹ ☺☺

การให้ข้อมูลป้อนกลับ (feedback)

1. สิ่งทีนักเรียนทำได้ดี: การให้ข้อมูลป้อนกลับแบบการระบุความสำเร็จ (achievement feedback)

ประเด็น	สิ่งที่ทำได้ดี		
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
คำศัพท์ (vocabulary)			
การสร้างประโยค (syntax)			
การเชื่อมโยงประโยค (cohesion)			
การออกเสียง (pronunciation)			
บรรลุมิติวัตถุประสงค์ (ideational function)			
ความคล่องแคล่ว (fluency)			
ประเด็นอื่นๆ			

2. สิ่งทีนักเรียนยังต้องปรับปรุง: การให้ข้อมูลย้อนกลับเพื่อปรับปรุงแก้ไขให้ถูกต้อง (corrective feedback)

ประเด็น	สิ่งที่ควรปรับปรุงและแนวทางปรับปรุง		
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
คำศัพท์ (vocabulary)			
การสร้างประโยค (syntax)			
การเชื่อมโยงประโยค (cohesion)			
การออกเสียง (pronunciation)			
บรรลุมิติประสงค์ (ideational function)			
ความคล่องแคล่ว (fluency)			
ประเด็นอื่นๆ			

สรุปความก้าวหน้าของนักเรียนแยกตามองค์ประกอบ พบว่า

1. คำศัพท์ (vocabulary)

.....

.....

.....

2. การสร้างประโยค (syntax)

.....

.....

.....

3. การเชื่อมโยงเนื้อหา (cohesion)

.....

.....

.....

4. การออกเสียง (pronunciation)

.....

.....

.....

5. บรรลุวัตถุประสงค์ (ideational function)

.....

.....

.....

6. ความคล่องแคล่ว (fluency)

.....

.....

.....

ข้อสังเกต

สิ่งที่นักเรียนทำได้ดี

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

สิ่งที่นักเรียนควรได้รับการพัฒนา

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

แนวทางการพัฒนานักเรียนในกิจกรรมต่อไป

.....

.....

.....

.....

.....

.....

สรุปความก้าวหน้าของนักเรียนแยกตามองค์ประกอบ (ในภาพรวม) พบว่า

1. คำศัพท์ (vocabulary)

.....

.....

.....

2. การสร้างประโยค (syntax)

.....

.....

.....

3. การออกเสียง (pronunciation)

.....

.....

.....

4. การเชื่อมโยงเนื้อหา (cohesion)

.....

.....

.....

5. บรรลู่วัตถุประสงค์ (ideational function)

.....

.....

.....

6. ความคล่องแคล่ว (fluency)

.....

.....

.....

ภาคผนวก

ภาระงานการพูดภาษาอังกฤษ

ภาระงาน (tasks) ที่ใช้ในการประเมินการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 ประกอบด้วย 4 ภาระงาน โดยเป็นภาระงานที่ใช้ในการประเมินระหว่างเรียน จำนวน 3 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า และการเล่านิทาน และเป็นภาระงานที่ใช้ประเมินสรุปรวม จำนวน 1 ภาระงาน คือ การบรรยายภาพ ซึ่งแต่ละภาระงานมีรายละเอียด ดังนี้

การประเมินระหว่างเรียน ประกอบด้วย

1. การสัมภาษณ์ (interview) เป็นการสัมภาษณ์โดยครูเป็นผู้สัมภาษณ์เกี่ยวกับเรื่องใกล้ตัว เช่น กิจกรรมที่ทำในตอนเช้าก่อนมาโรงเรียน กิจกรรมที่ทำในห้องเรียน กีฬาที่ชอบ และเหตุผลที่ชอบ เป็นต้น

2. การนำเสนอปากเปล่า (oral presentation) เป็นกิจกรรมการพูดที่ให้นักเรียนได้พูดนำเสนอเรื่องราวเกี่ยวกับตนเอง โดยครูให้นักเรียนเตรียมตัวมาก่อนล่วงหน้า เช่น ให้ค้นหาคำศัพท์ที่เกี่ยวข้อง ลองแต่งประโยค ปรึกษากับครูในชั้นเรียน หรือผู้ปกครอง เนื้อหาที่พูด เช่น ประวัติโดยย่อ งานอดิเรก กิจกรรมที่สนใจ เป็นต้น

3. การเล่านิทาน (storytelling) เป็นกิจกรรมที่ครูให้นักเรียนเตรียมตัวก่อนล่วงหน้า โดยเตรียมนิทานมาเล่าคนละ 1 เรื่อง หรือเลือกเรื่องจากที่ครูเตรียมให้ ซึ่งอาจนำหนังสือนิทาน หรืออุปกรณ์มาประกอบการเล่า โดยนิทานที่จะใช้ในการเล่าครูจะเป็นผู้พิจารณาความเหมาะสม

การประเมินสรุปรวม คือ การบรรยายภาพ ซึ่งมีรายละเอียด ดังนี้

การบรรยายภาพ (picture description) เป็นกิจกรรมที่ครูให้นักเรียนพูดบรรยายภาพ โดยนักเรียนจะได้เตรียมตัวก่อนล่วงหน้า คือ ศึกษารายละเอียดต่างๆ ของภาพ ชักถามรายละเอียดของภาพจากครูผู้สอน ค้นหาคำศัพท์ที่เกี่ยวข้อง ลองแต่งประโยค ปรึกษาการพูดกับครูในชั้นเรียน หรือผู้ปกครอง เป็นต้น โดยภาพประกอบที่ใช้เป็นภาพที่นักเรียนสนใจนำมาเอง หรือที่ครูเตรียมไว้ให้ แล้วให้นักเรียนบรรยายเรื่องราวในภาพของตนเอง โดยภาพที่จะนำมาบรรยายครูจะเป็นผู้พิจารณาความเหมาะสม

เกณฑ์การให้คะแนนทักษะการพูดภาษาอังกฤษ

ภาระงานการพูดทั้ง 4 ภาระงานมีเกณฑ์การให้คะแนนแบบแยกส่วน (analytic rating scale) มีขอบเขตเนื้อหาในการประเมิน ดังนี้

1. คำศัพท์ (vocabulary) คือ การให้คะแนนด้านคำศัพท์ จะพิจารณาจากการพูดของนักเรียนที่สะท้อน 2 สิ่ง คือ ขอบเขต (range) ของคำศัพท์ที่ใช้ และ ความถูกต้อง (accuracy) กล่าวคือ ขอบเขต (range) ของการใช้คำศัพท์ที่มีความหลากหลาย ความสามารถในการใช้คำศัพท์ได้อย่างถูกต้อง (accuracy) และเหมาะสม (appropriation) ตรงตามความหมายที่ต้องการสื่อสาร

2. การสร้างประโยค (syntax) คือ การให้คะแนนด้านการสร้างประโยค พิจารณาจากความสามารถของนักเรียนในสร้างประโยคได้อย่างมีความถูกต้อง (accuracy) ตามที่ต้องการสื่อสาร เช่น มีประธาน กริยา กรรม ที่เหมาะสม เป็นต้น

3. การเชื่อมโยงเนื้อหา (cohesion) การให้คะแนนด้านการเชื่อมโยงเนื้อหา พิจารณาจากการที่นักเรียนสามารถพูดเชื่อมโยงประโยคแต่ละประโยคให้เป็นเรื่องราวที่มีความสัมพันธ์กันได้อย่างถูกต้องเหมาะสม

4. การออกเสียง (pronunciation) การให้คะแนนด้านการออกเสียง พิจารณาจากการที่นักเรียนสามารถเน้นเสียงหนักเบา (stress) เว้นวรรคการหยุด (segmental sound) และการใช้เสียงสูงต่ำ (intonation) ได้อย่างถูกต้อง เหมาะสม และทำให้ผู้ฟังเข้าใจในสิ่งที่ต้องการสื่อสารได้

5. บรรลุวัตถุประสงค์ (ideational function) คือ การให้คะแนนด้านบรรลุวัตถุประสงค์ พิจารณาจากการที่นักเรียนสามารถพูดสื่อสารได้บรรลุตามวัตถุประสงค์ที่ต้องการจะสื่อสาร โดยการพูดนั้นทำให้ผู้ฟังเข้าใจว่าต้องการสื่อสารอะไร เช่น พูดได้ตรงกับหัวเรื่องที่ระบุ ตอบตรงคำถาม ตอบคำถามได้อย่างเหมาะสมกับลักษณะคำถาม เป็นต้น

6. ความคล่องแคล่ว (fluency) คือ การใช้ภาษาที่มีความคล่องแคล่ว ลื่นไหล ซึ่งพิจารณาจากการที่นักเรียนสามารถพูดภาษาอังกฤษได้อย่างคล่องแคล่ว ราบรื่น และมีความลื่นไหลในระหว่างที่พูด และไม่สะดุดหรือกระตุก หยุดชะงัก ระยะเวลาที่พูด

ให้คะแนนจาก 1-4 คะแนน ได้แก่ อ่อนมาก อ่อน พอใช้ และดี ตามลำดับ) ซึ่งมีรายละเอียด ดังนี้

1. คำศัพท์ (vocabulary)

ระดับความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● ความสามารถในการใช้คำศัพท์ยังจำกัดอยู่มาก ● มักใช้คำศัพท์คำเดิมซ้ำๆ <u>ทุกประโยค หรือเกือบทุกประโยค</u> ● ใช้คำศัพท์ไม่ถูกต้อง <u>ไม่เหมาะสม</u>
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● ความสามารถในการใช้คำศัพท์ยังจำกัด ใช้คำศัพท์ไม่หลากหลาย ● ยังพบการใช้คำศัพท์คำเดิมซ้ำๆ <u>อยู่บ่อยครั้ง</u> ● มีการใช้คำศัพท์ที่ไม่ถูกต้อง <u>ไม่เหมาะสมอยู่บ่อยครั้ง</u>
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● สามารถใช้คำศัพท์ได้หลากหลาย ● แต่ยังใช้คำศัพท์คำเดิมซ้ำบ้างเป็นบางครั้ง ● สามารถใช้คำศัพท์ได้อย่างถูกต้องและเหมาะสม อาจยังใช้คำศัพท์ผิดพลาดบ้างเป็นบางครั้ง
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● สามารถใช้คำศัพท์ได้อย่างหลากหลาย ● การใช้คำศัพท์ส่วนใหญ่สามารถใช้ได้ถูกต้องเหมาะสม และผิดพลาดน้อยมาก <u>หรือ ถูกต้องทั้งหมด ไม่มีความผิดพลาดเลย</u>

2. การสร้างประโยค (syntax)

ระดับความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● การพูดไม่มีการเชื่อมโยงให้เป็นประโยค พูดเป็นคำๆ ● มีการใช้โครงสร้างประโยคที่ไม่ถูกต้องทุกครั้ง
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการเชื่อมโยงให้เป็นประโยค ● การใช้โครงสร้างประโยคที่ไม่ถูกต้องอยู่บ่อยครั้ง ทำให้ผู้ฟังไม่สามารถเข้าใจได้อย่างกระจ่าง
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการเชื่อมโยงให้เป็นประโยค ● การใช้โครงสร้างประโยคที่ไม่ถูกต้องบ้างเป็นบางครั้ง แต่ผู้ฟังสามารถเข้าใจได้
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการเชื่อมโยงให้เป็นประโยค ● ส่วนใหญ่มีการใช้โครงสร้างประโยคที่ถูกต้อง และผิดพลาดน้อยมาก <u>หรือไม่มีความผิดพลาดเลย</u>

3. การเชื่อมโยงเนื้อหา (cohesion)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อยมาก หรือไม่ใช่เลย ● การใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคเกือบทั้งหมดไม่ถูกต้อง
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยคน้อย ● การใช้คำเชื่อมโยงทำให้ความสัมพันธ์ระหว่างประโยคมีความสับสนอยู่บ่อยครั้ง และทำให้เข้าใจคลาดเคลื่อน
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยค ● ความสัมพันธ์ระหว่างประโยคอาจมีความสับสนอยู่บ้าง แต่ไม่ทำให้ความเข้าใจคลาดเคลื่อน
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● ใช้คำเชื่อมโยงเนื้อหาระหว่างประโยค ● ถูกต้อง ชัดเจน เป็นส่วนใหญ่ และไม่ทำให้เข้าใจคลาดเคลื่อน ผิดพลาดน้อยมาก หรือ ไม่มีความผิดพลาดเลย

4. การออกเสียง (pronunciation)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงเกือบทั้งหมด มีความผิดพลาด ทำให้ไม่สามารถเข้าใจได้ ● นักเรียนพูดน้อยมาก หรือไม่พูดเลย
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงยังผิดพลาดอยู่บ่อยครั้ง ● การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดอยู่มาก ● ทำให้ไม่สามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้ หรือทำให้เข้าใจคลาดเคลื่อน
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงยังมีความผิดอยู่บ้าง ● การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดอยู่บ้าง ● แต่ผู้ฟังยังคงสามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● การออกเสียงส่วนใหญ่ถูกต้อง หรือไม่มีความผิดพลาดเลย ● การออกเสียงเน้นหนักเบา เว้นวรรคการหยุด เสียงสูงต่ำ ผิดพลาดน้อยมาก หรือไม่มีความผิดพลาดเลย ● ผู้ฟังสามารถเข้าใจในสิ่งที่ต้องการสื่อสารได้โดยง่าย

5. บรรลุวัตถุประสงค์ (ideational function)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● พูดไม่ตรงประเด็น หรือไม่พูดเลย ● การพูดเกือบทั้งหมดไม่สามารถเข้าใจได้ว่าต้องการสื่อสารอะไร
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● เนื้อหาในการพูดไม่ค่อยสอดคล้องเหมาะสมกับหัวเรื่องหรือสิ่งที่ต้องการสื่อสาร ● ไม่สามารถเข้าใจอย่างชัดเจนได้ว่าวัตถุประสงค์ของการพูดคืออะไร

ระดับ ความสามารถ	คำอธิบาย
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● เนื้อหาในการพูดสอดคล้องเหมาะสม ถึงแม้จะมีความผิดพลาดในการพูดในบางครั้ง ● สามารถทำให้เข้าใจวัตถุประสงค์ที่ต้องการสื่อสารได้
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● เนื้อหาในการพูดสอดคล้องเหมาะสม มีความผิดพลาดในการพูดน้อยมาก หรือไม่ผิดพลาดเลย ● สามารถเข้าใจวัตถุประสงค์ในสิ่งที่ต้องการสื่อสารได้โดยง่าย

6. ความคล่องแคล่ว (fluency)

ระดับ ความสามารถ	คำอธิบาย
อ่อนมาก (1 คะแนน)	<ul style="list-style-type: none"> ● การพูดยังตะกุกตะกัก พูดเป็นคำๆ ● ไม่เป็นประโยค หรือไม่พูดเลย
อ่อน (2 คะแนน)	<ul style="list-style-type: none"> ● การพูดบ่อยครั้งยังช้ามาก และไม่สิ้นไหล ● หยุดเป็นระยะ หยุดคิดนาน พูดไม่จบประโยค ทำให้ใจความขาดหาย
พอใช้ (3 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีการหยุดบ้าง ไม่สิ้นไหลเป็นบางครั้ง ● หยุดชะงักบ้าง หยุดคิดบ้างเป็นบางครั้ง ● สามารถพูดได้จบประโยค ใจความครบถ้วน
ดี (4 คะแนน)	<ul style="list-style-type: none"> ● การพูดมีความเร็วสม่ำเสมอ เป็นธรรมชาติ พูดได้สิ้นไหล ไม่ตะกุกตะกัก ● ไม่หยุดชะงัก ● สามารถพูดได้จบประโยค ใจความครบถ้วน

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวดารุวรรณ ศรีแก้ว เกิดเมื่อวันที่ 11 พฤศจิกายน 2520 สำเร็จการศึกษาศึกษาศาสตรบัณฑิต จากภาควิชาหลักสูตรและการสอน วิชาเอกการประถมศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ในปีการศึกษา 2543 สำเร็จการศึกษาวิทยาศาสตรมหาบัณฑิต สาขาวิชาการวิจัยพฤติกรรมศาสตร์ประยุกต์ สถาบันวิจัยพฤติกรรมศาสตร์ ระดับปริญญาโท มหาวิทยาลัยศรีนครินทรวิโรฒ ในปีการศึกษา 2549 และศึกษาต่อในระดับดุษฎีบัณฑิต ภาควิชาวิจัยและจิตวิทยาการศึกษา สาขาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2553

