

รายงานการวิจัย

เรื่อง

การสร้างและพัฒนามาตรวัดประเมินสัมฤทธิผล
ของการพูดแบบนำเสนอสำหรับรายวิชาการ
สนทนาภาษาอังกฤษชั้นสูง

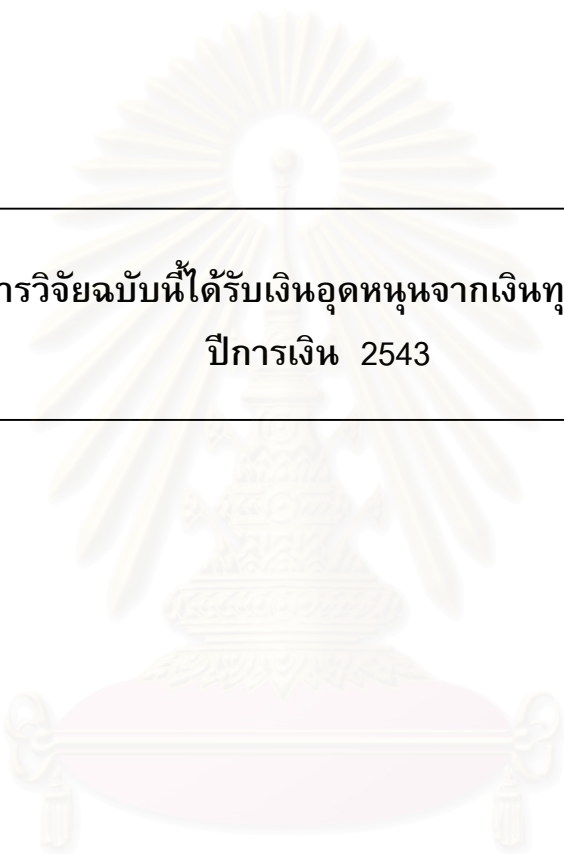
โดย

กัลยาณี เรืองมานะมงคล

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สถาบันภาษา

จุฬาลงกรณ์มหาวิทยาลัย พ.ศ. 2545



รายงานการวิจัยฉบับนี้ได้รับเงินอุดหนุนจากเงินทุนสถาบันภาษา
ปีการเงิน 2543

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A RESEARCH REPORT

The Construction and Development of Rating Scales to Evaluate
Achievement in Oral Presentations for an “Advanced
Business Oral Communication” Course

By

Kalayanee Ruangmanamongkol



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย
Language Institute

Chulalongkorn University, 2002

กิตติกรรมประกาศ

รายงานการวิจัยฉบับนี้เสร็จสมบูรณ์ได้ด้วยอุปการะคุณจากหลายฝ่ายคือ
จุฬาลงกรณ์มหาวิทยาลัย และคณะกรรมการวิจัยสถาบันภาษาที่ได้จัดสรร
งบประมาณเงินแผ่นดินประจำปี 2543 เพื่อเป็นทุนในการดำเนินการวิจัย

รองศาสตราจารย์ ดร. อัจฉรา วงศ์โสธร ที่ได้กรุณาให้คำแนะนำในช่วง เริ่ม
ต้นโครงการ

รองศาสตราจารย์ ดร. สุวัฒน์ สุขมลสันต์ ที่ได้กรุณาให้คำแนะนำในการ
ปรับปรุงแก้ไข

รองศาสตราจารย์อัญทิการ์ โรงสะอาด และรองศาสตราจารย์สุชาดา
นิมมานนิตย์ ที่ได้กรุณาให้คำปรึกษา

ผู้ช่วยศาสตราจารย์สำหรับ กุลศิริสวัสดิ์ ที่ได้กรุณาตรวจแก้ไขและปรับปรุง
ภาษา

จึงขอขอบพระคุณเป็นอย่างสูงมา ณ โอกาสนี้ด้วย

ผู้จัดทำ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อภาษาไทย

การวิจัยเรื่องการสร้างและพัฒนามาตรวัดประเมินสัมฤทธิผลของการพูดแบบนำเสนอสำหรับรายวิชาการสนทนาภาษาอังกฤษขั้นสูงมีวัตถุประสงค์เพื่อแก้ไขปัญหาที่เกิดขึ้นในอดีตเกี่ยวกับการทดสอบ และวัดผลทักษะการพูดแบบนำเสนอ (Oral Presentation Skill) เนื่องจากในอดีตการทดสอบทักษะการพูดนั้นทำโดยการให้ผู้เรียนสอบข้อเขียน ซึ่งการทดสอบวิธีนี้ไม่มีความตรง (Validity) และความเหมาะสม หากแต่มีความเที่ยง (Reliability) สูงกว่าวิธีการทดสอบพูด

ดังนั้นจึงได้มีความพยายามสร้างมาตรฐานในการทดสอบสัมฤทธิผลของการพูดแบบนำเสนอให้มีทั้งความตรงและความเที่ยงในการประเมิน โดยการสร้างและพัฒนามาตรวัดประเมินค่า (Rating Scales) ขึ้นเพื่อใช้เป็นเกณฑ์ในการประเมินผลการทดสอบดังกล่าว และภายหลังจากที่ได้พัฒนามาตรวัดประเมินค่าแล้วได้มีการตรวจสอบความตรงของมาตรวัดในสามประเด็นคือ ความตรงเชิงประจักษ์ (Face Validity) ความตรงเชิงเนื้อหา (Content Validity) และความตรงร่วมสมัย (Concurrent Validity) และนอกจากนี้ยังได้มีการตรวจสอบความเที่ยงของผู้ประเมิน (Rater Reliability) ในด้านความเที่ยงของผู้ประเมินแบบ Intra-rater Reliability และความเที่ยงระหว่างผู้ประเมินหรือ Inter-rater Reliability โดยการให้ผลวิจัยจำนวนรวมทั้งสิ้น 270 คน ประกอบด้วยผู้สอนและประเมินจำนวน 16 คน และนิสิตจำนวน 254 คน ตอบแบบสอบถาม ทำการทดสอบการพูดแบบนำเสนอและประเมินผลโดยใช้มาตรวัดที่สร้างขึ้นเป็นเกณฑ์ รวมทั้งทำแบบทดสอบข้อเขียนจำนวนสองชุด

ผลการตรวจสอบปรากฏว่ามาตรวัดประเมินค่านี้มีความตรงเชิงประจักษ์และความตรงเชิงเนื้อหาสูงมาก แต่ไม่มีหลักฐานแสดงว่ามาตรวัดดังกล่าวมีความตรงร่วมสมัยสำหรับความเที่ยงของผู้ประเมินทั้งแบบ Intra-rater Reliability และความเที่ยงระหว่างผู้ประเมินหรือ Inter-rater Reliability ผลปรากฏว่า ผู้ประเมินแต่ละคนมีความเที่ยงมากน้อยต่างกันและความต่างนี้มีตั้งแต่ มีความเที่ยงมากที่สุด ไปจนถึงไม่มีความเที่ยงเลย

สรุปคือ จากการนำมาตรวัดประเมินค่าที่สร้างและพัฒนาขึ้นไปใช้ในสถานการณ์จริง ผลที่ได้เป็นที่น่าพอใจโดยเฉพาะในด้านคุณภาพของมาตรวัดและวิธีการทดสอบ หากในแง่ของผู้ประเมินซึ่งมีความเที่ยงต่างกัน จึงอาจจำเป็นต้องมีการฝึกอบรมผู้ประเมิน และอาจมีการปรับปรุงพัฒนาคุณภาพของมาตรวัดเพื่อให้การทดสอบและการประเมินผล การพูดแบบนำเสนอมีมาตรฐานสูงขึ้นต่อไป

ABSTRACT

For years, a conflict between the validity and reliability of the evaluation of students' oral presentation skills (one of the main focuses of the Advanced Business Oral Communication course) has been a major concern. Therefore, the Rating Scales have been constructed and developed in an attempt to find a solution to this problem. Details in each band include the course's contents as the aim of this Rating Scales is to evaluate students' achievement and not proficiency. After the construction of the scales, the assessing teachers and the specialist helped develop them, and finally, the validity of the scales and the reliability of the raters were examined.

This research covers details and data regarding the construction and development of the Rating Scales, a survey of the validity of the scales and a survey of raters' reliability. The findings will be used for further development of evaluation procedures and criteria in the future.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

กิตติกรรมประกาศ	ก
บทคัดย่อภาษาไทย	ข
บทคัดย่อภาษาอังกฤษ	ค
บทที่ 1	
บทนำ	
1. ความเป็นมาและความสำคัญของปัญหา	1
2. ประเด็นปัญหาที่จะทำการวิจัย	2
3. วัตถุประสงค์ของการวิจัย	2
4. ประโยชน์ที่คาดว่าจะได้รับ	2
5. ขอบเขตของการวิจัย	3
6. ข้อจำกัดของการวิจัย	3
7. นิยามศัพท์เฉพาะ	3
บทที่ 2	
วรรณคดีและงานวิจัยที่เกี่ยวข้อง	
1. วรรณคดี	6
1.1 ประวัติความเป็นมาของการทดสอบทักษะการพูด (Oral Testing)	6
1.2 วิวัฒนาการของการทดสอบการพูด	6
1.2.1 Foreign Service Institute (FSI)	6
1.2.2 Interagency Language Roundtable (ILR)	7
1.2.3 The American Council on the Teaching of Foreign Languages (ACTFL) / Educational Testing Service (ETS)	7
1.3 การสร้างเกณฑ์ในการประเมินสมรรถภาพทางการพูด (Oral Proficiency Test)	8

สารบัญ (ต่อ)

	หน้า
1.4 ความตรงของการประเมิน (Validity)	9
1.4.1 ความตรงเชิงประจักษ์ (Face Validity)	10
1.4.2 ความตรงเชิงเนื้อหา (Content Validity)	10
1.4.3 ความตรงร่วมสมัย (Concurrent Validity)	10
1.5 ความเที่ยงของผู้ประเมิน (Rater Reliability)	10
1.5.1 ความเที่ยงของผู้ประเมิน (Intra-rater Reliability)	11
1.5.2 ความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability)	12
1.6 การอบรมผู้ประเมิน (Rater Training)	12
2. งานวิจัยที่เกี่ยวข้อง	13
2.1 ผลการศึกษาความตรง (Validity) และความเที่ยง (Reliability) ของการประเมินสมิทิตภาพทางการพูด (Oral Proficiency Test)	13
2.2 งานวิจัยที่เกี่ยวกับการประเมินการพูดแบบนำเสนอ	15
บทที่ 3	
วิธีดำเนินการวิจัย	
1. กลุ่มประชากรและพลวิจัย	18
2. เครื่องมือวิจัย	18
3. การสร้างและพัฒนามาตรวัดประเมินค่า	19
4. การรวบรวมข้อมูล	28
5. การวิเคราะห์ข้อมูล	29

สารบัญ (ต่อ)

	หน้า
บทที่ 4	
ผลการวิจัย	
1. ผลการตรวจสอบความตรง (Validity)	30
1.1 ความตรงเชิงประจักษ์ (Face Validity)	30
1.2 ความตรงเชิงเนื้อหา (Content Validity)	33
1.3 ความตรงร่วมสมัย (Concurrent Validity)	37
2. ผลการตรวจสอบความสอดคล้องของผู้ประเมิน (Rater Reliability)	40
2.1 ความเที่ยงของผู้ประเมิน (intra-rater Reliability)	40
2.2 ความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability)	43
บทที่ 5	
สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ	
1. สรุปผลการวิจัยและอภิปรายผล	54
2. ข้อเสนอแนะ	59
บรรณานุกรม	61
ภาคผนวก ก	65
ภาคผนวก ข	67
ภาคผนวก ค	68

สารบัญ (ต่อ)

	หน้า
ภาคผนวก ง	72
ภาคผนวก จ	78
ภาคผนวก ฉ	80
ภาคผนวก ช	84
ภาคผนวก ซ	85



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

		หน้า
ตารางที่ 1	เกณฑ์ในมาตราส่วนประเมินค่า : หัวข้อ และประเด็นในแต่ละหัวข้อ	22
ตารางที่ 2	รายละเอียดของเกณฑ์ในแต่ละช่วงคะแนน	26
ตารางที่ 3	ผลการตรวจสอบความตรงเชิงประจักษ์ ของวิธีการประเมินจากนิสิต	31
ตารางที่ 4	ผลการตรวจสอบความพอใจของนิสิต ต่อวิธีประเมินผล	31
ตารางที่ 5	ผลการตรวจสอบความตรงเชิงประจักษ์ของเกณฑ์ ประเมินผลจากนิสิต	32
ตารางที่ 6	ผลการตรวจสอบความตรงเชิงประจักษ์ของวิธีการ ประเมินจากผู้ประเมิน	32
ตารางที่ 7	ผลการตรวจสอบความตรงเชิงประจักษ์ของเกณฑ์ จากผู้ประเมิน	33
ตารางที่ 8	ผลการตรวจสอบความตรงเชิงเนื้อหา	34
ตารางที่ 9	ผลการตรวจสอบความเหมาะสมสำหรับรายวิชา	34
ตารางที่ 10	ผลการตรวจสอบความเหมาะสมและครอบคลุม เนื้อหาของภาษาที่ใช้บรรยายเกณฑ์ (descriptors)	35
ตารางที่ 11	ผลการตรวจสอบความสะดวกและง่ายต่อการ ใช้เกณฑ์	36

สารบัญตาราง (ต่อ)

		หน้า
ตารางที่ 12	ผลการตรวจสอบความตรงร่วมสมัยในด้าน คุณลักษณะในการใช้ภาษา (Language)	39
ตารางที่ 13	ผลการตรวจสอบความตรงร่วมสมัยในด้าน คุณลักษณะด้านโครงสร้าง (Organization)	40
ตารางที่ 14	ผลการตรวจสอบความเที่ยงของผู้ประเมิน โดยรวม 5 หัวข้อ	41
ตารางที่ 15	ผลการตรวจสอบความเที่ยงของผู้ประเมิน ในแต่ละหัวข้อ	42
ตารางที่ 16	ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน โดยรวม 5 หัวข้อ	44
ตารางที่ 17	ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน ในด้านโครงสร้าง	46
ตารางที่ 18	สถิติเชิงพรรณนาด้านโครงสร้าง	47
ตารางที่ 19	ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน ในด้านเนื้อหา	47
ตารางที่ 20	สถิติเชิงพรรณนาด้านเนื้อหา	48
ตารางที่ 21	ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน ในด้านการใช้ภาษา	49

สารบัญตาราง (ต่อ)

		หน้า
ตารางที่ 22	สถิติเชิงพรรณนาด้านการใช้ภาษา	50
ตารางที่ 23	ผลการตรวจสอบความเที่ยงระหว่างของผู้ประเมิน การนำเสนอ	50 ในด้าน
ตารางที่ 24	สถิติเชิงพรรณนาด้านการนำเสนอ	51
ตารางที่ 25	ผลการตรวจสอบความเที่ยงระหว่างของผู้ประเมิน ในด้านการใช้สื่อทัศนูปกรณ์	52
ตารางที่ 26	สถิติเชิงพรรณนาด้านการใช้สื่อประกอบ	53
ตารางที่ 27	ความเที่ยงระหว่างผู้ประเมิน	58

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญแผนภาพ

		หน้า
แผนภาพที่ 1	วิธีการวัดคุณลักษณะด้านการใช้ภาษา	38
แผนภาพที่ 2	วิธีการวัดคุณลักษณะด้านโครงสร้าง	38



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1. ความเป็นมาและความสำคัญของปัญหา (Rationale)

ทักษะการพูดแบบนำเสนอ (Oral Presentation Skill) มีบทบาทสำคัญ ทั้งในวง การการศึกษา และในวงการวิชาชีพในปัจจุบัน ดังนั้น สถาบันภาษา จุฬาลงกรณ์ มหาวิทยาลัย จึงได้จัดการเรียนการสอนรายวิชาการสนทนาภาษาอังกฤษขั้นสูง (Advanced Business Oral Communication) ให้แก่นิสิตคณะพาณิชยศาสตร์และการบัญชี ชั้นปีที่ 3 เป็นวิชาบังคับ ซึ่งเน้นทักษะการพูดแบบนำเสนอ (Oral Presentations) เป็นเนื้อหาหลักของ รายวิชา และการประเมินผลของรายวิชานี้ มิได้ใช้แบบทดสอบข้อเขียนแต่ใช้การประเมินแบบ เหมือนจริง (Authentic Assessment) โดยให้ผู้เรียนพูดแบบนำเสนอ ดังนั้นจึงจำเป็นต้องมี เครื่องมือหรือแบบทดสอบที่มี คุณภาพ เพื่อการวัดผลที่เที่ยงตรงเชื่อถือได้ เช่นเดียวกับการ สอบข้อเขียน รวมทั้งมีความสะดวกในทางปฏิบัติ ทั้งต่อผู้ประเมินเอง และต่อผู้เรียนด้วย

โดยก่อนหน้าการวัดผลรายวิชาดังกล่าว มีปัญหาในด้านความตรง เชิง ประจักษ์ (Face Validity) เนื่องจากการวัดผลโดยการสอบข้อเขียน ไม่ตรงกับ วัดดู ประสงค์ของรายวิชาซึ่งเน้นทักษะการพูด แต่ทั้งนี้การทดสอบนิตจำนวน 570 คน โดยผู้ ประเมิน 13-16 คน โดยให้นิสิตพูดนั้น มีค่าความน่าเชื่อถือน้อย เนื่องจากผู้ประเมินทั้ง หมด มีความต่างกันทั้งด้านหลักเกณฑ์ และมาตรฐานในการประเมินนิต ถึงแม้ว่าจะมีการ ตกลงหลักเกณฑ์การปฏิบัติให้มีมาตรฐานเดียวกันในเบื้องต้นก็ตาม จึงปรากฏปัญหาในด้าน ความน่าเชื่อถือของการประเมินและความ ยุติธรรมในการประเมินในกรณีนี้นิตต่างกลุ่ม ได้รับการประเมินจากผู้ประเมินต่างคนกันซึ่งมีความเข้มงวด และมาตรฐานในการตัดสินที่ แตกต่างกัน

ดังนั้น จึงได้มีการสร้างและพัฒนาแบบประเมินสัมฤทธิ์ผลของการพูดแบบนำ เสนอ สำหรับรายวิชาการสนทนาภาษาอังกฤษขั้นสูงนี้ขึ้น โดยการสร้างเครื่องมือกำหนด คะแนนที่เป็นมาตรวัดประเมินค่า (Rating Scales) ให้มีคุณภาพในด้านความตรงเชิง ประจักษ์ (Face Validity) ความตรงเชิงเนื้อหา (Content Validity) ความตรงร่วมสมัย (Concurrent Validity) และความสะดวกในทางปฏิบัติ (Practicality) ทั้งนี้ ได้มีการวัดความ เที่ยงของผลการทดสอบการใช้แบบทดสอบดังกล่าว โดยการวัดความเที่ยงของผู้ประเมิน (Intra-rater Reliability) และการวัดความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability)

ในการประเมินผู้เรียนคนเดียวหรือกลุ่มเดียวกัน เพื่อตรวจสอบคุณภาพของแบบประเมินอีก
ชั้นหนึ่งด้วย

2. ประเด็นปัญหาที่จะทำการวิจัย (Statement of Problems) ประเด็นปัญหาที่จะทำ การวิจัยคือ

2.1 คุณภาพแบบประเมินสัมฤทธิ์ผล สำหรับรายวิชาการสนทนาภาษาอังกฤษ ชั้นสูง
ในรูปแบบของมาตรวัดประเมินค่า (Rating Scales) ในด้านความตรงเชิงประจักษ์ (Face
Validity) ความตรงเชิงเนื้อหา (Content Validity) และความตรงร่วมสมัย (Concurrent
Validity)

2.2 ความเที่ยงของผู้ประเมิน ทั้งความสอดคล้องภายในของผู้ประเมิน (Intra-rater
Reliability) และความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) โดยใช้หลักเกณฑ์ใน
มาตราส่วนประเมินค่าที่สร้างขึ้น

3. วัตถุประสงค์ของการวิจัย (Objectives) งานวิจัยนี้มีจุดประสงค์หลักสอง ประการคือ

3.1 เพื่อสร้างและพัฒนามาตรวัดประเมินค่าเพื่อทดสอบสัมฤทธิ์ผลในการพูดแบบนำ
เสนอให้มีความตรงเชิงประจักษ์ ความตรงเชิงเนื้อหาและความตรงร่วมสมัยให้เป็นที่ยอมรับ
ในกลุ่มบุคคลที่ใช้เกณฑ์นี้คือ ผู้ประเมินผลและนิสิต

3.2 เพื่อตรวจสอบผลของการนำมาตราวัดดังกล่าวไปใช้ในสถานการณ์จริงในด้านความ
เที่ยงของผู้ประเมิน (Intra-rater Reliability) และความเที่ยงระหว่างผู้ประเมิน (Inter-rater
Reliability)

4. ประโยชน์ที่คาดว่าจะได้รับ (Significance of Study)

4.1 ได้มาตราวัดประเมินค่าเพื่อทดสอบสัมฤทธิ์ผลที่มีคุณภาพ ครอบคลุมเนื้อหา แม่น
ตรง และง่ายต่อการใช้

4.2 ได้ตรวจสอบผลของการนำมาตราวัดประเมินค่าไปใช้ในสถานการณ์จริงและความ
สอดคล้องของผู้ใช้แบบประเมินเพื่อทราบปัญหาและเพื่อการปรับปรุงวิธีการประเมินต่อไปใน
อนาคต

5. ขอบเขตของการวิจัย (Scope of the Study) ขอบเขตของงานวิจัยนี้คือ

เครื่องมือที่เป็นแบบประเมินกำหนดคะแนนแบบมาตรวัดประเมินค่านี้ใช้เพื่อทดสอบสัมฤทธิ์ผล (Achievement Test) ในการพูดแบบนำเสนอ มีใช้การวัด สมรรถภาพ (Proficiency Test) จึงเหมาะที่จะใช้วัดผลผู้เรียนวิชาที่เน้นการพูดแบบ นำเสนอเนื่องจากแบบประเมินนี้ครอบคลุมเนื้อหาของรายวิชาดังกล่าว และผู้ใช้แบบประเมินคือ ผู้สอนวิชานั้นเท่านั้น

6. ข้อจำกัดของการวิจัย (Limitation)

6.1 ผู้ประเมินที่เข้าร่วมโครงการไม่สามารถประเมินการพูดแบบนำเสนอของนิสิตในเวลาเดียวกันได้ เนื่องจากมีข้อจำกัดในเรื่องเวลา

6.2 การประเมินนิสิตเป็นการดูข้อมูลจากวิดีโอซึ่งไม่สามารถสะท้อนให้เห็น ภาพรวมทั้งหมด นอกจากนี้คุณภาพของเสียงและภาพไม่ชัดเจนเหมือนดูจากเหตุการณ์จริง ทำให้การประเมินจากเทปอาจไม่ได้ผลเช่นเดียวกับการประเมินสด อีกทั้งวิดีโอไม่สามารถครอบคลุมข้อมูลที่สำคัญบางส่วน เช่นการออกแบบแผ่นใส และข้อความบนแผ่นใสทั้งหมด

7. นิยามศัพท์เฉพาะ (Definition of Terms)

7.1 รายวิชาการสนทนาภาษาอังกฤษขั้นสูง (Advanced Business Oral Communication) หมายถึง รายวิชาบังคับรายวิชาหนึ่งสำหรับนิสิตชั้นปีที่ 3 คณะพาณิชยศาสตร์และการบัญชี มีเนื้อหาหลักสูตรที่เน้นการพูดเพื่อสื่อสารทางธุรกิจ

7.2 การพูดแบบนำเสนอ (Oral Presentations) หมายถึง การพูดในเชิงนำเสนอที่มีรูปแบบตามเนื้อหาของหลักสูตร รายวิชาการสนทนาภาษาอังกฤษขั้นสูง

7.3 แบบทดสอบสัมฤทธิ์ผล (Achievement Test) หมายถึง การทดสอบเพื่อ วัดผลของการเรียน เป็นการทดสอบที่มีเนื้อหาตรงกับเนื้อหาและวัตถุประสงค์ของ หลักสูตร

7.4 แบบทดสอบสมรรถภาพ (Proficiency Test) หมายถึง การวัดความสามารถทั่วไปทางภาษา โดยไม่มีการเน้นเนื้อหาเฉพาะเจาะจงในหลักสูตรใดหลักสูตรหนึ่ง

7.5 แบบประเมินหรือมาตรวัดประเมินค่า (Rating Scales) หมายถึง เกณฑ์ในการประเมินสัมฤทธิ์ผลของรายวิชา ABOC ถูกสร้างขึ้นโดยเน้นเนื้อหาของหลักสูตรเป็นประเด็นหลัก มีลักษณะเป็นแบบช่วงคะแนน (Band Scores)

7.6 โครงสร้าง (Organization) หมายถึง โครงสร้างของการพูดแบบนำเสนอ หมายถึง ลำดับขั้นตอน ซึ่งเป็นองค์ประกอบของการนำเสนอ เช่น มีการกล่าวนำ การให้ข้อมูล การสรุป และจบการนำเสนอ เป็นต้น

7.7 เนื้อหา (Content) หมายถึง เนื้อหาและข้อมูลรายละเอียดรวมถึงตัวอย่างต่างๆ ซึ่งเกี่ยวข้องกับหัวข้อที่นำเสนอ รวมทั้งความน่าสนใจของรายละเอียด

7.8 การใช้ภาษา (Language) หมายถึง การใช้ภาษาในการนำเสนอ เช่น จะต้องเป็นภาษาพูด และเข้าใจง่าย รวมทั้งถูกต้องในแง่ของ ไวยากรณ์ การออกเสียง การใช้ศัพท์ ความคล่อง และความเหมาะสม

7.9 การนำเสนอ (Delivery) หมายถึง การนำเสนอที่รวมทั้งการใช้เสียง ภาษาท่าทาง รวมไปถึงการใช้สายตา และสีหน้า ขณะนำเสนอข้อมูล

7.10 การใช้สื่อทัศนูปกรณ์ (Visual Aids) หมายถึง ด้วยการใช้สื่อต่างๆ เพื่อประกอบการนำเสนอ เช่น แผ่นภาพ แผ่นใส สไลด์ วีดิทัศน์ LCD หรือ Visualizer เป็นต้น

7.11 ความตรงเชิงประจักษ์ (Face Validity) หมายถึง ความตรงที่ปรากฏให้เห็นตามสายตาของผู้สอบว่าแบบทดสอบสามารถวัดสิ่งที่ต้องการจะวัดได้ เช่น แบบทดสอบการพูด ผู้สอบก็รู้สึกว่าการทดสอบนั้นสามารถวัดความสามารถในการพูดของตนได้

7.12 ความตรงเชิงเนื้อหา (Content Validity) หมายถึง ความตรงของแบบทดสอบในด้านเนื้อหาที่ครอบคลุมและตรงกับเนื้อหาที่ต้องการวัด และเนื้อหาที่ระบุอยู่ในหลักสูตรและวัตถุประสงค์ของรายวิชา

7.13 ความตรงร่วมสมัย (Concurrent Validity) หมายถึง แบบทดสอบที่ใช้วัดความสามารถซึ่งได้ผลตรงกับแบบทดสอบมาตรฐานที่มุ่งทดสอบความรู้และความสามารถเดียวกัน เมื่อทำการสอบในเวลาใกล้เคียงกัน

7.14 ความเที่ยงของผู้ประเมิน (Intra-rater Reliability) หมายถึง ความคงที่และแม่นยำในการประเมินของผู้ประเมินแต่ละคน จากการประเมิน 2 ครั้ง และ เปรียบเทียบผลการประเมินระหว่างครั้งที่ 1 และครั้งที่ 2

7.15 ความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) หมายถึง ความคล้ายคลึงกัน ของผลการประเมินระหว่างผู้ประเมิน 2 คน ซึ่งวัดได้จากการนำผลที่ได้จากผู้ประเมิน 2 คน มาเปรียบเทียบกัน



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

1. วรรณกรรม

1.1 ประวัติความเป็นมาของการทดสอบทักษะการพูด (Oral Testing)

การทดสอบทักษะการพูด (Oral Testing) เริ่มขึ้นเมื่อสมัยสงครามโลกครั้งที่สองโดย FSI (Foreign Service Institute) ในประเทศสหรัฐอเมริกา (อ้างจาก Fulcher, 1993 : 8) สาเหตุเกิดจากการที่นายทหารชาวอเมริกัน และนักการทูตมีความจำเป็นต้องใช้ทักษะทางภาษาเพื่อการปฏิบัติหน้าที่อย่างมีประสิทธิภาพ ดังนั้นจึงมีการพัฒนาการทดสอบความสามารถในการพูดภาษาต่างประเทศขึ้นมา เริ่มจากในประเทศสหรัฐ แล้วจึงขยายไปยังทวีปยุโรป และประเทศอื่นๆ ต่อมา

ในปี ค.ศ. 1952 ซึ่งเป็นปีที่มีการวัดและประเมินความสามารถในการพูดของทหารอเมริกันเป็นครั้งแรก ได้มีการตั้งคณะกรรมการในการประเมินขึ้นมา และได้สร้างมาตราส่วนประเมินค่า (Rating Scales) ซึ่งแบ่งเป็นช่วงคะแนน (Bands) 6 ช่วง เริ่มจากช่วงที่ 1 (Band 1) ซึ่งไม่มีความสามารถในการพูดเพื่อสื่อสารได้เลย ไปจนถึงคะแนนช่วงที่ 6 คือพูดได้เหมือนเจ้าของภาษา ซึ่งเกณฑ์ที่สร้างขึ้นเป็นแบบประเมินโดยรวม (Holistic) ไม่มีการแยกหัวข้อ และไม่มีการละเอียดเกี่ยวกับความแตกต่างหรือเนื้อหาในช่วงคะแนนแต่ละช่วง มีเพียงคำอธิบายสั้นๆ ในช่วงคะแนนที่ 1 และ 6 เท่านั้น

ในปี 1958 FSI Testing Unit ได้เพิ่มเติมองค์ประกอบในการประเมินขึ้น 5 องค์ประกอบ คือ การออกเสียง (Accent) ไวยากรณ์ (Grammar) คำศัพท์ (Vocabulary) ความคล่อง (Fluency) และความเข้าใจ (Comprehension) (โปรดดูรายละเอียดที่ภาคผนวก ก)

1.2 วิวัฒนาการของการทดสอบทักษะการพูด

1.2.1 Foreign Service Institute (FSI)

กระบวนการทดสอบทักษะการพูด ซึ่งพัฒนาโดย FSI ได้รับความสนใจและความเชื่อถือสูงมากซึ่ง Fulcher (1993) อธิบายถึงสาเหตุว่าเป็นเพราะ FSI ได้รับการยอมรับ

รับว่ามีความตรงเชิงประจักษ์สูง และมีความสอดคล้องระหว่างผู้ประเมินสูงเช่นกัน

1.2.2 Interagency Language Roundtable (ILR)

Fulcher (1993) กล่าวว่า ในปี ค.ศ. 1960 สถาบันภาษาของกระทรวงกลาโหม สหรัฐฯ (Defense Language Institute) หน่วยข่าวกรอง (The Central Intelligence Agency หรือ CIA) และองค์กรอาสาสมัครอเมริกัน (The Peace Corps) ได้นำเกณฑ์และวิธีการประเมินความสามารถทางการพูดตามแบบของ FSI ไปปรับและประยุกต์ใช้ และได้มีความพยายามเพิ่มเติมความเชื่อถือได้ในการประเมิน โดยการให้ผู้ประเมินหลายคนเฉลี่ยคะแนนกัน และองค์กรของรัฐบาลสหรัฐฯ เหล่านี้ได้ร่วมกันสร้างมาตรฐานให้กับกระบวนการวัดผลนี้ โดยสร้างเกณฑ์ประเมินทักษะการพูดขึ้น มีช่วงคะแนน 6 ระดับเริ่มจาก 0 0+ 1 2 3 4 5 (โปรดดูรายละเอียดที่ภาคผนวก ข)

1.2.3 The American Council on the Teaching of Foreign Languages (ACTFL) / Educational Testing Service (ETS)

ในปี ค.ศ. 1979 ประธานาธิบดี Carter ของสหรัฐฯ ได้มอบหมายให้ ACTFL สร้างมาตรฐานการประเมินในระดับชาติเพื่อพัฒนาการทดสอบทางภาษา และการประเมินผลการเรียนภาษาในสหรัฐฯ โดยมีผลงานของ FSI เป็นก้าวที่สำคัญของการประเมินสมรรถภาพทางภาษา (Language Proficiency) แต่เกณฑ์ของ ACTFL (ภาคผนวก ค) มีคำบรรยายเกณฑ์ (Descriptors) ที่ยาว และมีการผสมผสานระหว่างเกณฑ์ทางภาษา (Linguistic Criteria) และเกณฑ์ที่ไม่เกี่ยวข้องกับภาษา (Non-linguistic Criteria) ซึ่ง Fulcher กล่าวว่า เกณฑ์ของ ACTFL มีความสับสนและถูกวิพากษ์วิจารณ์อย่างหนัก โดย Bachman and Savignon (1986) และ Matthews (1990) นอกจากนี้เกณฑ์ของ ACTFL ทำให้การศึกษาเกี่ยวกับความตรง (Validity) เป็นเรื่องยากมาก

การพัฒนาการทดสอบทักษะการพูดในยุคต้นจาก FSI สู่อินเตอร์แลงจิวเอจ ลังวีกส์ (ILR) และ ACTFL นี้เอง ที่เป็นรากฐานของการทดสอบการพูดในแบบอื่นๆ และทำให้เกิดความสนใจเกี่ยวกับการประเมินในประเด็นต่างๆ เช่น การประเมินโดยรวม (Holistic Rating) และการประเมินโดยดูส่วนประกอบ (Componential Rating) ช่วงคะแนน (Bands) และคำบรรยายช่วงคะแนน (Descriptors) ความไม่เป็นกลางในการประเมิน (Bias) ความแตกต่างระหว่าง

เกณฑ์ทางภาษา และเกณฑ์เกี่ยวกับการสื่อสาร (Linguistic and Communicative Criteria) และความสอดคล้อง (Reliability)

1.3 การสร้างเกณฑ์ในการประเมินสมรรถภาพทางการพูด (Oral Proficiency Test)

1.3.1 Alderson (1991) ได้กล่าวถึงข้อดีของการใช้มาตรวัดประเมินค่า (Rating Scales) แบบช่วงคะแนน (Band Scores) เป็นเกณฑ์ในการประเมินทักษะการพูดว่า สามารถลดปัญหาเกี่ยวกับความแม่นยำของการให้คะแนนแบบเป็นตัวเลข กล่าวคือ การประเมินแบบช่วงคะแนนสามารถลดความคลาดเคลื่อน และช่วยเพิ่มความสอดคล้องในการประเมินทักษะการพูดหรือการเขียน แต่จะเกิดปัญหาเรื่องการตีความของเกณฑ์ ผู้ประเมินแต่ละคนอาจมีความเห็นต่างกันในเรื่องความหมายของคำบรรยายเกณฑ์ ว่าแต่ละเกณฑ์มีความแตกต่างกันอย่างไร

1.3.2 Fulcher (1993) กล่าวว่ามีการเขียนถึงกระบวนการสร้างเกณฑ์เพื่อให้คะแนนความสามารถในการพูดน้อยมาก และการสร้างและพัฒนาเกณฑ์นั้น เกิดจากความเชื่อของผู้พัฒนาเกณฑ์ว่าอะไรคือแก่นของสมรรถภาพทางภาษา ซึ่ง ทศนคติของผู้พัฒนาเกณฑ์นี้มีรากฐานมาจากประสบการณ์ที่ได้จากผู้เรียนภาษาที่สองและโดยปกติ คำบรรยายเกณฑ์จะถูกตรวจสอบโดยผู้เชี่ยวชาญในสาขา ดังนั้นการสร้างเกณฑ์จึงไม่ใช่ประเด็นหลักที่สำคัญ แต่เป็นความสอดคล้องของผู้ใช้เกณฑ์ เมื่อเกณฑ์ได้รับการพัฒนาแล้ว ซึ่ง Fulcher กล่าวว่านี่อาจเป็นสาเหตุของการขาดความตรง (Invalidity) ในกระบวนการทดสอบได้ นอกจากนี้เขายังได้กล่าวถึงวิธีการสร้างคำบรรยายเกณฑ์ของ IELTS ว่าเป็นวิธีการพัฒนาเกณฑ์แบบใช้สาเหตุทำนายผล (a-priori methods of development) ซึ่งประกอบด้วย 3 วิธีคือ

- (1) ร่างคำบรรยายส่งไปให้กับครู หรือผู้ให้คะแนนที่มีประสบการณ์ หรือผู้ดำเนินการถามคำถาม ฯลฯ เพื่อขอความเห็นและนำกลับมาปรับปรุงเกณฑ์
- (2) ผู้ให้คะแนนที่มีประสบการณ์ จะพยายามจับคู่ (Match) ตัวอย่างภาษาพูด (Speech Samples) กับระดับต่างๆ ในเกณฑ์ เพื่อระบุคุณลักษณะหลัก (Key Features) ของตัวอย่าง
- (3) กระบวนการของเกณฑ์ที่ใช้ในการประเมินผล และข้อมูลป้อนกลับจากผู้ประเมิน (Raters) จะให้ข้อมูลอย่างต่อเนื่องสำหรับการปรับปรุงภาษาที่ใช้บรรยายเกณฑ์

จากตัวอย่างของ IELTS Fulcher เห็นว่าความหมายของคำบรรยายเกณฑ์ ขึ้นอยู่กับการใช้ความหรือความรู้สึกของผู้ประเมิน เมื่อได้ทำการพบปะพูดคุย (Socialize) ในหมู่ผู้ใช้เกณฑ์ ซึ่งเป็นสิ่งที่ดูเหมือนจะหลีกเลี่ยงไม่ได้ トラบเท่าที่เกณฑ์นั้นถูกสร้างขึ้นบนพื้นฐานของเหตุไปสู่อผล (a-priori) เช่น เกณฑ์ของ IELTS, ACTFL, ILR เป็นต้น

ส่วนในด้านความสั้นยาวของมาตราส่วนประเมินค่า หรือเกณฑ์นั้น Fulcher กล่าวถึงผู้ที่ให้ความเห็นที่ต่างกัน เช่น Lado (1961) Baker (1989), Hughes (1989) และ Matthews (1990) ซึ่งกล่าวว่ามาตราส่วนประเมินค่าจำเป็นต้องครอบคลุมรายละเอียดที่เฉพาะเจาะจงให้มากที่สุดเท่าที่จะทำได้ ในขณะที่ Porter (1991) เห็นว่าคำบรรยายเกณฑ์ควรจะสั้นและชัดเจน เพื่อความสะดวกและเพื่อลดปัญหาความยุ่งยากในการใช้เกณฑ์ การประเมินจะดูจากความสามารถที่ขาดไป มากกว่าที่จะดูจากสิ่งที่สามารถทำได้ Fulcher สรุปว่ายังไม่มียานวิจัยที่เกี่ยวกับความยาวของมาตราส่วนประเมินค่าที่เหมาะสม และสามารถนำไปใช้ได้อย่างสะดวกและง่ายมีเพียงความเห็นส่วนตัวของผู้เขียนแต่ละคนจากประสบการณ์ส่วนตัวที่มี เท่านั้น

1.4 ความตรงของการประเมิน (Validity)

ความตรง (Validity) ตามความหมายของ Henning (1992) คือ ความเหมาะสมของการทดสอบ หรือส่วนประกอบของการทดสอบ ในการวัดสิ่งที่ต้องการจะวัด แบบทดสอบนั้นจะถือว่ามีความตรง หากสามารถวัดสิ่งที่ต้องการวัดได้ แบบทดสอบนั้นอาจมีความตรงกับวัตถุประสงค์หนึ่ง แต่ไม่ตรงกับอีกวัตถุประสงค์หนึ่ง ดังนั้น การทดสอบจะขาดความตรงหากไม่สามารถวัดสิ่งที่ตั้งใจจะวัดได้ ตัวอย่างเช่น แบบทดสอบสมิททิภาพทางภาษาที่มีความตรงสูง อาจจะไม่มีความตรงเลยหากนำมาวัดหรือประเมินสัมฤทธิ์ผลในห้องเรียนซึ่งมีวัตถุประสงค์เฉพาะสำหรับรายวิชา เพราะแบบทดสอบนั้นไม่ได้ครอบคลุมเนื้อหาที่ระบุอยู่ในหลักสูตร หรือในบทเรียน แต่เป็นแบบทดสอบที่วัดความสามารถในด้านอื่น Hughes (1987) และ Henning (1992) อธิบายความหมายของความตรงเชิงประจักษ์ (Face Validity) ความตรงเชิงเนื้อหา (Content Validity) และความตรงร่วมสมัย (Concurrent Validity) ไว้ดังนี้

1.4.1 ความตรงเชิงประจักษ์ (Face Validity) การทดสอบที่มีความตรงเชิงประจักษ์ คือ การทดสอบที่ปรากฏชัดเจนว่าสามารถวัดสิ่งที่ต้องการจะวัด เช่น หากจะวัดความสามารถในการพูด ก็จะต้องให้ผู้ถูกทดสอบพูดออกเสียง ถ้าหากผู้ถูกทดสอบไม่ได้พูดเลยแต่ใช้วิธีเขียน การทดสอบนี้ก็จะขาดความตรงเชิงประจักษ์ ซึ่งเป็นสิ่งที่สำคัญ หากขาดความตรงเชิงประจักษ์ การทดสอบนั้นก็จะเป็นที่ยอมรับสำหรับผู้สอบ ครู หรือนักการศึกษา ฯลฯ และผู้สอบอาจจะไม่ได้แสดงความสามารถที่แท้จริงออกมา Hughes และ Henning กล่าวว่า มีผู้เชี่ยวชาญด้านการทดสอบหลายคนอธิบายเรื่องความตรงเชิงประจักษ์ว่า สามารถตัดสินได้จากการซักถามผู้เรียนว่าวิธีการทดสอบนั้นเหมาะสมหรือตรงกับสิ่งที่พวกเขาคาดหวังไว้หรือไม่

1.4.2 ความตรงเชิงเนื้อหา (Content Validity) การทดสอบที่มีความตรงเชิงเนื้อหา คือ การทดสอบที่มีเนื้อหาและรายละเอียดตรงกับเนื้อหาในบทเรียน หรือหลักสูตร ซึ่งถูกกำหนดอยู่ในวัตถุประสงค์ของการเรียนการสอน

1.4.3 ความตรงร่วมสมัย (Concurrent Validity) การทดสอบที่มีความตรงร่วมสมัย คือ การทดสอบความสามารถในทักษะใดทักษะหนึ่งโดยใช้แบบทดสอบที่สร้างขึ้น และแบบทดสอบที่มีมาตรฐาน โดยเปรียบเทียบผลที่ได้จากแบบทดสอบที่ต่างกันเหล่านี้ หากผลจากการทดสอบที่ได้มีค่าสหสัมพันธ์สูง ถือว่าแบบทดสอบนั้นมีความตรงร่วมสมัย หากแต่ต้องทำการทดสอบภายในระยะเวลาเดียวกัน หรือใกล้เคียงกันเท่านั้น

1.5 ความเที่ยงของผู้ประเมิน (Rater Reliability)

ความเที่ยงของผู้ประเมิน (Rater Reliability) หมายถึง การประเมินที่มีความเที่ยงคือ ได้ผลคะแนนที่ตรงกัน ยิ่งผลคะแนนใกล้เคียงกันมากเท่าใด ก็ยิ่งมีความเที่ยงมากเท่านั้น ไม่ว่าจะทำการวัด หรือตรวจสอบผลการประเมินกี่ครั้ง แต่คะแนนที่ได้จะต้องไม่มีความแตกต่าง หรือมีน้อยมาก จึงจะถือว่ามีความเที่ยงในการประเมินสูง Hughes (1987) กล่าวว่า ความเที่ยง คือ ความถูกต้อง แม่นยำ คงเส้นคงวา เชื่อถือได้ และยุติธรรมในการวัดผล ความเที่ยงจะถูกประเมินออกมาเป็นตัวเลข เพื่อตรวจสอบความผิดพลาดในการวัด และความเชื่อถือได้ในผลคะแนนที่ได้จากการทดสอบ เขากล่าวว่าการทดสอบที่มีความตรง (Validity) ก็จะมี ความเที่ยง (Reliability) รวมอยู่ด้วย แต่การทดสอบที่มีความเที่ยงอาจไม่มีความตรงก็ได้ ซึ่ง คำกล่าวนี้สอดคล้องกับคำกล่าวของ Bachman

และ Palmer (1982) ที่ว่า ความเที่ยงจะต้องมาก่อนความตรงเสมอ ตามหลักความจริงของการทดสอบ

Huges และ Underhill ได้กล่าวถึงความสัมพันธ์ระหว่างความตรงและความเที่ยงว่า มีความขัดแย้งกันในแง่ที่ว่า ใต้อย่าง ก็จะต้องเสียอย่าง เช่น การทดสอบที่มีความตรงก็อาจจะขาดความเที่ยง หรือในทางกลับกันการทดสอบที่มีความเที่ยงสูง ก็อาจไม่ตรงกับวัตถุประสงค์ของวิชาหรือหลักสูตร หากจะยกตัวอย่าง ก็อาจกล่าวได้ว่า การทดสอบโดยการให้เลือกคำตอบที่ถูกต้องเพียงคำตอบเดียว (Multiple Choice) หรือการจับคู่ (Matching) หรือการทดสอบแบบอื่นที่ใกล้เคียงกัน ถือว่ามีความเที่ยงสูงมาก เพราะไม่ว่าจะทำการตรวจสอบผลคะแนนกี่ครั้งก็จะได้ผลเท่าเดิม แต่การทดสอบแบบนี้จะขาดความตรง (โดยเฉพาะความตรงเชิงประจักษ์) หากนำมาวัดความสามารถในการพูด และในทางกลับกัน การวัดความสามารถในการพูด จะมีความตรงสูงหากให้ผู้สอบพูด แต่อาจจะขาดความเที่ยง เนื่องจากผู้ประเมินตัดสินจากความเห็นของตนเอง ปัญหาเรื่องความเที่ยงจึงเป็นปัญหาที่สำคัญในการประเมินทักษะการพูด ทั้งความเที่ยงภายในของผู้ประเมิน (Intra-rater Reliability) และความเที่ยงของผู้ประเมิน (Inter-rater Reliability)

1.5.1 ความเที่ยงของผู้ประเมิน (Intra-rater Reliability)

ความเที่ยงของผู้ประเมิน (Intra-rater Reliability) คือการที่ผู้ประเมินสามารถประเมินผลคะแนนได้ใกล้เคียงกับการประเมินครั้งก่อน ยิ่งผลคะแนนใกล้เคียงกับผลคะแนนเดิมมากเท่าใด ก็ยิ่งมีความเที่ยงสูง ความแม่นยำ และความคงเส้น คงวาในการประเมินสองครั้ง ในระยะเวลาที่ห่างกันนี้คือ ความเที่ยงของผู้ประเมิน Henning (1992) ชี้ว่าสาเหตุที่ทำให้ผู้ประเมินขาดความเที่ยง อาจเป็นเพราะความเหนื่อยล้า เจตคติ หรือปัจจัยอื่นๆ ที่มีผลต่ออารมณ์ในขณะนั้น ลักษณะนิสัย หรือการที่ผู้ประเมินมีความคุ้นเคยกับผู้ถูกประเมิน ก็อาจมีผลต่อการประเมินได้เช่นกัน

1.5.2 ความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability)

ความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) คือผลการประเมินของผู้ประเมินสองคนขึ้นไป ยิ่งผลการประเมินระหว่างผู้ประเมินมีความแตกต่างกันน้อยเท่าใด ก็ถือว่ามีความเที่ยงระหว่างมากเท่านั้น Hughes (1987) อ้างถึงคำกล่าวของ Lado (1961) ว่าการทดสอบการพูดนั้น ควรมีค่าความเที่ยง (reliability coefficient) ในระดับ

.70 ถึง .79 และหากตัวเลขแสดงความสอดคล้องขึ้นไปถึง .85 ก็นับว่ามีความเที่ยงสูงมาก สำหรับทักษะการพูด แต่นับว่าอยู่ในระดับต่ำหากเป็นการทดสอบการอ่าน (Reading) ซึ่งตัวเลขความเที่ยงควรอยู่ระหว่าง .90 ถึง .99 Mullen (1978) Fulcher, Shohamy (1983) Hugh, Underhill (1987) Weir (1988) Alderson (1991) Henning (1992) Douglas (1994) กล่าวว่าโดยทั่วไปยิ่งมีจำนวนผู้ประเมินมากเท่าใด ความเที่ยงระหว่างของผู้ประเมินก็จะเพิ่มมากขึ้น ดังนั้นจึงควรให้มีผู้ประเมินมากกว่าหนึ่ง

1.6 การอบรมผู้ประเมิน (Rater Training)

Underhill (1987) กล่าวว่า ปัญหาหลักของการประเมินผลโดยครูผู้สอน คือ ความเชื่อถือได้ของผลการประเมิน และยังมีจำนวนผู้ประเมินในหลักสูตรหนึ่งมาก เท่าใด ปัญหาด้านความเที่ยงและเชื่อถือได้ของผลการประเมินก็จะยิ่งมากขึ้น ดังนั้น จึงควรต้องมีการฝึกอบรมผู้ประเมิน (Rater training) เพื่อลดความแตกต่าง หรือความไม่สอดคล้องกันของผู้ประเมิน ประเด็นการอบรมผู้ประเมินนี้ Hughes (1987) และ Fulcher (1993) มีความเห็นที่สอดคล้องกับ Underhill แต่ Fulcher เห็นว่า การฝึกอบรมผู้ประเมิน เป็นการยอมรับโดยปริยายว่าเกณฑ์นั้นมีความตรง (Validity) และจะทำให้การตรวจสอบความตรงของเกณฑ์เป็นเรื่องยาก ดังนั้นจึงจำเป็นต้องสร้างและตรวจสอบความตรงของเกณฑ์ก่อนที่จะอบรมผู้ประเมิน และไม่ควรฝึกอบรม ผู้ประเมินในขณะที่เกณฑ์และกระบวนการทดสอบยังอยู่ในขั้นตอนที่กำลังพัฒนาหรือตรวจสอบความตรง แต่ควรอบรมผู้ประเมินเพื่อใช้เกณฑ์ หลังจากได้ตรวจสอบความตรงแล้ว และควรใช้เกณฑ์นั้นหลังการอบรมทันที

Alderson (1991) เรียกกระบวนการอบรมผู้ประเมินซึ่งมีผลทำให้เกิดทรรณะคติที่ตรงกันเกี่ยวกับภาษาพูดว่าเป็นการ “ถ่ายสำเนามนุษย์” (Cloning)

อย่างไรก็ดีการอบรมผู้ประเมินนี้ดูจะเป็นสิ่งสำคัญในการวัดผลการพูด (Oral Testing) ตามที่ Fulcher ได้อ้างถึงผู้สนับสนุนความคิดนี้ เช่น Wilds (1975) Lowe (1987) Adams (1980) Ingram (1985) และ Barnwell (1989)

นอกจากนี้ Lumley & McNamara (1995) ได้กล่าวถึงผลการอบรม ผู้ประเมินว่า ไม่สามารถทำให้ความเข้มงวดของผู้ประเมินแต่ละคนหมดไปได้ ความแตกต่างในการประเมินจะยังคงมีอยู่ และการอบรมแต่ละครั้งนั้น ผลของมันจะไม่ยาวนานนัก จึงจำเป็นต้องมีการอบรมซ้ำก่อนการประเมินทุกครั้ง

2. งานวิจัยที่เกี่ยวข้อง

2.1 ผลการศึกษาความตรง (Validity) และความเที่ยง (Reliability) ของการทดสอบสมรรถภาพทางการพูด (Oral Proficiency Test)

2.1.1 ผลการศึกษาความตรง (Validity)

2.1.1.1 Fulcher (1993) ได้กล่าวถึงคำกล่าวของ Lowe (1987) ว่ามีการศึกษาเรื่องความตรง และความเที่ยงของการวัดสมรรถภาพทางการพูดน้อยมาก และเขาได้ทำการศึกษาเกี่ยวกับความตรงร่วมสมัย (Concurrent Validity) ของมาตรวัดประเมินค่าของ FCE (the University of Cambridge Local Examinations Syndicate's First Certificate in English) เปรียบเทียบกับมาตรวัดประเมินค่าของ ELTS (English Language Testing Services) โดยทำการทดสอบทักษะการพูดของนักศึกษาจำนวน 47 คน และประเมินโดยมาตรวัดทั้งสองแบบ ผลการศึกษาปรากฏว่า ยังไม่มีหลักฐานเพียงพอที่จะสรุปได้ว่า มาตรวัดประเมินค่าที่เขาทำการศึกษานั้นมีความตรงร่วมสมัย เขากล่าวว่าการศึกษาความตรงร่วมสมัยของมาตรวัดประเมินค่าที่ใช้วัดความสามารถในการพูด ยังมีความซับซ้อนในประเด็นและปัจจัยต่างๆ ซึ่งอาจมีผลกระทบกับการให้คะแนน ซึ่งประเด็นและปัจจัยต่างๆ เหล่านี้ยังไม่เป็นที่เข้าใจถ่องแท้

2.1.1.2 Fulcher (1993) ยังได้กล่าวถึงผลการศึกษาของ Levine, Haus และ Cort (1987) ในเรื่องความตรงร่วมสมัยของมาตรวัดประเมินค่าของ ACTFL ที่ใช้ประเมินผู้เรียนภาษาฝรั่งเศส และสเปน โดยเปรียบเทียบผลการประเมินโดยครู ผู้สอน (โดยใช้มาตรวัดของ ACTFL) กับผลการประเมินโดย ACTFL ปรากฏว่าผลของการประเมินโดยครู ผู้สอนแตกต่างกันอย่างมีนัยสำคัญจากผลที่ได้จากการประเมินโดย ACTFL โดยมีความต่างอยู่ที่ $P < .01$ จากผลที่ได้นี้ Levine และคณะสรุปว่าครูผู้ประเมินควรได้รับการฝึกฝนในการประเมินโดยใช้มาตรวัดของ ACTFL และจากผลการศึกษาของ Levine และคณะ Fulcher สรุปว่ายังไม่มีหลักฐานแสดงว่ามีความตรงร่วมสมัยในการวัดความสามารถในการพูด จากการประเมินโดยครูผู้สอน เปรียบเทียบกับการประเมินโดยแบบสอบมาตรฐาน เช่น ของ ACTFL

2.1.2 ผลการศึกษาความเที่ยง (Reliability)

2.1.2.1 Fulcher (1993) กล่าวถึงงานวิจัยที่ได้รับการเผยแพร่เริ่มปรากฏในช่วงปี 1978 และได้อ้างถึง Adams (1978) เกี่ยวกับผลการศึกษา ความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) ว่ามีความเที่ยงระหว่างผู้ประเมินสูงมากถึง .87 หรือสูงกว่า เช่นเดียวกับ Barnwell (1987) ซึ่งสรุปว่าผู้ประเมินที่ไม่ได้รับการฝึกอบรมสามารถมีความสอดคล้องกันในระดับที่ยอมรับได้ ผู้ที่ไม่เห็นด้วยกับ Adams และ Barnwell คือ Liskin-Gasparro (1984) และ Harrison (1982)

นอกจากนี้ Fulcher ยังได้กล่าวถึงผู้ที่ศึกษาเกี่ยวกับประเด็นเดียวกันคือ Mullen (1978) Clark (1975) และ Clark and Swinton (1980) ซึ่งมีความเห็นว่าความเที่ยงระหว่างของผู้ประเมิน (Inter-rater Reliability) จะสูงมากเมื่อมีผู้ตัดสิน 2 คน แต่ Fulcher ไม่เห็นด้วยกับ Clark and Swinton เนื่องจากผลการศึกษา ปรากฏว่า ค่าสหสัมพันธ์ต่ำในทุกหัวข้อ ยกเว้นไวยากรณ์ซึ่งมีค่าสหสัมพันธ์สูง

2.1.2.2 Hendricks et. al. (1980) สรุปผลการศึกษาความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) โดยใช้กระบวนการตามแบบ FSI ผลปรากฏว่าค่าสหสัมพันธ์อยู่ในระดับที่ยอมรับได้ ยกเว้นหัวข้อ การออกเสียง (accent) (Fulcher 1993)

2.1.2.3 Fulcher กล่าวถึงผลการศึกษา ความเที่ยงระหว่างของผู้ประเมิน (Inter-rater Reliability) ของ Lowe and Liskin-Gasparro (1986) และ Shohamy (1983) พบว่ามีความสอดคล้องกันสูงมาก

2.1.2.4 Morrison and Lee (1985) รายงานการทดสอบการพูด ณ Hong Kong University โดยถ่ายวีดิทัศน์และให้เจ้าของภาษา และผู้ที่ไม่ใช่เจ้าของภาษาทำการประเมินโดยใช้เกณฑ์แบบ 5 ช่วงคะแนน (5 – band scale) ในหัวข้อต่อไปนี้ : สมรรถภาพทั่วไป (General proficiency) ไวยากรณ์ (Grammar) การออกเสียง (Pronunciation) ความคล่อง (Fluency) ความสามารถในการสื่อสาร (Communicative ability) และความสามารถเชิงวิชาการ (Academic potential) ผลปรากฏว่าความสอดคล้องกันของผู้ประเมินอยู่ในระดับ .85 โดยไม่ปรากฏว่ามีความแตกต่างในการประเมินระหว่างเจ้าของภาษา และผู้ที่ไม่ใช่เจ้าของภาษา (Fulcher (1993))

2.1.2.5 Jafarpur (1985) ทำการศึกษาความเที่ยงระหว่างของผู้ประเมิน (Inter-rater Reliability) โดยใช้เกณฑ์ตามแบบ FSI โดยมีผู้ตัดสิน 3 คน ทำการประเมิน นักศึกษา 58 คน พบว่าตัวเลขความสอดคล้องกันของผู้ประเมินอยู่ที่ .58 และ .65 Jafarpur สรุปว่าผู้ตัดสินทั้ง 3 คนนั้น มีความแตกต่างกันอย่างมาก ซึ่ง ผลลัพธ์นี้ขัดแย้งกับผลของ Morrison and Lee

2.2 งานวิจัยที่เกี่ยวกับการประเมินการพูดแบบนำเสนอ (Oral Presentations)

2.2.1 Stephen Hall (1996) กล่าวถึงความยากในการประเมินทักษะที่ไม่ได้เกี่ยวกับการใช้คำพูดอื่นๆ เช่น อวจนภาษา และทักษะรอง (Non-verbal skills และ sub-skills) ในการประเมินการพูดแบบนำเสนอโดยเขาได้ทำการวิจัยเกี่ยวกับการประเมินทักษะการพูดแบบนำเสนอ ในภาคพื้นเอเชียตะวันออกเฉียงใต้ Hall กล่าวว่า การพูดแบบนำเสนอต่อหน้าผู้ฟังเป็นครั้งแรกของผู้เรียนนั้น ไม่อาจแสดงถึงทักษะและความสามารถที่แท้จริงของผู้เรียนได้ เนื่องจากผู้เรียนจะเกิดอาการตื่นตกใจ (Stage fright) ไม่ว่าจะพูดเป็นภาษาของตนเอง หรือพูดภาษาที่สองก็ตาม ดังนั้น จึงควรมีการฝึกฝน และให้ข้อมูลป้อนกลับ (feedback) เพื่อเป็นประโยชน์แก่ผู้เรียน เพื่อ ผู้เรียนจะได้พัฒนาทักษะในการสื่อสารด้วยภาษาทำทางและอื่นๆ นอกเหนือจากภาษาพูด Hall ยังได้กล่าวถึงแบบประเมินการพูดแบบนำเสนอของ Wood (1982) ว่าเขาได้ใช้มาตราส่วนประเมินค่าแบบช่วงคะแนน 1-5 โดยพิจารณาเกณฑ์ดังต่อไปนี้ :

- บทนำ (Introduction) การพัฒนาเนื้อเรื่องและบทสรุป (Development and conclusion)
- ข้อมูลและแหล่งที่มา (Data and sources)
- โครงสร้าง (Organization)
- เนื้อหาข้อมูลและการตอบสนอง (Information and response)
- ประสิทธิภาพของภาษาพูดและภาษาทำทาง (Vocal and non-verbal effectiveness)
- ลีลา (style)

นอกจากนี้ Hall ได้กล่าวถึงเกณฑ์ซึ่งเขากำหนดขึ้นจากประสบการณ์ และจากการปรับแต่งเครื่องมือการใช้ข้อมูลป้อนกลับของ Mendelsohn ในปี ค.ศ. 1992 เกณฑ์ดังกล่าวคือ : เนื้อหา (Content) โครงสร้าง (Organization) ภาษาทำทาง หรืออวจนภาษา

(Non-verbal) สื่อทัศนูปกรณ์ (Visual aids) ความคล่องในการพูด (Fluency) การออกเสียง (Pronunciation) คำศัพท์ และไวยากรณ์ (Vocabulary and grammar)

2.2.2. Laina Ho (1997) ได้ทำการวิจัยเรื่องการประเมินทักษะการพูดแบบนำเสนอของนักศึกษาวิชาวิทยาศาสตร์คอมพิวเตอร์ (Computer Science) ในปี 1995 ณ มหาวิทยาลัยแห่งชาติ ประเทศสิงคโปร์ (National University of Singapore) ซึ่งเป็นส่วนหนึ่งของวิชา “Business and Technical Communication” และการพูดแบบนำเสนอซึ่งเป็นแกนสำคัญของรายวิชานี้มีส่วนคล้ายกับวิชา Advanced Business Oral Communication วิธีประเมินผลก็คล้ายกัน ตัวอย่างเช่น ให้นักศึกษาพูดแบบนำเสนอคนละ 5-8 นาที แต่ต่างกันตรงที่ให้นักศึกษาพูดในหัวข้ออะไรก็ได้โดยไม่กำหนด แต่จะต้องมีการวิเคราะห์เกี่ยวกับหัวข้อที่เลือกนั้น โดยไม่ต้องมีการค้นคว้า ข้อมูล หรือเตรียมสื่อทัศนูปกรณ์เพื่อประกอบการนำเสนอ (Visual aids) วัตถุประสงค์คือเพื่อศึกษาว่าการให้ผู้เรียนนำเสนอถึง 3 ครั้ง จะสามารถพัฒนาทักษะของผู้เรียนได้หรือไม่ และเพื่อปรับปรุงแก้ไขงานที่มอบหมายให้แก่ผู้เรียน ดังนั้นงานวิจัยของ Ho จึงมุ่งที่พัฒนาการของผู้เรียน มิใช่ความเที่ยงตรงของผู้ประเมินหรือวิธีการประเมินผล แต่ทั้งนี้ประเด็นหรือหัวข้อในการประเมินก็มีความคล้ายคลึงกัน เช่น ประเด็นด้านเนื้อหา (Content) การนำเสนอ (Delivery) ซึ่งรวมถึงการใช้สายตา (Eye Contact) ท่าทาง (Gesture / Posture) และการใช้เสียง (Voice) ส่วนในด้านภาษา (Language) โครงสร้าง (Organization) และการใช้สื่อทัศนูปกรณ์ (Visual aids) Ho มิได้กล่าวถึงเนื่องจากภาษาอังกฤษเป็นภาษาที่ผู้เรียนในมหาวิทยาลัยในสิงคโปร์ใช้ได้ อย่างคล่องแคล่วอยู่แล้ว แต่จะมีการประเมินด้านการออกเสียง (Enunciation) ตาม มาตรฐานของชาวสิงคโปร์

ผลการศึกษารายงานว่า การให้ผู้เรียนนำเสนอถึง 3 ครั้งนั้น ไม่ได้ช่วยทำให้ผู้เรียนส่วนใหญ่ (69%) พัฒนาทักษะการพูดได้ และมีการแนะนำในตอนท้ายว่าควรให้มีการเน้นการใช้อุปกรณ์ด้านสื่อประกอบ และการเตรียมสื่อทัศนูปกรณ์ที่มีความชัดเจน และมีประสิทธิภาพ

2.2.3 Janet Lim Hong Eng (1997) ได้ศึกษาเกี่ยวกับข้อดีของการประเมินการพูดแบบนำเสนอโดยให้เพื่อนในชั้นเป็นผู้ประเมิน เพื่อนำมาเป็นประโยชน์ในการเรียนการสอน โดยใช้ชั้นศึกษาปีที่ 3 คณะวิทยาศาสตร์ มหาวิทยาลัยแห่งชาติประเทศสิงคโปร์

ในระหว่างปี ค.ศ. 1992 –1994 ในฐานะครูและผู้ประเมิน Eng พบว่าผู้เรียนจะสนใจผล การประเมินในด้านไวยากรณ์ (Grammar) และการออกเสียง (Pronunciation) เท่านั้น ส่วนองค์ประกอบอื่นๆ เช่น หัวข้อ (Topics) การใช้เสียง (Voices) กิริยาท่าทาง (Gestures / Manner) เนื้อหาหรือรายละเอียด (Content / Information) การใช้สื่อประกอบ (Visual aids) ผู้เรียนไม่ให้ความสำคัญนัก Eng เชื่อว่าการประเมินของตนทำให้ผู้เรียนไม่มี พัฒนาการในประเด็นดังกล่าว และได้ กล่าวถึง Gronlund (1976) ซึ่งเสนอวิธีการแก้ปัญหา ที่เกิดขึ้นโดยให้เพื่อนในชั้น ร่วมประเมินด้วยเพื่อช่วยเสริมผลการประเมินของครู Eng ได้ทำการศึกษาว่า การให้เพื่อนร่วมชั้นประเมินด้วยนั้น จะทำให้ผู้เรียนมีความสนใจใน ประเด็นอื่นๆ นอกจากไวยากรณ์ และการออกเสียงหรือไม่ ผลคือการทำเพื่อนร่วมชั้น ประเมินการพูดแบบนำเสนอ สามารถแสดงจุดแข็งและจุดอ่อนของผู้เรียนได้ แต่ความ สำเร็จนั้นต้องขึ้นอยู่กับข้อกำหนดวิธีการในการประเมิน เพื่อที่นักศึกษาจะได้เข้าใจชัดเจน ว่าจะต้องประเมินอย่างไร และต้องอธิบายแก่ผู้ที่ตนประเมินได้ว่าทำไมจึงประเมินเช่นนั้น การประเมินด้วยวิธีดังกล่าวจึงจะประสบผลที่น่าพอใจ

จากผลการศึกษาความตรง และความเที่ยงในงานวิจัยต่างๆ ที่กล่าวมา ข้างต้น จะเห็นว่ามีความเห็นที่ขัดแย้งและสอดคล้องในประเด็นต่างๆ เช่น ด้านความตรงและ ความเที่ยงของการประเมิน ด้านความสำคัญหรือความจำเป็นในการ อบรมผู้ประเมิน และเป็นที่น่าสนใจว่าผลการศึกษาความตรงร่วมสมัยทั้งของ Fulcher และของ Levine และ คณะได้ผลที่สอดคล้องกันกับผลการวิจัยในฉบับนี้ในเรื่องความตรงร่วมสมัยดังจะได้กล่าวถึง ในบทที่ 4

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

วิธีดำเนินการวิจัย (Research Methods)

1. กลุ่มประชากรและพลวิจัย (Population and Subjects) แบ่งเป็น 2 กลุ่มใหญ่ ได้แก่ ประชากร และพลวิจัย

1.1 ประชากรมี 2 กลุ่ม ได้แก่

1.1.1 อาจารย์ชาวไทยและชาวต่างประเทศผู้สอนและประเมินวิชาการสนทนาภาษาอังกฤษชั้นสูงจำนวน 16 คน และผู้เชี่ยวชาญ 1 คน รวมเป็น 17 คน

1.1.2 นิสิตคณะพาณิชยศาสตร์และการบัญชีชั้นปีที่ 3 จำนวน 573 คน (ได้แก่นิสิตทั้งหมด)

1.2 พลวิจัย (Subject) มี 3 กลุ่ม ได้แก่

1.2.1 อาจารย์ชาวไทย และชาวต่างประเทศ ผู้สอนและประเมินวิชาการสนทนาภาษาอังกฤษชั้นสูง จำนวน 16 คน

1.2.2 นิสิตคณะพาณิชยศาสตร์และการบัญชีชั้นปีที่ 3 จำนวน 254 คน จำนวนของพลวิจัยกลุ่มนี้ได้มาจากการคำนวณตามสูตรของ Yamane (1967) เมื่อความเชื่อถือได้ 95% และความคลาดเคลื่อนไม่เกิน 5% และจากการสุ่มตัวอย่างแบบ Stratified Random Sampling

1.2.3 นิสิตคณะพาณิชยศาสตร์และการบัญชีชั้นปีที่ 3 จำนวน 30 คน (คิดเป็นร้อยละ 5.24 ของนิสิตทั้งหมด) พลวิจัยกลุ่มนี้ได้มาจากการสุ่มตัวอย่างแบบ Stratified Random Sampling

2. เครื่องมือวิจัย (Research Instruments) เครื่องมือวิจัย ประกอบด้วย

2.1 แบบสอบถาม (Questionnaires) จำนวน 2 ชุด คือ

2.1.1 แบบสอบถามสำหรับนิสิต เพื่อตรวจสอบความตรงเชิงประจักษ์ (Face Validity) 254 ชุด (ภาคผนวก จ)

2.1.2 แบบสอบถามสำหรับอาจารย์ผู้สอนและประเมิน เพื่อตรวจสอบความตรงเชิงประจักษ์ (Face Validity) และความตรงเชิงเนื้อหา (Content Validity) 16 ชุด (ภาคผนวก ฉ)

2.2 วิธีทัศนัยบันทึกการพูดแบบนำเสนอของกลุ่มตัวอย่าง คือนิสิตคณะ พาณิชยศาสตร์และการบัญชีชั้นปีที่ 3 จำนวน 30 คน คนละ 5 นาทีโดยประมาณ เพื่อตรวจสอบความเที่ยงของผู้ประเมิน (Rater Reliability)

2.3 ข้อสอบข้อเขียน จำนวน 2 ชุด เพื่อตรวจสอบความตรงเชิงทฤษฎี ข้อสอบทั้ง 2 ชุด ประกอบด้วย

2.3.1 ข้อสอบรายวิชาการสนทนาภาษาอังกฤษชั้นสูงภาคปลายปีการศึกษา 2542 (ภาคผนวก ข) ซึ่งเป็นข้อสอบอัตนัยความยาว 30 นาที

2.3.2 ข้อสอบ CU-TEP จุฬาลงกรณ์มหาวิทยาลัย

3. การสร้างและพัฒนามาตรวัดประเมินค่า (Development of Rating Scales)

เป็นเวลากว่า 40 ปี ที่มีการใช้มาตรวัดประเมินค่า (Rating Scales) ซึ่งมีรูปแบบเป็นช่วงคะแนน (Band Scores) เพื่อใช้ในการประเมินทักษะการพูดเพื่อให้การประเมินมีความถูกต้อง เที่ยงตรง และแม่นยำ มาตรวัดประเมินค่านี้ต้องมีความเหมาะสม ถูกต้อง ครบคลุมเนื้อหา ชัดเจน และสะดวกต่อการใช้ จึงมีการสร้างและพัฒนามาตรวัดนี้ ตามกระบวนการดังต่อไปนี้

3.1 สร้างมาตรวัดประเมินค่า โดยใช้วิธีกำหนดช่วงคะแนนแบบ Band Scores เน้นเนื้อหาที่ต้องการวัด และครอบคลุมประเด็นที่สำคัญ ตามวัตถุประสงค์ของรายวิชา โดยมีรายละเอียดดังต่อไปนี้

เพื่อให้สอดคล้องกับวัตถุประสงค์ของรายวิชา รายละเอียดการกำหนดเกณฑ์ และการสร้างมาตรวัดประเมินค่าเพื่อใช้สำหรับการทดสอบสัมฤทธิ์ผลรายวิชาการสนทนาภาษาอังกฤษชั้นสูงจึงประกอบด้วย 5 หัวข้อหลักคือ

3.1.1 โครงสร้าง (Organization)

3.1.2 เนื้อหา (Content)

3.1.3 การใช้ภาษา (Language)

3.1.4 การนำเสนอ (Delivery)

3.1.5 การใช้สื่อทัศนูปกรณ์ (Visual aids)

ประเด็นต่าง ๆ ในแต่ละหัวข้อมีดังนี้ (โปรดดูตารางที่ 1 ประกอบ)

3.1.1 โครงสร้าง (Organization) ประกอบด้วย

- บทนำ (Introduction)
 - : หัวข้อ (Topic)
 - : วัตถุประสงค์ (Purpose)

- : ระยะเวลาในการนำเสนอ (Duration)
- : ลำดับของประเด็นที่จะนำเสนอ (Outline of main parts)
- : โอกาสในการซักถาม (Questions)
- เนื้อเรื่อง (Body)
 - : รายละเอียดของประเด็นหลัก (Main parts)
- บทจบ (Ending)
 - : เรื่องย่อ (Summary)
 - : บทสรุป / คำแนะนำ (Conclusion / recommendation)
 - : ช่วงตอบคำถาม (Questions)
 - : การจบการนำเสนอ (Closing)

3.1.2 เนื้อหา (Content)

- ความเชื่อมโยงของเนื้อหา (Transitions)
- ความเกี่ยวเนื่องสัมพันธ์กันของรายละเอียดต่างๆ รวมทั้ง ตัวอย่างที่น่าสนใจ (Coherence)

3.1.3 การใช้ภาษา (Language)

- เหมาะสม ชัดเจน เข้าใจง่าย และใช้ภาษาได้อย่างคล่องแคล่ว (Appropriate, clear, simple and fluent)
- เป็นภาษาพูดที่ฟังดูเป็นธรรมชาติ (Spoken language)
- คำศัพท์ที่ใช้ (Vocabulary)
- ไวยากรณ์ (Grammar)
- การออกเสียง (Pronunciation)

3.1.4 การนำเสนอ (Delivery)

- การสบตาผู้ฟัง (Eye contact)
- การใช้เสียง (Voice)
- ความสามารถในการถ่ายทอดความคิดออกมาเป็นคำพูด (Articulation)
- การใช้ท่าทางประกอบ (Manner and gestures)

3.1.5 การใช้สื่อทัศนูปกรณ์ (Visual aids)

- การออกแบบ (Design)
- ประโยชน์และวิธีการใช้ (Use and handling)
- ความต่อเนื่องของเนื้อหาและรายละเอียด (Relevance)



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 1 แสดงเกณฑ์ในมาตราส่วนประเมินค่า: หัวข้อและประเด็นในแต่ละหัวข้อ

Assessment Areas

Organisation	Content	Language	Delivery	Visual Aids
<ul style="list-style-type: none"> - Introduction <ul style="list-style-type: none"> : Topic : Purpose : Duration of talk : Outline of main parts : Questions - Body <ul style="list-style-type: none"> : Main parts - Ending <ul style="list-style-type: none"> : Summary : Conclusion/ recommendation : Questions : Closing 	<ul style="list-style-type: none"> - Coherence (Details and examples are relevant, interesting and coherent) 	<ul style="list-style-type: none"> - Appropriate, clear and fluent - Natural spoken language - Vocabulary - Grammar - Pronunciation 	<ul style="list-style-type: none"> - Eye contact - Voice - Articulation - Manner and gestures 	<ul style="list-style-type: none"> - Design - Use / Handling - Relevance

ในการประเมินผล มีการแบ่งระดับคะแนนออกเป็นช่วงๆ หรือที่เรียกกันว่า Band Scores มีทั้งสิ้น 5 ระดับ (5-point scale) รายละเอียดของเกณฑ์ในแต่ละช่วงคะแนนมีดังนี้ (โปรดดูตารางที่ 2 ประกอบ)

3.1.1 โครงสร้าง (Organization)

ช่วงคะแนนที่ 5

ครอบคลุมทุกประเด็นเกี่ยวกับโครงสร้าง ทั้งประเด็นหลัก และประเด็นย่อย มีการเชื่อมโยงข้อมูล และรายละเอียดอย่างชัดเจน

ช่วงคะแนนที่ 4

ครอบคลุมทุกประเด็นหลัก และมีโครงสร้างที่ชัดเจน มีการเชื่อมโยงรายละเอียด

ช่วงคะแนนที่ 3

ประเด็นหลักไม่ชัดเจน เนื่องจากขาดการเชื่อมโยงที่ดี

ช่วงคะแนนที่ 2

โครงสร้างของการนำเสนอไม่ชัดเจนเนื่องจากขาดส่วนประกอบสำคัญเช่น บทนำ หรือเนื้อเรื่อง หรือบทสรุป

ช่วงคะแนนที่ 1

ขาดโครงสร้างของการนำเสนอ และขาดความเชื่อมโยงในประเด็นต่าง

3.1.2 เนื้อหา (Content)

ช่วงคะแนนที่ 5

เนื้อหาที่น่าสนใจมากและรายละเอียดทุกส่วนมีความเกี่ยวเนื่องสัมพันธ์กัน และมีการใช้คำเชื่อม(Transitions) อย่างถูกต้อง สมบูรณ์ ตัวอย่างและรายละเอียดต่างๆ มีความเชื่อมโยง และให้ภาพที่ชัดเจน

ช่วงคะแนนที่ 4

เนื้อหาที่น่าสนใจ ข้อมูลมีความเชื่อมโยงกัน และมีการใช้คำเชื่อมที่เหมาะสม ตัวอย่างและรายละเอียดมีเนื้อหาที่สัมพันธ์กัน

ช่วงคะแนนที่ 3

เนื้อหาไม่ค่อยน่าสนใจ มีการใช้คำเชื่อมบ้างเป็นครั้งคราว ตัวอย่างและรายละเอียดต่างๆ ต่อเนื่องกัน

ช่วงคะแนนที่ 2

มีการใช้คำเชื่อมต่างๆ น้อยมาก ตัวอย่างและรายละเอียดไม่เกี่ยวเนื่องสัมพันธ์กัน และไม่น่าสนใจ

ช่วงคะแนนที่ 1

ไม่มีการใช้คำเชื่อม รายละเอียดต่างๆ ไม่เกี่ยวเนื่องสัมพันธ์กัน ไม่มีการยกตัวอย่าง หรือให้ข้อมูลใดๆ เป็นการนำเสนอที่ขาดความน่า สนใจ

3.1.3 การใช้ภาษา (Language)

ช่วงคะแนนที่ 5

ใช้ภาษาพูดที่เป็นธรรมชาติ และชัดเจนมาก ถูกต้องตามหลักไวยากรณ์ การออกเสียงถูกต้องชัดเจน (แต่ไม่จำเป็นต้องพูดด้วยสำเนียงเหมือนเจ้าของภาษาโดยไม่ผิดเพี้ยน)

ช่วงคะแนนที่ 4

ใช้ภาษาพูดที่ชัดเจน เข้าใจง่าย และฟังดูเป็นธรรมชาติ อาจมีข้อผิดพลาดเล็กน้อยด้านไวยากรณ์ การออกเสียงถูกต้องและชัดเจน

ช่วงคะแนนที่ 3

พูดโดยอาศัยการท่องจำ มีข้อผิดพลาดด้านไวยากรณ์ และการออกเสียง ซึ่งบางครั้งทำให้เกิดความเข้าใจที่คลาดเคลื่อน

ช่วงคะแนนที่ 2

พูดโดยการท่องจำ มีข้อผิดพลาดมากด้านไวยากรณ์ และการออกเสียง ทำให้เป็นอุปสรรคมากในการทำความเข้าใจ

ช่วงคะแนนที่ 1

พูดตะกุกตะกัก ผิดหลักไวยากรณ์ และการออกเสียงไม่ถูกต้อง ทำให้เกิดความไม่เข้าใจ

3.1.4 การนำเสนอ (Delivery)

ช่วงคะแนนที่ 5

การใช้เสียงชัดเจนมาก มีการมองผู้ฟังในห้องอย่างทั่วถึง มีการใช้ท่าทางประกอบที่ดีมาก

ช่วงคะแนนที่ 4

พูดชัดชัดเจน มองผู้ฟังค่อนข้างจะทั่วถึง ใช้ท่าทางประกอบ ได้ดี

ช่วงคะแนนที่ 3

มีการใช้ท่าทางประกอบบ้าง และมองผู้ฟังเป็นครั้งคราว แต่การใช้เสียงขาดความชัดเจน

ช่วงคะแนนที่ 2

ใช้ท่าทางประกอบน้อยมาก พูดไม่ชัดเจนและยากต่อการเข้าใจ มีการสบตาผู้ฟังน้อยมาก

ช่วงคะแนนที่ 1

ไม่มีการมองสบตาผู้ฟัง พูดไม่ชัดเจนและเบา

3.1.5 การใช้สื่อทัศนูปกรณ์ (Visual aids)

ช่วงคะแนนที่ 5

สื่อประกอบที่ใช้มีความชัดเจนมาก ง่ายต่อการเข้าใจ ใช้สื่อประกอบการนำเสนอได้อย่างมีประสิทธิภาพมาก

ช่วงคะแนนที่ 4

สื่อประกอบชัดเจน และง่ายต่อการเข้าใจ ใช้สื่อประกอบการนำเสนอได้อย่างมีประสิทธิภาพ

ช่วงคะแนนที่ 3

สื่อประกอบชัดเจนพอควร แต่อาจมีข้อผิดพลาด หรือมีรายละเอียดมากเกินไป วิธีการใช้สื่อไม่ค่อยมีประสิทธิภาพนัก

ช่วงคะแนนที่ 2

สื่อประกอบไม่ชัดเจน และไม่สนับสนุนสิ่งที่กำลังนำเสนอ มีข้อผิดพลาดมาก และ/หรือ เข้าใจยาก วิธีการใช้สื่อค่อนข้างขาดประสิทธิภาพ

ช่วงคะแนนที่ 1

ไม่สามารถเข้าใจสื่อประกอบที่ใช้ วิธีการใช้สื่อขาดประสิทธิภาพ เช่น ผู้พูดยืนบังจอ OHP เป็นต้น สื่อที่ใช้ไม่เกิดประโยชน์ต่อการนำเสนอ

ตารางที่ 2 แสดงรายละเอียดของเกณฑ์ในแต่ละช่วงคะแนน

5 = excellent; 4 = good; 3 = moderate; 2 = poor; 1 = very poor

Rate	Organisation	Content	Language	Delivery	Visual Aids
5	A very well organized presentation with all the components clearly stated with the effective use of linking words to connect supporting details in each part.	Very interesting talk with substantial details and examples which are not only relevant but also highly illustrative and coherent.	Very clear, simple and natural spoken language. Free of grammatical errors. Very good pronunciation (no need to sound like a native speaker though).	Maintains very good eye contact with audience in all areas of the room. Loud, clear voice and very articulate. Effective use of gestures and body language to support the ideas expressed in the presentation.	Very clear & creative visuals. Very effective handling without distraction so that visuals fully support the presentation.
4	A clear structure of the presentation. The three main components are clearly stated except for the linking of some supporting details.	Examples and details are coherent. Interesting talk with necessary details and examples which are relevant and interesting.	Clear, simple and natural use of spoken language. A few unimportant grammatical errors. Good pronunciation.	Maintains good eye contact with almost all of the audience. Articulate with loud, clear voice. Uses gestures and body language to support the ideas expressed.	Clear and simple visuals. Effective handling visuals (perhaps with minor distractions.) Visuals support the talk.

Rate	Organisation	Content	Language	Delivery	Visual Aids
3	Ideas are vaguely organised into three main parts: Introduction body and conclusion due to ineffective use of linking words.	Average presentation. Details and examples are not quite relevant or interesting.	Language is unnatural (sounds like a rehearsed text). Noticeable grammatical and pronunciation errors, some of which lead to miscomprehension.	Only occasional eye contact and use of gestures. Articulation and voice lack clarity.	Reasonably clear visuals but wordy and/or contain some mistakes. Handling visuals with some distractions.
2	Unclear structure due to lack of crucial components; for example, the introduction, body or conclusion.	Uninteresting presentation with insufficient details and examples. Details and examples are not relevant or coherent.	A memorized text. Pronunciation is hardly comprehensible . Lots of grammatical mistakes which seriously <u>impair</u> comprehension.	Little use of gestures Utterances frequently unclear and quite incomprehensible at times.	Unclear visuals which hardly support the talk. A number of mistakes. Ineffective handling with many distractions.
1	Badly structured with many necessary components mixed up or left out. No organization of ideas, making it difficult for listeners to identify the topic and the main points of the presentation.	Very uninteresting presentation. Very few details and examples provided.	Jerky. Full of grammatical errors. Pronunciation is incomprehensible.	No eye contact. Very inarticulate and incomprehensible.	Visuals are too cramped or incomprehensible. Very poor handling. Visuals do not support the talk.

3.2 อาจารย์ผู้สอนวิชานี้ทั้งหมดจำนวน 16 คน พิจารณาเกณฑ์ที่กำหนดขึ้น ใน มาตรฐานประเมินค่า และให้ข้อมูลป้อนกลับเพื่อปรับปรุง ครั้งที่ 1 (ภาคผนวก ง)

3.3 เชิญผู้เชี่ยวชาญคือ Professor Alan Maley ตรวจสอบมาตรฐานประเมินค่าเพื่อปรับปรุงครั้งที่ 2 (ภาคผนวก ง)

3.4 นำมาตรฐานประเมินค่าที่ปรับปรุงแล้ว ให้อาจารย์ผู้สอน และผู้เชี่ยวชาญตรวจสอบอีกครั้ง และให้ข้อมูลป้อนกลับ โดยการกรอกแบบประเมินความเที่ยง ระหว่างมาตรฐานประเมินค่าและคุณลักษณะที่ต้องการจะวัด เพื่อทำการปรับปรุงครั้งสุดท้าย

4. การรวบรวมข้อมูล (Data Collection) มีการรวบรวมข้อมูลจากพลวิจัย จำนวน 3 กลุ่ม โดยมีรายละเอียดของขั้นตอนและวิธีการดังนี้

4.1 พลวิจัย ซึ่งประกอบด้วยอาจารย์ผู้สอนวิชาการสนทนาภาษาอังกฤษชั้นสูง 16 คน และพลวิจัยที่เป็นนิสิต 254 คน ตอบแบบสอบถามเพื่อหาค่าความตรงเชิงประจักษ์ของวิธีการวัดสัมฤทธิ์ผลของผู้เรียนวิชาการสนทนาภาษาอังกฤษชั้นสูงโดยวิธีการพูดแบบนำเสนอ และใช้เกณฑ์การให้คะแนนตามมาตรฐานประเมินค่า (Rating Scales) ที่สร้างขึ้นเพื่อตรวจสอบความตรงเชิงประจักษ์ เมื่อวันที่ 15-19 กุมภาพันธ์ 2543

4.2 พลวิจัยซึ่งเป็นอาจารย์ผู้สอน 16 คน (กลุ่มเดียวกับในข้อ 4.1) ตอบแบบสอบถาม เพื่อตรวจสอบความตรงเชิงเนื้อหาของมาตรฐานประเมินค่า ระหว่างวันที่ 22-26 กุมภาพันธ์ 2543

4.3 พลวิจัย ที่เป็นนิสิตคณะพาณิชยศาสตร์และการบัญชี ชั้นปีที่ 3 จำนวน 30 คน สอบข้อเขียน 2 ชุด เพื่อตรวจสอบความตรงร่วมสมัย

ข้อสอบชุดที่ 1 คือ ข้อสอบปลายภาค วิชาการสนทนาภาษาอังกฤษ ชั้นสูง สอบเมื่อวันที่ 22 กุมภาพันธ์ 2543

ข้อสอบชุดที่ 2 คือ CU-TEP สอบระหว่างวันที่ 14-21 มีนาคม 2543

4.4 พลวิจัยที่เป็นนิสิตคณะพาณิชยศาสตร์ และการบัญชี ชั้นปีที่ 3 จำนวน 30 คน (ซึ่งเป็นกลุ่มเดียวกับพลวิจัยในข้อ 4.3) ทำการสอบโดยการพูดแบบนำเสนอในหัวข้อที่เกี่ยวกับธุรกิจความยาวประมาณคนละ 5 นาที โดยบันทึกเทปวีดิทัศน์ในช่วงระหว่างวันที่ 14-21 มีนาคม 2543

4.5 พลวิชัยที่เป็นอาจารย์ผู้ประเมินชาวไทย 2 คน และชาวต่างประเทศ 2 คน ทำการทดสอบสัมฤทธิ์ผลของนิสิตที่เป็นพลวิชัย 30 คน โดยประเมินจากวีดิทัศน์และใช้เกณฑ์ตามมาตรวัดประเมินค่า เพื่อตรวจสอบความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) การประเมินรอบแรกใช้เวลาประมาณ 1 เดือน โดยผู้ประเมินทั้ง 4 คน นำเทปกลับไปดูนับจากวันที่ 22 มีนาคม 2543 ถึง 25 เมษายน 2543

4.6 พลวิชัยซึ่งเป็นอาจารย์ผู้ประเมินชาวไทย และชาวต่างประเทศใน 4.5 ทำการประเมินสัมฤทธิ์ผลนิสิตกลุ่มตัวอย่างเดิมจากวีดิทัศน์ซ้ำอีกครั้งหลังการประเมินครั้งแรกประมาณ 1 เดือน คือในช่วงระหว่างวันที่ 23 พฤษภาคม 2543 ถึง 23 มิถุนายน 2543 เพื่อตรวจสอบความเที่ยงของผู้ประเมิน (Intra-rater Reliability)

5. การวิเคราะห์ข้อมูล (Data Analyses)

5.1 การตรวจสอบความตรงเชิงประจักษ์ (Face Validity) ของการวัดผล Oral presentations โดยการพูดแบบนำเสนอ โดยวิธีประมาณค่าจากมาตรวัดแบบ Likert จากแบบสอบถามของนิสิต 254 คน และของอาจารย์ผู้ประเมิน 16 คน

5.2 การตรวจสอบความตรงเชิงเนื้อหา (Content Validity) ของการวัดผลการพูดแบบนำเสนอโดยอาศัยเกณฑ์จากมาตรวัด (Rating Scales) ที่สร้างขึ้น โดยวิธีหาค่าแบบ Item-objective Congruence Index จากแบบสอบถามของอาจารย์ จำนวน 16 ชุด

5.3 การตรวจสอบความตรงร่วมสมัย (Concurrent Validity) โดยวิธีหาค่า Spearman Correlation Coefficients จากวิธีทดสอบ 3 แบบ คือ

5.3.1 การพูดแบบนำเสนอ (Oral Presentations)

5.3.2 ข้อสอบ CU-TEP

5.3.3 ข้อสอบปลายภาค วิชาการสนทนาภาษาอังกฤษชั้นสูง

5.4 การตรวจสอบความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) และความเที่ยงของผู้ประเมิน (Intra-rater Reliability) โดยวิธีหาค่า Spearman Correlation Coefficients และ Cronbach's Alpha Coefficients

บทที่ 4

ผลการวิจัย (Research Results)

เนื่องจากวัตถุประสงค์หลักสองประการของงานวิจัยนี้คือ

1. สร้างและพัฒนามาตรวัดประเมินค่าเพื่อทดสอบสัมฤทธิ์ผลของการพูดแบบนำเสนอให้มีความตรง (Validity) ในด้านความตรงเชิงประจักษ์ (Face Validity) ความตรงเชิงเนื้อหา (Content Validity) และความตรงร่วมสมัย (Concurrent Validity) และ
 2. ตรวจสอบผลของการนำมาตรวัดดังกล่าวไปใช้ในสถานการณ์จริงในด้านความเที่ยงของผู้ประเมิน (Rater Reliability) ทั้งความเที่ยงของผู้ประเมิน (Intra-rater Reliability) และความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) จึงได้มีการตรวจสอบใน 2 ประเด็นหลักคือ 1. ความตรง (Validity) ของเกณฑ์ประเมิน และ 2. ความเที่ยงของผู้ประเมิน (Rater Reliability) เมื่อนำเกณฑ์ไปใช้ในสถานการณ์จริงโดยมีผลดังต่อไปนี้
1. ผลการตรวจสอบความตรง (Validity) ในสามประเด็นคือ 1. ความตรงเชิงประจักษ์ (Face Validity) 2. ความตรงเชิงเนื้อหา (Content Validity) และ 3. ความตรงร่วมสมัย (Concurrent Validity)
- 1.1 ความตรงเชิงประจักษ์ (Face validity)

จากการตรวจสอบความตรงเชิงประจักษ์จากผู้ที่ใช้เกณฑ์คือ นิสิตผู้ถูกประเมินจำนวน 254 คน และอาจารย์ผู้ประเมินผลจำนวน 16 คน โดยวิธี Likert Scale Rating ในประเด็นต่างๆ ดังต่อไปนี้ 1. ความตรงเชิงประจักษ์ของวิธีในการประเมินตามความเห็นของนิสิต 2. ความพอใจของนิสิตต่อวิธีการประเมิน 3. ความตรงเชิงประจักษ์ของเกณฑ์ประเมินในทรรศนะของนิสิต 4. ความตรงเชิงประจักษ์ของวิธีการประเมินตามความเห็นของผู้ประเมิน 5. ความตรงเชิงประจักษ์ของเกณฑ์ประเมินในทรรศนะของผู้ประเมิน ผลคือ

1.1.1 ความตรงเชิงประจักษ์ของวิธีการประเมินตามความเห็นของนิสิต

นิสิตมีความเห็นว่า วิธีประเมินผลโดยให้นิสิตพูดแบบนำเสนอ มีความตรงเชิงประจักษ์สูง โดยมีค่าเฉลี่ยที่ 4.186 ดังตารางที่ 3

ตารางที่ 3 ผลการตรวจสอบความตรงเชิงประจักษ์ของวิธีการประเมินจากนิสิตโดยวิธีประมาณค่าจากมาตรวัดแบบ Likert Scale Rating

Mean	S.D.	Min	Max	N
4.186	.724	2	5	254

1.1.2 ความพอใจของนิสิตต่อวิธีการประเมิน

นอกจากนิสิตจะมีความเห็นว่าวิธีประเมินผลดังกล่าวมีความตรงเชิงประจักษ์สูงแล้ว ในด้านความพอใจของผู้ถูกประเมิน ผลปรากฏว่านิสิตพอใจวิธีประเมินในลักษณะนี้สูงเช่นกัน โดยมีค่าเฉลี่ยที่ 4.031 ดังในตารางที่ 4

ตารางที่ 4 ผลการตรวจสอบความพอใจของนิสิตต่อวิธีประเมินผล โดยวิธี Likert Scale Rating

Mean	S.D.	Min	Max	N
4.031	.743	2	5	254

1.1.3 ความตรงเชิงประจักษ์ของเกณฑ์ประเมินในทรรศนะของนิสิต

ด้านความตรงเชิงประจักษ์ของเกณฑ์ประเมิน นิสิตเห็นว่าเกณฑ์มีความตรงเชิงประจักษ์สูงเช่นกัน โดยมีค่าเฉลี่ยที่ 4.157 ดังในตารางที่ 5

ตารางที่ 5 ผลการตรวจสอบความตรงเชิงประจักษ์ของเกณฑ์ประเมินจากนิสิตโดย
วิธีประมาณค่าจากมาตรวัดแบบ Likert Scale Rating

Mean	S.D.	Min	Max	N
4.157	.688	2	5	254

1.1.4 ความตรงเชิงประจักษ์ของวิธีการประเมินตามความเห็นของผู้
ประเมิน

ผลการตรวจสอบความตรงเชิงประจักษ์ ในการวัดสัมฤทธิ์ผลของการพูด
แบบนำเสนอจากแบบสอบถามของผู้ประเมินจำนวน 16 ชุด ผลคือ ผู้ประเมินมีความเห็น
ว่า การประเมินผลโดยวิธีนี้มีความตรงเชิงประจักษ์สูงมากกว่าการสอบข้อเขียน โดยมีค่า
เฉลี่ยที่ 4.88 ดังรายละเอียดในตารางที่ 6

ตารางที่ 6 ผลการตรวจสอบความตรงเชิงประจักษ์ของวิธีการประเมินจากผู้ประเมิน
โดยวิธีประมาณค่าจากมาตรวัดแบบ Likert Scale Rating

Mean	S.D.	Min	Max	N
4.88	.34	4	5	16

1.1.5 ความตรงเชิงประจักษ์ของเกณฑ์ประเมินในทรรศนะของผู้ประเมิน

สำหรับเกณฑ์การประเมินซึ่งประกอบด้วย 5 หัวข้อ ผู้ประเมินเห็นว่ามีความ
ตรงเชิงประจักษ์สูงมาก โดยมีค่าเฉลี่ยในแต่ละหัวข้ออยู่ระหว่าง 4.12 – 4.69 ดังราย
ละเอียดในตารางที่ 7

ตารางที่ 7 ผลการตรวจสอบความตรงเชิงประจักษ์ของเกณฑ์ประเมินจากผู้ประเมิน โดยวิธีประมาณค่าจากมาตรวัดแบบ Likert Scale Rating

หัวข้อ	Mean	S.D.	Min	Max	N
1. โครงสร้าง	4.44	.73	3	5	16
2. เนื้อหา	4.31	.87	3	5	16
3. การใช้ภาษา	4.69	.48	4	5	16
4. การนำเสนอ	4.56	.63	3	5	16
5. การใช้สื่อประกอบ	4.12	1.09	1	5	16

1.2 ความตรงเชิงเนื้อหา (Content Validity)

จากการตรวจสอบความตรงเชิงเนื้อหาของแบบประเมินสัมฤทธิ์ผลของการพูดแบบนำเสนอจากแบบสอบถามของอาจารย์ผู้สอนและประเมิน 16 คน โดยวิธี Item – Objective Congruence Index (โดยแสดงค่า 1 = เห็นด้วย 0 = ไม่แน่ใจ -1 = ไม่เห็นด้วย) ในประเด็นต่อไปนี้ คือ 1. ตรงตามเนื้อหาของหลักสูตรและ วัตถุประสงค์รายวิชา 2. เหมาะสมสำหรับรายวิชา 3. ภาษาที่ใช้บรรยายเกณฑ์ (descriptors) เหมาะสมและครอบคลุมเนื้อหา 4. สามารถนำเกณฑ์ประเมินไปใช้ได้อย่างสะดวกและง่ายต่อการใช้ผลปรากฏว่า

1.2.1 ตรงตามเนื้อหาของหลักสูตรและวัตถุประสงค์รายวิชา

ผู้ประเมินมีความเห็นว่า เกณฑ์ที่สร้างขึ้นมีความตรงตามเนื้อหาของ หลักสูตร และวัตถุประสงค์ของรายวิชา ดังตัวเลขแสดงค่าเฉลี่ยในตารางที่ 8 ซึ่งแสดงรายละเอียดของแต่ละหัวข้อ คือ โครงสร้าง (Organization) เนื้อหา (Content) การใช้ภาษา (Language) การนำเสนอ (Delivery) และการใช้สื่อทัศนูปกรณ์ (Visual aids) โดยเฉพาะด้านโครงสร้างค่าเฉลี่ยสูงมากถึง .94

ตารางที่ 8 ผลการตรวจสอบความตรงเชิงเนื้อหาตามหลักสูตรและวัตถุประสงค์รายวิชา โดยวิธีหาค่าแบบ Item – Objective Congruence Index

หัวข้อ	Mean	S.D.	Min	Max	N
โครงสร้าง (Organization)	.94	.25	0	1	16
เนื้อหา (Content)	.87	.34	0	1	16
การใช้ภาษา (Language)	.81	.40	0	1	16
การนำเสนอ (Delivery)	.87	.34	0	1	16
การใช้สื่อประกอบ (Visual aids)	.87	.34	0	1	16

1.2.2 ความเหมาะสมสำหรับรายวิชา

ผู้ประเมินเห็นว่าเกณฑ์นี้มีความเหมาะสมสำหรับรายวิชาสูง โดยได้ค่าเฉลี่ย .87 ดังตารางที่ 9

ตารางที่ 9 ผลการตรวจสอบความเหมาะสมสำหรับรายวิชา โดยวิธีหาค่าแบบ Item – Objective Congruence Index

Mean	S.D.	Min	Max	N
.87	.34	0	1	16

1.2.3 ภาษาที่ใช้บรรยายเกณฑ์ (Descriptors) มีความเหมาะสม และครอบคลุมเนื้อหา

ผู้ประเมินมีความเห็นว่าภาษาที่ใช้บรรยายเกณฑ์ มีความเหมาะสมและครอบคลุมเนื้อหาในระดับสูง ดังตารางที่ 10 ซึ่งแสดงรายละเอียดของแต่ละช่วงคะแนน (band) ตั้งแต่ 1-5 ในแต่ละหัวข้อคือ โครงสร้าง เนื้อหา การใช้ภาษา การนำเสนอ และการใช้สื่อประกอบ โดยเฉพาะหัวข้อ เนื้อหา จะได้ค่าเฉลี่ยสูงมากถึง .94 ในช่วงคะแนนที่ 1 2 4 และ 5 และ .88 ในช่วงคะแนนที่ 3

ตารางที่ 10 ผลการตรวจสอบความเหมาะสมและครอบคลุมเนื้อหาของภาษาที่ใช้
บรรยายเกณฑ์ (descriptors) โดยวิธีหาค่าแบบ Item – Objective
Congruence Index

	Mean	S.D.	Min	Max	N
ด้านโครงสร้าง (Organization)					
Band 1	.94	.25	0	1	16
Band 2	.87	.34	0	1	16
Band 3	.87	.40	0	1	16
Band 4	.87	.34	0	1	16
Band 5	.87	.34	0	1	16
	Mean	S.D.	Min	Max	N
ด้านเนื้อหา (Content)					
Band 1	.94	.25	0	1	16
Band 2	.94	.25	0	1	16
Band 3	.88	.34	0	1	16
Band 4	.94	.25	0	1	16
Band 5	.94	.25	0	1	16
	Mean	S.D.	Min	Max	N
ด้านการใช้ภาษา (Language)					
Band 1	.81	.25	0	1	16
Band 2	.87	.34	0	1	16
Band 3	.87	.34	0	1	16
Band 4	.87	.34	0	1	16
Band 5	.87	.34	0	1	16

	Mean	S.D.	Min	Max	N
ด้านการนำเสนอ (Delivery)					
Band 1	.94	.25	0	1	16
Band 2	.94	.25	0	1	16
Band 3	.87	.34	0	1	16
Band 4	.87	.34	0	1	16
Band 5	.94	.25	0	1	16
	Mean	S.D.	Min	Max	N
ด้านการใช้สื่อประกอบ (Visual aids)					
Band 1	.87	.34	0	1	16
Band 2	.87	.34	0	1	16
Band 3	.81	.40	0	1	16
Band 4	.87	.34	0	1	16
Band 5	.87	.34	0	1	16

1.2.4 ความสะดวกและง่ายต่อการใช้

ผู้ประเมินมีความเห็นว่า เกณฑ์ที่สร้างขึ้นมีความสะดวกและง่ายต่อการใช้ใน ระดับสูงปานกลาง คือ มีค่าเฉลี่ยอยู่ที่ .69 ดังในตารางที่ 11

ตารางที่ 11 ผลการตรวจสอบความสะดวกและง่ายต่อการใช้เกณฑ์ โดยวิธีหาค่าแบบ Item – Objective Congruence Index

Mean	S.D.	Min	Max	N
.69	.48	0	1	16

1.3 ความตรงร่วมสมัย (Concurrent Validity)

มีการสำรวจความตรงร่วมสมัยของมาตรวัดประเมินค่าเมื่อนำไปทดสอบ สมรรถิ ผลการพูดแบบนำเสนอ โดยการศึกษาผลการวัดคุณลักษณะโดยการใช้เกณฑ์ที่กำหนดใน มาตราส่วนประเมินค่าที่สร้างขึ้น คุณลักษณะที่ต้องการวัดคือ 1. คุณลักษณะ ด้านการใช้ภาษา ในเชิงการพูดแบบนำเสนอ (Language) และ 2. คุณลักษณะด้าน โครงสร้าง (Organization) โดยใช้การทดสอบสามแบบ และศึกษาค่าสหสัมพันธ์ของผลการ ทดสอบทั้งสามแบบซึ่งประกอบด้วย การพูดแบบนำเสนอ (Oral presentations) ข้อสอบ CU-TEP และข้อสอบปลายภาครายวิชาการสนทนาภาษาอังกฤษชั้นสูง โดยข้อสอบ CU- TEP มีลักษณะเป็น multiple-choice และข้อสอบปลายภาครายวิชาการสนทนาภาษา ภาษาอังกฤษชั้นสูงมีลักษณะเป็นข้อสอบอัตนัยโดยให้ผลิตเขียนสิ่งที่ต้องการนำเสนอ

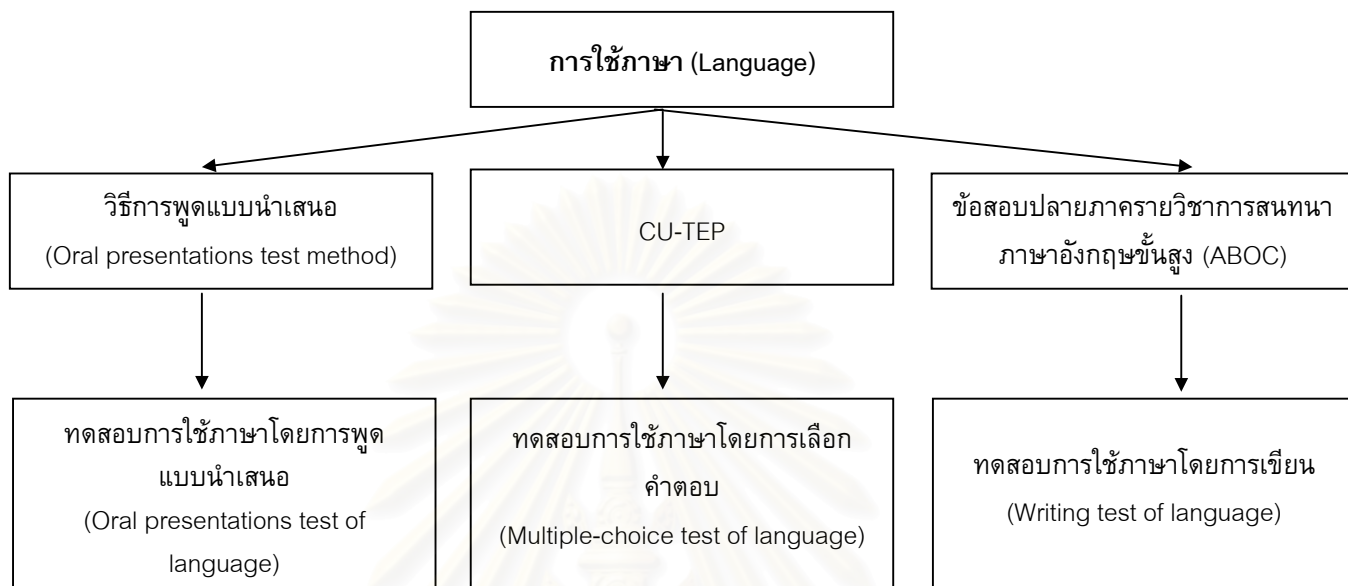
จากการทดสอบทั้งสามแบบนี้ สามารถวัดผลความสามารถด้านการใช้ภาษา (Language) (โปรดดูแผนภาพที่ 1 ประกอบ) ส่วนความสามารถด้านโครงสร้าง (Organization) จะวัดได้จากการทดสอบโดยการพูดแบบนำเสนอ (Oral presentations) และจากข้อสอบปลายภาครายวิชาการสนทนาภาษาอังกฤษชั้นสูง (โปรดดูแผนภาพที่ 2 ประกอบ) ส่วนข้อสอบ CU-TEP นั้น ไม่สามารถวัดความรู้ด้านโครงสร้าง (Organization) ได้

แผนภาพที่ 1 แสดงการวัดคุณลักษณะด้านการใช้ภาษา (Language) ซึ่ง สามารถวัดได้โดยวิธีการวัดสามวิธีคือ การพูดแบบนำเสนอ (Oral presentations) ข้อ สอบ CU-TEP และข้อสอบปลายภาครายวิชาการสนทนาภาษาอังกฤษชั้นสูง (ABOC)

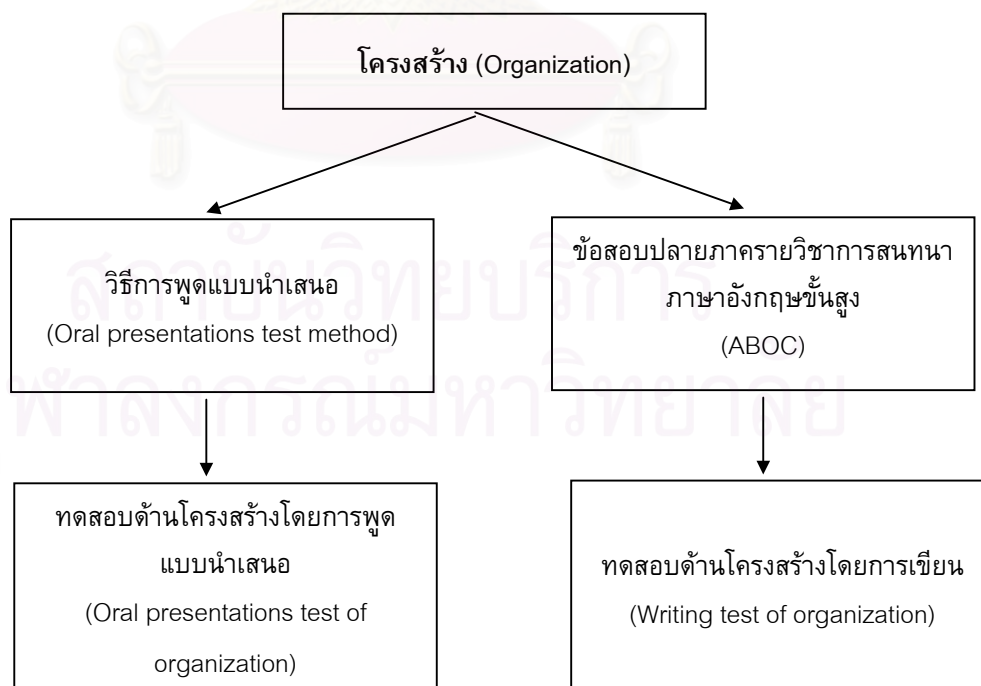
และแผนภาพที่ 2 แสดงการวัดคุณลักษณะด้านโครงสร้าง (Organization) ซึ่ง ทำได้โดยการพูดแบบนำเสนอ (Oral presentations) และข้อสอบปลายภาครายวิชาการ สนทนาภาษาอังกฤษชั้นสูง (ABOC)

จุฬาลงกรณ์มหาวิทยาลัย

แผนภาพที่ 1 วิธีการวัดคุณลักษณะด้านการใช้ภาษาโดยวิธีทดสอบ 3 วิธี



แผนภาพที่ 2 วิธีการวัดคุณลักษณะด้านโครงสร้างโดยวิธีทดสอบ 2 วิธี



1.3.1 จากการวัดคุณลักษณะด้านการใช้ภาษาโดยใช้แบบทดสอบข้อเขียนสองชุดคือ CU-TEP และข้อสอบ ABOC ค่าสหสัมพันธ์อยู่ในระดับสูงคือ .5250 ในขณะที่ค่าสัมประสิทธิ์สหสัมพันธ์ของการทดสอบโดยการพูดแบบนำเสนอและจากการสอบ CU-TEP อยู่ในระดับต่ำคือ .3638 และค่าสัมประสิทธิ์สหสัมพันธ์ของการทดสอบโดยการพูดแบบนำเสนอและข้อสอบ ABOC อยู่ในระดับต่ำมากคือ .1089 ดังในตารางที่ 12 ดังนั้นจึงไม่อาจกล่าวได้ว่าการวัดคุณลักษณะด้านการใช้ภาษาโดยวิธีทดสอบการพูด (OP) และการเขียน (CU-TEP และ ABOC) มีหลักฐานชี้ให้เห็นว่ามีความตรงร่วมสมัย

ตารางที่ 12 ผลการตรวจสอบความตรงร่วมสมัยในด้านคุณลักษณะในการใช้ภาษา (Language) โดยวิธีหาค่า Spearman Correlation Coefficients

	OP	CU-TEP
CU-TEP	.3638 (p = .048)	
ABOC	.1089 (p = .567)	.5250 (p = .003)

1.3.2 การวัดคุณลักษณะด้านโครงสร้างโดยวิธีทดสอบสองวิธีคือการพูดและการเขียนมีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ในระดับต่ำมากคือ .1089 จากตารางที่ 13 ซึ่งชี้ให้เห็นว่าวิธีทดสอบโดยการพูด มิได้สัมพันธ์กับวิธีทดสอบโดยการเขียน ดังนั้นจึงอาจกล่าวได้ว่า การวัดความสามารถของผู้เรียนในด้านทักษะการพูดแบบนำเสนอ โดยวิธีสอบข้อเขียนไม่มีความตรงร่วมสมัย หรืออีกนัยหนึ่งคือ ความสามารถในการพูดของนิสิต ไม่อาจวัดได้โดยวิธีสอบข้อเขียน

ตารางที่ 13 ผลการตรวจสอบความตรงร่วมสมัยในด้านคุณลักษณะด้านโครงสร้าง (Organization) โดยวิธีหาค่า Spearman Correlation Coefficients

	OP
ABOC	.2295
	(p = .223)

หมายเหตุ	OP	=	การพูดแบบนำเสนอ
	CU-TEP	=	ข้อสอบ CU-TEP
	ABOC	=	ข้อสอบปลายภาครายวิชาการสนทนาภาษาอังกฤษ ชั้นสูง

2. ผลการตรวจสอบความเที่ยงของผู้ประเมิน (Rater Reliability) ใน 2 ประเด็นคือ ความเที่ยงของผู้ประเมิน (Intra – rater Reliability) และความเที่ยงระหว่าง ผู้ประเมิน (Inter – rater Reliability)

2.1 ความเที่ยงของผู้ประเมิน (Intra-rater Reliability)

ได้มีการตรวจสอบความเที่ยงของผู้ประเมิน ซึ่งทำโดยการให้ผู้ประเมินจำนวน 4 คน ประกอบด้วยชาวไทย 2 คน และชาวต่างประเทศ 2 คน (รายละเอียดโปรดดูภาคผนวก ข) ทำการประเมินสัมฤทธิ์ผลของนิสิตที่เป็นพลวิชัย 30 คน จาก วิดีทัศน์โดยใช้เกณฑ์ตามมาตราส่วนประเมินค่าที่สร้างขึ้น ผู้ประเมินทั้ง 4 ทำการประเมิน 2 ครั้ง โดยระยะเวลาประเมินแต่ละครั้งห่างกัน 1 เดือน จากการวิเคราะห์โดยวิธีหาค่า Cronbach's Alpha Scale ปรากฏผลตรวจสอบความเที่ยงของผู้ประเมินใน 2 ลักษณะคือ

- ความเที่ยงของผู้ประเมินโดยรวมทั้ง 5 หัวข้อ และ
- ความเที่ยงของผู้ประเมินในแต่ละหัวข้อ

2.1.1 ความเที่ยงของผู้ประเมินโดยรวมทั้ง 5 หัวข้อ คือ โครงสร้าง (Organization) เนื้อหา (Content) การใช้ภาษา (Language) การนำเสนอ (Delivery) และการใช้สื่อทัศนูปกรณ์ (Visual aids) ปรากฏผลดังในตารางที่ 14

ตารางที่ 14 ผลการตรวจสอบความเที่ยงของผู้ประเมิน (Intra-rater Reliability) โดยรวม 5 หัวข้อ โดยวิธีหาค่า Cronbach's Alpha Scale

	ผู้ประเมิน 1	ผู้ประเมิน 2	ผู้ประเมิน 3	ผู้ประเมิน 4
ครั้งที่ 1	.9126	.9340	.9434	.6264
ครั้งที่ 2	.8932	.8797	.9490	.7712
ความแตกต่าง	(.0194)	(.0543)	(.0056)	(.1448)

จากตาราง จะเห็นว่าผู้ประเมินที่มีความเที่ยงสูงสุดคือ ผู้ประเมิน 3 ซึ่งความแตกต่างระหว่างการประเมินครั้งที่ 1 และ 2 อยู่ที่ .0056 หรือต่างกันเพียง 0.59%

ผู้ประเมินซึ่งมีความเที่ยงสูงเป็นอันดับที่สองคือ ผู้ประเมิน 1 โดยมีความแตกต่างในการประเมิน 2 ครั้งอยู่ที่ .0194 หรือ 2.13%

ผู้ประเมินซึ่งมีความเที่ยงอยู่ในอันดับสามคือ ผู้ประเมิน 2 โดยมีผลความแตกต่างอยู่ที่ .0543 หรือ 5.81%

ผู้ประเมินที่มีความเที่ยงอันดับสี่คือ ผู้ประเมิน 4 โดยมีผลความแตกต่างในการประเมินอยู่ที่ .1448 หรือ 23.12%

จากตารางอาจสรุปได้ว่าผู้ประเมินแต่ละคนนั้นมีความเที่ยงในการประเมินต่างกัน ตัวอย่างเช่น ผู้ประเมิน 3 เป็นผู้มีความเที่ยงในการประเมินสูงมาก ในขณะที่ผู้ประเมิน 4 มีความเที่ยงในการประเมินน้อยมาก

2.1.2 ความเที่ยงของผู้ประเมินในแต่ละหัวข้อมีรายละเอียดดังในตาราง ที่ 15

ตารางที่ 15 ผลการตรวจสอบความเที่ยงของผู้ประเมิน (Intra-rater Reliability) ในแต่ละหัวข้อโดยวิธี Cronbach's Alpha Scale

	ผู้ประเมิน 1	(p=)	ผู้ประเมิน 2	(p=)	ผู้ประเมิน 3	(p=)	ผู้ประเมิน 4	(p=)
โครงสร้าง	.5175	.003	.0900	.636	.8981	.000	.2397	.202
เนื้อหา	.5218	.003	.4493	.013	.8736	.000	.4415	.015
การใช้ ภาษา	.5583	.001	.7073	.000	.8840	.000	.0992	.602
การนำ เสนอ	.3528	.056	.7281	.000	.8336	.000	.4355	.016
การใช้สื่อ ทัศนูปกร ณ์	.1330	.484	.2818	.131	.8613	.000	.4980	.005

จากตาราง จะเห็นรายละเอียดของการประเมินในแต่ละหัวข้อ ซึ่งแสดงให้เห็นถึงจุดอ่อนของผู้ประเมินในการประเมินหัวข้อต่างๆ ดังต่อไปนี้

2.1.2.1 โครงสร้าง (Organization)

ผู้ประเมิน 2 มีความเที่ยงในการประเมินโครงสร้างต่ำมากโดยมีค่าสหสัมพันธ์อยู่ที่ .0900 และผู้ประเมิน 4 อยู่ในระดับต่ำ คือ .2397

2.1.2.2 เนื้อหา (Content)

ผู้ประเมิน 1 2 และ 4 มีความเที่ยงในการประเมินเนื้อหาอยู่ในระดับปานกลาง คือ .5218 .4493 และ .4415

2.1.2.3 การใช้ภาษา (Language)

ผู้ประเมิน 4 มีความเที่ยงในการประเมินการใช้ภาษาในระดับต่ำมากโดยมีค่าสหสัมพันธ์อยู่ที่ .0992

2.1.2.4 การนำเสนอ (Delivery)

ผู้ประเมิน 1 มีความเที่ยงในการประเมินการนำเสนอในระดับค่อนข้างต่ำ โดยมีค่าสหสัมพันธ์ของความสอดคล้องอยู่ที่ .3528

2.1.2.5 การใช้สื่อทัศนูปกรณ์ (Visual aids)

ผู้ประเมิน 1 และ 2 มีความเที่ยงในการประเมินการใช้สื่อทัศนูปกรณ์ ในระดับต่ำ คือ .1330 และ .2818

สรุปคือผู้ประเมิน 1 มีความเที่ยงในการประเมินด้านการใช้สื่อทัศนูปกรณ์ต่ำที่สุดจากการประเมินทั้ง 5 หัวข้อ และผู้ประเมิน 2 จะมีความเที่ยงต่ำที่สุดในหัวข้อโครงสร้าง สำหรับผู้ประเมิน 3 จะมีความเที่ยงสูงมากในทุกหัวข้อ และผู้ประเมิน 4 จะมีความเที่ยงในการประเมินด้านการใช้ภาษาต่ำที่สุด จากผลที่ได้นี้ผู้ประเมินสามารถนำไปปรับปรุงการประเมินในครั้งต่อไป

2.2 ความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability)

ได้มีการตรวจสอบความเที่ยงระหว่างผู้ประเมิน โดยนำผลการทดสอบ สัมฤทธิ์ผลของนิสิตจากวิชิตทัศน์ ซึ่งเป็นเครื่องมือวิจัยเดียวกับ 2.1 ได้ทำการสำรวจความเที่ยงระหว่างผู้ประเมินทั้ง 4 โดยการจับคู่ รวมทั้งสิ้นผู้ประเมิน 6 คู่ คือ

คู่ที่ 1	ประกอบด้วย	ผู้ประเมิน 1 +	ผู้ประเมิน 2
คู่ที่ 2	ประกอบด้วย	ผู้ประเมิน 1 +	ผู้ประเมิน 3
คู่ที่ 3	ประกอบด้วย	ผู้ประเมิน 1 +	ผู้ประเมิน 4
คู่ที่ 4	ประกอบด้วย	ผู้ประเมิน 2 +	ผู้ประเมิน 3
คู่ที่ 5	ประกอบด้วย	ผู้ประเมิน 2 +	ผู้ประเมิน 4
คู่ที่ 6	ประกอบด้วย	ผู้ประเมิน 3 +	ผู้ประเมิน 4

จากการวิเคราะห์โดยวิธีหาค่า Spearman Correlation Coefficients ผลที่ได้ผู้วิจัยนำเสนอใน 2 ลักษณะคือ ความเที่ยงระหว่างผู้ประเมิน โดยรวมทั้ง 5 หัวข้อ และความเที่ยงระหว่างผู้ประเมินในแต่ละหัวข้อ

2.2.1 ความเที่ยงระหว่างผู้ประเมินโดยรวมทั้ง 5 หัวข้อคือ โครงสร้าง (Organization) เนื้อหา (Content) การใช้ภาษา (Language) การนำเสนอ (Delivery) และ การใช้สื่อทัศนูปกรณ์ (Visual aids) มีรายละเอียดดังในตารางที่ 16

ตารางที่ 16 ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) โดยรวม 5 หัวข้อ โดยวิธีหาค่า Spearman Correlation Coefficients

	ครั้งที่ 1		ครั้งที่ 2	
		(Sig.)		(Sig.)
คู่ที่ 1	.3893	.033	.4738	.008
คู่ที่ 2	.3058	.100	.6969	.000
คู่ที่ 3	-.0955	.622	.6307	.000
คู่ที่ 4	.5381	.002	.4301	.018
คู่ที่ 5	.1865	.333	.0371	.846
คู่ที่ 6	.3698	.048	.5312	.003

จากตาราง จะเห็นว่าผลการประเมินของแต่ละคู่ระหว่างครั้งที่ 1 และครั้งที่ 2 แตกต่างกัน แต่เป็นที่น่าสังเกตว่า คู่ที่ผลการประเมินระหว่างครั้งที่ 1 และครั้งที่ 2 ต่างกันมากที่สุดคือ คู่ที่ 3 จากการประเมิน 2 ครั้ง มีผลการประเมินที่ต่างกันอยู่ถึง 0.7262 โดยการประเมินครั้งที่ 1 มีความเที่ยงระหว่างผู้ประเมินน้อยที่สุดคือ $-.0955$ ในขณะที่การประเมินครั้งที่ 2 มีความเที่ยงระหว่างผู้ประเมินสูงมากคือ $.6307$ ในขณะที่คู่ที่ 1, 2, 4, 5 และ 6 ต่างกัน $.0845$ $.3911$ $.1080$ $.1494$ และ $.1614$ ตามลำดับ

เมื่อพิจารณาในแต่ละคู่จะพบว่า

คู่ที่ 1 มีความเที่ยงระหว่างผู้ประเมินในการประเมินทั้ง 2 รอบ ในระดับค่อนข้างสูง คือ $.3893$ (Sig. = $.033$) และ $.4738$ (Sig. = $.008$)

คู่ที่ 2 ความเที่ยงระหว่างผู้ประเมินในการประเมินครั้งแรก ไม่สูงนักอยู่ที่ $.3058$ (Sig. $.100$) ในขณะที่การประเมินครั้งที่สองมีความเที่ยงสูงมากถึง $.6969$ (Sig. $.000$)

คู่ที่ 3 เป็นคู่ที่มีความแตกต่างในการประเมินระหว่างรอบแรก และรอบสองมากที่สุด ในรอบแรกค่าสัมประสิทธิ์สหสัมพันธ์อยู่ที่ $-.0955$ (Sig. = $.622$) ซึ่งถือว่ามีความเที่ยงระหว่างผู้ประเมินในระดับต่ำมาก แต่การประเมินรอบสองกลับให้ผลตรงกันข้ามคือค่าสัมประสิทธิ์สหสัมพันธ์ในรอบที่สองสูงมากอยู่ที่ $.6307$ (Sig. = $.000$)

คู่ที่ 4 ผลการประเมินทั้งสองรอบ มีความเที่ยงระหว่างผู้ประเมินสูง คือ $.5381$ (Sig. = $.002$) และ $.4301$ (Sig. = $.018$) ตามลำดับ

คู่ที่ 5 ผลการประเมินทั้งสองรอบ มีค่าสัมประสิทธิ์สหสัมพันธ์ต่ำ คือ $.1865$ (Sig. = $.333$) และ $.0371$ (Sig. = $.846$)

คู่ที่ 6 ค่าสัมประสิทธิ์สหสัมพันธ์ของผลการประเมินรอบแรกค่อนข้างสูง คือ $.3698$ (Sig. $.048$) ในรอบที่สองผลการประเมินมีค่าสัมประสิทธิ์สหสัมพันธ์สูงคือ $.5312$ (Sig. $.003$)

จากผลที่ได้สรุปได้ว่าความเที่ยงระหว่างผู้ประเมินไม่คงที่อีกทั้งความแตกต่างของความเที่ยงระหว่างผู้ประเมินอาจมีมากถึงขนาดติดลบในการประเมินครั้งแรก แต่กลับสูงมากในการประเมินครั้งที่ 2

2.2.2 ความเที่ยงระหว่างผู้ประเมินในแต่ละหัวข้อ คือ 1. โครงสร้าง (Organization) 2. เนื้อหา (Content) 3. การใช้ภาษา (Language) 4. การนำเสนอ (Delivery) 5. การใช้สื่อทัศนูปกรณ์ (Visual aids) มีรายละเอียดดังใน ตารางที่ 17-26

2.2.2.1 โครงสร้าง (Organization)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 17 ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) ใน
ด้านโครงสร้าง ครั้งที่ 1 และครั้งที่ 2 โดยวิธีหาค่า Spearman Correlation Coefficients

	ครั้งที่ 1		ครั้งที่ 2	
		(Sig.)		(Sig.)
คู่ที่ 1	.4310	.017	.5765	.001
คู่ที่ 2	.5804	.001	.5657	.001
คู่ที่ 3	.3483	.059	.5167	.003
คู่ที่ 4	.1781	.346	.3535	.055
คู่ที่ 5	.2956	.113	.2799	.134
คู่ที่ 6	.4076	.025	.5134	.004

จากตารางแสดงผลการประเมิน ในหัวข้อโครงสร้าง (Organization) คู่ที่มีความ
เที่ยงระหว่างกันสูงที่สุด

ครั้งที่ 1 คือ คู่ที่ 2

ครั้งที่ 2 คือ คู่ที่ 1

และคู่ที่มีความเที่ยงระหว่างกันน้อยที่สุด

ครั้งที่ 1 คือ คู่ที่ 4

ครั้งที่ 2 คือ คู่ที่ 5

ตารางที่ 18 แสดงสถิติเชิงพรรณนา (Descriptive Statistics) ด้านโครงสร้าง
(Organization)

ตารางที่ 18 สถิติเชิงพรรณนาด้านโครงสร้าง ครั้งที่ 1 และครั้งที่ 2

	ครั้งที่ 1				ครั้งที่ 2			
	X	S.D.	Min	Max	X	S.D.	Min	Max
ผู้ประเมิน 1	3.70	.75	2.00	5.00	3.78	.45	3.00	4.50
ผู้ประเมิน 2	3.08	.74	1.00	4.50	3.63	.47	3.00	4.00
ผู้ประเมิน 3	4.02	.60	3.00	5.00	3.95	.62	3.00	5.00
ผู้ประเมิน 4	3.00	.26	2.00	4.00	3.10	.55	2.00	4.00

สรุปผลการ ตรวจสอบความเที่ยงระหว่างผู้ประเมินในด้านโครงสร้าง ครั้งที่ 1 และครั้งที่ 2 คือ มีความต่างกันระหว่างผลการประเมินทั้ง 2 ครั้ง แต่เมื่อดูจาก ตารางที่ 17 จะเห็นว่ามี ความแตกต่างในผลการประเมินด้านโครงสร้างน้อย เนื่องจากตัวเลขจะค่อนข้างใกล้เคียงกัน หรือต่างกันสูงสุดไม่เกิน .1754

2.2.2.2 เนื้อหา (Content)

ตารางที่ 19 ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) ในด้านเนื้อหา ครั้งที่ 1 และครั้งที่ 2 โดยวิธีหาค่า Spearman Correlation Coefficients

	ครั้งที่ 1		ครั้งที่ 2	
		(Sig.)		(Sig.)
คู่ที่ 1	.4008	.028	.3771	.040
คู่ที่ 2	.4531	.012	.4943	.005
คู่ที่ 3	.5068	.004	.4072	.026
คู่ที่ 4	.1275	.502	.2826	.130
คู่ที่ 5	.2625	.161	-.2610	.749
คู่ที่ 6	.6722	.000	.2740	.143

จากตารางแสดงผลการประเมินในหัวข้อเนื้อหา (Content) ครั้งที่ 1 คู่ที่มีความ
เที่ยงระหว่างกันสูงสุดคือ

ครั้งที่ 1 คือ คู่ที่ 6

ครั้งที่ 2 คือ คู่ที่ 2

และคู่ที่มีความเที่ยงระหว่างกันน้อยที่สุด

ครั้งที่ 1 คือ คู่ที่ 4

ครั้งที่ 2 คือ คู่ที่ 5

ตารางที่ 20 แสดงค่าสถิติเชิงพรรณนาในด้านเนื้อหา (Content)

ตารางที่ 20 สถิติเชิงพรรณนาด้านเนื้อหา ครั้งที่ 1 และครั้งที่ 2

	ครั้งที่ 1				ครั้งที่ 2			
	\bar{X}	S.D.	Min	Max	\bar{X}	S.D.	Min	Max
ผู้ประเมิน 1	3.27	.64	2.00	5.00	3.45	.66	1.00	4.00
ผู้ประเมิน 2	2.87	.83	1.00	4.50	2.97	.68	2.00	4.00
ผู้ประเมิน 3	3.43	.56	2.50	4.50	3.48	.53	2.50	4.50
ผู้ประเมิน 4	2.73	.58	2.00	4.00	2.53	.57	1.00	3.00

สรุปผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินในด้านเนื้อหาครั้งที่ 1 และ
ครั้งที่ 2 ปรากฏว่า ความเที่ยงระหว่างผู้ประเมินค่อนข้างใกล้เคียงกันในคู่ที่ 1, 2, 3 และ 4
แต่ในกรณีคู่ที่ 5 และ 6 ผลการประเมินทั้ง 2 ครั้ง จะต่างกันค่อนข้างมากคือ ประมาณ .4 -
.5

จุฬาลงกรณ์มหาวิทยาลัย

2.2.2.3 การใช้ภาษา (Language)

ตารางที่ 21 ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) ในด้านการใช้ภาษา ครั้งที่ 1 และครั้งที่ 2 โดยวิธีหาค่า Spearman Correlation Coefficients

	ครั้งที่ 1		ครั้งที่ 2	
		(Sig.)		(Sig.)
คู่ที่ 1	.5484	.002	.4772	.008
คู่ที่ 2	.3091	.097	.6625	.000
คู่ที่ 3	-.1931	.307	.6155	.000
คู่ที่ 4	.2045	.278	.5725	.001
คู่ที่ 5	.2053	.277	.1692	.371
คู่ที่ 6	.1046	.582	.5569	.001

จากตารางแสดงผลการประเมินในหัวข้อการใช้ภาษาในเชิงนำเสนอ (Language) คู่ที่มีความเที่ยงระหว่างกันสูงสุดคือ

ครั้งที่ 1 คือ คู่ที่ 1

ครั้งที่ 2 คือ คู่ที่ 2

และคู่ที่มีความเที่ยงระหว่างกันน้อยที่สุด

ครั้งที่ 1 คือ คู่ที่ 3

ครั้งที่ 2 คือ คู่ที่ 5

ตารางที่ 22 แสดงสถิติเชิงพรรณนาด้านการใช้ภาษา

ตารางที่ 22 สถิติเชิงพรรณนาด้านการใช้ภาษา ครั้งที่ 1 และครั้งที่ 2

	ครั้งที่ 1				ครั้งที่ 2			
	\bar{X}	S.D.	Min	Max	\bar{X}	S.D.	Min	Max
ผู้ประเมิน 1	2.70	.88	1.00	4.00	2.60	.80	1.00	4.00
ผู้ประเมิน 2	2.55	.75	1.00	4.50	2.57	.68	2.00	4.00
ผู้ประเมิน 3	2.98	.67	2.00	4.50	3.07	.68	2.50	4.50
ผู้ประเมิน 4	2.62	.76	.50	4.00	2.17	.87	.00	4.00

สรุปผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินในด้านการใช้ภาษาครั้งที่ 1 และครั้งที่ 2 คือ มีความแตกต่างกันค่อนข้างมาก โดยเฉพาะคู่ที่ 3 ซึ่งต่างกันมากถึง .8086 และเป็นที่น่าสังเกตว่าผลการประเมินในหัวข้อการใช้ภาษานี้ มีความแตกต่างระหว่างการประเมินครั้งที่ 1 และ 2 มากกว่าการประเมินในหัวข้อโครงสร้างและเนื้อหา

2.2.2.4 การนำเสนอ (Delivery)

ตารางที่ 23 ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) ในด้านการนำเสนอ ครั้งที่ 1 และครั้งที่ 2 โดยวิธีหาค่า Spearman Correlation Coefficients

	ครั้งที่ 1		ครั้งที่ 2	
		(Sig.)		(Sig.)
คู่ที่ 1	.2890	.121	.4146	.023
คู่ที่ 2	.2334	.214	.5742	.001
คู่ที่ 3	.2177	.248	.4341	.017
คู่ที่ 4	.4565	.011	.4160	.022
คู่ที่ 5	.5214	.003	-.0920	.629
คู่ที่ 6	.7314	.000	.3817	.037

จากตารางแสดงผลการประเมินในหัวข้อการนำเสนอ (Delivery) คู่ที่มีความเที่ยง
 ระหว่างกันสูงสุดคือ ครั้งที่ 1 คือ คู่ที่ 6
 ครั้งที่ 2 คือ คู่ที่ 2
 และคู่ที่มีความเที่ยงระหว่างกันน้อยที่สุด
 ครั้งที่ 1 คือ คู่ที่ 3
 ครั้งที่ 2 คือ คู่ที่ 5

ตารางที่ 24 แสดงสถิติเชิงพรรณนาด้านการนำเสนอ

ตารางที่ 24 สถิติเชิงพรรณนาด้านการนำเสนอ ครั้งที่ 1 และครั้งที่ 2

	ครั้งที่ 1				ครั้งที่ 2			
	\bar{X}	S.D.	Min	Max	\bar{X}	S.D.	Min	Max
ผู้ประเมิน 1	2.70	.79	1.00	4.00	2.70	.84	.50	4.00
ผู้ประเมิน 2	2.52	.88	1.00	4.00	2.80	.68	2.00	4.00
ผู้ประเมิน 3	3.03	.55	2.50	4.50	3.10	.65	2.00	4.50
ผู้ประเมิน 4	2.70	.92	2.00	5.00	2.40	1.00	1.00	4.00

สรุปผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินในด้านการนำเสนอครั้งที่ 1
 และครั้งที่ 2 จะเห็นว่ามีความต่างกันตั้งแต่ .1256 - .6134 ดังนั้นในด้านการนำเสนอผล
 ส้ารวจความเที่ยงระหว่างผู้ประเมินทั้ง 2 ครั้ง มีความแตกต่างกันมากกว่าในด้านโครงสร้าง
 และเนื้อหา

จุฬาลงกรณ์มหาวิทยาลัย

2.2.2.5 การใช้สื่อทัศนูปกรณ์ (Visual Aids)

ตารางที่ 25 ผลการตรวจสอบความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability) ในด้านการใช้สื่อทัศนูปกรณ์ ครั้งที่ 1 และครั้งที่ 2 โดยวิธีหาค่า Spearman Correlation Coefficients

	ครั้งที่ 1		ครั้งที่ 2	
		(Sig.)		(Sig.)
คู่ที่ 1	.1889	.318	.4087	.025
คู่ที่ 2	-.0797	.676	.6238	.000
คู่ที่ 3	.2251	.232	.3350	.070
คู่ที่ 4	.3081	.098	.2992	.108
คู่ที่ 5	.2869	.124	.3799	.038
คู่ที่ 6	.2734	.144	.2865	.125

จากตาราง แสดงผลการประเมินในหัวข้อการใช้สื่อประกอบ (Visual aids) คู่ที่มีความเที่ยงระหว่างกันสูงที่สุดคือ

ครั้งที่ 1 คือ คู่ที่ 4

ครั้งที่ 2 คือ คู่ที่ 2

และคู่ที่มีความสอดคล้องระหว่างกันน้อยที่สุดคือ

ครั้งที่ 1 คือ คู่ที่ 2

ครั้งที่ 2 คือ คู่ที่ 6

ตารางที่ 26 แสดงสถิติเชิงพรรณนาด้านการใช้สื่อทัศนูปกรณ์

ตารางที่ 26 สถิติเชิงพรรณนาด้านการใช้สื่อประกอบ ครั้งที่ 1 และครั้งที่ 2

	ครั้งที่ 1				ครั้งที่ 2			
	\bar{X}	S.D.	Min	Max	\bar{X}	S.D.	Min	Max
ผู้ประเมิน 1	3.07	1.07	1.00	5.00	3.20	.66	2.00	4.00
ผู้ประเมิน 2	2.77	1.08	1.00	4.50	2.82	.68	2.00	4.00
ผู้ประเมิน 3	3.16	.60	2.00	4.50	3.15	.71	2.00	4.50
ผู้ประเมิน 4	2.47	.63	2.00	4.00	2.67	.84	1.00	5.00

สรุปผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินในด้านการใช้สื่อทัศนูปกรณ์ทั้ง 2 ครั้ง ปรากฏว่ามีความแตกต่างกันไม่มากนัก จากตัวเลขในตารางความแตกต่างจะอยู่ระหว่าง .0131 - .2198 ยกเว้นในคู่ที่ 2 ที่มีความแตกต่างกันมากถึง .7035 เป็นที่น่าสังเกตว่า ในการประเมินครั้งที่ 1 คู่ที่ 2 นี้ มีความเที่ยงระหว่างกันน้อยที่สุด คือ - .0797 แต่ในการประเมินครั้งที่ 2 กลับมีความเที่ยงระหว่างกันสูงที่สุดคือ .6238

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ (Conclusion and Recommendations)

1. สรุปผลการวิจัยและอภิปรายผล

จากปัญหาในการวัดสัมฤทธิ์ผลการพูดแบบนำเสนอสำหรับรายวิชาการสนทนาภาษาอังกฤษขั้นสูงในอดีต ซึ่งไม่มีเกณฑ์ในการทดสอบที่ชัดเจน จึงได้มีการสร้างและพัฒนามาตรวัดประเมินค่าสำหรับรายวิชานี้ขึ้น โดยมีรูปแบบเป็นมาตรวัดประเมินค่า (Rating Scales) ซึ่งมีองค์ประกอบเป็นช่วงคะแนน (Band Scores) ตั้งแต่ 1-5 และมีการแยกเป็น 5 หัวข้อ เพื่อให้ครอบคลุมวัตถุประสงค์และเนื้อหา ของรายวิชา หัวข้อทั้ง 5 ประกอบด้วย โครงสร้าง (Organization) เนื้อหา (Content) การใช้ภาษา (Language) การนำเสนอ (Delivery) และการใช้สื่อประกอบ (Visual aids)

ภายหลังจากการสร้างและพัฒนามาตรวัดประเมินค่าขึ้นแล้วได้มีการตรวจสอบในสองประเด็นหลักคือ

1. ตรวจสอบความตรง (Validity) ของเกณฑ์ประเมินอันประกอบด้วย
 - 1.1 ความตรงเชิงประจักษ์ (Face Validity)
 - 1.2 ความตรงเชิงเนื้อหา (Content Validity)
 - 1.3 ความตรงร่วมสมัย (Concurrent Validity)
2. ตรวจสอบความเที่ยงของผู้ประเมิน (Rater Reliability) เมื่อนำเกณฑ์ประเมินไปใช้ในสถานการณ์จริง โดยการตรวจสอบ
 - 2.1 ความเที่ยงของผู้ประเมิน (Intra-rater Reliability)
 - 2.2 ความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability)

สรุปผลการตรวจสอบในประเด็นต่างๆ ได้ดังนี้

1. ผลการตรวจสอบความตรง (Validity) ของมาตราส่วนประเมินค่าที่สร้างขึ้นใน 3 ประเด็น คือ ความตรงเชิงประจักษ์ ความตรงเชิงเนื้อหา และความตรงร่วมสมัย

1.1 ความตรงเชิงประจักษ์ (Face Validity)

ผลการตรวจสอบความตรงเชิงประจักษ์ พบว่าวิธีการวัดสัมฤทธิ์ผลการพูดแบบนำเสนอ สำหรับรายวิชาการสนทนาภาษาอังกฤษชั้นสูง โดยกระบวนการประเมินผลที่ให้นิสิตผู้ถูกประเมิน พูดแบบนำเสนอคนละประมาณ 5 นาที และอาจารย์ผู้ประเมินใช้เกณฑ์การประเมินที่สร้างขึ้น มีความตรงเชิงประจักษ์สูงมาก ตามเจตคติของผู้ประเมินคือ อาจารย์ และผู้ถูกประเมินคือ นิสิต กล่าวคือ ผู้ประเมินมีความเห็นว่า วิธีการประเมินนี้มีความตรงเชิงประจักษ์สูงมากโดยมีค่าเฉลี่ยสูงถึง 4.88 จาก 5.00 และในด้านเกณฑ์การประเมินที่สร้างขึ้นในรูปแบบมาตราส่วนประเมินค่านั้น ผู้ประเมินเห็นว่ามี ความตรงเชิงประจักษ์สูงเหมือนกัน โดยเฉพาะอย่างยิ่งใน หัวข้อ การใช้ภาษา (Language) (4.69) สำหรับนิสิตผู้ถูกประเมิน ก็เห็นว่าวิธีการประเมินนี้มีความตรงเชิงประจักษ์สูงเช่นกัน (4.186) และมีความพอใจต่อวิธีการประเมินนี้ (4.031) ส่วนในด้านเกณฑ์การประเมินที่สร้างขึ้น นิสิตเห็นว่ามี ความตรงเชิงประจักษ์สูงเช่นกัน (4.157)

1.2 ความตรงเชิงเนื้อหา (Content)

ผลการตรวจสอบความตรงเชิงเนื้อหาพบว่า ผู้ประเมินเห็นว่าเกณฑ์ที่สร้างขึ้น มีความตรงตามเนื้อหาของหลักสูตร และวัตถุประสงค์ของรายวิชาสูงมากโดยเฉพาะอย่างยิ่ง ความตรงตามเนื้อหา ในหัวข้อโครงสร้าง (Organization) โดยมีค่าเฉลี่ยสูงถึง .94 โดยวิธี Item – Objective Congruence Index ส่วนในหัวข้ออื่นๆ คือ เนื้อหา (Content) การใช้ภาษา (Language) การนำเสนอ (Delivery) และการใช้สื่อประกอบ (Visual aids) ก็มีความตรงเชิงเนื้อหาสูงเช่นกัน โดยมีค่าเฉลี่ยที่ .87 .81 .87 และ .87 ตามลำดับ

โดยสรุปแล้วเมื่อผู้ประเมินเห็นว่าเกณฑ์ประเมินมีความตรงเชิงเนื้อหาสูง ความเหมาะสมสำหรับรายวิชาที่สูงด้วยเช่นกัน โดยมีค่าเฉลี่ยที่ .871 และเมื่อตรวจสอบ คำบรรยายเกณฑ์ (descriptors) ในทุกช่วงคะแนน (band) ว่ามีความเหมาะสม และครอบคลุมเนื้อหาในแต่ละหัวข้อหรือไม่ ผลปรากฏว่า เนื้อหาในคำบรรยายเกณฑ์ตรงกับเนื้อหา ของหลักสูตร และวัตถุประสงค์ในทุกช่วงคะแนน (band)

แต่จากการตรวจสอบความสะดวกและง่ายต่อการใช้ของเกณฑ์ผู้ประเมินเห็นว่า มีความสะดวกและง่ายต่อการใช้ในระดับสูงปานกลาง โดยดูจากค่าเฉลี่ยที่ .69

1.3 ความตรงร่วมสมัย (Concurrent Validity)

การตรวจสอบความตรงร่วมสมัยในสองประเด็นคือ คุณลักษณะในการใช้ภาษา (Language) และคุณลักษณะด้านโครงสร้าง (Organization) จากการพูดแบบนำเสนอโดยการเปรียบเทียบผลที่ได้จากวิธีทดสอบสามวิธี คือ วิธีพูดแบบนำเสนอ (Oral Presentations) วิธีสอบข้อเขียน ซึ่งเป็นข้อสอบปลายภาคของรายวิชาการสนทนาภาษาอังกฤษชั้นสูง (ABOC) และวิธีสอบข้อเขียน ซึ่งเป็นข้อสอบ CU-TEP ปรากฏผลดังนี้

1.3.1 จากการวัดคุณลักษณะด้านการใช้ภาษา (Language) โดยวิธีทดสอบ 3 วิธีคือ พูดแบบนำเสนอและสอบข้อเขียน 2 ชุด ปรากฏว่าค่าสัมประสิทธิ์ สหสัมพันธ์ของวิธีการทดสอบโดยการพูด และโดยการเขียนอยู่ในระดับต่ำ แต่หากใช้วิธีทดสอบโดยการเขียน (ข้อสอบ CU-TEP และข้อสอบปลายภาครายวิชา ABOC) ค่า สัมประสิทธิ์สหสัมพันธ์ จะอยู่ในระดับสูงปานกลาง จึงสรุปได้ว่าการวัดคุณลักษณะด้านการใช้ภาษาโดยวิธีพูด และเขียนไม่มีความตรงร่วมสมัย แต่หากวัดโดยวิธีเขียนอาจมีความตรงร่วมสมัย

1.3.2 ผลจากการวัดคุณลักษณะด้านโครงสร้าง (Organization) โดยวิธีทดสอบ 2 วิธี คือการพูดแบบนำเสนอ และการสอบปลายภาครายวิชา ABOC ปรากฏว่าการวัดคุณลักษณะด้านโครงสร้างของรายวิชานี้โดยการสอบพูด และเขียน มีค่าสัมประสิทธิ์สหสัมพันธ์ต่ำคือ .1089 ($p = .567$) ดังนั้นจึงสรุปได้ว่าการวัด คุณลักษณะของผู้เรียนในด้านโครงสร้างโดยวิธีพูดและเขียนไม่มีความตรงร่วมสมัย

2. ผลการตรวจสอบความเที่ยงของผู้ประเมิน (Rater Reliability) ใน 2 ประเด็น คือ ความเที่ยงของผู้ประเมิน (Intra-rater Reliability) และความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability)

2.1 ความเที่ยงของผู้ประเมิน (Intra-rater Reliability)

ผลจากการตรวจสอบความสอดคล้องภายในของผู้ประเมินพบว่า ผู้ประเมินทั้ง 4 คน มีความเที่ยงต่างกัน และไม่อาจสรุปได้ว่า ผู้ประเมินที่เป็นชาวไทยกับผู้ประเมินชาวต่างประเทศผู้เป็นเจ้าของภาษา จะมีความแม่นยำตรง และ สอดคล้องต่างกันอย่างไร หรืออายุ เพศ ประสบการณ์ เจตคติ ฯลฯ จะมีผลกระทบในการประเมินมากน้อยอย่างไร สถานที่ เวลา หรือปัจจัยแวดล้อมอื่นๆ ปัญหาส่วนตัว อาจมีผลกระทบด้วยหรือไม่ ประเด็นเรื่องความสอดคล้องภายในของผู้ประเมิน จึงเป็นปัญหาที่สำคัญในการประเมินทักษะการพูด ตั้งแต่อดีตมาจนถึงปัจจุบัน

จากรายละเอียดและข้อมูลส่วนตัวของผู้ประเมินในภาคผนวก ข จะเห็นว่า ผู้ประเมินที่มีความเที่ยงสูงที่สุดเป็นชาย แต่ผู้ประเมินที่มีความเที่ยงต่ำที่สุด (หรือ อาจกล่าวได้ว่า ไม่มีความเที่ยง) ก็เป็นชายเช่นเดียวกัน จึงไม่อาจกล่าวได้ว่า ผู้ประเมินชายมีความเที่ยงสูงกว่าผู้ประเมินหญิง ในด้านอายุ และความเป็นเจ้าของภาษา ก็ไม่อาจนำมาเป็นข้อสรุปได้เช่นเดียวกัน และหากสังเกตประสบการณ์การสอนและประเมินผล (โดยมิได้ระบุว่าเป็นการประเมินทักษะการพูด) จะเห็นว่า ผู้ที่มีความเที่ยงสูงที่สุดนั้น มีประสบการณ์ที่ยาวนาน แม้จะไม่ยาวนานที่สุดก็ตาม แต่ ผู้ประเมินที่มีความเที่ยงน้อยที่สุดนั้น มีประสบการณ์น้อยที่สุดเมื่อเปรียบเทียบกับ ผู้ประเมินที่เหลืออีก 3 คน จึงอาจมีความเป็นไปได้ว่าประสบการณ์นั้นมีผลต่อความเที่ยงของผู้ประเมิน

2.2 ความเที่ยงระหว่างผู้ประเมิน (Inter-rater Reliability)

มีการตรวจสอบความเที่ยงของผู้ประเมิน 2 ครั้ง เพื่อดูว่าผู้ประเมินคู่ใด จากทั้งหมด 6 คู่ มีความเที่ยงระหว่างกันสูงที่สุด เมื่อเปรียบเทียบผลการตรวจสอบการประเมิน 2 ครั้ง ปรากฏว่า ความเที่ยงระหว่างผู้ประเมิน ไม่ตรงกัน ตัวอย่างเช่น จากการประเมินครั้งที่ 1 คู่ที่ 3 เป็นคู่ที่มีความเที่ยงระหว่างกันน้อยที่สุด หรืออาจกล่าวได้ว่าไม่มีความเที่ยงเลย จากตัวเลข -0.0955 (Sig. = $.622$) แต่ในการประเมินครั้งที่ 2 คู่นี้กลับมีความเที่ยงสูงมาก คือ $.6307$ (Sig. = $.000$) และสำหรับคู่ที่ 1 2 4 5 และ 6 ค่าความเที่ยงกัน ระหว่างการประเมินครั้งที่ 1 และครั้งที่ 2 ก็ต่างกันมากบ้างน้อยบ้าง จึงอาจสรุปได้ว่า

ความเที่ยงระหว่างผู้ประเมินนั้นไม่คงที่ เมื่อทำการประเมินซ้ำครั้งที่ 2 จากตารางที่ 27 จะเห็นว่า ความเที่ยงระหว่างผู้ประเมินแต่ละคู่ ในการประเมิน 2 ครั้ง นั้น ได้ผลแตกต่างกัน

ตารางที่ 27 ความเที่ยงระหว่างผู้ประเมิน

คู่ที่มีความเที่ยงระหว่างผู้ประเมินมากที่สุด		คู่ที่มีความเที่ยงระหว่างผู้ประเมินน้อย ที่สุด	
ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 1	ครั้งที่ 2
คู่ที่ 4 ผู้ประเมิน 2+3	คู่ที่ 2 ผู้ประเมิน 1+3	คู่ที่ 3 ผู้ประเมิน 1+4	คู่ที่ 5 ผู้ประเมิน 2+4

อย่างไรก็ดี จากผลการตรวจสอบความเที่ยงระหว่างผู้ประเมินจะเห็นว่า ผู้ประเมินที่มีความเที่ยงสูงที่สุด คือ ผู้ประเมิน 3 จะทำให้เกิดความเที่ยงระหว่าง ผู้ประเมินสูงตามไปด้วย โดยคู่ที่ 4 และ คู่ที่ 2 ต่างก็มีผู้ประเมิน 3 ประกอบอยู่ด้วย

และในทางกลับกัน ผู้ประเมินที่มีความเที่ยงต่ำที่สุดคือ ผู้ประเมิน 4 ก็ทำให้เกิดความเที่ยงระหว่างผู้ประเมินต่ำลงไป กล่าวคือ ผู้ประเมินคู่ที่ 3 และคู่ที่ 5 ซึ่งมีความเที่ยงระหว่างกันน้อยที่สุดมีผู้ประเมิน 4 อยู่ด้วย

ดังนั้นจึงอาจสรุปได้ว่า จากผลของการประเมินโดยใช้มาตรวัดประเมินค่าที่สร้างและพัฒนาขึ้น พบว่าผู้ประเมินมีความเที่ยงต่างกัน และเมื่อนำผลของการประเมินมาเปรียบเทียบความเที่ยงระหว่างผู้ประเมิน ผลที่ได้จึงต่างกันมากบ้างน้อยบ้าง ทั้งนี้ขึ้นอยู่กับความเที่ยงของผู้ประเมินแต่ละคน

ดังนั้น ผลจากการตรวจสอบนี้จะถูกนำไปประกอบการพิจารณาในการจับคู่ผู้ประเมิน ในการประเมินสัมฤทธิ์ผลรายวิชาในครั้งต่อไป

2. ข้อเสนอแนะ (Recommendations)

จากการตรวจสอบมาตรวัดประเมินค่า (Rating Scales) ที่ใช้ในการวัดสัมฤทธิ์ผล การพูดแบบนำเสนอ สำหรับรายวิชาการสนทนาภาษาอังกฤษชั้นสูง (Advanced Business Oral Communication) ในด้านความตรง (Validity) และความเที่ยงของผู้ประเมิน (Rater Reliability) ในประเด็นต่างๆ ผลสรุปที่ได้นำไปสู่ข้อเสนอแนะ 4 ประการคือ

2.1 ประเมินผลโดยผู้ประเมินอย่างน้อย 2 คน

ควรมีผู้ประเมินอย่างน้อย 2 คน เพื่อเฉลี่ยคะแนนโดยเฉพาะในกรณีที่ ผู้ประเมิน มีความเข้มงวดมากน้อยต่างกัน การเฉลี่ยคะแนนระหว่างผู้ประเมินที่มีความต่างกันมากๆ จะช่วยให้ผลคะแนนเฉลี่ยมีความยุติธรรม และเชื่อถือได้เพิ่มขึ้น

2.2 การพิจารณาคัดเลือกผู้ประเมินคู่

ควรมีการพิจารณาคัดเลือกผู้ประเมินคู่อย่างรอบคอบเพื่อเพิ่มความ สอดคล้อง และความถูกต้อง แม่นยำในการประเมินซึ่งจะทำให้เกิดความยุติธรรม และน่าเชื่อถือ

2.3 การอบรมผู้ประเมิน

จำเป็นต้องมีการอบรมผู้ประเมินในกรณีประเมินสัมฤทธิ์ผลการพูดแบบ นำเสนอ สำหรับรายวิชานี้ เพื่อประโยชน์ในการสร้างความเข้าใจเกี่ยวกับเกณฑ์ และการตีความ คำบรรยายเกณฑ์ที่ตรงกัน โดยเฉพาะผู้ประเมินหน้าใหม่ ซึ่งยังไม่คุ้นเคยกับเนื้อหา และวัตถุประสงค์ของหลักสูตร

2.4 การปรับปรุงและพัฒนาคำบรรยายเกณฑ์

จะต้องมีการปรับปรุงและพัฒนาคำบรรยายเกณฑ์ให้มีความชัดเจนมากขึ้น และเข้าใจได้ง่ายขึ้น เพื่อความสะดวกในการใช้ แต่ทั้งนี้จะต้องไม่เพิ่มความยาวของคำบรรยายเกณฑ์ มิฉะนั้นอาจสร้างความสับสน และไม่สะดวก เนื่องจากผู้ประเมินต้องเสียเวลาในการอ่านคำบรรยายเกณฑ์ที่ยาว โดยเฉพาะในขณะที่กำลังทำการประเมิน

ประเด็นในการทำวิจัยต่อไป

ประเด็นที่ควรทำการวิจัยต่อไป คือ ศึกษาตัวแปรและปัจจัย หรือภาวะต่างๆ ที่มีผลต่อการประเมิน เช่น การศึกษา ประสบการณ์ วุฒิภาวะ เพศ วัย และสภาพแวดล้อมต่างๆ ซึ่งมีงานวิจัยหลายชิ้นที่กล่าวถึงสาเหตุและปัจจัยต่างๆ เหล่านี้ว่ามีผลกระทบต่อ การประเมิน ทั้งนี้เพื่อศึกษา และหาวิธีการพัฒนาการประเมินผลสำหรับ รายวิชา การ สอนทนาภาษาอังกฤษในระดับต่างๆ ให้มีประสิทธิภาพยิ่งขึ้น



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บรรณานุกรม

สุมิตรา อังวัฒน์กุล. (2535). การวิจัยทางการเรียนการสอนภาษาอังกฤษ กรุงเทพฯ:

สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย

อัจฉรา วงศ์โสธร. (2539) การทดสอบและประเมินผลการเรียนการสอนภาษา

อังกฤษ กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Bibliography

- Alderson, J.C. (1991). "Bands and scores." In C. Alderson, & B. North. (Eds.), **Language Testing in the 1990s.** (pp. 71-86). London: Macmillan Publishers Ltd.
- Bachman, L.F. (1990). **Fundamental Considerations in Language Testing.** Oxford: Oxford University Press.
- Bachman, L.F. & Palmer, A.S. (1996). **Language Testing in Practice.** Oxford: Oxford University Press.
- Baker, D. (1989). **Language Testing : A Critical Survey and Practical Guide.** Edward Arnold.
- Brown, Anne. (1995). "The effect of rater variables in the development of an occupation-specific language performance test" **Language Testing** (12), 1-15.
- Brown, J.D. (1996). **Testing in Language Programs.** New Jersey: Prentice Hall.
- Calderbank, M. & Awwad, M. (1988). "Testing Oral Communication: Why and How?" **System** 16 (1) 49-59.
- Fulcher, G. (1993). **The Construction and Validation of Rating Scales for Oral Tests in English as a Foreign Language.** Unpublished Ph.D. Thesis. University of Lancaster.
- Douglas, D. (1994). "Quantity and quality in speaking test performance," **Language Testing** 11 (12), 125-138.
- Hall, S. J. (1996). "I'd like to present but how are you going to score me?" **Pasaa** (26), 20-25.
- Harris, D.P. (1969). **Testing English as a Second Language.** New York: McGraw-Hill.
- Henning, G. (1987). **A Guide to Language Testing: Development Evaluation Research.** Cambridge : Newbury House.

- Henning, G. (1992). "The ACTFL oral proficiency interview: validity evidence." *System* 20 (3) 365-372.
- Ho, Laina. (1997). "Evaluation the Oral Presentation Skills of Computer Science Students." *Working Papers on Language*. (pp. 19-25). Singapore: National University of Singapore.
- Hockje, B. & Linnell, K. (1994). "Authenticity" in Language testing: evaluating spoken language tests for international teaching assistants." *TESOL Quarterly* 28 (1), 103-125.
- Hughes, A. (1987). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Jafapur, A. (1988). "Non-native raters determining the oral proficiency of EFL learners" *System* 16 (1), 61-67.
- Lim Hong Eng, Janet. (1997). "The Value of Peer Evaluations: Its Implications for Teaching." *Working Papers on Language*. (pp. 26-33). Singapore: National University of Singapore.
- Liskin – Gasparrrd, J. (1987). *Testing & Teaching for Oral Proficiency*. Baston: Heinle & Heinle Publishers.
- Lumley, Tom & McNamara, T.F. (1995). "Rater characteristics and rater bias : implications for traning." *Language Testing* (12), 54-71.
- McNamara, Tim. (1996). "Raters and ratings: introduction to multi-faceted measurement." *Measuring Second Language Performance*. (pp. 117 – 148). London: Longman.
- Morrow, K. (1991). "Evaluating Communication Tests." In S. Anivan (Ed.), *Current Developments in Language Testing*. (pp. 92-102). Singapore: SEAMEO Regional Language Centre.
- Mullen, K.A. (1980). "Rater reliability and oral proficiency evaluations." In J.W. Oller & K. Perkins (Eds.), *Research in Language Testing*. (pp. 91-101). Rowley: Newbury House.

- Pathak, A. & Keng, C.S. (2001). "Peer Assessment of Oral Presentations: What These Two Cases Tell Us." *Pasaa* (31), 13-26.
- Pollitt, A. & Murray, N.L. (1996). "What raters really pay attention to." In M. Milanovic & N. Saville (Eds.), *Studies in Language Testing 3* (pp. 74-91). Cambridge: Cambridge University Press.
- Porter, D. (1991). "Affective factors in the assessment of oral interaction: gender and status." In S. Anivan (Ed.). *Current Developments in Language Testing* (pp. 92-102). Singapore: SEAMEO Regional Language Centre.
- Seliger, H. W. & Shohamy, E. (1989). *Second Language Research Methods*. Oxford: Oxford University Press.
- Shohamy, E. (1983). "The stability of oral proficiency assessment on the oral interview testing procedures." *Language Learning*. 33 (4), 527-540.
- Underhill, N. (1987). *Testing Spoken Language*. Cambridge: Cambridge University Press.
- Underhill, N. (1991). "The Great Reliability Validity Trade- Off: Problems in Assessing the Productive Skills." In B. Heaton (Ed.), *Language Testing*. (pp. 17-23). Hays, Middx: Modern English Publications Limited.
- Upshur, J.A. & Turner, C.E. (1995). "Constructing rating scales for second language tests." *ELT Journal* (49), 3-12.
- Weir, E. (1988). *Communicative Language Testing*. Exeter: Prentice Hall International.
- Yamane, T. (1967). *Elementary Sampling Theory*. New York: Prentice-Hall, Inc.

ภาคผนวก ก
FSI Rating Scale 1958

Accent

1. Pronunciation frequently unintelligible.
2. Frequent gross errors and a very heavy accent make understanding difficult, requires frequent repetition.
3. "Foreign accent" requires concentrated listening and mispronunciations lead to occasional misunderstanding and apparent errors in grammar or vocabulary.
4. Marked "foreign accent" and occasional mispronunciations which do not interfere with understanding.
5. No conspicuous mispronunciations, but would not be taken for a native speaker.
6. Native pronunciation, with no trace of "foreign accent".

Grammar

1. Grammar almost entirely inaccurate except in stock phrases.
2. Constant errors showing control of very few major patterns and frequently preventing communication.
3. Frequent errors showing some major patterns uncontrolled and causing occasional irritation and misunderstanding.
4. Occasional errors showing imperfect control of some patterns but no weakness that causes misunderstanding.
5. Few errors, with no patterns of failure.
6. No more than two errors during the interview.

Fluency

1. Speech is so halting and fragmentary that conversation is virtually impossible.

2. Speech is very slow and uneven except for short or routine sentences.
3. Speech is frequently hesitant and jerky; sentences may be left uncompleted.
4. Speech is occasionally hesitant, with some unevenness caused by rephrasing and groping for words.
5. Speech is effortless and smooth, but perceptibly non-native in speed and evenness.
6. Speech on all professional and general topics as effortless and as smooth as a native speaker's.

Comprehension

1. Understands too little for the simplest type of conversation.
2. Understands only slow very simple speech on common social and touristic topics: requires constant repetition and rephrasing.
3. Understands careful, somewhat simplified speech directed to him, with considerable repetition and rephrasing.
4. Understands quite well normal educated speech directed to him, but requires occasional repetition and rephrasing.
5. Understands everything in normal educated conversation except for very colloquial or low-frequency items, or exceptionally rapid or slurred speech.
6. Understands everything in both formal and colloquial speech to be expected of an educated native speaker.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข
ILR Rating Scale

- Level 0:** No functional ability
- Level 0+:** Operated with memorized material, and short lists.
- Level 1:** Survival. Consistently creates with the language. Tends to operate in the present time. Uses short sentences.
- Level 2:** Concrete. Operates in past and future as well as present time. Can use paragraphs (combining sentences into limited connected discourse).
- Level 3:** Abstract. Thinks in the target language. Supports opinions and hypothesizes. Handles unfamiliar situations and topics. Can use organized discourse. General vocabulary is controlled with only sporadic errors in basic grammar, but with some errors in frequent complex structures and more errors in low-frequency complex structures. Can use colloquial and careful formal speech.
- Level 4:** Representation. Tailors language to suit the audience. Discourse is well organized and a controlled use of synonyms.
- Level 5:** Educated Native. Functions equivalent to a well-educated native speaker. Absolutely appropriate discourse in a wide range of contexts from colloquial to careful formal speech.

ภาคผนวก ค
ACTFL Levels

Novice: The Novice level is characterized by an ability to communicate minimally with learned material.

Novice Low: Oral production consists of isolated words and perhaps a few high-frequency phrases. Essentially no functional communicative ability.

Novice Mid: Oral production continues to consist of isolated words and learned phrases within very predictable areas of need, although quantity is increased. Vocabulary is sufficient only for handling simple, elementary needs and expressing basic courtesies. Utterances rarely consist of more than two or three words and show frequent long pauses and repetition of interlocutor's words. Speaker may have some difficulty producing even the simplest utterances. Some Novice-Mid speakers will be understood only with great difficulty.

Novice High: Able to satisfy partially the requirements of basic communicative exchanges by relying heavily on learned utterances but occasionally expanding these through simple recombinations of their elements. Can ask questions or make statements involving learned material. Shows signs of spontaneity, although this falls short of real autonomy of expression. Speech continues to consist of learned utterances rather than of personalized, situationally adapted ones. Vocabulary centers on areas such as basic objects, places, and most common kinship terms. Pronunciation may still be strongly influenced by first language. Errors are frequent and, in spite of repetition, some Novice-High speakers will have difficulty in being understood even by sympathetic interlocutors.

Intermediate: The intermediate level is characterized by an ability to

- create with the language by combining and recombining learned elements, though primarily in a reactive mode;

- initiate, minimally sustain, and close in a simple way basic communicative tasks; and
- ask and answer questions.

Intermediate Low: Able to handle successfully a limited number of interactive, task-oriented social situations. Can ask and answer questions, initiate and respond to simple statements, and maintain face-to-face conversation, although in a highly restricted manner and with much linguistic inaccuracy. Within these limitations, can perform such tasks as introducing self, ordering a meal, asking directions, and making purchases. Vocabulary is adequate to express only the most elementary needs. Strong interference from native language may occur. Misunderstandings frequently arise, but with repetition, the Intermediate-Low speaker can generally be understood by sympathetic interlocutors.

Intermediate-Mid: Able to handle successfully a variety of uncomplicated, basic communicative tasks and social situations. Can talk simply about self and family members. Can ask and answer questions and participate in simple conversations on topics beyond the most immediate needs; e.g., personal history and leisure-time activities. Utterance length increases slightly, but speech may continue to be characterized by frequent long pauses, since the smooth incorporation of even basic conversational strategies is often hindered as the speaker struggles to create appropriate language forms. Pronunciation may continue to be strongly influenced by first language and fluency may still be strained. Although misunderstandings still arise, the Intermediate-Mid speaker can generally be understood by sympathetic interlocutors.

Intermediate High: Able to handle successfully most uncomplicated communicative tasks and social situations. Can initiate, sustain, and close a general conversation with a number of strategies appropriate to a range of circumstances and topics, but errors are evident. Limited vocabulary still necessitates hesitation and may bring about slightly unexpected circumlocution. There is emerging evidence of connected discourse, particularly for simple narration and/or description. The Intermediate-High speaker can generally be understood even by

interlocutors not accustomed to dealing with speaking at this level, but repetition may still be required.

Advanced: The Advanced level is characterized by an ability to

- converse in a clearly participatory fashion;
- initiate, sustain, and bring to closure a wide variety of communicative tasks, including those that require an increased ability to convey meaning with diverse language strategies due to a complication or an unforeseen turn of events;
- satisfy the requirements of school and work situations; and
- narrate and describe with paragraph-length connected discourse.

Advanced: Able to satisfy the requirements of everyday situations and routine school and work requirements. Can handle with confidence but not facility complicated tasks and social situations, such as elaborating, complaining, and apologizing. Can narrate and describe with some details, linking sentences together smoothly. Can communicate facts and talk casually about topics of current public and personal interest, using general vocabulary. Shortcomings can often be smoothed over by communicative strategies, such as pause fillers, stalling devices, and different rates of speech. Circumlocution which arises from vocabulary or syntactic limitations very often is quite successful, though some groping for words may still be evident. The Advanced level speaker can be understood without difficulty by native interlocutors.

Advanced Plus: Able to satisfy the requirements of a broad variety of everyday, school and work situations. Can discuss concrete topics relating to particular interests and special fields of competence. There is emerging evidence of ability to support opinions, explain in detail, and hypothesize. The Advance-Plus speaker often shows a well-developed ability to compensate for an imperfect grasp of some forms with confident use of communicative strategies, such as paraphrasing and circumlocution. Differentiated vocabulary and intonation are effectively used to communicate fine shades of meaning. The Advanced-Plus speaker often shows remarkable fluency and ease of speech, but under the demands of Superior-Level, complex tasks, language may break down or prove inadequate.

Superior: The Superior level is characterized by an ability to

- participate effectively in most formal and informal conversations on practical, social, professional, and abstract topics; and
- support opinions and hypothesize using native-like discourse strategies.

Superior: Able to speak the language with sufficient accuracy to participate effectively in most formal and informal conversations on practical, social, professional, and abstract topics. Can discuss special fields of competence and interest with ease. Can support opinions and hypothesize, but may not be able to tailor language to audience or discuss in depth highly abstract or unfamiliar topics. Usually the Superior level speaker is only partially familiar with regional or other dialectical variants. The Superior level speaker commands a wide variety of interactive strategies and shows good awareness of discourse strategies. The latter involves the ability to distinguish main ideas from supporting information through syntactic, lexical, and suprasegmental features (pitch, stress, intonation). Sporadic errors may occur, particularly in low-frequency structures and some complex high-frequency structures more common to formal writing, but no patterns of error are evident. Errors do not disturb the native speaker or interfere with communication.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Assessment criteria

Table 1

Organisation	Content	Language	Delivery	Visual Aids
<ul style="list-style-type: none"> - Introduction <ul style="list-style-type: none"> : Purpose : Outline of main parts : Questions - Body <ul style="list-style-type: none"> : Main parts - Ending <ul style="list-style-type: none"> : Summary : Conclusion/ recommendation : Closing 	<ul style="list-style-type: none"> - Transitions (all parts are well linked) - Coherence (details and examples are relevant and interesting) 	<ul style="list-style-type: none"> - Appropriate, clear, simple and fluent - Natural spoken language - Vocabulary - Grammar - Pronunciation 	<ul style="list-style-type: none"> - Eye contact - Voice - Articulation - Manner and gestures 	<ul style="list-style-type: none"> - Design - Use - Relevance

Table 2

Rate: 5 point scale

5 = excellent; 4 = good; 3 = moderate; 2 = poor; 1 = very weak

Rate	Organisation	Content	Language	Delivery	Visual Aids
5	A very clear structure of the presentation with all the necessary components included.	All parts are very well linked with a perfect use of transitions. Examples and details are highly illustrative and coherent.	Very clear, simple and natural spoken language. Free of grammatical errors. Very good pronunciation (no need to sound like a native speaker though).	Very good eye contact and gesture. Highly articulate with loud, clear voice.	Very clear & simple with minimum use of figures and very effective use of visuals.
4	A clear structure of the talk with nearly all the components except for one or two unimportant elements. (e.g. questions)	All parts are well linked together with appropriate use of transitions. Examples and details are coherent.	Clear, simple and natural use of spoken language. Hardly noticeable grammatical errors. Good pronunciation.	Good eye contact and gesture. Articulate with loud, clear voice.	Clear and simple. Easy to comprehend. An effective use of visual.

Rate	Organisation	Content	Language	Delivery	Visual Aids
3	The structure of the talk does not include certain components such as purpose or outline of main parts, etc.	The parts are linked by occasional use of transitions. Examples and details are not quite coherent or interesting.	Language is unnatural or sounds like a rehearsed text. Noticeable grammatical errors. Pronunciation is comprehensible.	Fair eye contact and gesture. Articulation and use of voice are moderate.	Not very clear with too many letters and number.
2	Unclear structure due to lack of crucial components; for example, the introduction, body and conclusion.	Rarely use transitions to link parts. Examples and details are irrelevant and uninteresting.	A memorized text. Lots of grammatical mistakes. Hardly comprehensible pronunciation.	Rare eye contact. Utterances unclear and quite incomprehensible.	Unclear due to extremely small letters and numbers.
1	Badly structured with many crucial components being left out.	No use of transitions. The parts are not coherent. No details or examples provided.	Jerky. Tons of grammatical errors. Pronunciation is incomprehensible.	No eye contact. Totally inarticulate.	Visuals totally incompressible.

ปรับปรุงครั้งที่ 2

Assessment criteria

Table 1

Organisation	Content	Language	Delivery	Visual Aids
<ul style="list-style-type: none"> - Introduction <ul style="list-style-type: none"> : Purpose : Outline of main parts : Questions - Body <ul style="list-style-type: none"> : Main parts - Ending <ul style="list-style-type: none"> : Summary : Conclusion/ recommendation : Closing 	<ul style="list-style-type: none"> - Transitions (all parts are well linked) - Coherence (details and examples are relevant and interesting) 	<ul style="list-style-type: none"> - Appropriate, clear, simple and fluent - Natural spoken language - Vocabulary - Grammar - Pronunciation 	<ul style="list-style-type: none"> - Eye contact - Voice - Articulation - Manner and gestures 	<ul style="list-style-type: none"> - Design - Use / Handling - Relevance

Table 2

Rate: 5 point scale


5 = excellent; 4 = good; 3 = moderate; 2 = poor; 1 = very weak

Rate	Organisation	Content	Language	Delivery	Visual Aids
5	A very clear structure of the presentation with all the necessary components included (e.g. Introduction: topic, purpose, outline, time, questions. Body : parts 1, 2, 3, etc. Finishing off: summary, conclusion, questions).	All parts are very well linked with excellent use of transitions. Examples and details are highly illustrative and coherent. Very interesting talk.	Very clear, simple and natural spoken language. Free of grammatical errors. Very good pronunciation (no need to sound like a native speaker though).	Maintains very good eye contact with audience in all areas of the room. Loud, clear voice.	Very clear & simple visuals. Handles the visuals very effectively without distraction so that visuals fully support the presentation.
4	A clear structure of the talk with nearly all the components except for one or two unimportant elements.	All parts are well linked together with appropriate use of transitions. Examples and details are coherent. Interesting talk.	Clear, simple and natural use of spoken language. A few unimportant grammatical errors. Good pronunciation.	Maintains good eye contact with almost all of the audience. Articulate with loud, clear voice.	Clear and simple visuals. Handles visuals effectively (perhaps with minor distractions.) Visuals support the talk.

Rate	Organisation	Content	Language	Delivery	Visual Aids
3	The structure of the talk does not include certain components such as the purpose, outline of main parts, etc.	Most parts are linked by occasional use of transitions. Examples and details are not quite coherent or interesting.	Language is unnatural (sounds like a rehearsed text) Noticeable grammatical and pronunciation errors some of which lead to miscomprehension.	Only occasional eye contact and use of gestures. Articulation and voice lack clarity.	Reasonably clear visuals but wordy and/or contain some mistakes. Handles visuals with some distractions.
2	Unclear structure due to lack of crucial components; for example, the introduction, body and conclusion.	Rarely use transitions to link parts. Examples and details are irrelevant and uninteresting.	A memorized text. Lots of grammatical mistakes. Pronunciation is hardly comprehensible. Lots of grammatical mistakes which seriously <u>impair</u> comprehension.	Little use of gestures Utterances frequently unclear and quite incomprehensible.	Unclear visuals which hardly support the talk. Lots of mistakes and/or difficult to make out words. Handles visuals with many distractions.
1	Badly structured with many necessary components being left out.	No use of transitions. The parts are not coherent. No details or examples provided. Very uninteresting talk.	Jerky. Full of grammatical errors. Pronunciation is incomprehensible.	No eye contact. Very inarticulate and inaudible.	Visuals are incomprehensible. Very poor handling of visuals e.g. presenter talks to the screen or OHP. Visuals do not support the talk.

ภาคผนวก จ
Advanced Business Oral Communication
(5500314)

วัตถุประสงค์ของแบบสอบถามนี้ คือ เก็บข้อมูลจากนิสิตเกี่ยวกับการวัดและประเมินผล รายวิชาการสนทนาภาษาอังกฤษชั้นสูง (5500314)

กรุณาวางกลม  รอบหมายเลขที่ตรงกับความเห็นของนิสิตมากที่สุด/ เห็นด้วยอย่างยิ่ง ← → น้อยที่สุด/ไม่เห็นด้วยอย่างยิ่ง

5 4 3 2 1

นิสิตได้รับข้อมูลที่ชัดเจนเกี่ยวกับกระบวนการ และวิธีการพูดแบบนำเสนอ (oral presentations) ในด้าน

1. ขั้นตอนการเรียงลำดับข้อมูล (organization)	5	4	3	2	1
2. ความสมบูรณ์และต่อเนื่องของเนื้อหา (content)	5	4	3	2	1
3. การใช้ภาษาที่ถูกต้อง และเหมาะสม (language)	5	4	3	2	1
4. วิธีการนำเสนอ (delivery) เช่น การใช้ภาษาท่าทาง และการใช้เสียง	5	4	3	2	1
5. การใช้สื่อประกอบการนำเสนอ (visual aids)	5	4	3	2	1
6. นิสิตได้รับการฝึกปฏิบัติเกี่ยวกับขั้นตอนการเรียงลำดับข้อมูล (organization) อันประกอบด้วย ส่วน ที่เป็นบทนำ (introduction) ส่วนที่เป็นเนื้อหาข้อมูล (body) และส่วนสรุปจบ (finishing off)	5	4	3	2	1
7. นิสิตได้รับการฝึกปฏิบัติ ในเรื่องความสมบูรณ์ต่อเนื่องของเนื้อหา (content)	5	4	3	2	1
8. นิสิตได้รับการฝึกปฏิบัติในเรื่องการใช้ภาษาที่ถูกต้องและเหมาะสม (language)	5	4	3	2	1
9. นิสิตได้รับการฝึกปฏิบัติในด้านวิธีการนำเสนอ (delivery) เช่น การใช้ภาษาท่าทาง และการใช้เสียง	5	4	3	2	1

10. นิสิตได้รับการฝึกปฏิบัติในด้านการใช้สื่อประกอบ การนำเสนอ (visual aids)	5	4	3	2	1
11. นิสิตได้รับการข้อมูลป้อนกลับ (feedback) เช่น คำอธิบายหรือคำชี้แจงที่ชัดเจน และเป็นประโยชน์หลังจากที่ได้ฝึกปฏิบัติแล้ว เพื่อให้สามารถนำไปปรับปรุงแก้ไข	5	4	3	2	1
12. นิสิตได้นำข้อมูลป้อนกลับจากอาจารย์ผู้สอนมาปรับปรุงแก้ไขการพูดแบบนำเสนอในครั้งต่อไป	5	4	3	2	1
13. วิธีการประเมินผลโดยให้นิสิตพูดแบบนำเสนอในเชิงธุรกิจคนละ 5 นาที ตรงกับวัตถุประสงค์รายวิชา	5	4	3	2	1
14. นิสิตได้รับข้อมูลเกี่ยวกับเกณฑ์การประเมินผลที่ชัดเจน (ตารางแสดงเกณฑ์การประเมิน)	5	4	3	2	1
15. นิสิตเห็นว่าเกณฑ์การประเมิน โดยดูจากหัวข้อหลัก 5 หัวข้อ (organization, content, language, delivery, visual aids) มีความเหมาะสม	5	4	3	2	1
16. นิสิตเห็นว่าวิธีการประเมินโดยอาจารย์ 2 คน (อาจารย์ผู้สอนและอาจารย์จากกลุ่มอื่น) มีความเหมาะสม	5	4	3	2	1
17. นิสิตเห็นว่าวิธีการประเมินผลการพูดแบบนำเสนอ โดยให้นิสิตพูดนั้นเหมาะสมกว่าการวัดผล โดยการสอบข้อเขียน	5	4	3	2	1
18. นิสิตพอใจวิธีการประเมินผลโดยให้นิสิตพูดแบบนำเสนอหน้าชั้นเรียนคนละ 5 นาที	5	4	3	2	1

เหตุผล: _____

ข้อคิดเห็น/เสนอแนะ: _____

ภาคผนวก จ

Advanced Business Oral Communication (ABOC) Teachers' Evaluation of Assessment Criteria

Part I

This part of the questionnaires aims to find the face validity of the test.

Please ○ around the number of your choice.

1. How well do you think your students in presentation class understand the following points:

	most	←————→		least
1.1 Organization	5	4	3	2 1
1.2 Content	5	4	3	2 1
1.3 Language	5	4	3	2 1
1.4 Delivery	5	4	3	2 1
1.5 Visual aids	5	4	3	2 1

2. To what extent do you think students have practiced in

2.1 Organization	5	4	3	2	1
2.2 Content	5	4	3	2	1
2.3 Language	5	4	3	2	1
2.4 Delivery	5	4	3	2	1
2.5 Visual aids	5	4	3	2	1

3. To what extent do your students understand your feedback? 5 4 3 2 1
4. To what extent have students applied the feedback you gave to improve their subsequent presentations? 5 4 3 2 1
5. Do you agree that students' achievement test (a 5-minute presentation) is a valid means of evaluation? 5 4 3 2 1

6. How well do you think your students understand the assessment criteria?	5	4	3	2	1
7. To measure students' presentation skills, a 5-minute oral presentation is more valid than a written presentation.	5	4	3	2	1
8. You prefer listening to students' oral report than marking their written paper.	5	4	3	2	1
9. You think the assessment of five criteria is appropriate.					
9.1 Organization	5	4	3	2	1
9.2 Content	5	4	3	2	1
9.3 Language	5	4	3	2	1
9.4 Delivery	5	4	3	2	1
9.5 Visual aids	5	4	3	2	1
10. You think each descriptor in the above criteria is valid.	5	4	3	2	1

Part II

This part of the questionnaires aims to find the content validity of the rating scales.

Please around your choices.

1. You think the rating scales cover the following areas:

1.1 Organization	agree	questionable	disagree
	1	2	3
1.2 Content	agree	questionable	disagree
	1	2	3
1.3 Language	agree	questionable	disagree
	1	2	3
1.4 Delivery	agree	questionable	disagree
	1	2	3
1.5 Visual aids	agree	questionable	disagree
	1	2	3

- | | | | |
|---|------------|-------------------|---------------|
| 2. The 5-point scales (1-5) are appropriate. | agree
1 | questionable
2 | disagree
3 |
| 3. The length of the descriptors is appropriate (not too long). | agree
1 | questionable
2 | disagree
3 |
| 4. The language used in the descriptors is clear and understandable. | agree
1 | questionable
2 | disagree
3 |
| 5. The rating scales are practical for the achievement test of this course.
(Module 1 : presentations only) | agree
1 | questionable
2 | disagree
3 |
| 6. The rating scales are easy to use for both native and non-native teachers. | agree
1 | questionable
2 | disagree
3 |
| 7. You think each band of the following criteria is appropriate | | | |

Organization

- | | | | |
|------------|------------|-------------------|---------------|
| 7.1 band 1 | agree
1 | questionable
2 | disagree
3 |
| 7.2 band 2 | agree
1 | questionable
2 | disagree
3 |
| 7.3 band 3 | agree
1 | questionable
2 | disagree
3 |
| 7.4 band 4 | agree
1 | questionable
2 | disagree
3 |
| 7.5 band 5 | agree
1 | questionable
2 | disagree
3 |

Content

- | | | | |
|-------------|------------|-------------------|---------------|
| 7.6 band 1 | agree
1 | questionable
2 | disagree
3 |
| 7.7 band 2 | agree
1 | questionable
2 | disagree
3 |
| 7.8 band 3 | agree
1 | questionable
2 | disagree
3 |
| 7.9 band 4 | agree
1 | questionable
2 | disagree
3 |
| 7.10 band 5 | agree
1 | questionable
2 | disagree
3 |

Language

7.11	band 1	agree 1	questionable 2	disagree 3
7.12	band 2	agree 1	questionable 2	disagree 3
7.13	band 3	agree 1	questionable 2	disagree 3
7.14	band 4	agree 1	questionable 2	disagree 3
7.15	band 5	agree 1	questionable 2	disagree 3

Delivery

7.16	band 1	agree 1	questionable 2	disagree 3
7.17	band 2	agree 1	questionable 2	disagree 3
7.18	band 3	agree 1	questionable 2	disagree 3
7.19	band 4	agree 1	questionable 2	disagree 3
7.20	band 5	agree 1	questionable 2	disagree 3

Visual aids

7.21	band 1	agree 1	questionable 2	disagree 3
7.22	band 2	agree 1	questionable 2	disagree 3
7.23	band 3	agree 1	questionable 2	disagree 3
7.24	band 4	agree 1	questionable 2	disagree 3
7.25	band 5	agree 1	questionable 2	disagree 3

Comments / suggestions: _____

ภาคผนวก ข

Write up a complete presentation on AAT using the information provided. (30 points)

84

Auto Alliance (Thailand) Co., Ltd. (AAT)

History & Background	Production Process and Technology	Vision for the Future
<ul style="list-style-type: none"> ● (establish) 28 November 1995 ● a joint venture between Ford Motor Company and Mazda Motor Corporation ● (produce) pick-up trucks, cars, components for local and overseas markets ● Ford, major shareholder, (own) 48% of the company's shares. ● Mazda (own) 45% of the company's shares. ● Mazda Sales (Thailand) Co., Ltd. and the RPN Group (own) 7% of the company's shares. ● (invest) over US\$500 million on building the first state-of-the-art integrated vehicle automobile manufacturing plant in Thailand ● (locate) in the Eastern Seaboard Industrial Estate, Rayong Province ● (grant) promotional privileges by the Board of Investment (BOI) 	<ul style="list-style-type: none"> ● (use) advanced production technology, high – precision robots, a computerized quality inspection system to produce international standard vehicles. ● (conduct) quality inspection at every step of production process: <ul style="list-style-type: none"> - the receipt of raw materials - the process of stamping - body and engine assembly - painting - trimming - final processing to vehicle – performance examination Kd packaging ● (utilise) just-in-time manufacturing plan, vehicle management program, standard delivery and packaging system ● (ensure) that zero-defect trucks and cars (deliver) on time to its customers worldwide 	<ul style="list-style-type: none"> ● confident of a bright future ● a key participant in both the local and global automotive industry ● (exceed) customer expectations ● (achieve) financial strength, long-term growth ● a preferred employer ● (focus) on social and environmental responsibilities ● (meet) the expanding needs for customers in Thailand and around the world

ภาคผนวก ซ

รายละเอียด และข้อมูลเกี่ยวกับผู้ประเมิน

ความสอดคล้องภายใน ของผู้ประเมินจากมาก ไปหาน้อย	ภาษาของ ผู้ประเมิน	วัยวุฒิ	เพศ	ประสบการณ์ การสอนและ การประเมินผล
1. ผู้ประเมิน 3	อังกฤษ	51	ชาย	17 ปี
2. ผู้ประเมิน 1	ไทย	37	หญิง	11 ปี
3. ผู้ประเมิน 2	ไทย	47	หญิง	20 ปี
4. ผู้ประเมิน 4	อังกฤษ	44	ชาย	10 ปี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย