

การศึกษาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ
ของพนักงานต้อนรับบนเครื่องบิน บริษัท การบินไทย จำกัด (มหาชน)

สารนิพนธ์
ของ
วัชรพงษ์ กริบุตร

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต สาขาวิชาการสอนภาษาอังกฤษในฐานะภาษาต่างประเทศ
ตุลาคม 2549

การศึกษาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ
ของพนักงานต้อนรับบนเครื่องบิน บริษัท การบินไทย จำกัด (มหาชน)

สารนิพนธ์
ของ
วัชรพงษ์ กรีบุตร

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต สาขาวิชาการสอนภาษาอังกฤษในฐานะภาษาต่างประเทศ
ตุลาคม 2549
ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การศึกษาคูณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ
ของพนักงานต้อนรับบนเครื่องบิน บริษัท การบินไทย จำกัด (มหาชน)

บทคัดย่อ
ของ
วัชรพงษ์ กริบุตร

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต สาขาวิชาการสอนภาษาอังกฤษในฐานะภาษาต่างประเทศ
ตุลาคม 2549

วัชรพงษ์ กริบุตร. (2549). การศึกษาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบิน บริษัท การบินไทย จำกัด (มหาชน). สารนิพนธ์ ศศ.ม. (การสอนภาษาอังกฤษในฐานะภาษาต่างประเทศ). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อาจารย์ที่ปรึกษาสารนิพนธ์: ผู้ช่วยศาสตราจารย์ เนลิยวศรี พิบูลชล.

การศึกษาค้นคว้าครั้งนี้ มีจุดมุ่งหมายเพื่อศึกษาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ (Placement test) ของพนักงานต้อนรับบนเครื่องบิน บริษัท การบินไทย จำกัด (มหาชน) กลุ่มตัวอย่างเป็นพนักงานต้อนรับบนเครื่องบินที่เริ่มปฏิบัติงานปี พ.ศ. 2546 ทั้งหญิงและชายที่เข้ารับการทดสอบความสามารถทางภาษาอังกฤษ จากกองฝึกอบรมภาษา บริษัทการบินไทย จำกัด (มหาชน) จำนวน 489 คน ซึ่งเป็นกลุ่มตัวอย่างที่ถูกเลือกแบบเจาะจง เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ ประเภทเลือกตอบ (Multiple choices) จำนวน 90 ข้อ โดยแบ่งเป็น 3 ด้าน คือ แบบทดสอบแบบ การทดสอบการใช้ภาษาเพื่อการสื่อสาร ประเภทเลือกตอบ แบบทดสอบความเข้าใจในการอ่าน และ คำศัพท์ และแบบทดสอบโครงสร้างทางไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทาง ไวยากรณ์ ประเภทเลือกเติมคำ นำมาศึกษาคุณภาพของแบบทดสอบด้านความเที่ยงตรง ด้านค่า อำนาจจำแนก ค่าความยากง่าย ความเชื่อมั่น ตลอดจนการพิจารณาความสอดคล้องของ แบบทดสอบจัดระดับกับแบบทดสอบ TOEIC และสร้างเกณฑ์ปกติของแบบทดสอบจัดระดับในรูป คะแนนที่-ปกติ โดยมีผลการวิจัยดังนี้

1. ค่าความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบจัดระดับความสามารถในการใช้ ภาษาอังกฤษ ที่ได้จากการหาค่าสัมประสิทธิ์ความสอดคล้องของความสัมพันธ์ระหว่างคะแนนราย ข้อกับคะแนนรวมทั้งฉบับ โดยใช้สูตรการหาค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple correlation) ของเพียร์สัน (Pearson product moment correlation) มีค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง -.230 ถึง .395

2. ค่าความยากง่ายของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ มีค่า อยู่ระหว่าง .11 ถึง .99

3. ค่าอำนาจจำแนกของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ โดยใช้เทคนิค 27% ของ จุง เทห์ ฟาน (Chung The Fan) มีค่าอำนาจจำแนกอยู่ระหว่าง -.26 ถึง .66

4. ค่าความเชื่อมั่นของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ จากค่า สัมประสิทธิ์สหสัมพันธ์ของความเชื่อมั่น (Reliability coefficient) โดยใช้สูตร K-R 20 ของคูเดอร์และ ริชาร์ดสัน (Kuder-Richardson) มีค่าความเชื่อมั่นเท่ากับ .80

5. ค่าความสอดคล้องของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษกับ แบบทดสอบ TOEIC โดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson product moment correlation) มีค่าเท่ากับ .585 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

6. เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษมีเกณฑ์
คะแนนที่ปกติอยู่ระหว่าง T10 ถึง T82

A STUDY OF THE ENGLISH PLACEMENT TEST FOR FLIGHT ATTENDANTS,
THAI AIRWAYS INTERNATIONAL PUBLIC COMPANY LIMITED

AN ABSTRACT
BY
WACHARAPONG KRIBUTR

Presented in Partial Fulfillment of the Requirements
for the Master of Arts in Teaching English as a Foreign Language
at Srinakharinwirot University

October 2006

Wacharapong Kributr. (2006). *A Study of The English Placement Test for Flight Attendants, Thai Airways International Public Company Limited*. Master's Project, M.A. (Master of Arts). Bangkok: Graduate School, Srinakharinwirot University. Project Advisor: Assist. Prof. Chaleosri Pibulchol.

The purpose of this research was to study the quality of an English placement test for flight attendants of Thai Airways International (PLC). The sample was consisted of 489 flights attendants, both male and female, who started working since 2003 and took the test from Language Training Department. This sample was selected by the purposive sampling technique. The instrument of this research was the placement test consisting of 90 multiple-choice items. The test was divided into 3 sections: Language communication test (Service language) 20 items, Reading comprehension and vocabulary test 40 items, and Grammar with error recognition test 30 items.

The quality of the test was examined by construct validity, discrimination power, difficulty, reliability as well as the correlation between the placement test and the TOEIC test. Moreover, normalized T-score was used to construct test norms.

The results of the research indicated that:

1. The construct validity of the placement test for flight attendants was calculated by the Pearson product moment correlation formula. The correlation coefficient between each test item score and the total test score ranged from -.230 to .395
2. The difficulty of the placement test for flight attendants ranged from .11 to .99
3. The discrimination power of the placement test for flight attendants using 27% of Chung The Fan's technique ranged from -.26 to .66
4. The reliability of the placement test for flight attendants using the Kuder-Richardson formula 20 (K-R 20) was .80
5. The correlation between the placement test for flight attendants and the TOEIC test using the Pearson correlation coefficient was .585 at .01 level of significance.
6. The norm of the placement test for flight attendants had the normalized T-score ranged from T10 to T82

อาจารย์ที่ปรึกษาสารนิพนธ์ ประธานคณะกรรมการบริหารหลักสูตร และคณะกรรมการ
สอบได้พิจารณาสารนิพนธ์เรื่อง การศึกษาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้
ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบิน บริษัท การบินไทย จำกัด (มหาชน) ของ วัชรพงษ์ กรีบุตร
ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรมหาบัณฑิต
สาขาวิชาการสอนภาษาอังกฤษในฐานะภาษาต่างประเทศ ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

อาจารย์ที่ปรึกษาสารนิพนธ์

.....
(ผู้ช่วยศาสตราจารย์ เฉลียวศรี พิบูลชล)

ประธานคณะกรรมการบริหารหลักสูตร

.....
(ดร. วัลลภา ไทยจินดา)

คณะกรรมการสอบ

..... ประธาน
(ผู้ช่วยศาสตราจารย์ เฉลียวศรี พิบูลชล)

..... กรรมการสอบสารนิพนธ์
(อาจารย์ เตือนใจ เฉลิมกิจ)

..... กรรมการสอบสารนิพนธ์
(ผู้ช่วยศาสตราจารย์ จิราภรณ์ บุญส่ง)

อนุมัติให้รับสารนิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
ศิลปศาสตรมหาบัณฑิต สาขาวิชาการสอนภาษาอังกฤษในฐานะภาษาต่างประเทศ ของ
มหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีคณะมนุษยศาสตร์
(ผู้ช่วยศาสตราจารย์ เฉลียวศรี พิบูลชล)
วันที่ เดือน ตุลาคม พ.ศ. 2549

ประกาศคุณูปการ

สารนิพนธ์ฉบับนี้สำเร็จลุล่วงสมบูรณ์ได้ด้วยความกรุณาอย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ เฉลียวศรี พิบูลชล ประธานกรรมการควบคุมสารนิพนธ์ อาจารย์ เตือนใจ เฉลิมกิจ กรรมการควบคุมสารนิพนธ์ ท่านทั้งสองได้ให้คำปรึกษาแนะนำชี้แนะในทุกขั้นตอนของการจัดทำวิจัยครั้งนี้เพื่อให้ได้เนื้อหาสาระที่ถูกต้องสมบูรณ์ รวมทั้งคณาจารย์ภาควิชาภาษาตะวันตกทุกท่านที่ได้ประสิทธิประสาทวิชาความรู้ตลอดระยะเวลาของการศึกษา จนผู้วิจัยสามารถนำความรู้ที่ได้รับมาปรับและประยุกต์ใช้ในการดำเนินการทำสารนิพนธ์จนสำเร็จ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง ณ โอกาสนี้

นอกจากนี้ผู้วิจัยขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ จิราภรณ์ บุญส่ง ที่กรุณาให้ความช่วยเหลือ แนะนำในการศึกษาคุณภาพของเครื่องมือ ทำให้ผู้วิจัยได้เพิ่มพูนความรู้ ความเข้าใจ ตลอดจนขั้นตอนและวิธีการต่างๆทางสถิติที่ใช้ในงานวิจัย

การอนุเคราะห์ข้อมูลต่างๆเพื่อใช้ในการศึกษาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบิน บริษัทการบินไทย จำกัด (มหาชน) และการอำนวยความสะดวกในการเก็บรวบรวมข้อมูล ผู้วิจัยได้รับความช่วยเหลือสนับสนุน และความอนุเคราะห์จาก อาจารย์ เพ็ญสุข วัฒนวงศ์ ผู้อำนวยการสถาบันภาษา บริษัทการบินไทย จำกัด (มหาชน) และคณะครูทุกท่านที่อนุเคราะห์ข้อมูลและข้อเสนอแนะที่มีคุณค่าในงานวิจัยครั้งนี้

ผู้วิจัยขอโน้มรำลึกและกราบขอบพระคุณทุกท่านที่มีส่วนช่วยให้สารนิพนธ์ฉบับนี้สำเร็จลุล่วงอย่างสมบูรณ์ตามจุดหมายทุกประการ

วัชรพงษ์ กรีบุตร

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง.....	1
ความมุ่งหมายของการวิจัย.....	3
ความสำคัญของการวิจัย.....	4
ขอบเขตของการวิจัย.....	4
นิยามศัพท์เฉพาะ.....	5
กรอบแนวคิดในการวิจัย.....	6
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	7
การวัดและการประเมินผล.....	7
ความหมายของการวัด และการประเมินผล.....	7
จุดประสงค์ของการวัด และการประเมินผล.....	8
เครื่องมือวัดผลการศึกษา.....	9
การทดสอบ.....	9
ความหมายของการทดสอบ.....	9
แบบทดสอบทางภาษา.....	10
ประเภทของแบบทดสอบทางภาษา.....	10
การสร้างแบบทดสอบความสามารถทางภาษา.....	15
แบบทดสอบมาตรฐาน.....	22
ความหมายของแบบทดสอบมาตรฐาน.....	22
คุณภาพของแบบทดสอบ.....	23
เกณฑ์ปกติ.....	33
คะแนนมาตรฐาน.....	33
วิธีสร้างเกณฑ์ปกติคะแนนที่.....	37
งานวิจัยที่เกี่ยวข้อง.....	39
3 วิธีการดำเนินการวิจัย.....	45
การกำหนดประชากรและการเลือกกลุ่มตัวอย่าง.....	45
เครื่องมือที่ใช้ในการวิจัย.....	45

สารบัญ (ต่อ)

บทที่	หน้า
3 (ต่อ)	
การเก็บรวบรวมข้อมูล.....	46
การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล.....	47
4 ผลการวิเคราะห์ข้อมูล.....	49
สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล.....	49
การเสนอผลการวิเคราะห์ข้อมูล.....	49
ผลการวิเคราะห์ข้อมูล.....	50
5 สรุปผล อภิปรายผล และข้อเสนอแนะ	65
ความมุ่งหมายของการศึกษาค้นคว้า.....	65
การกำหนดประชากรและการเลือกกลุ่มตัวอย่าง.....	65
เครื่องมือที่ใช้ในการวิจัย.....	65
การเก็บรวบรวมข้อมูล.....	66
การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล.....	66
สรุปผลการวิจัย.....	67
อภิปรายผลการวิจัย.....	68
ข้อเสนอแนะทั่วไป.....	72
ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป.....	72
บรรณานุกรม.....	73
ภาคผนวก.....	79
ประวัติย่อผู้ทำสารนิพนธ์.....	84

บัญชีภาพประกอบ

ภาพประกอบ

หน้า

1	การกระจายปกติ.....	13
2	ขั้นตอนของการสร้างข้อทดสอบทางภาษา.....	16
3	กราฟแสดงการสร้างเกณฑ์ปกติ และการขยายคะแนนที่ปกติ.....	38
4	กราฟแสดงความสัมพันธ์ระหว่างคะแนนดิบและคะแนนมาตรฐานที่-ปกติ แบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ.....	80
5	กราฟแสดงความสัมพันธ์ระหว่างคะแนนดิบและคะแนนมาตรฐานที่-ปกติ แบบทดสอบการใช้ภาษาเพื่อการสื่อสารประเภทเลือกตอบ.....	81
6	กราฟแสดงความสัมพันธ์ระหว่างคะแนนดิบและคะแนนมาตรฐานที่-ปกติ แบบทดสอบความเข้าใจในการอ่าน และคำศัพท์.....	82
7	กราฟแสดงความสัมพันธ์ระหว่างคะแนนดิบและคะแนนมาตรฐานที่-ปกติ แบบทดสอบด้านโครงสร้างทางไวยากรณ์ทางภาษาประเภท การหา ข้อบกพร่องทางไวยากรณ์ และประเภทเลือกเติมคำ.....	83

บัญชีตาราง

ตาราง		หน้า
1	เกณฑ์ประเมินระดับความสามารถของคะแนนมาตรฐานที่ปกติ	5
2	ค่าและความหมายความยากง่าย	31
3	เกณฑ์การพิจารณาค่าอำนาจจำแนก	32
4	ตัวอย่างการแปลงคะแนนดิบให้เป็นเปอร์เซ็นต์ไทล์ และคะแนนที่ปกติ.....	35
5	ตารางสำหรับแปลงคะแนนที่-ปกติ ให้เป็นจำนวนร้อยละ ที่อยู่เหนือกว่าผู้อื่น.....	36
6	ผลการวิเคราะห์จำนวนข้อของแบบทดสอบที่มีความสัมพันธ์ระหว่างคะแนน รายข้อกับคะแนนรวมทั้งฉบับ.....	50
7	ค่าความยากง่ายและค่าอำนาจจำแนกของแบบทดสอบความสามารถ ในการใช้ภาษาอังกฤษของพนักงานต้อนรับ บริษัท การบินไทย จำกัด (มหาชน)	52
8	สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรวมทั้งฉบับกับคะแนนรายข้อ ค่าความยากง่ายและค่าอำนาจจำแนกเป็นรายข้อ.....	54
9	ค่าสถิติพื้นฐานและค่าความเชื่อมั่นของแบบทดสอบแต่ละฉบับ และแบบทดสอบรวมทุกฉบับของแบบทดสอบความสามารถในการใช้ ภาษาอังกฤษ (Placement test).....	57
10	ค่าสัมประสิทธิ์สหสัมพันธ์ของแบบทดสอบแต่ละฉบับ กับแบบทดสอบ TOEIC.....	58
11	เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของ พนักงานต้อนรับบนเครื่องบิน ในรูปของคะแนนมาตรฐานที่-ปกติ.....	60
12	เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของ พนักงานต้อนรับบนเครื่องบิน ด้านการใช้ภาษาเพื่อการสื่อสาร ประเภทเลือกตอบ ในรูปของคะแนนมาตรฐานที่-ปกติ.....	61
13	เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของ พนักงานต้อนรับบนเครื่องบิน ด้านความเข้าใจในการอ่าน และคำศัพท์ ในรูปของคะแนนมาตรฐานที่-ปกติ.....	62

บัญชีตาราง (ต่อ)

ตาราง

หน้า

14	เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของ พนักงานต้อนรับบนเครื่องบิน ด้านโครงสร้างไวยากรณ์ทางภาษา ประเภทการหาข้อบกพร่องทางไวยากรณ์ และประเภทเลือกเติมคำ ในรูปของคะแนนมาตรฐานที่-ปกติ.....	63
----	--	----

บทที่ 1

บทนำ

ภูมิหลัง

บริษัท การบินไทย จำกัด (มหาชน) เป็นหน่วยงานรัฐวิสาหกิจที่ก่อตั้งเพื่อดำเนินธุรกิจการบินพาณิชย์ ในฐานะสายการบินแห่งชาติ ให้การบริการผู้โดยสารทั้งชาวไทย และชาวต่างประเทศ ตลอดจนให้การบริการธุรกิจเกี่ยวเนื่องที่เกี่ยวข้องกับการบิน ทำการติดต่อสื่อสารกับบุคคลต่างๆ ทั่วโลกที่เกี่ยวข้องทั้งทางตรงและทางอ้อมตลอด 24 ชั่วโมง ดังนั้นภาษาจึงเป็นเครื่องมือที่สำคัญที่สุดในการติดต่อสื่อสารระหว่างกันเพื่อให้ธุรกิจประสบความสำเร็จลุล่วงตามวัตถุประสงค์ที่ตั้งไว้ นำรายได้และความเจริญสู่ประเทศ

ภาษาอังกฤษเป็นภาษาสากลที่ใช้กันแพร่หลายทั่วโลก และเป็นภาษาหลักที่บริษัทใช้ติดต่อสื่อสารกับหน่วยงานต่างๆ และบุคคลต่างๆ ทั่วโลก ดังนั้นบุคลากรที่ปฏิบัติงานกับบริษัทการบินไทยจึงต้องเป็นบุคคลที่มีความรู้ ความสามารถที่จะติดต่อสื่อสารกับบุคคลต่างๆ ได้อย่างมีประสิทธิภาพ โดยเฉพาะการใช้ภาษาอังกฤษซึ่งเป็นภาษาหลักที่ใช้ในการติดต่อสื่อสาร

พนักงานต้อนรับบนเครื่องบิน (Flight attendant) เป็นบุคลากรกลุ่มหนึ่งของบริษัทที่ปฏิบัติหน้าที่บนเครื่องบิน และต้องเดินทางไปตามเส้นทางการบินของบริษัทเกือบทั่วโลก บุคลากรเหล่านี้จำเป็นต้องใช้ภาษาอังกฤษได้อย่างมีประสิทธิภาพในการติดต่อสื่อสารตลอดเวลาทั้งกับเจ้าหน้าที่ภาคพื้นดินในประเทศต่างๆ ทั่วโลก ตลอดจนผู้โดยสาร พนักงานโรงแรม และบุคคลต่างๆ ที่เกี่ยวข้อง เป็นต้น

ความสามารถในการใช้ภาษาอังกฤษได้อย่างดีนั้นเป็นคุณสมบัติประการหนึ่งที่บริษัทการบินไทยได้กำหนดขึ้นเพื่อให้พนักงานต้อนรับบนเครื่องบินมีความสามารถในการใช้ภาษาอังกฤษอยู่ในระดับมาตรฐาน โดยกำหนดว่าจะต้องจบการศึกษาระดับปริญญาตรี และมีผลคะแนนสอบ TOEIC ตั้งแต่ 600 คะแนนขึ้นไป (Thai Airways International PLC. 2003: Online) ซึ่งในระดับนี้ผู้รับการทดสอบจะมีความสามารถในการใช้ภาษาอังกฤษได้มากเป็นที่น่าสนใจ ในการตอบคำถามไม่ว่าการฟัง หรือการอ่านอยู่ในเกณฑ์ที่ดีทั้งหมด (สุทิน พูลสวัสดิ์. 2546: 16) นอกจากนี้บริษัทได้ใช้การทดสอบสัมภาษณ์ภาษาอังกฤษมาเป็นเกณฑ์ในการคัดเลือกพนักงานต้อนรับบนเครื่องบินด้วย

จากสภาพการณ์ที่เป็นมาพบว่า พนักงานต้อนรับบนเครื่องบินที่ผ่านการทดสอบบางคนยังไม่สามารถใช้ภาษาอังกฤษได้ดีตามความต้องการของงาน ซึ่งอาจเกิดจากหลายสาเหตุดังต่อไปนี้ (เพิ่มสุข วิษณุวงศ์. 2547 : สัมภาษณ์)

1. คุณวุฒิทางการศึกษาที่แตกต่างกัน บริษัท การบินไทย จำกัด (มหาชน) ได้กำหนดคุณวุฒิทางการศึกษาของพนักงานต้อนรับบนเครื่องบินที่เข้าก่อนปี พ.ศ. 2543 ไว้เพียงมัธยมศึกษาตอนปลายหรือเทียบเท่า และเพิ่งเริ่มรับบัณฑิตวุฒิปริญญาตรีมาปฏิบัติหน้าที่นี้ครั้งแรกในปี 2546

2. การกำหนดใช้คะแนนภาษาอังกฤษเป็นตัวคัดเลือก ปรับเปลี่ยนมาตลอด กลุ่มพนักงานต้อนรับบนเครื่องบินในรุ่นแรกๆ ที่ยังปฏิบัติหน้าที่อยู่ในขณะนี้บ้างนั้น ใช้คะแนนสอบสัมภาษณ์อย่างเดียว ในขณะที่รุ่นต่อมาเริ่มใช้คะแนนข้อสอบสากลเช่น TOEIC, IELTS และ/หรือ TOEFL มาเป็นตัวกำหนด โดยใช้ระดับคะแนนคือ ในระยะแรก TOEIC 450 คะแนน, TOEFL 400 คะแนน และ IELTS ระดับ 5 จากระดับคะแนนที่ไม่คำนึงถึงระดับความยากง่ายและความแตกต่างในเนื้อหาของข้อสอบแต่ละชุดนี้เอง ผู้สมัครส่วนใหญ่จะเข้ามาด้วยคะแนน TOEIC ซึ่งผลการปฏิบัติงานแสดงให้เห็นเด่นชัดว่า ผู้ได้รับการคัดเลือกเข้ามาโดยใช้คะแนน TOEIC มีปัญหาในการใช้ภาษาอังกฤษเป็นอย่างมาก ฝ่ายบุคลากรของบริษัทฯ จึงได้ปรับระดับคะแนน TOEIC ที่ต้องการให้สูงขึ้นมาถึง 3 ครั้ง คือ 500 เป็น 550 และ 600 คะแนน สำหรับรุ่นสุดท้าย คือรุ่นปี 2546

จากการสัมภาษณ์ เพิ่มสุข วิษณุวงศ์ ซึ่งเป็นผู้จัดการกองฝึกอบรมภาษา บริษัทการบินไทย จำกัด (มหาชน) ได้ให้ความเห็นเพิ่มเติมว่า ผลของการปรับระดับคะแนน TOEIC เพื่อการคัดเลือกให้สูงขึ้นยังไม่ใช่วิธีแก้ปัญหาการขาดความพร้อมของการสื่อสารที่ใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบิน มีผู้สมัครหลายคนสอบผ่านเข้ามาด้วยคะแนนที่สูงกว่าคะแนนที่กำหนด และมีเป็นจำนวนมากที่เข้ามาด้วยคะแนน TOEIC ตามที่กำหนดหรือสูงกว่าเล็กน้อย แต่พบว่าพนักงานต้อนรับบนเครื่องบินเกินกว่า 60% ที่ผ่านคะแนน TOEIC ตามความต้องการของบริษัทไม่สามารถใช้ภาษาอังกฤษได้ดีพอกับความต้องการของงาน จึงเป็นปัญหาหลักของหน่วยงาน และยังผลให้ยากที่จะพัฒนางานให้ได้ตามจุดมุ่งหมาย ทางกองฝึกอบรมภาษาจึงได้พิจารณาถึงสาเหตุของปัญหาที่เกิดขึ้นและเห็นว่า การใช้คะแนนข้อทดสอบของ TOEIC นั้น น่าจะไม่ได้เป็นตัวชี้วัดถึงความสามารถที่แท้จริงของผู้รับการทดสอบ มีพนักงานต้อนรับบนเครื่องบินจำนวนมากที่สอบได้คะแนนตามข้อกำหนดด้วยการฝึกฝน (Coaching) และผ่านการสอบเป็นจำนวนหลายๆ ครั้งเพื่อให้ได้คะแนนตามเกณฑ์ที่กำหนด

บริษัท การบินไทย จำกัด (มหาชน) มีนโยบายให้พนักงานต้อนรับบนเครื่องบินรักษาระดับมาตรฐานความสามารถในการใช้ภาษาอังกฤษให้อยู่ในเกณฑ์ที่น่าพอใจตลอดไป จึงได้กำหนดให้บุคลากรดังกล่าวที่ปฏิบัติการบินต้องเข้ารับการอบรมภาษาอังกฤษอย่างต่อเนื่องโดยมีกองฝึกอบรมภาษา (Flight Crew Language Training Department) เป็นหน่วยงานรับผิดชอบฝึกฝนและอบรมการใช้ภาษาอังกฤษที่เกี่ยวข้องกับงานอาชีพโดยตรง การกิจหลักของกองอบรมภาษาฝ่ายปฏิบัติการบริษัทการบินไทยคือ การจัดอบรมภาษาอังกฤษให้แก่พนักงานต้อนรับบนเครื่องบินโดยหลักสูตรที่จัดทำจะเน้นเนื้อหาการเรียนการสอนที่เป็นเรื่องเฉพาะอาชีพหรือที่เรียกว่า English for Specific Purposes (ESP) รวมถึงภาษาอังกฤษทั่วไปที่จะต้องใช้ในชีวิตประจำวัน (Everyday English) ที่เหมาะสมกับระดับความสามารถ และประสบการณ์ที่แท้จริงของผู้เรียนทั้งทักษะ การฟัง การพูด การอ่าน และการเขียน หลักสูตรที่กองฝึกอบรมภาษาจัดทำขึ้นมีเนื้อหาและจุดมุ่งหมายให้พนักงานต้อนรับบนเครื่องบินได้พัฒนาศักยภาพการใช้ภาษาอังกฤษในทุกทักษะ โดยคำนึงถึงการใช้ภาษาอังกฤษในการสื่อสารผ่านสถานการณ์และเหตุการณ์ต่างๆ ที่เกิดขึ้นจริงบนเครื่องบิน (Situational based) ตลอดจนธุรกิจที่เกี่ยวข้องกับการบิน (Airline business) ตั้งแต่ระดับพื้นฐาน ระดับกลาง และระดับสูง

ในการจัดพนักงานเข้ารับการอบรมตามหลักสูตรต่างๆ กองฝึกอบรมภาษามีความจำเป็นที่จะต้องจัดระดับความสามารถทางภาษาอังกฤษของพนักงานต้อนรับอีกครั้งแม้ว่าพนักงานเหล่านี้จะผ่านการสอบ TOEIC ซึ่งถือว่าเป็นแบบทดสอบมาตรฐานแล้วก็ตาม โดยพบว่าพนักงานยังไม่สามารถใช้ภาษาเพื่อสื่อสารได้ตามความต้องการของงาน ทั้งนี้อาจเป็นเพราะว่าเนื้อหาส่วนใหญ่ของแบบทดสอบ TOEIC ไม่เน้นไปในงานอาชีพสาขาใดสาขาหนึ่งโดยเฉพาะ เป็นการวัดความสามารถในการใช้ภาษาเพื่อการปฏิบัติงานโดยทั่วไป

จากเหตุผลข้างต้นกองฝึกอบรมภาษาจึงได้จัดทำแบบทดสอบจัดระดับภาษาอังกฤษ (Placement test) ขึ้น และได้ทดสอบพนักงานต้อนรับบนเครื่องบินรุ่นใหม่ที่ใช้ปฏิบัติงานในปี พ.ศ. 2546 ตั้งแต่วันที่ 1 ตุลาคม พ.ศ. 2546 – 27 มกราคม พ.ศ. 2547 เพื่อจัดระดับของผู้เรียนให้เข้าเรียนตามหลักสูตรที่เหมาะสมกับความสามารถและความต้องการทางภาษาของแต่ละบุคคล อย่างไรก็ตามแบบทดสอบดังกล่าวยังมิได้เคยมีการศึกษาหาคุณภาพของแบบทดสอบเพื่อคัดเลือกข้อทดสอบที่เหมาะสม มีความเที่ยงตรงเชื่อถือได้ เนื่องจากกลุ่มตัวอย่างที่ใช้ในการศึกษาที่เป็นพนักงานต้อนรับบนเครื่องบินรุ่นใหม่มีเพียงพนักงานที่ปฏิบัติงานในปี พ.ศ. 2546 เท่านั้น ผู้วิจัยจึงได้มีความสนใจที่จะศึกษาหาคุณภาพของแบบทดสอบดังกล่าวเพื่อคัดเลือกข้อทดสอบที่มีคุณภาพ พร้อมทั้งศึกษาแนวทางในการสร้างแบบทดสอบที่มีคุณภาพขึ้นใช้เองภายในองค์กร ตลอดจนเกณฑ์ในการจัดระดับความสามารถของพนักงานต้อนรับบนเครื่องบินที่มีจำนวนเพิ่มขึ้นได้อย่างเหมาะสม รวดเร็วและมีประสิทธิภาพ อันจะเป็นประโยชน์สูงสุด และสนองจุดมุ่งหมายในการฝึกอบรมให้กับพนักงานต้อนรับบนเครื่องบินรุ่นใหม่ในอนาคตต่อไป

ความมุ่งหมายของการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยได้ตั้งความมุ่งหมายไว้ดังนี้

1. เพื่อคัดเลือกข้อทดสอบจัดระดับความสามารถภาษาอังกฤษ (Placement test) สำหรับพนักงานต้อนรับบนเครื่องบิน บริษัทการบินไทยจำกัด (มหาชน) ที่มีคุณภาพจัดเก็บไว้ในคลังข้อสอบ (Test bank)
2. เพื่อสร้างเกณฑ์ปกติ (Norm) ของแบบทดสอบสำหรับเป็นเกณฑ์ในการกำหนดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบินว่ามีความสามารถอยู่ในระดับใด อันจะนำไปเป็นแนวทางในการจัดกลุ่มของพนักงานต้อนรับบนเครื่องบินเข้ารับการอบรมในหลักสูตรภาษาอังกฤษได้อย่างเหมาะสมต่อไป

ความสำคัญของการวิจัย

ผลจากการวิจัยที่จัดทำโดยกองฝึกอบรมภาษา บริษัท การบินไทย จำกัด (มหาชน) จะเป็นประโยชน์ในการสร้างแบบทดสอบจัดระดับภาษาอังกฤษ (Placement test) ที่มีคุณภาพและใช้จัดระดับพนักงานต้อนรับบนเครื่องบินตลอดจนให้เข้ารับการอบรมในหลักสูตรที่เหมาะสมตรงตามความสามารถทางภาษาในการนำไปปฏิบัติงานจริงอีกทั้งทางกองฝึกอบรมภาษาสามารถนำผลการวิจัยดังกล่าวไปเป็นเกณฑ์ในการสร้างแบบทดสอบจัดระดับภาษาอังกฤษในทักษะด้านอื่นๆ ให้กับพนักงานต้อนรับบนเครื่องบินของบริษัทอย่างมีประสิทธิภาพ

ในอนาคตกองฝึกอบรมภาษายังจะต้องมีหน้าที่ในการจัดการฝึกอบรมภาษาอังกฤษให้กับพนักงานในส่วนอื่นๆของบริษัท เช่น พนักงานจากกองสำรองที่นั่งและบัตรโดยสาร (Reservation and Ticketing Department) ฝ่ายช่าง (Technical Department) และพนักงานในส่วนอื่นๆที่มีความต้องการที่จะพัฒนาความสามารถในการใช้ภาษาอังกฤษของตน แนวทางในการพัฒนาแบบทดสอบดังกล่าวจะสามารถทำให้กองฝึกอบรมภาษาใช้เป็นแนวทางและเกณฑ์ในการสร้างแบบทดสอบที่มีคุณภาพ ความเที่ยงตรงและเชื่อมั่นให้เหมาะสมกับความต้องการของหน่วยงานในบริษัทต่อไป

ขอบเขตของการวิจัย

ประชากรที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นพนักงานต้อนรับบนเครื่องบินของ บริษัท การบินไทย จำกัด (มหาชน)

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นกลุ่มตัวอย่างที่ถูกเลือกแบบเจาะจง จากพนักงานต้อนรับบนเครื่องบินรุ่นใหม่ของบริษัท การบินไทย จำกัด (มหาชน) ที่เริ่มปฏิบัติงานในปี พ.ศ. 2546 ทั้งหญิงและชายที่เข้าทดสอบความสามารถทางภาษาอังกฤษ (Placement test) จากกองฝึกอบรมภาษา บริษัท การบินไทย จำกัด (มหาชน) จำนวน 489 คน

เนื้อหาที่ใช้ในการทดลอง

แบบทดสอบที่สร้างขึ้นเป็นแบบทดสอบจัดระดับความสามารถทางภาษาอังกฤษที่มีเนื้อหาครอบคลุมตามแนวหลักสูตรของกองฝึกอบรมภาษาในด้านต่างๆ เช่น การใช้ภาษาในการสื่อสารเพื่อให้บริการบนเครื่องบิน ภาษาที่เกี่ยวข้องกับธุรกิจการบินและความรู้ด้านวัฒนธรรมต่างๆที่สำคัญ แบบทดสอบดังกล่าวเป็นข้อสอบแบบปรนัย (Multiple choices) มีความยากง่ายในการใช้ภาษาทั้งในระดับเบื้องต้น ระดับกลาง และระดับสูง

นิยามศัพท์เฉพาะ

1. พนักงานต้อนรับบนเครื่องบิน หมายถึง พนักงานของบริษัท การบินไทย จำกัด (มหาชน) ที่ปฏิบัติหน้าที่บนเครื่องบิน ในปี พ.ศ. 2546 ทั้งหญิงและชายทำหน้าที่ให้บริการ และอำนวยความสะดวกตลอดจนดูแลด้านความปลอดภัยให้กับผู้โดยสารตามที่บริษัทได้กำหนดไว้
2. แบบทดสอบจัดระดับความสามารถทางภาษาอังกฤษ (English placement test) หมายถึงแบบทดสอบความสามารถทางภาษาอังกฤษของกองฝึกอบรมภาษา (Language Training Department) บริษัทการบินไทยจำกัด (มหาชน) ที่จัดทำขึ้นเพื่อใช้จัดกลุ่มพนักงานต้อนรับบนเครื่องบินตามความสามารถในการใช้ภาษาอังกฤษเพื่อการสื่อสาร (Communicative English)
3. คุณภาพของแบบทดสอบ หมายถึง คุณลักษณะของแบบทดสอบจัดระดับความสามารถทางภาษาอังกฤษของกองฝึกอบรมภาษา บริษัท การบินไทย จำกัด (มหาชน) ที่มีความเที่ยงตรงทั้งด้านโครงสร้างและเนื้อหา มีความเชื่อมั่นในการวัด มีความยากง่ายพอเหมาะ (ค่าความยากง่ายระหว่าง 0.2-0.8) และสามารถจำแนกพนักงานต้อนรับบนเครื่องบินได้ (ค่าอำนาจจำแนกตั้งแต่ 0.2 ขึ้นไป)
4. เกณฑ์ปกติ (Norm) หมายถึง ตารางความสัมพันธ์ระหว่างคะแนนดิบกับคะแนนมาตรฐานที่ (T-score) เพื่อใช้เป็นเกณฑ์ในการแปลงคะแนนดิบที่ได้จากการสอบของผู้รับการทดสอบเป็นคะแนนมาตรฐานที่ เพื่อนำไปใช้ในการจัดระดับความสามารถของผู้รับการทดสอบตามเกณฑ์ประเมินในตาราง 1

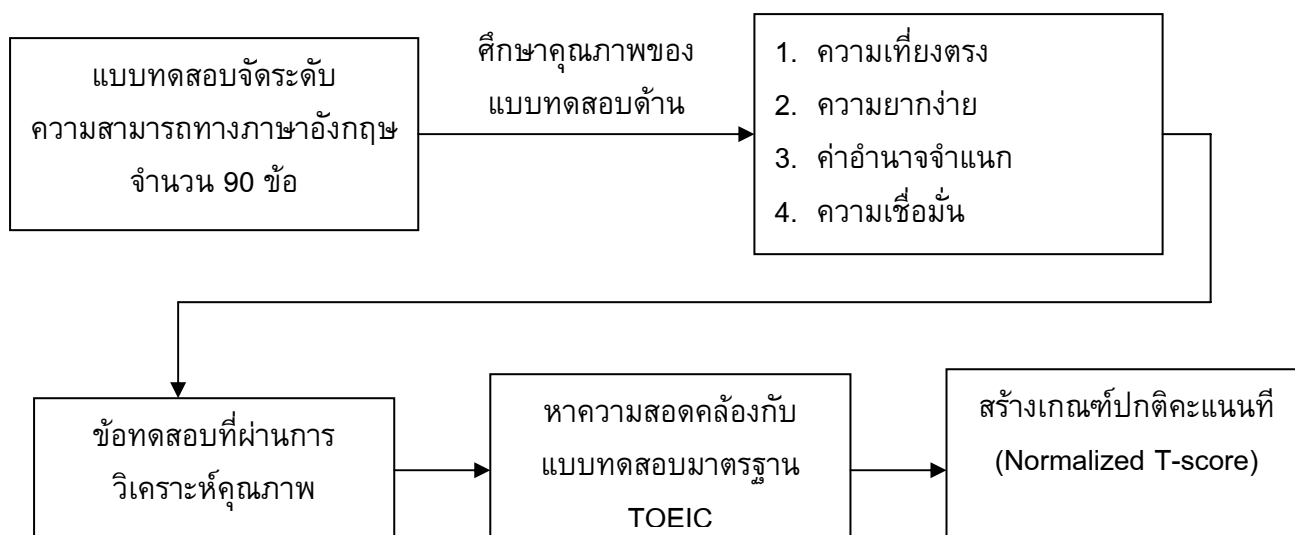
ตาราง 1 เกณฑ์ประเมินระดับความสามารถจำแนกตามคะแนนมาตรฐานที่ปกติ

คะแนนที่ปกติ	ระดับความสามารถ
ตั้งแต่ 65 และสูงกว่า	ดีมาก
ตั้งแต่ 55-65	ดี
ตั้งแต่ 45-55	พอใช้
ตั้งแต่ 35-45	อ่อน
ตั้งแต่ 35 และต่ำกว่า	อ่อนมาก

ที่มา: สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ. ม.ป.ป. : ไม่ปรากฏเลขหน้า.

กรอบแนวคิดในการวิจัย

การศึกษารอบแนวคิดในการวิจัย ผู้วิจัยได้ศึกษาแนวทางในการศึกษาคุณภาพของแบบทดสอบความสามารถทางภาษาอังกฤษดังนี้



บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง และได้นำเสนอตามหัวข้อต่อไปนี้

1. การวัดและการประเมินผล
 - 1.1 ความหมายของการวัด และการประเมินผล
 - 1.2 จุดประสงค์ของการวัด และการประเมินผล
 - 1.3 เครื่องมือวัดผลการศึกษา
2. การทดสอบ
 - 2.1 ความหมายของการทดสอบ
 - 2.2 แบบทดสอบทางภาษา
 - 2.2.1 ประเภทของแบบทดสอบทางภาษา
 - 2.2.2 การสร้างแบบทดสอบทางภาษา
3. แบบทดสอบมาตรฐาน
 - 3.1 ความหมายของแบบทดสอบมาตรฐาน
 - 3.2 คุณภาพของแบบทดสอบ
 - 3.3 เกณฑ์ปกติ
 - 3.4 คะแนนมาตรฐาน
 - 3.5 วิธีสร้างเกณฑ์ปกติ
4. งานวิจัยที่เกี่ยวข้องกับคุณภาพของแบบทดสอบ

1. การวัด และการประเมินผล

1.1 ความหมายของการวัดและการประเมินผล

อนันต์ ศรีโสภ (2548: 197) ให้ความหมายของการวัดและการประเมินผลว่า การวัด (Measurement) หมายถึงการกำหนดตัวเลข หรือสัญลักษณ์ลงบนวัตถุ สิ่งของ หรือพฤติกรรมต่างๆ ที่เป็นไปตามกฎ การวัดเป็นการพิจารณาสถานภาพของแต่ละสิ่งที่ถูกวัดเท่านั้น ไม่ได้มีการตีความหมายว่า เป็นอย่างไร เช่นผู้รับการทดสอบที่ได้คะแนน 65 คะแนนจะไม่สามารถกล่าวได้ว่าผู้นั้นเก่งหรืออ่อน ส่วนการประเมินผล (Evaluation) เป็นการพิจารณาคูณค่าของสิ่งที่จะประเมินโดยการนำไปเปรียบเทียบกับเกณฑ์ (Criterion) เช่นผู้รับการทดสอบที่ทำคะแนนได้ 65 คะแนน เมื่อมีการนำไปเปรียบเทียบกับคะแนนเกณฑ์ซึ่งเท่ากับ 50 คะแนน แสดงว่าผู้รับการทดสอบมีความสามารถสูงกว่าความสามารถเฉลี่ยของผู้รับการทดสอบทั้งหมดในชั้นเรียนเดียวกัน

1.2 จุดประสงค์ของการวัด การประเมินผลการศึกษา และการทดสอบ

การวัดและประเมินผลเป็นองค์ประกอบที่สำคัญอย่างหนึ่งในกระบวนการเรียนการสอน ซึ่งเริ่มจากการกำหนดจุดมุ่งหมายของการเรียนการสอน นำไปสู่การเรียนการสอน การที่จะทราบว่าคุณเรียนเกิดการเรียนรู้และเป็นไปตามจุดมุ่งหมายที่ตั้งไว้เพียงใดนั้นต้องมีการวัดและประเมินผล ทั้งนี้การวัดและประเมินผลต้องทำอย่างเป็นระบบเพื่อจะได้นำผลไปปรับปรุงการเรียนการสอนให้มีประสิทธิภาพมากยิ่งขึ้น (เฉลิมเกียรติ เฟื่องแก้ว, 2539: 10) หน้าที่และจุดประสงค์ของการประเมินหรือการทดสอบทางภาษาของผู้เรียนอาจมีจุดประสงค์ที่แตกต่างกัน สมิตรา อังวัฒนกุล (2539: 3) ได้กล่าวว่าในการที่จะสร้างแบบทดสอบเพื่อประเมินความสามารถทางภาษาได้อย่างเหมาะสมนั้นจำเป็นที่จะต้องทราบถึงจุดประสงค์ที่ชัดเจนว่าต้องการเน้นการนำผลการประเมินไปใช้ในประโยชน์ด้านใด วิรัช วรรณรัตน์ (2547: 25-28) ; และ อัจฉรา วงศ์โสธร (2544: 8) ได้เพิ่มเติมรายละเอียดในเรื่องจุดประสงค์ของการทดสอบสรุปได้ดังนี้

1. เพื่อจัดการตำแหน่งหรือการจัดกลุ่ม (Placement) โดยการนำผลการทดสอบของแต่ละคนมาเปรียบเทียบกัน เพื่อจะได้ทราบว่าผู้ทดสอบมีความรู้ความสามารถในระดับใดของกลุ่ม เพื่อจะจัดประเภทและตำแหน่งให้ถูกต้อง สามารถแบ่งผู้เรียนออกเป็นกลุ่มได้ นอกจากนี้ยังสามารถคัดเลือก (Selection) ผู้เข้าทดสอบในตำแหน่งที่ต้องการ หรือจำแนก (Classification) ผู้ทดสอบว่าใครเก่งหรืออ่อนกว่ากันเพียงใด

2. เพื่อการวินิจฉัย (Diagnosis) ผลจากการทดสอบจะทำให้ทราบว่า การที่ผู้เรียนเก่งหรืออ่อนเป็นเพราะเหตุใด และเก่งหรืออ่อนตรงไหน เรื่องอะไร ทั้งนี้เพื่อหาสาเหตุในการพัฒนาและปรับปรุงการเรียนการสอนให้ดีขึ้น นอกจากนี้ยังสามารถวินิจฉัยปัญหาการเรียน (Prognosis) ได้อีกด้วย

3. เพื่อการเปรียบเทียบ (Assessment) การทดสอบแบบนี้เป็นการทดสอบเพื่อดูพัฒนาการเกี่ยวกับการเรียนของผู้เรียน หรือความก้าวหน้า (Progress) จากเดิมหรือไม่ เท่าไร โดยการเปรียบเทียบกับความสามารถของตัวเองในช่วงเวลาหนึ่ง

4. เพื่อการพยากรณ์ (Prediction) การทดสอบชนิดนี้มีวัตถุประสงค์ที่ต้องการนำเอาผลการทดสอบที่ได้ไปพยากรณ์ล่วงหน้าว่าในอนาคต ผู้เรียนสามารถเรียนต่อไปได้ในระดับใด จะเรียน จะเรียนสำเร็จหรือไม่ และควรเรียนอะไร

5. เพื่อการประเมินค่า (Evaluation) เป็นการนำผลการทดสอบมาประเมินคุณภาพการศึกษาโดยรวมว่า กระบวนการศึกษาที่จัดทำขึ้นเหมาะสมหรือไม่ ควรปรับปรุงแก้ไขที่ใด หลักสูตรที่ใช้เหมาะสมหรือยัง ตลอดจนการวัดผลสัมฤทธิ์ในการเรียน (Achievement) และการประเมินเพื่อปรับปรุงการสอน และเพื่อตัดสินผลการเรียน

6. เพื่อความเป็นมาตรฐาน (Standardized tests) คูมบี้ (Coombe, 2005: 2) เพิ่มเติมว่า จุดประสงค์ของการทดสอบยังเป็นการเตรียมการวัดความสามารถของผู้เรียนให้เป็นมาตรฐานเมื่อมีการใช้แบบทดสอบที่เป็นสากล

ดังนั้นจะเห็นได้ว่าจุดประสงค์ของการทดสอบมีหลายประการ มีประโยชน์ทั้งในด้านการบริหาร การเรียน การสอน ด้านการวิจัย ประโยชน์ที่ได้รับดังกล่าวจะนำมาช่วยปรับปรุงแก้ปัญหา การเตรียมตัวของผู้เรียน คุณภาพของการเรียนการสอน ตลอดจนแสดงถึงความสามารถของผู้เรียน ในการปฏิบัติงานในอนาคตได้อีกด้วย

1.3 เครื่องมือวัดผลการศึกษา

ในการวัดผลจะให้ได้สิ่งที่ต้องการวัดจะต้องมีเครื่องมือที่ดีมีคุณภาพ สิ่งที่ต้องการวัดนั้นเครื่องมือบางอย่างวัดได้บางอย่างอาจวัดไม่ได้ เครื่องมือวัดผลแบบต่างๆมีดังนี้ (ล้วน สายยศ ; และ อังคณา สายยศ. 2543: 21)

1. แบบทดสอบ (Test)
2. การสังเกต (Observation)
3. การสัมภาษณ์ (Interview)
4. แบบทดสอบถาม (Questionnaire)
5. การจัดอันดับคุณภาพ (Rating scale)

2. การทดสอบ

2.1 ความหมายของการทดสอบ

การทดสอบ (Testing) เป็นคำที่ใช้กันแพร่หลายในวงการศึกษา การประเมิน และการวัดผล มีการใช้กันเป็นประจำ เช่น การทดสอบเก็บคะแนน การทดสอบกลางภาคเรียน การทดสอบปลายภาคเรียน เป็นต้น ความหมายที่นักการศึกษา และนักวัดผลการศึกษาได้ให้ไว้ มีดังนี้

ซวาล แพร์ตกุล (2518: 100) ได้กล่าวว่าการทดสอบ หมายถึง การนำชุดของคำถามหรือกลุ่มงานใด ๆ ที่เราสร้างขึ้น เพื่อจะชักนำให้ผู้ถูกสอบแสดงพฤติกรรมอย่างใดอย่างหนึ่งออกมา ให้ผู้รับการทดสอบสังเกตได้และวัดได้

ล้วน สายยศ ; และ อังคณา สายยศ (2539: 14) กล่าวว่า การทดสอบ หมายถึง การนำแบบทดสอบซึ่งสร้างขึ้นอย่างเป็นกระบวนการและมีระบบไปตรวจสอบตัวอย่างของคุณลักษณะของสิ่งที่ต้องการจัดการทดสอบจึงใช้กฎเกณฑ์หนึ่งของการวัดผลด้วยการทดสอบจะเกิดขึ้นได้ก็ต้องอาศัยข้อทดสอบหรือแบบทดสอบเป็นเครื่องมือ

เยวดี วิบูลย์ศรี (2540: 5) กล่าวว่า การทดสอบ หมายถึง การเสนอสิ่งเร้าชุดใดชุดหนึ่งให้กับบุคคลที่เกี่ยวข้อง ตอบสนองตามวิธีที่กำหนดไว้เพื่อนำผลการตอบสนองมากำหนดเป็นคะแนน

เพลินพิศ ธรรมรัตน์ (2542: 3) กล่าวว่า การทดสอบ หมายถึง การจัดสถานการณ์หรือ เครื่องมือต่างๆ กระตุ้นให้ผู้รับการทดสอบแสดงพฤติกรรมตอบสนอง เพื่อดันหาคุณลักษณะหรือความคิดบางอย่างที่ต้องการ

คูบิชซินส์ และ โบริช (Kubiszyn ; & Borich. 1999: 17) กล่าวว่า การทดสอบคือ การดำเนินการและการใช้ผลของการทดสอบเพื่อที่จะใช้ตัดสินใจทางการศึกษา

พิชิต ฤทธิ์จัญญ (2544: 2) กล่าวว่า การทดสอบ หมายถึง เทคนิคอย่างหนึ่งของ การวัดผลซึ่งเครื่องมือที่ใช้วัดคือแบบทดสอบ

ไวร์สมา และ เจอร์ส (Wiersma ; & Jurs อ้างถึงใน พิชิต ฤทธิ์จัญญ, 2544: 1) กล่าวว่า การทดสอบเป็นกระบวนการบริหารแบบทดสอบซึ่งเป็นส่วนหนึ่งของการวัดผลโดยใช้แบบทดสอบเป็นสถานการณ์ไปเร้าให้เกิดการตอบสนอง

จากที่กล่าวมาแล้วสรุปได้ว่า การทดสอบ หมายถึง การนำแบบทดสอบ กิจกรรม งาน ต่างๆ ที่ผู้รับการทดสอบหรือครูผู้สอนได้สร้างขึ้น ไปให้ผู้ถูกสอบหรือนักเรียนได้ตอบหรือแสดง พฤติกรรมออกมา ซึ่งอาจจะเป็นการเขียน การพูด การกระทำ หรืออื่นๆ ที่ผู้รับการทดสอบสังเกต และวัดได้

2.2 แบบทดสอบทางภาษา

2.2.1 ประเภทของแบบทดสอบทางภาษา (Test types)

ในการประเมินหรือวัดความสามารถทางภาษาของผู้เรียนได้มีแนวคิดในการทดสอบโดยจัดทำแบบทดสอบแบ่งเป็นประเภทต่างๆ อาทิ คูมบ์ และ ฮับเลย์ (Coombe ; & Hubley. 2005: 7-9) ฮิวส์ (Huges. 2003: 11-17) และ แฮริสัน (Harison. 1986: 5-9) ได้แบ่งประเภทของแบบทดสอบไว้ดังนี้

1. แบบทดสอบวัดความสามารถทางภาษา (Proficiency test)
2. แบบทดสอบสอบวัดผลสัมฤทธิ์ทางภาษา (Achievement test)
3. แบบทดสอบวินิจฉัยความสามารถทางภาษา (Diagnostic test)
4. แบบทดสอบเพื่อจัดระดับความสามารถทางภาษา (Placement test)
5. แบบทดสอบอื่นๆ (Other test types)

5.1 แบบทดสอบอิงกลุ่มและแบบทดสอบอิงเกณฑ์ (Norm-referenced and Criterion-referenced test)

5.2 การทดสอบแบบปรนัยและการทดสอบแบบอัตนัย (Objective and Subjective Tests)

รายละเอียดของแบบทดสอบทางภาษาทั้ง 5 ประเภทมีดังนี้

1. แบบทดสอบวัดความสามารถทางภาษา (Proficiency test) เป็นแบบทดสอบที่สร้างขึ้นเพื่อที่จะวัดความสามารถทางภาษาของผู้เข้ารับการทดสอบ โดยไม่คำนึงถึงว่าผู้รับการทดสอบจะได้รับการฝึกหรือเรียนภาษานั้นมากน้อยเพียงใด ดังนั้นเนื้อหาของแบบทดสอบจึงไม่ขึ้นอยู่กับเนื้อหาหรือจุดประสงค์ทางภาษาในรายวิชาต่างๆ แต่จะขึ้นอยู่กับความสามารถทางภาษาของผู้เรียน เพื่อตัดสินว่าผู้เข้ารับการทดสอบนั้นมีความสามารถทางภาษา

เพียงพอที่จะใช้เพื่อการใดการหนึ่ง ซึ่งแบบทดสอบดังกล่าวนี้จะต้องพิจารณาถึงระดับของภาษา และประเภทของภาษาที่จะใช้ในการศึกษาแต่ละสาขาวิชาด้วย (สุมิตรา อังวัฒนกุล. 2539: 3-4)

อัจฉรา วงศ์โสธร (2544: 114-115) ; และ แฮริสัน (Harrison. 1986: 7-9) เพิ่มเติมรายละเอียดสรุปได้ว่า แบบทดสอบความสามารถทางภาษามีลักษณะดังต่อไปนี้

1. ใช้วัดความสามารถในการใช้ภาษาในสถานการณ์ที่จำเป็น
2. ไม่สัมพันธ์กับหลักสูตรใดหลักสูตรหนึ่ง เพราะมุ่งที่จะวัดความสามารถในปัจจุบันของผู้รับการทดสอบและความสามารถในการใช้ความสามารถนั้นตามความจำเป็นในอนาคต
3. เป็นการวัดมุ่งไปข้างหน้า คือความสามารถในอนาคต
4. เนื้อหาภาษาที่ทดสอบเป็นภาษาที่ใช้โดยแท้จริง (Authentic)
5. เหมาะที่จะใช้วัดการใช้ภาษาอังกฤษเฉพาะกิจ (English for Specific Purposes)

ตัวอย่างของแบบทดสอบความสามารถทางภาษาอังกฤษที่ใช้กันแพร่หลายได้แก่แบบทดสอบ TOEIC (Test of English for International Communication) เป็นแบบทดสอบความสามารถทางภาษาอังกฤษ (English proficiency test) เป็นที่ยอมรับว่าเป็นแบบทดสอบภาษาอังกฤษมาตรฐาน (English standardized test) ที่ใช้ทดสอบความรู้แก่เจ้าหน้าที่พนักงาน ทั้งในภาครัฐและเอกชน ภาคธุรกิจต่างๆ หน่วยงานการศึกษา บริษัทการบิน ธนาคาร บริษัทการเงิน และบริษัทข้ามชาติ โดยแบบทดสอบจะครอบคลุมการใช้ภาษาอังกฤษระดับนานาชาติที่เกี่ยวข้องกับภาษาอังกฤษเพื่อการสื่อสารธุรกิจ ทั้งด้านการฟัง การพูด และการเขียน แบบทดสอบ TOEIC เป็นแบบทดสอบประเภทเลือกตอบ (Multiple choices) ประกอบด้วยแบบทดสอบการฟัง (Listening section) จำนวน 100 ข้อ และแบบทดสอบการอ่าน (Reading section) จำนวน 100 ข้อ มีคะแนนรวมตั้งแต่ 10 ถึง 990 คะแนน (สุทิน พูลสวัสดิ์. 2546: 6-7) ส่วนแบบทดสอบความสามารถทางภาษาอังกฤษอื่นๆ ได้แก่ แบบทดสอบ TOEFL, IELTS และ Michigan เป็นต้น

2. แบบทดสอบวัดผลสัมฤทธิ์ทางภาษา (Achievement test) เป็นเครื่องมืออย่างหนึ่งที่ใช้ประเมินการเรียนการสอนโดยเฉพาะการวัดสมรรถภาพทางสมอง ได้มีนักวิชาการให้ความหมายไว้ดังนี้

สุมิตรา อังวัฒนกุล (2539: 4) อธิบายว่า แบบทดสอบผลสัมฤทธิ์ทางภาษา มีลักษณะตรงข้ามกับแบบทดสอบวัดความสามารถทางภาษา มีจุดประสงค์ที่จะวัดว่าผู้เรียนแต่ละคนหรือแต่ละกลุ่มที่เรียนรายวิชานั้นบรรลุถึงวัตถุประสงค์ในการเรียนหรือไม่ แบบทดสอบดังกล่าวนี้แบ่งได้เป็น 2 ประเภทย่อยคือ แบบทดสอบปลายภาค และแบบทดสอบวัดความก้าวหน้า

วิสตัน (Whiston. 2000: 18) ได้ให้ความหมายว่า แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนเป็นเครื่องมือที่ใช้วัดความรู้ (Knowledge) หรือ ความคล่องแคล่ว (Proficiency) โดยที่ผลสัมฤทธิ์นั้นเป็นความรอบรู้ของผู้เรียนอย่างแท้จริง หลังจากที่ได้รับการเรียนรู้ หรือ ได้รับการฝึกฝนแล้ว

พัคเคทท์ และ แบล็ค (Puckett ; & Black. 2000: 211) ได้กล่าวว่า แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน เป็นแบบทดสอบที่ใช้วัดสิ่งที่ผู้เรียนได้เรียนมาแล้ว หรือสิ่งที่ เป็นทักษะที่ผู้เรียนได้รับการสอน

อัจฉรา วงศ์โสธร (2544: 320) กล่าวว่า แบบทดสอบผลสัมฤทธิ์เป็น แบบทดสอบที่มีเนื้อหาสัมพันธ์กับหลักสูตร และใช้วิธีการเดียวกับวิธีที่ใช้ในการเรียนการสอน ใช้ เพื่อกำหนดระดับความสำเร็จในการเรียนโดยพิจารณากับเกณฑ์ที่กำหนดไว้หรือกับกลุ่มผู้เรียน

สรุปได้ว่าแบบทดสอบวัดผลสัมฤทธิ์ทางภาษาเป็นแบบทดสอบที่วัด ความสามารถของผู้เรียนแต่ละคน หลังจากที่ได้รับความรู้จากการเรียนและการฝึกฝนมาแล้ว เพื่อใช้ วัดว่าผู้เรียนสามารถบรรลุตามวัตถุประสงค์ในการเรียนมากน้อยเพียงใด

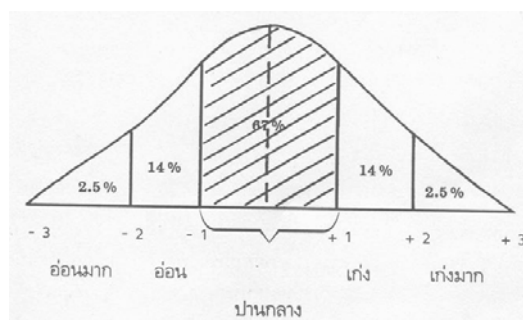
3. แบบทดสอบวินิจฉัยความสามารถทางภาษา (Diagnostic test) เป็น แบบทดสอบที่ใช้วัดว่าผู้เรียนมีความรู้หรือมีจุดอ่อนทางภาษาอะไรบ้าง เพื่อช่วยให้ผู้สอนทราบว่า ควรจะสอนอะไรต่อไปให้แก่ผู้เรียน แบบทดสอบชนิดนี้ต้องมีรายละเอียดมาก จึงจะสามารถบอกได้ ว่าผู้เรียนมีจุดอ่อนตรงไหน จึงเป็นเรื่องยากในการสร้างแบบทดสอบนี้ เพราะข้อทดสอบเพียงหนึ่ง หรือสองข้อ ไม่สามารถจะวินิจฉัยความสามารถทางภาษาของผู้เรียนได้ ต้องใช้เวลาในการทดสอบ มากกว่าปกติ จึงไม่ค่อยมีผู้นิยมใช้นัก แบบทดสอบประเภทนี้จึงมีน้อยและที่มีอยู่ก็ไม่ค่อยมี รายละเอียดมากนัก อย่างไรก็ตามแบบทดสอบประเภทนี้มีประโยชน์มาก โดยเฉพาะในการสอนแบบ การเรียนด้วยตนเอง เพราะผู้เรียนจะสามารถทราบจุดอ่อนในการใช้ภาษาของตนเอง และสามารถ แก้ไขได้โดยดูจากตัวอย่างและการฝึกหัด

4. แบบทดสอบเพื่อจัดระดับความสามารถทางภาษา (Placement test) หมายถึงแบบทดสอบที่ใช้จัดกลุ่มตามความสามารถของผู้เรียน มีการใช้วิธีการทดสอบทักษะรวม หรือทักษะสัมพันธ์แบบต่างๆ (Integrated skills) แบบทดสอบเพื่อจัดระดับความสามารถทางภาษา นี้มักจะสร้างขึ้นสำหรับสถานการณ์เฉพาะ ขึ้นอยู่กับลักษณะและระดับการเรียนการสอนของแต่ละ สถาบัน ส่วนใหญ่จึงเป็นแบบทดสอบที่แต่ละสถาบันการศึกษาสร้างขึ้นเอง สำหรับประโยชน์ในการ จัดกลุ่มผู้เรียน กำหนดผู้ที่จำเป็นต้องเรียนเพิ่มเติม แบบทดสอบชนิดนี้อาจใช้เป็นแบบทดสอบ ความสามารถ (Proficiency test) ซึ่งวัดพื้นความรู้ในปัจจุบันเพื่อใช้ทำนายความสามารถในอนาคต หรือแบบทดสอบสัมฤทธิ์ (Achievement test) ซึ่งอิงหลักสูตรที่ยังไม่เรียนก็ได้ และเพื่อให้เกิด ประสิทธิภาพในการจัดระดับความสามารถทางภาษาควรมีการสัมภาษณ์เพื่อสังเกตพฤติกรรมการใช้ภาษาในการสื่อสารของผู้รับการทดสอบด้วย (อัจฉรา วงศ์โสธร. 2544: 115) และการทดสอบ ชนิดนี้จะไม่ผล ได้-ตก เนื่องจากผู้เรียนจะต้องเข้ารับการอบรมเพิ่มเติมในอนาคต (แฮริส. 1969: 3)

5. แบบทดสอบอื่นๆ (Other test types) ได้แก่

5.1 แบบทดสอบอิงกลุ่มและแบบทดสอบอิงเกณฑ์

5.1.1 แบบทดสอบอิงกลุ่ม (Norm-referenced test) อัจฉรา วงศ์โสธร (2544: 94-95) กล่าวว่า การทดสอบดังกล่าวเป็นการทดสอบที่มุ่งเน้นผู้เรียนส่วนใหญ่ในชั้น โดยยึดถือความก้าวหน้าในการเรียนของผู้เรียนว่าเป็นไปพร้อมๆ กัน โดยพิจารณาผู้เรียนทั้งกลุ่มว่าเป็นกลุ่มที่ส่วนใหญ่จะมีความสามารถปานกลาง จะมีเพียงส่วนน้อยที่เก่งมากหรืออ่อนมาก คือ ดูความสามารถเป็นแบบการกระจายปกติ (Normal distribution) ในการทดสอบผู้เรียนทุกคนจะสอบพร้อมกัน ข้อทดสอบจะมีความยากง่ายในระดับปานกลาง ค่าความยากง่ายจะอยู่ระหว่าง 45% - 55% ตามการวิเคราะห์แบบประเพณีนิยม ข้อทดสอบประเภทนี้จะสามารถแยกนักเรียนเก่งออกจากนักเรียนอ่อนได้ คือ นักเรียนเก่งควรทำข้อทดสอบนั้นๆ ได้มากกว่านักเรียนอ่อน โดยควรมีค่าอำนาจจำแนก .30 ขึ้นไป นอกจากนี้การทดสอบแบบอิงกลุ่มจะทำให้ทราบว่าผู้รับการทดสอบทำแบบทดสอบนั้นได้ดีเพียงไร ซึ่งอาจจะได้คำตอบว่า ผู้รับการทดสอบได้คะแนนอยู่ 10 เปอร์เซนต์ ในลำดับต้นๆ ของผู้เข้าสอบทั้งหมด หรืออยู่ 5 เปอร์เซนต์ในลำดับท้ายๆ หรือทำให้ดีกว่า 60 เปอร์เซนต์ ของผู้เข้าสอบเป็นต้น (สุมิตรา อังวัฒนกุล. 2539: 6)



แผนภูมิ การกระจายปกติ

ภาพประกอบ 1 การกระจายปกติ

ที่มา: อัจฉรา วงศ์โสธร. (2544). *แนวทางการสร้างข้อทดสอบภาษา*. หน้า 96.

5.1.2 แบบทดสอบอิงเกณฑ์ (Criterion-referenced test) เป็นการทดสอบรายบุคคล ผู้เรียนแต่ละคนจะสอบไม่พร้อมกัน และจะทำเมื่อผู้เรียนพร้อมที่จะรับการทดสอบ ข้อทดสอบประเภทนี้ไม่จำเป็นต้องมีระดับความยากง่าย ปานกลาง เหมือนข้อทดสอบอิงกลุ่ม โดยความยากง่ายควรเป็นระดับ 80% ขึ้นไป ซึ่งชี้ว่าผู้เรียนแต่ละคนสามารถทำข้อทดสอบได้เป็นส่วนใหญ่ นอกจากนี้ข้อทดสอบไม่จำเป็นต้องแยกผู้เรียนเก่งออกจากผู้เรียนอ่อน เนื่องจากผู้เรียนไม่ว่าเก่งหรืออ่อนก็ตามจะสอบเมื่อตนพร้อม ดังนั้นค่าอำนาจจำแนกของแบบทดสอบอิงเกณฑ์จึงควรเป็น 0.0 ซึ่งหมายถึงว่า ไม่สามารถแยกผู้เรียนได้ ทั้งนี้เพราะผู้เรียนทุกคนเรียนรู้ตามวัตถุประสงค์แล้ว (อัจฉรา วงศ์โสธร. 2544: 96)

สมิตรา อังวัฒนกุล (2539: 6) สรุปว่า การสร้างแบบทดสอบแบบอิงเกณฑ์นั้น อาจแตกต่างจากการสร้างแบบสอบอิงกลุ่ม ทั้งที่สร้างขึ้นโดยมีวัตถุประสงค์เดียวกัน เช่น การประเมินความสามารถทางภาษาอังกฤษของผู้เรียนในระดับอุดมศึกษา แบบทดสอบแบบอิงเกณฑ์จะต้องอยู่บนพื้นฐานการวิเคราะห์ว่านักศึกษาสามารถใช้ภาษาอังกฤษที่เรียนทำอะไรได้บ้าง ในทางตรงกันข้ามแบบทดสอบแบบอิงกลุ่ม อาจจะมีเนื้อหาจากการวิเคราะห์แบบเดียวกัน แต่งานที่กำหนดในแบบทดสอบอาจต่างออกไป อาจจะเป็นการวัดองค์ประกอบต่างๆ ทางภาษา เช่น โครงสร้างไวยากรณ์ ศัพท์ และความเข้าใจในการอ่าน โดยใช้แบบทดสอบแบบปรนัย คะแนนที่ได้จากการทำแบบทดสอบแบบนี้ไม่ได้บอกโดยตรงว่า ความสามารถทางภาษาของผู้เรียนมีความสัมพันธ์อย่างไรกับความต้องการของมหาวิทยาลัย ในการทดสอบทางภาษาส่วนใหญ่มักจะกล่าวถึงการทดสอบอิงกลุ่มมากกว่า การทดสอบแบบอิงเกณฑ์ยังไม่เป็นที่นิยมมากนัก

5.2 การทดสอบแบบปรนัยและการทดสอบแบบอัตนัย (Objective and Subjective Tests) เยาวดี วิบูลย์ศรี (2540: 216–228) ได้อธิบายถึงการทดสอบแบบปรนัยว่าแบบทดสอบปรนัยโดยทั่วไปหมายถึงแบบทดสอบที่ผู้ตรวจแต่ละคนสามารถให้คะแนนผลการทดสอบของผู้รับการทดสอบได้ตรงกันไม่ว่าจะตรวจซ้ำกี่ครั้ง หรือถ้าให้ผู้ตรวจต่างคนกันตรวจให้คะแนนผู้รับการทดสอบคนเดียวกันก็จะให้คะแนนเท่ากัน แบบทดสอบปรนัยสามารถแบ่งเป็นประเภทต่างๆ ได้ดังนี้

1. แบบทดสอบประเภทเติมคำให้สมบูรณ์ เป็นแบบทดสอบที่ให้ผู้ตอบเติมคำ วลี หรือประโยคที่ถูกต้อง ต่อจากข้อความที่เว้นว่างไว้ ส่วนแบบทดสอบแบบตอบสั้นมุ่งให้ผู้ตอบเขียนคำ วลี หรือจากประโยคคำถามที่กำหนดไว้
2. แบบทดสอบประเภทถูก-ผิด เป็นการทดสอบหรับใช้วัดความจำที่เกี่ยวข้องกับข้อเท็จจริง โดยผู้ตอบต้องพิจารณาข้อความของคำถามให้ได้ว่า ถูก หรือ ผิด
3. แบบทดสอบประเภทจับคู่ เป็นแบบทดสอบที่ใช้วัดความรู้เกี่ยวกับความหมายของคำ ความรู้ที่เป็นข้อเท็จจริง และความสัมพันธ์ระหว่างสิ่งต่างๆ ข้อทดสอบประเภทนี้จะมีข้อความสองกลุ่ม ผู้รับการทดสอบจะต้องเลือกคำหรือข้อความของอีกกลุ่มหนึ่งที่มีความสอดคล้องและสัมพันธ์กัน
4. แบบทดสอบประเภทหลายตัวเลือก จะประกอบด้วยคำถามและตัวเลือก ซึ่งเป็นคำตอบ 1 ตัว และตัวลวง 3-4 ตัว แล้วแต่กรณี ผู้ตอบจะต้องเลือกตัวเลือกที่กำหนดไว้ให้เพียงตัวเดียว แบบทดสอบประเภทนี้สามารถวัดระดับการเรียนรู้ชั้นสูง เช่น การวิเคราะห์ การสังเคราะห์ และการประเมินผล

ส่วนแบบทดสอบแบบอัตนัย (Subjective test) เป็นแบบทดสอบที่ให้ผู้ตอบได้แสดงความคิดเห็น ความจำ และความเข้าใจ เหมาะสำหรับวัดความรู้ชั้นสูง (อนุวัติ คุณแก้ว. 2004: ออนไลน์)

เยาวดี วิบูลย์ศรี (2540: 216–228) อธิบายเพิ่มเติมเพิ่มเติมว่า ความแตกต่างระหว่างแบบทดสอบปรนัยและอัตนัยอยู่ที่วิธีการให้คะแนน ถ้าการให้คะแนนไม่ต้องใช้การ

พินิจพิจารณาเข้าช่วยในการให้คะแนน การให้คะแนนนั้นถือว่ามีความเป็นปรนัย แบบทดสอบและเลือกตอบซึ่งมีคำตอบที่ถูกต้องชัดเจน ไม่คลุมเครือ ถือว่าเป็นแบบทดสอบแบบปรนัย แต่ถ้าวัดการให้คะแนนต้องมีการพินิจพิจารณาในการให้คะแนนถือว่าเป็นแบบทดสอบแบบอัตนัย แต่อย่างไรก็ตามยังมีระดับของความแตกต่างในการทดสอบแบบอัตนัย การให้คะแนนการเขียนเรียงความอาจจัดว่ามีความเป็นอัตนัยมากกว่าการให้คะแนนคำตอบสั้นๆ ในการตอบคำถามเพื่อประเมินความสามารถในการอ่านเพื่อความเข้าใจ

แบบทดสอบปรนัยเป็นที่นิยมใช้กันมาก เพราะตรวจข้อสอบได้รวดเร็วและมีค่าความเที่ยงตรงสูง สำหรับการทดสอบแบบอัตนัยโดยทั่วไปแล้วถ้าวัดการให้คะแนนมีความเป็นอัตนัยน้อยเท่าใด จะทำให้ผู้ให้คะแนนสองคนมีความเห็นตรงกันได้มากขึ้น ซึ่งมีวิธีการที่จะหาความเที่ยงตรงของการให้คะแนนการทดสอบแบบอัตนัยได้เช่นกัน

สรุปแบบทดสอบทางภาษามีอยู่ด้วยกันหลายประเภทได้แก่ แบบทดสอบวัดความสามารถทางภาษา แบบทดสอบวัดผลสัมฤทธิ์ทางภาษา แบบทดสอบวินิจฉัยความสามารถทางภาษา แบบทดสอบจัดระดับความสามารถทางภาษา และแบบทดสอบอื่นๆ ซึ่งได้แก่แบบทดสอบอิงกลุ่ม แบบทดสอบอิงเกณฑ์ รวมถึงแบบทดสอบแบบปรนัย และแบบทดสอบอัตนัยด้วย ในการวิจัยครั้งนี้ผู้วิจัยได้ศึกษาแบบทดสอบจัดระดับความสามารถทางภาษา (Proficiency test) เพื่อใช้ในการจัดกลุ่มผู้เรียนให้เหมาะสมกับหลักสูตรที่จัดทำขึ้น โดยใช้แบบทดสอบแบบปรนัยซึ่งเป็นแบบทดสอบประเภทเลือกตอบแบบหลายตัวเลือก แบบทดสอบดังกล่าวเป็นแบบทดสอบที่นิยมใช้กันอย่างแพร่หลาย เนื่องจากมีความเที่ยงตรงสูง ช่วยประหยัดเวลาในการตรวจ ทำให้ทราบผลการทดสอบได้อย่างรวดเร็ว เหมาะสมกับการจัดการทดสอบภายในองค์กรที่มีผู้รับการทดสอบเป็นจำนวนมากได้เป็นอย่างดี

2.2.2 การสร้างแบบทดสอบความสามารถทางภาษา

ในการสร้างแบบทดสอบแต่ละครั้ง ผู้สอนจะต้องพยายามสร้างให้แบบทดสอบเป็นตัวแทนของเนื้อหาบทเรียนที่ต้องการทดสอบได้จริงๆ ทั้งในแง่ของเนื้อหาและจุดมุ่งหมายของการเรียนรู้ในการสร้างแบบทดสอบ (อนันท์. ม.ป.ป. : ไม่ปรากฏเลขหน้า ; อ้างอิงมาจาก ศศิธร แม้นสงวน. 2531: 203)

สุมิตรา อังวัฒนกุล (2539: 19) กล่าวถึงการสร้างแบบทดสอบความสามารถทางภาษาว่ามีขั้นตอนการสร้างและเทคนิคการออกข้อทดสอบดังนี้

1. ผู้สร้างแบบทดสอบจะต้องมีความเข้าใจอย่างชัดเจนก่อนว่า ต้องการทราบอะไร และนำผลการทดสอบไปทำอะไร โดยมีรายละเอียดดังนี้

1.1. วัตถุประสงค์ที่เฉพาะเจาะจงในการทดสอบมีอะไรบ้าง

1.2 แบบทดสอบที่ต้องการเป็นแบบทดสอบประเภทใด เช่น เป็นการวัดผลสัมฤทธิ์ทางภาษา วัดความสามารถทางภาษา วินิจฉัยความสามารถทางภาษา หรือเพื่อจัดระดับความสามารถทางภาษา

1.3 ต้องการทดสอบความสามารถอะไรบ้าง

1.4 ผลที่ได้ควรมีรายละเอียดเพียงใด

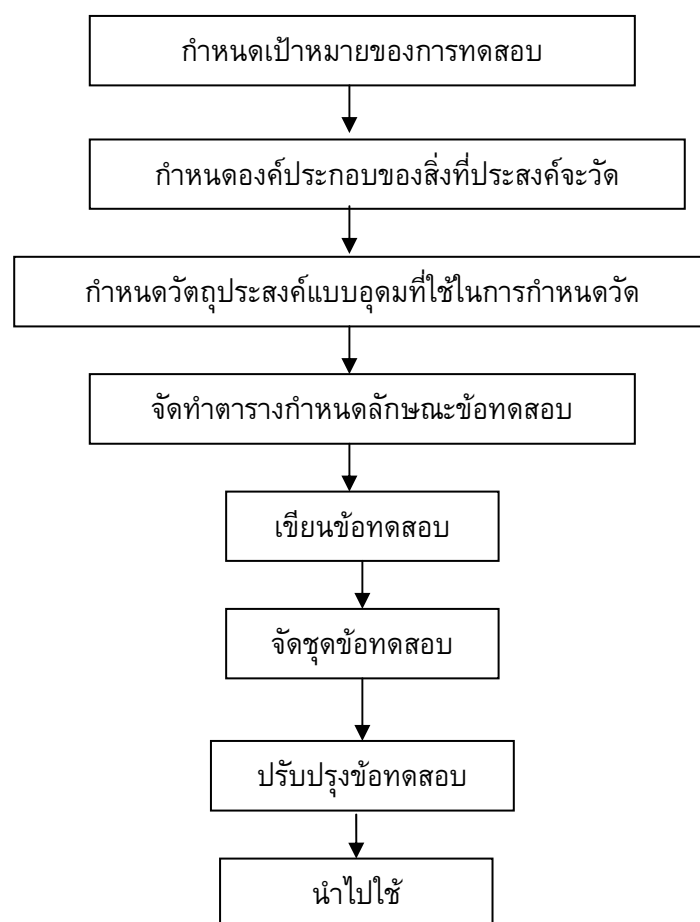
1.5 ผลที่ได้ควรมีความแม่นยำเพียงใด

1.6 ผลการประเมินมีความสำคัญเพียงใด

1.7 ข้อจำกัดอะไรบ้าง ที่อาจจะเกิดขึ้นจากการสร้างแบบทดสอบ การนำแบบทดสอบไปใช้ และการให้คะแนน

2. จัดทำตารางวิเคราะห์เนื้อหาสำหรับการทดสอบ ซึ่งประกอบไปด้วยเนื้อหา รูปแบบ และเวลาที่ใช้ ระดับความสามารถ และการให้คะแนน

อัจฉรา วงศ์โสธร (2544: 175) ได้เสนอขั้นตอนของการสร้างข้อทดสอบทางภาษา โดยการดัดแปลงมาจากแนวความคิดของ Brown สรุปเป็นแผนภูมิได้ดังนี้



ขั้นตอนของการสร้างข้อทดสอบทางภาษา

ภาพประกอบ 2 ขั้นตอนของการสร้างข้อทดสอบทางภาษา

ที่มา: อัจฉรา วงศ์โสธร. (2544). การพัฒนาข้อทดสอบอิงวัตถุประสงค์แบบอุดม: วัดความสามารถสูงสุดและการกระทำปกติทางภาษาอังกฤษ. หน้า 175.

การสร้างข้อทดสอบที่ดีควรเน้นให้ผู้รับการทดสอบได้แสดงพฤติกรรมที่บ่งบอกความสามารถทางภาษาของเขาออกมาโดยต้องแสดงถึงพฤติกรรมที่เที่ยงและตรง สามารถให้คะแนนได้อย่างเชื่อถือได้ ใช้เวลาและความพยายามน้อย และให้ผลต่อการเรียนการสอนอย่างมีประสิทธิภาพสำหรับเทคนิคในการออกข้อทดสอบที่นิยมใช้ในการทดสอบความสามารถทางภาษาที่เป็นที่รู้จักกันดี ได้แก่ ข้อทดสอบแบบเลือกตอบ (Multiple choice) ข้อทดสอบแบบโคลซ (Cloze) แบบทดสอบซี (C-test) และการเขียนตามคำบอก (Dictation) โดยมีรายละเอียดดังนี้

1. ข้อทดสอบแบบเลือกตอบ (Multiple choice) ดิลก ดิลกานนท์ (2547: 7) กล่าวถึงลักษณะของข้อทดสอบแบบเลือกตอบว่า ข้อทดสอบแต่ละข้อประกอบด้วยสองส่วนที่สำคัญคือส่วนที่เป็นข้อความ (Stem) ซึ่งเป็นส่วนที่ถามตามจุดประสงค์ที่ต้องการจะตรวจสอบว่าผู้รับการทดสอบเกิดความรู้ ความเข้าใจในประเด็นที่ถามนั้นหรือไม่ ซึ่งข้อความที่ดีต้องชัดเจนไม่คลุมเครือ อ่านแล้วต้องเข้าใจประเด็นในการถามตรงกัน ส่วนที่เป็นตัวเลือก (Choices of alternatives) หรือคำตอบ (Response) ของข้อความข้างต้น โดยทั่วไปคำถามแต่ละข้อจะประกอบด้วยคำตอบ 3-5 ตัวเลือก ให้ผู้รับการทดสอบเลือกหาคำตอบที่ถูกต้องตรงกับประเด็นคำถามเพียงข้อเดียวเท่านั้น ดังนั้นในแต่ละชุดของคำตอบข้อหนึ่งๆ นั้น จึงมีคำตอบที่ถูกต้องจริงๆ เพียงหนึ่งคำตอบเท่านั้น ส่วนที่เหลือจะเป็นตัวลวง ตัวลวงที่ดีในคำถามข้อหนึ่งๆ นั้นต้องไม่เด่นชัดจนเกินไปจนผู้รับการทดสอบสามารถเดาได้ ผู้รับการทดสอบจะต้องเลือกคำตอบที่ถูกต้องหรือเหมาะสมที่สุด ข้อทดสอบแบบเลือกตอบมีข้อดีในการให้คะแนนที่มีความเที่ยงตรง ตรวจให้คะแนนได้ง่ายและรวดเร็ว และผู้รับการทดสอบสามารถทำข้อทดสอบได้รวดเร็ว

สรุปได้ว่าข้อทดสอบแบบเลือกตอบประกอบด้วย ส่วนที่เป็นข้อความหรือส่วนที่เป็นคำถาม และส่วนที่เป็นตัวเลือกซึ่งมักจะประกอบด้วยคำตอบ 3-5 ตัวเลือก โดยที่将有คำตอบที่ถูกต้องเพียงข้อเดียว แบบทดสอบชนิดนี้จะมีคะแนนเที่ยงตรง สามารถตรวจให้คะแนนง่ายและรวดเร็ว เหมาะกับการที่มีผู้เข้ารับการทดสอบเป็นจำนวนมาก

2. ข้อทดสอบแบบโคลซ (Cloze) เบิลีย์ (Bailey. 1998: 60-73) กล่าวว่า ข้อทดสอบแบบโคลซเป็นที่ยอมรับกันอย่างแพร่หลายโดยเฉพาะอย่างยิ่ง การใช้วัดทักษะความสามารถในการอ่าน (Reading skill) และความสามารถทางภาษาโดยรวม ข้อทดสอบแบบโคลซเป็นข้อทดสอบประเภทเติมคำ ประกอบด้วยข้อความที่ยาวมากกว่า 1 ประโยค อย่างน้อยจะมีความยาวประมาณ 1 ย่อหน้า โดยในประโยคแรกจะไม่เว้นคำไว้ให้เติมเพื่อช่วยในการสร้างบริบท จากนั้นจะพบว่าการเว้นคำให้เติมอย่างเป็นระบบซึ่งปกติจะเป็นทุกๆ คำที่ 5, 6, หรือ 7 ผู้รับการทดสอบจะต้องเติมคำที่เหมาะสมและถูกต้องตามไวยากรณ์ลงไปในแต่ละช่องว่างซึ่งผู้รับการทดสอบจะได้คะแนนจากคำที่เติมได้ถูกต้อง ในแต่ละเรื่องที่จะอ่านจะมีประมาณ 30-50 คำที่เว้นช่องว่างให้เติม ข้อทดสอบโคลซนี้เป็นเครื่องมือในการวัดความสามารถทั่วไปในการใช้ภาษาที่ดี ผู้ที่มีคะแนนสูงจากการทดสอบประเภทนี้ก็จะเป็นผู้ที่มีความรู้ความสามารถในการใช้ภาษาได้ดีเช่นกัน

แบบทดสอบโคลซได้พัฒนาออกเป็นหลายรูปแบบ วาเล็ทท์ (Valette. 1997: 212-215) ได้ให้รูปแบบของการทดสอบแบบโคลซไว้ 3 ลักษณะดังนี้

1. แบบทดสอบโคลซมาตรฐาน เป็นการเว้นคำให้ผู้เรียนเติมอย่างมีระบบ ดังที่ได้กล่าวไว้ข้างต้น เช่น

Cloze tests consist of texts of at least one paragraph in length (usually longer). The first sentence is left intact, _____ thereafter every "nth" word is systematically _____ (every fifth, seventh, or ninth word), _____ the students' task is to write _____ appropriate and grammatical word in each _____. The ability to do so depends _____ the learners' syntagmatic competence and paradigmatic _____, as well as their discourse competence. _____ are usually thirty or fifty blanks _____ each cloze passage.

ที่มา: Kathleen M. Bailey. (1998). *Learning About language Assessment*. p. 60.

2. แบบทดสอบโคลซชนิดให้เลือกคำที่ถูกต้องเพียงคำเดียวจาก 2 คำ ที่ให้มา เช่น

Circle the word that most appropriately completes the thought expressed in the passage. (Note : Distracters were randomly selected from the text)

The great murder wave of the 1970s appears to have ebbed at last in big-city America. (First/Have) reports from police departments (last/in) twelve selected cities show (from/that) in nine of them, (the/at) number of homicides dropped (markedly/cities) and in some cases (swiftly/nine) last year.

ที่มา: Rebecca M. Valette. (1997). *Modern Language Testing*. p. 213.

3. แบบทดสอบโคลซชนิดที่มีลักษณะคล้ายกับแบบทดสอบชนิดให้
เลือกตอบ (Multiple choices) เช่น

Crossing the Pacific

The Pacific Ocean is the largest ocean 1 the world. In fact, 2 occupies almost one quarter of 3 whole surfaces of the 4 . It 5 early explorers a long time to sail across it.

- | | | |
|-------------|--------------|---------|
| 1. a. at | 2. a. it | 3. a. a |
| b. on | b. she | b. all |
| c. in | c. Pacific | c. most |
| d. by | d. world | d. the |
| 4. a. world | 5. a. took | |
| b. earth | b. takes | |
| c. water | c. taken | |
| d. land | d. was taken | |

ที่มา: สุมิตรา อังวัฒนกุล. (2539). *การทดสอบความสามารถทางภาษา*. หน้า 26.

สุมิตรา อังวัฒนกุล (2539: 26) แสดงความคิดเห็นว่าแบบทดสอบโคลซมีข้อดีหลายประการ มีการสร้าง การดำเนินการทดสอบ และการตรวจให้คะแนนที่ง่าย นอกจากนี้ในการใช้แบบทดสอบประเภทนี้เพื่อช่วยวัดความสามารถทางภาษา ไม่ว่าผู้ออกข้อสอบจะเลือกเรื่องใดมาใช้ทดสอบ หรือไม่ว่าจะเลือกตัดคำใดออกไป ผลที่ได้ก็ยังคงมีความเที่ยงตรงอยู่ สำหรับหลักในเรื่องการสร้างข้อทดสอบแบบโคลซมีดังนี้

1. ข้อความที่นำมาใช้ ควรมีความยากเหมาะสมกับระดับความสามารถของผู้รับการทดสอบหากไม่แน่ใจในเรื่องความยาก ควรนำบทความหลายๆเรื่องมาทดลองสอบก่อนใช้จริง
2. เนื้อหาควรมีลักษณะที่เหมาะสมแก่การทดสอบทางภาษาโดยรวม
3. ไม่ควรตัดคำที่ต้นประโยค จากนั้นจึงตัดทุกๆคำที่ 8 หรือ 10
4. ควรนำความเรียงที่ตัดคำแล้วนั้นไปทดลองใช้กับเจ้าของภาษาเพื่อกำหนดคำตอบที่อาจเป็นไปได้
5. ควรเขียนคำสั่งให้ชัดเจน

6. การออกแบบในการจัดวางรูปเรื่องที่ตี ช่วยให้การตรวจให้คะแนนเป็นไปได้ง่ายขึ้น ผู้ตรวจควรมีบัตรแสดงคำตอบที่เป็นไปได้ ซึ่งสามารถนำมาเทียบในตำแหน่งเดียวกับคำตอบของผู้รับการทดสอบ

7. ผู้ที่รับการทดสอบควรมีความคุ้นเคยกับแบบทดสอบโคลซ หากผู้รับการทดสอบมีประสบการณ์ในแบบทดสอบประเภทนี้มากเท่าใด คะแนนที่ได้ก็จะยิ่งเป็นตัวแทนของความสามารถทางภาษาที่แท้จริงของผู้รับการทดสอบมากเท่านั้น

8. คะแนนจากการทำแบบทดสอบโคลซเพียงอย่างเดียว ไม่อาจนำมาประเมินค่าความสามารถของนักเรียนได้โดยตรง ในกรณีของการทดสอบเพื่อจัดชั้นเรียน คะแนนที่ได้ควรนำมาเปรียบเทียบกับระดับชั้นของผู้เรียน นอกจากนั้นยังประเมินควบคู่ไปกับข้อมูลจากครูและตัวผู้เรียนเอง และเมื่อกำหนดระดับคะแนนของผู้เรียนและระดับชั้นได้อย่างลงตัวแล้วจึงจะสามารถนำแบบทดสอบโคลซมาเป็นส่วนหนึ่งของการทดสอบเพื่อจัดชั้นเรียนได้

สรุปแบบทดสอบโคลซ เป็นแบบทดสอบที่มีความเหมาะสมในการวัดความสามารถทั่วไปในการใช้ภาษา มีหลายลักษณะทั้งประเภทแบบทดสอบที่เว้นคำไว้ให้เติมอย่างเป็นระบบ แบบชนิดให้เลือกคำที่ถูกต้อง ตลอดจนแบบทดสอบที่มีลักษณะให้เลือกตอบ และเพื่อให้ได้ผลการวัดที่มีประสิทธิภาพ คะแนนที่ได้จากการทดสอบด้วยแบบทดสอบโคลซควรจะมีการประเมินควบคู่ไปกับข้อมูลด้านอื่นๆจากผู้สอบและผู้เรียนด้วย

3. แบบทดสอบซี (C-test) เป็นแบบทดสอบโคลซประเภทหนึ่ง โดยแทนที่จะตัดคำออกไปทั้งคำกลับตัดเพียงครึ่งคำและตัดแบบคำเว้นคำดังตัวอย่างต่อไปนี้

A C-test is a type of language test in which the students read a brief paragraph in the target language. The first two sentences are left intact. There _ _ _ , every ot _ _ word i _ printed int _ _ , but f _ each alte _ _ _ _ word, on _ the fi _ _ half o _ the wo _ is wri _ _ _ , and t _ second ha _ is indi _ _ _ _ by a bl _ _ space repres _ _ _ _ each let _ _ . T _ students' abi _ _ _ to fi _ in t _ blank spa _ _ is tho _ _ _ to b _ a mea _ _ _ of th _ _ language profi _ _ _ _ .

ที่มา: Kathleen M. Bailey. (1998). *Learning About language Assessment*. p. 63.

ราส และ บราเลย์ (Raatz ; & Braley. 2004: Online) กล่าวว่าแบบทดสอบซี (C-test) เป็นแบบทดสอบความสามารถทางภาษาชนิดหนึ่ง สามารถใช้ได้กับผู้เรียนภาษาทั่วไป รวมถึงผู้เรียนที่เป็นเจ้าของภาษาด้วย สำหรับผู้เรียนที่เป็นผู้ใหญ่เจ้าของภาษาหากต้องการที่จะใช้วิธีการทดสอบประเภทนี้จะต้องมีการปรับปรุงแบบทดสอบให้มีความยากเพิ่มขึ้นเพื่อความเหมาะสม ก่อน ซึ่งผู้รับการทดสอบที่เป็นเจ้าของภาษาจะต้องทำผลการทดสอบให้ได้ตั้งแต่ 90% ขึ้นไป จึงจะ

สามารถใช้กับผู้รับการทดสอบที่ไม่ใช่เป็นเจ้าของภาษาได้ คะแนนที่ได้จากการทดสอบแบบนี้สามารถนำไปใช้วัดและประเมินผลได้กับผู้เรียนที่เป็นกลุ่ม และผู้เรียนรายบุคคล เหมาะสำหรับการทดสอบจัดระดับ (Placement test) หรือการทดสอบที่ต้องการคัดเลือกรุ่นที่มีความสามารถต่ำกว่าเกณฑ์ที่กำหนดออก ในบางกรณีจะใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ด้วย (Achievement test) และเนื่องจากแบบทดสอบชนิดนี้เป็นแบบทดสอบวัดความสามารถ ดังนั้นจึงไม่สามารถให้ข้อมูลรายละเอียดเกี่ยวกับ จุดแข็ง (Strength) หรือจุดอ่อน (Weakness) ของผู้เรียนได้

สมิทธา อังวัฒนกุล (2539: 27-28) ให้คำอธิบายว่าแบบทดสอบซึ่งสามารถตรวจให้คะแนนได้อย่างเที่ยงตรงแม่นยำ และสามารถทดสอบด้วยความเร็วที่สั้นกว่า สามารถทดสอบได้อย่างหลากหลายในด้านหัวเรื่อง ลีลาภาษาและระดับความสามารถ การตัดคำไม่หมดทั้งคำมีผลให้เกิดความเป็นตัวแทนของคำประเภทต่าง ๆ และประการสำคัญ เมื่อเทียบกับแบบทดสอบโคลซแล้ว แบบทดสอบซึ่งจำนวน 100 ข้อ ประหยัดเนื้อที่และให้เวลาในการทำน้อยกว่า และผู้รับการทดสอบไม่ต้องเสียเวลาอ่านเรื่องมากนัก ข้อจำกัดของแบบทดสอบซึ่ง คือ มีลักษณะคล้ายกับกิจกรรมเกมปริศนาและอ่านยากกว่าความเร็วของแบบทดสอบโคลซ

สรุปแบบทดสอบซึ่ง เป็นแบบทดสอบประเภทวัดความสามารถทางภาษา เป็นแบบทดสอบประเภทโคลซชนิดหนึ่งที่มีการตัดคำเพื่อให้ผู้รับการทดสอบเติมคำให้ได้ใจความที่ถูกต้อง สมบูรณ์ เหมาะสมกับผู้เรียนภาษาทั่วไป และเหมาะสมกับการนำไปใช้เป็นแบบทดสอบจัดระดับ

4. การเขียนตามคำตอบ (Dictation) ให้ผลที่ดีกว่าการทำแบบทดสอบโคลซคือ นอกจากจะสามารถวัดความสามารถโดยรวมแล้ว ยังสามารถทดสอบความสามารถในการฟังได้ด้วย ประโยชน์ที่เด่นชัดในการสร้างกิจกรรมการทดสอบประเภทนี้คือ สร้างได้ง่ายและจัดสอบค่อนข้างง่าย นักวิจัยหลายท่านยืนยันว่า การทดสอบด้วยการเขียนตามคำบอก (Dictation) เป็นแบบทดสอบที่มีความเที่ยงตรง และความน่าเชื่อถือ (Takeuchi. 2004: Online) นอกจากนี้ ออลเลอร์ (Oller.1979) ได้กล่าวว่า แบบทดสอบดังกล่าวเป็นแบบทดสอบแบบบูรณาการที่สามารถทดสอบทักษะต่างๆทางภาษาได้ในขณะเดียวกัน ฮีตัน (Heaton. 1988: 17) เพิ่มเติมว่า แบบทดสอบชนิดนี้เป็นแบบทดสอบที่สามารถบ่งชี้ หรือทำนายความสามารถทางภาษาโดยรวมได้เป็นอย่างดี

สรุปข้อทดสอบที่นิยมใช้ในการทดสอบความสามารถทางภาษามีหลายประเภทคือ ข้อทดสอบแบบเลือกตอบ ข้อทดสอบแบบโคลซ ข้อทดสอบแบบชี้ และข้อทดสอบประเภทการเขียนตามคำบอก เป็นต้น ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาข้อทดสอบที่สร้างจากข้อทดสอบแบบเลือกตอบ และข้อทดสอบแบบโคลซ เนื่องจากข้อทดสอบดังกล่าวมีความเที่ยงตรงสูง และนิยมใช้กันอย่างแพร่หลาย สามารถวัดความสามารถทางภาษาโดยรวมของผู้รับการทดสอบได้ดี ตลอดจนให้ความสะดวกและรวดเร็วในการประเมินผล

3. แบบทดสอบมาตรฐาน (Standardized tests)

3.1 ความหมายของแบบทดสอบมาตรฐาน

นักวิชาการที่ได้ให้ความหมายของแบบทดสอบมาตรฐานมีดังนี้

คูบิสซินส์ และ แกรี่ (Kubiszyn; & Gary. 1996: 345) กล่าวว่า แบบทดสอบมาตรฐานเป็นแบบทดสอบที่สร้างขึ้นโดยผู้เชี่ยวชาญทางด้านหลักสูตร เช่น ครู ผู้บริหารโรงเรียน โดยมีจุดประสงค์เพื่อวัดระดับความสามารถของผู้เรียน การดำเนินการทดสอบและการให้คะแนนเป็นไปตามขั้นตอนเป็นรูปแบบเดียวกัน มีความเป็นมาตรฐานสามารถนำไปใช้ได้ทุกสถานที่

ภัทรา นิคมานนท์ (2534) กล่าวถึงแบบทดสอบมาตรฐานว่า เป็นแบบทดสอบที่สร้างขึ้นด้วยกระบวนการหรือวิธีการที่ซับซ้อนมากกว่าแบบทดสอบที่ครูสร้างขึ้นเอง เมื่อสร้างขึ้นแล้วมีการนำไปทดลองสอบ แล้วนำผลมาวิเคราะห์ด้วยวิธีการทางสถิติหลายครั้งหลายหนเพื่อปรับปรุงให้มีคุณภาพดี มีความเป็นมาตรฐาน สามารถนำไปใช้ได้อย่างกว้างขวาง

แบบทดสอบมาตรฐานจะมีความเป็นมาตรฐานอยู่ 2 ประการ คือ

1. มาตรฐานในการดำเนินการทดสอบ ไม่ว่าผู้ใดจะใช้แบบทดสอบมาตรฐานเมื่อใดก็ตาม การดำเนินการทดสอบจะปฏิบัติเหมือนกันทุกขั้นตอน ตั้งแต่ขั้นการเตรียมตัวก่อนการทดสอบ วิธีปฏิบัติขณะสอบและวิธีปฏิบัติเมื่อสอบเสร็จแล้ว ทั้งนี้เพื่อควบคุมตัวแปรที่จะมีผลกระทบต่อคะแนนของผู้รับการทดสอบ ฉะนั้นข้อทดสอบมาตรฐานจึงจำเป็นต้องมีคู่มือดำเนินการทดสอบไว้เป็นแนวปฏิบัติสำหรับผู้ใช้ข้อทดสอบ

2. มาตรฐานในการแปลความหมายคะแนน ไม่ว่าข้อทดสอบนั้นจะใช้สอบที่ไหนเมื่อไรก็ตามก็ต้องแปลความหมายคะแนนได้ตรงกันว่า ใครเก่ง อ่อน เพียงไร ทั้งนี้เพราะข้อทดสอบมาตรฐานมีเกณฑ์สำหรับเปรียบเทียบคะแนนให้เป็นมาตรฐานเดียวกัน ซึ่งเรียกว่า เกณฑ์ปกติ (Norm)

อัจฉรา วงศ์โสธร (2544: 97-98) ให้ความหมายของแบบทดสอบมาตรฐานว่า เป็นแบบทดสอบที่ได้ผ่านขั้นตอนการวิเคราะห์และการปรับปรุงโดยมีขั้นตอนของการพัฒนาให้เป็นมาตรฐานโดยมีความเที่ยงตรง ความเชื่อมั่น มีความสะดวกในการใช้ ค่าความยากง่าย และค่าอำนาจจำแนกที่เป็นมาตรฐาน นอกจากนี้ได้เพิ่มเติมด้วยการเสนอขั้นตอนของการพัฒนาแบบทดสอบให้มีความเป็นมาตรฐานดังนี้

1. การพิจารณาร่างกำหนดเนื้อหา ทักษะภาษา และกำหนดวัตถุประสงค์ ของการทดสอบ
2. การสร้างตารางกำหนดลักษณะข้อทดสอบ หรือพิมพ์เขียว ซึ่งเป็นแม่แบบของการวัดโดยใช้ข้อทดสอบ ตารางนี้จะบรรจุวัตถุประสงค์ (Purpose) ของการทดสอบรูปแบบการถาม/ตอบ จำนวนข้อ คะแนนในการทดสอบ เป็นต้น
3. สร้างข้อทดสอบ เพื่อวัดแต่ละวัตถุประสงค์
4. ผู้เขียนข้อทดสอบตรวจทานข้อทดสอบที่สร้างขึ้น

5. มีการตรวจสอบข้อทดสอบแต่ละข้อว่าสอดคล้องกับวัตถุประสงค์ มีความเที่ยงตรงเชิงประจักษ์ (Face validity) และความเที่ยงตรงตามเนื้อหา (Content validity) นอกจากนี้ยังตรวจสอบข้อทดสอบว่า มีความเป็นตัวแทน (Representatives) คือเป็นกลุ่มตัวอย่างที่เหมาะสมของประชากร หรือเนื้อหาที่จะสอบหรือไม่

6. มีการตรวจทานเพิ่มเติม โดยมีการแลกเปลี่ยนข้อทดสอบระหว่างผู้เรียน เพื่อตรวจทาน ทดลองทำ และนำปัญหามาอภิปรายเพื่อการแก้ไขปรับปรุงข้อทดสอบ

7. รวบรวมข้อทดสอบเข้าเป็นชุด เรียงตามลำดับความยากง่าย โดยใช้ข้อทดสอบง่ายมาก่อนยากและจัดประเภทข้อทดสอบที่วัดแบบเดียวกัน โดยให้คำแนะนำในการทดสอบเดียวกันไว้ด้วยกัน ไม่ปะปนกัน เพื่อความสะดวกในการทำข้อทดสอบของผู้เรียน

8. ตั้งเกณฑ์มาตรฐานในการแปลความคะแนน

9. ทดสอบ

10. หาค่าสถิติเกี่ยวกับข้อทดสอบและแบบทดสอบ

11. ปรับปรุงข้อทดสอบทางเนื้อหา

12. ทดลองสอบอีกครั้งหนึ่ง

13. ตรวจสอบค่าสถิติ กำหนดค่าความเที่ยงตรง (Validity) ค่าความเชื่อมั่น (Reliability) และคะแนนบรรทัดฐาน (Norm)

14. มีการทดสอบเพื่อเก็บข้อมูลกับกลุ่มอื่นๆ เพื่อให้ได้ข้อมูลเกี่ยวกับคุณสมบัติของข้อทดสอบ และแบบทดสอบเพิ่มเติม

สรุปได้ว่าแบบทดสอบมาตรฐาน เป็นแบบทดสอบที่สร้างขึ้นด้วยกระบวนการหรือวิธีการที่ซับซ้อน เมื่อสร้างขึ้นแล้วมีการนำไปทดลองสอบ นำผลที่ได้มาวิเคราะห์ด้วยวิธีการทางสถิติหลายครั้งหลายหนเพื่อปรับปรุงให้มีคุณภาพเป็นมาตรฐาน โดยมีความเที่ยงตรง ความเชื่อมั่น ค่าความยากง่าย ค่าอำนาจจำแนก ตลอดจนมีเกณฑ์สำหรับเปรียบเทียบคะแนนให้เป็นมาตรฐานเดียวกัน หรือที่เรียกว่าเกณฑ์ปกติ (Norm)

3.2 คุณภาพของแบบทดสอบ

ไพศาล หวังพานิช (2547: 57) กล่าวถึง คุณภาพของแบบทดสอบหรือเครื่องมือที่ใช้ในการวัดว่า ขึ้นอยู่กับองค์ประกอบหลายอย่าง เช่น การเลือกใช้เครื่องมือได้ถูกต้อง การดำเนินการในการวัดเป็นไปอย่างถูกต้องยุติธรรม เป็นต้น เครื่องมือที่มีคุณภาพมีคุณสมบัติหลายประการ อาทิ เช่น มีความเที่ยงตรง ความเชื่อมั่น มีความเป็นปรนัย มีความไว และมีประสิทธิภาพ เป็นต้น

อัจฉรา วงศ์โสธร (2544: 229) เพิ่มเติมว่า คุณภาพของข้อทดสอบประกอบด้วย คุณภาพของภาษาที่ใช้ คือต้องชัดเจน สื่อความหมายได้ดี วัดได้อย่างคงที่และตรงตามวัตถุประสงค์ของการวัด ต้องไม่ยากหรือง่ายเกินไป และสามารถให้ข้อมูลเกี่ยวกับตัวแปรด้านความสามารถ ด้านเจตคติ บุคลิกภาพ ความสนใจของผู้เข้าสอบได้ ข้อทดสอบที่ดีต้องมีประสิทธิภาพในการวัด ใช้

ประโยชน์ได้ดีตามวัตถุประสงค์ ต้องมีความเที่ยงตรงในการตรวจให้คะแนน และในการวัดพฤติกรรม เป้าหมายอย่างสม่ำเสมอ และต้องมีความตรงซึ่งได้จากการทดสอบตรงตามวัตถุประสงค์ที่ตั้งไว้ ใช้เนื้อหา วิธีการและให้ความสำคัญต่อองค์ประกอบที่วัดตรงตามตารางกำหนดลักษณะของข้อทดสอบ หรือตามพิมพ์เขียวของข้อทดสอบ ความเที่ยงตรงที่สำคัญคือ ความตรงเชิงประจักษ์ (Face or content validity) ที่ทำให้ผู้เรียนเชื่อว่า ข้อทดสอบวัดองค์ประกอบจริงๆ และความตรงเชิงภาวะ สันนิษฐาน (Construct validity) ซึ่งแสดงถึงความสอดคล้องของการวัดกับพื้นฐานด้านคุณลักษณะ ทางภาษาและองค์ประกอบที่ถูกต้องแม่นยำ

คุณภาพของแบบทดสอบตามความเห็นของ พวงรัตน์ ทวีรัตน์ (2540: 115) สามารถ แบ่งได้เป็น 5 ประเภทดังนี้

1. ความเที่ยงตรง (Validity)
2. ความเชื่อมั่น (Reliability)
3. ค่าความยากง่าย (Difficulty)
4. ค่าอำนาจจำแนก (Discrimination)
5. ความเป็นปรนัย (Objectivity)

1. ความเที่ยงตรง หรือความตรงของแบบทดสอบ (Validity)

มีผู้ให้ความหมายของความเที่ยงตรงไว้ดังนี้

ไอเซนเบิร์ก (Eysenck. 1969: 97) กล่าวว่า ความเที่ยงตรงของแบบทดสอบคือ แบบทดสอบที่สามารถวัดได้ตรงตามสิ่งที่ต้องการวัด

กรอนรันด์ (Gronlund. 1976: 79) กล่าวว่า ความเที่ยงตรงคือ ผลการประเมินที่ สอดคล้องกับจุดประสงค์ที่ตั้งไว้

อนาสตาซี (Anastasi. 1982: 27) กล่าวว่า ความเที่ยงตรงของแบบทดสอบคือ แบบทดสอบนั้นใช้วัดได้ผลตามที่ตั้งไว้

เพอร์วิน (Pervin. 1993: 510) ได้ให้ความหมายว่า ความเที่ยงตรงคือ เนื้อหาที่ วัดในสิ่งที่เราสนใจ

เดวิส และฮักตัน (Davies; & Houghton. 1995: 255) กล่าวว่า ความเที่ยงตรงคือ แบบทดสอบที่สร้างขึ้นให้วัดได้ตามจุดมุ่งหมายที่ตั้งไว้

ดวาทสกี (Dwartzky. 1997: 457) กล่าวว่า ความเที่ยงตรงคือ วิสัยความสามารถของ เครื่องมือในการวัดให้ได้ตามจุดมุ่งหมายที่ตั้งไว้

บุญชม ศรีสะอาด (2520: 126) กล่าวว่า ความเที่ยงตรงคือ ความสามารถของ แบบทดสอบในการที่จะวัดได้ตามวัตถุประสงค์

ภัทรา นิคมานนท์ (2522: 49) กล่าวว่า เครื่องมือที่มีความเที่ยงตรงสูงสามารถทำ หน้าที่รวบรวมข้อมูลของสิ่งที่เราต้องการวัดได้ตามจุดมุ่งหมายที่ต้องการวัด

อนันต์ ศรีโสภา (2525: 43) กล่าวว่า แบบทดสอบใดที่มีความเที่ยงตรง หมายถึง แบบทดสอบนั้นสามารถวัดในสิ่งที่ต้องการจะให้วัดได้อย่างถูกต้อง

ล้วน สายยศ; และอังคณา สายยศ (2539: 246) กล่าวว่า ความเที่ยงตรงคือ แบบทดสอบที่สามารถวัดได้ตรงตามลักษณะหรือจุดประสงค์ที่ต้องการวัด

วิญญา วิศาลาภรณ์ (2540: 146) กล่าวว่า ความเที่ยงตรงคือ ความถูกต้องที่เครื่องมือวัดในสิ่งที่ต้องการวัด หรือความถูกต้องแม่นยำที่เครื่องมือวัดตามจุดประสงค์ที่วางไว้

ไพศาล หวังพานิช (2547: 57) สรุปว่า ความเที่ยงตรง หมายถึง มีความถูกต้องแม่นยำ (Accuracy) ในการวัดซึ่งทำให้ผลที่ได้จากการวัดตรงตามความต้องการหรือตรงตามวัตถุประสงค์ ดังนั้นเครื่องมือวัดที่มีความเที่ยงตรงจะสามารถวัดคุณลักษณะที่ต้องการจะวัดได้อย่างแท้จริง ถูกต้องตามความมุ่งหมายของการวัดนั้น ทำให้คะแนนหรือข้อมูลต่าง ๆ จากการวัดนั้น แทนจำนวนหรือลักษณะของสิ่งที่ถูกวัดได้อย่างแท้จริง และนำไปแปลผลได้ตรงตามเป้าหมายที่ต้องการ ความเที่ยงตรงถือว่าเป็นคุณลักษณะที่จำเป็นที่สุด สำคัญที่สุดของเครื่องมือวัดทุกชนิด การพิจารณาว่าเครื่องมือวัดใดมีความเที่ยงตรงหรือไม่นั้น จะต้องมีเกณฑ์หรือจุดมุ่งหมายที่ต้องการวัดมาเปรียบเทียบ จุดมุ่งหมายที่ต้องการวัดซึ่งใช้เป็นเกณฑ์มีหลายอย่าง อาจเป็นเนื้อหาวิชา พฤติกรรม สภาพหรือคุณลักษณะที่มีอยู่ในปัจจุบันหรือที่จะเกิดขึ้นในอนาคตก็ได้

สุมิตรา อังวัณณะกุล (2539: 29) สรุปเพิ่มเติมว่าแบบทดสอบจะมีความเที่ยงตรง ถ้าสามารถวัดสิ่งที่ต้องการวัดถูกต้องตรงตามความเป็นจริง อย่างไรก็ตามความตรงของแบบทดสอบมีหลายด้านด้วยกันดังนี้

1.1 ความเที่ยงตรงตามเนื้อหา (Content validity) พวงรัตน์ ทวีรัตน์ (2540: 115) อธิบายว่า ความเที่ยงตรงตามเนื้อหาเป็นความสอดคล้องระหว่างเนื้อหาสาระของเครื่องมือที่สร้างขึ้นกับเนื้อหาสาระของสิ่งที่ต้องการศึกษา ถ้าเครื่องมือใดสร้างได้สอดคล้องกับเนื้อหาสาระของสิ่งที่ต้องการศึกษา แสดงว่าเครื่องมือที่มีความเที่ยงตรงตามเนื้อหา ไพศาล หวังพานิช (2547: 57) เพิ่มเติมว่า ความเที่ยงตรงตามเนื้อหาเป็นสิ่งที่จำเป็นอย่างมากสำหรับข้อทดสอบวัดผลสัมฤทธิ์ เป็นคุณภาพของเครื่องมือที่สามารถวัดเนื้อหาวิชาหรือประสบการณ์การเรียนรู้ที่ต้องการจะวัดได้อย่างแท้จริงและครบถ้วน

วิธีการตรวจสอบเครื่องมือว่ามีความเที่ยงตรงตามเนื้อหากระทำดังนี้

1. ตรวจสอบความสอดคล้องของเนื้อหาของข้อทดสอบรายวิชาที่สร้างขึ้นกับเนื้อหาที่ระบุไว้ในตารางวิเคราะห์หลักสูตร รวมทั้งตรวจสอบสัดส่วนของจำนวนข้อคำถามในแต่ละเนื้อหาด้วย
2. ตรวจสอบโดยอาศัยดุลยพินิจของผู้เชี่ยวชาญซึ่งควรมีอย่างน้อย 3 ท่าน ถ้าผู้เชี่ยวชาญทุกท่านมีความเห็นตรงกันว่าใช้ได้ แสดงว่าแบบทดสอบนั้นมีความเที่ยงตรงตามเนื้อหา หากมีความเห็นไม่ตรงกันก็ควรปรับปรุงแก้ไข

1.2 ความเที่ยงตรงเชิงโครงสร้าง (Construct validity) พวงรัตน์ ทวีรัตน์. (2540: 116); และ ไพศาล หวังพานิช (2547: 57-58) ให้ความเห็นว่า ความเที่ยงตรงเชิงโครงสร้างนี้

เป็นคุณลักษณะที่สำคัญของเครื่องมือวัดทุกชนิด ที่จะต้องสามารถวัดพฤติกรรมหรือคุณลักษณะของ
สิ่งนั้นได้อย่างแท้จริง ครบถ้วนตามความต้องการ ตรงตามเป้าหมายของสิ่งที่ต้องการจะวัด

การตรวจสอบหาความเที่ยงตรงตามโครงสร้างตามความเห็นของ

พวงรัตน์ ทวีรัตน์. (2540: 116-119) สามารถกระทำได้ดังนี้

1. ตรวจสอบความสอดคล้องของข้อทดสอบกับมวลพฤติกรรมที่ข้อ
ทดสอบฉบับนั้นต้องการวัด
2. ตรวจสอบความสอดคล้องของลักษณะพฤติกรรมของข้อคำถาม
และสัดส่วนของจำนวนข้อคำถามในแต่ละพฤติกรรมกับตารางวิเคราะห์หลักสูตรตามรายวิชานั้นๆ
3. ตรวจสอบโดยอาศัยดุลยพินิจของผู้เชี่ยวชาญ หรือที่เรียกว่าการ
ตรวจสอบความเที่ยงตรงเชิงพินิจ (Face validity) โดยการตรวจสอบจะกระทำเช่นเดียวกับการ
ตรวจสอบความเที่ยงตรงตามเนื้อหา แต่จะเป็นการตรวจสอบที่มีลักษณะพฤติกรรมของข้อคำถาม
แทนเนื้อหา
4. ตรวจสอบความสอดคล้องระหว่างคะแนนรายข้อกับคะแนนทั้ง
ฉบับ โดยการหาค่าสหสัมพันธ์ (Item-test correlation) โดยถือว่าคะแนนรวมทั้งฉบับเป็นคะแนนที่
แทนโครงสร้างรวมของคุณลักษณะที่ต้องการวัด ถ้าข้อคำถามใดมีค่าสหสัมพันธ์กับคะแนนรวมสูง
(มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ) ข้อคำถามข้อนั้นก็จะมีค่าเที่ยงตรงเชิงโครงสร้างสูง
5. ตรวจสอบโดยนำเครื่องมือที่สร้างขึ้นไปทำการสอบวัดกับกลุ่ม
ตัวอย่างที่มีคุณลักษณะตรงกับคุณลักษณะที่ต้องการศึกษา แล้วนำผลที่ได้ไปเปรียบเทียบกับผลการ
วัดกับอีกกลุ่มหนึ่งซึ่งไม่มีคุณลักษณะตรงตามที่ต้องการศึกษา
6. ตรวจสอบความสัมพันธ์ระหว่างเครื่องมือที่สร้างขึ้นกับ
แบบทดสอบมาตรฐาน โดยนำเครื่องมือที่สร้างขึ้นกับแบบทดสอบมาตรฐานที่จะใช้เทียบไปสอบวัด
กับกลุ่มตัวอย่างซึ่งเป็นกลุ่มทดลอง แล้วนำคะแนนทั้ง 2 ชุดมาหาค่าสหสัมพันธ์กันด้วยวิธีของ
เพียร์สัน ถ้ามีค่าสหสัมพันธ์สูงก็แสดงว่าเครื่องมือที่สร้างขึ้นมีความเที่ยงตรงเชิงโครงสร้าง

การตรวจสอบความเที่ยงตรงตามเนื้อหา และความเที่ยงตรงตาม
โครงสร้างโดยอาศัยดุลยพินิจของผู้เชี่ยวชาญนี้สามารถตรวจสอบไปพร้อมกันได้ (พวงรัตน์ ทวีรัตน์.
2540: 117) และสามารถวิเคราะห์ออกมาในเชิงปริมาณหรือตัวเลขได้ ด้วยการหาค่าดัชนีความ
สอดคล้องระหว่างข้อคำถามกับลักษณะเฉพาะกลุ่มพฤติกรรม (Index of Consistency) โดยการนำ
เครื่องมือหรือแบบทดสอบที่สร้างขึ้นไปให้ผู้เชี่ยวชาญอย่างน้อย 3 ท่าน พิจารณาลงความเห็นและ
ให้เป็นคะแนนดังนี้

- +1 เมื่อแน่ใจว่าข้อคำถามนั้นเป็นตัวแทนลักษณะเฉพาะกลุ่มพฤติกรรม
นั้น
- 0 เมื่อไม่แน่ใจว่าข้อคำถามนั้นเป็นตัวแทนลักษณะเฉพาะกลุ่มพฤติกรรม
นั้นหรือไม่

-1 เมื่อแน่ใจว่าข้อคำถามไม่เป็นตัวแทนลักษณะเฉพาะของกลุ่มพฤติกรรมนั้น

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

เมื่อ IOC หมายถึง ดัชนีความสอดคล้องระหว่างข้อคำถามและพฤติกรรม

เมื่อ R หมายถึง ผลรวมของคะแนนความคิดเห็นของผู้เชี่ยวชาญ

เมื่อ N หมายถึง จำนวนผู้เชี่ยวชาญ

ค่าดัชนี IOC ที่คำนวณได้มากกว่าหรือเท่ากับ 0.5 แสดงว่าข้อคำถามนั้นเป็นตัวแทนลักษณะเฉพาะของกลุ่มพฤติกรรมนั้น ถ้าหากข้อคำถามใดมีค่าดัชนีต่ำกว่า 0.5 ข้อคำถามนั้นก็จะถูกตัดออกไป หรือนำไปปรับปรุงแก้ไขใหม่ให้ดีขึ้นต่อไป

1.3 ความเที่ยงตรงตามสภาพ (Concurrent validity) หมายถึงคุณภาพของแบบทดสอบที่สามารถบ่งชี้ได้อย่างถูกต้องตามสภาพที่เป็นจริงของผู้ถูกสอบ เช่น ผู้ที่ตอบข้อทดสอบได้ถูกต้องมากๆ แสดงว่าเป็นคนเก่ง และในสภาพชีวิตจริงของผู้รับการทดสอบก็เป็นเช่นนั้น แสดงว่าแบบทดสอบดังกล่าวมีความเที่ยงตรงตามสภาพ

1.4 ความเที่ยงตรงเชิงพยากรณ์ (Predictive validity) หมายถึงคุณภาพในการวัดของแบบทดสอบที่สามารถนำผลจากการวัดไปพยากรณ์ความสำเร็จในการเรียนวิชาต่างๆ หรือการประกอบอาชีพต่างๆในอนาคตได้อย่างถูกต้อง สำเร็จ บุญเรืองรัตน์ (2548: 111-112) กล่าวว่า ความเที่ยงตรงเชิงพยากรณ์สามารถคำนวณเป็นค่าตัวเลขทางสถิติได้ หรือบอกคุณภาพของความเที่ยงตรงเป็นตัวเลขได้ ดรรชนีที่นิยมใช้คือ สัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient) ของคะแนนจากผลการทดสอบกับคะแนนที่วัดได้จากเกณฑ์ที่กำหนดไว้

ดังนั้นจะเห็นได้ว่าความเที่ยงตรงของแบบทดสอบมีอยู่หลายด้านเช่น ความเที่ยงตรงตามเนื้อหา ความเที่ยงตรงเชิงโครงสร้าง ความเที่ยงตรงตามสภาพ และความเที่ยงตรงเชิงพยากรณ์ ซึ่งเป็นคุณภาพที่สำคัญของเครื่องมือ สามารถตรวจสอบได้จากเนื้อหาของแบบทดสอบที่สร้างขึ้นให้มีความสอดคล้องกับหลักสูตร สามารถวัดพฤติกรรมของสิ่งที่ต้องการวัดได้และตรวจสอบจากดุลยพินิจของผู้เชี่ยวชาญ

2. ความเชื่อมั่นของแบบทดสอบ (Reliability) นักการศึกษา และนักจิตวิทยาได้ให้ความหมายของความเชื่อมั่นต่างๆดังนี้

อนาสตาซี (Anastasi. 1968: 105) ให้ความหมายว่าความเชื่อมั่นเป็นความคงที่ของคะแนนที่ได้จากการทดสอบบุคคลเดียวกันแต่ต่างเวลาและโอกาสกัน

บราวน์ (Brown. 1976: 54) ได้ให้ความหมายเชิงปฏิบัติว่า ความเชื่อมั่นเท่ากับ อัตราส่วนของความแปรปรวนที่ได้จากคะแนนจริง (True score) กับความแปรปรวนของคะแนนที่สังเกตได้ (Observed score)

อีเบล และพริสบี (Ebel; & Prisbie. 1986: 71) ให้ความหมายว่าความเชื่อมั่นเป็นค่าสัมประสิทธิ์ระหว่างคะแนนชุดหนึ่งกับคะแนนอีกชุดหนึ่งของแบบทดสอบ ที่มีลักษณะเหมือนกันสองชุดและเป็นอิสระจากกัน ซึ่งได้จากนักเรียนหรือผู้ตอบกลุ่มเดียวกัน

บุญเชิด ภิญโญนนท์พงษ์ (2521: 269) ให้ความหมายว่า ความเชื่อมั่นหมายถึงความคงที่ของคะแนนซึ่งได้จากการวัดนักเรียนกลุ่มเดียวกันด้วยแบบทดสอบฉบับเดียวกันหลายๆ ครั้ง หรือด้วยแบบทดสอบสองฉบับที่มีลักษณะเหมือนกัน หรือภายใต้เงื่อนไขของตัวแปรอื่นๆ ในการวัดนั้น

เยาวดี วิบูลย์ศรี (2535: 31) กล่าวว่าแบบทดสอบที่มีความเชื่อมั่นนั้น เป็นแบบทดสอบที่ผู้รับการทดสอบจะต้องได้คะแนนในการทำแบบทดสอบในเวลาที่แตกต่างกัน คงเส้นคงวาหรือไม่แตกต่างกัน ถ้าคะแนนมีความใกล้เคียงกันมากเท่าไรก็แสดงว่าแบบทดสอบมีความเชื่อมั่นมากเท่านั้น เราสามารถที่จะแสดงให้เห็นความเชื่อมั่นของแบบทดสอบ โดยการหาค่าสัมประสิทธิ์แห่งความเชื่อมั่น (Reliability coefficient) ค่าสัมประสิทธิ์แห่งความเชื่อมั่นมีค่าเท่ากับ 1 แสดงว่าผู้ทำทดสอบจะได้คะแนนเช่นเดิมไม่ว่าจะทำแบบทดสอบนี้เมื่อไหร่ โดยทั่วไปความเชื่อมั่นของแบบทดสอบจะต้องไม่ต่ำกว่า .50 พอใช้ได้อยู่ระดับ .60, .70 แต่ถ้าจะให้ดีควรสูงกว่า .80 ขึ้นไป

ล้วน สายยศ; และอังคณา สายยศ (2539: 209) ให้ความหมายว่า ความเชื่อมั่นของแบบทดสอบหมายถึง ความคงที่ของคะแนนที่ได้จากการทดสอบนักเรียนคนเดียวกันหลายครั้งในแบบทดสอบชุดเดิม

จากความหมายข้างต้นสามารถสรุปได้ว่า ความเชื่อมั่นแสดงถึงความคงที่ของคะแนนจากการทดสอบผู้ตอบกลุ่มเดียวกัน ด้วยแบบทดสอบฉบับเดียวกัน หรือแบบทดสอบที่มีลักษณะคู่ขนานกัน

นอกจากนี้ ล้วน สายยศ; และอังคณา สายยศ (2539: 210 - 244) อธิบายถึงความเชื่อมั่นว่า โดยทั่วไปแล้วความเชื่อมั่นแบ่งเป็น 2 ประเภท คือ ความเชื่อมั่นของแบบทดสอบอิงกลุ่ม (Reliability of norm-referenced test) และความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ (Reliability of criterion-referenced test) ในแต่ละประเภทของการคำนวณหาความเชื่อมั่นนั้นจะมีสูตรที่ใช้ในการประมาณค่าความเชื่อมั่นอยู่ด้วยกันหลายสูตรด้วยกันซึ่งมีวิธีคำนวณจำแนกตามลักษณะของแบบทดสอบอิงกลุ่ม และอิงเกณฑ์ ดังนี้

1. ความเชื่อมั่นของแบบทดสอบอิงกลุ่ม มีวิธีคำนวณหาค่าสัมประสิทธิ์อยู่ 3 วิธี คือ

1.1 ความเชื่อมั่นแบบความคงที่ของคะแนน (Stability reliability) เป็นการประมาณค่าความเชื่อมั่น โดยพิจารณาจากคะแนนที่ได้จากการทดสอบ 2 ครั้งในเครื่องมือวัดชุดเดียวกัน วิธีนี้เป็นวิธีการทดสอบซ้ำ (Test-retest Method) จุดอ่อนของวิธีนี้อยู่ตรงช่วงระยะเวลาของการทดสอบซ้ำ กล่าวคือ ถ้าช่วงเวลาห่างกันมากเกินไป ผู้รับการทดสอบอาจลืม หรือถ้าช่วงเวลาสั้นเกินไป ผู้รับการทดสอบอาจจำแบบทดสอบได้

1.2 ความเชื่อมั่นโดยใช้แบบทดสอบที่เหมือนกันสองฉบับ (Equivalent-form reliability) เป็นการประมาณค่าความเชื่อมั่นโดยใช้แบบทดสอบคู่ขนานกัน (Parallel form) ไปทดสอบกับนักเรียนกลุ่มเดียวกัน จุดอ่อนของวิธีนี้คือ ปัญหาการสร้างเครื่องมือวัดเนื้อหาเดียวกัน และต้องมีคะแนนเฉลี่ย ความแปรปรวน และค่าสถิติอื่นเท่ากันด้วย จึงเป็นการยากที่จะสร้างเครื่องมือให้มีเนื้อหาเดียวกัน หรือมีความคู่ขนานกัน

1.3 ความเชื่อมั่นโดยใช้ความสอดคล้องภายใน (Internal consistency reliability) การประมาณค่าความเชื่อมั่นจากแบบทดสอบฉบับเดียวทดสอบเพียงครั้งเดียวโดยใช้วิธีดังนี้

1.3.1 วิธีแบ่งครึ่งทดสอบ (Split-half method) วิธีนี้จะแบ่งแบบทดสอบเป็นสองส่วนโดยแบ่งให้แต่ละส่วนมีลักษณะเป็นคู่ขนานกัน ดังนั้นจึงนิยมแบ่งเป็นฉบับข้อคู่กับฉบับข้อคี่ โดยตรวจให้คะแนนแยกเป็นคะแนนข้อคู่กับคะแนนข้อคี่ โดยใช้สูตรของเพียร์สัน (Pearson product-moment coefficient correlation) จะได้ค่าความเชื่อมั่นเพียงครึ่งฉบับ จากนั้นจึงนำไปหาค่าความเชื่อมั่นของทั้งฉบับ โดยใช้สูตรของสเปียร์แมนบราวน์ (Spearman-Brown formula)

1.3.2 วิธีของคูเดอร์-ริชาร์ดสัน (Kuder-Richardson procedure) ในปี ค.ศ. 1937 คูเดอร์และริชาร์ดสันได้พัฒนาสูตรที่หาความเชื่อมั่นให้ง่ายขึ้น โดยที่คะแนนในแต่ละข้อจะต้องอยู่ในลักษณะที่ถูกทำได้ 1 คะแนน ทำผิดได้ 0 คะแนนเท่านั้น วิธีนี้มีสูตรที่ใช้หาค่าความเชื่อมั่นอยู่ 2 สูตรคือ สูตร K-R 20 กับ K-R 21

1.3.3 วิธีของครอนบัค (Cronbach Alpha procedure) ครอนบัคได้พัฒนาสูตรหาความเชื่อมั่นในรูปสัมประสิทธิ์แอลฟา (α - Coefficient) ในปี ค.ศ.1951 โดยพัฒนามาจากสูตร K-R 20 ทั้งนี้เป็นเพราะจะได้ใช้หาค่าความเชื่อมั่นกับเครื่องมือที่ไม่ได้ตรวจให้คะแนนเป็น 1 กับ 0 โดยจะตรวจให้คะแนนในลักษณะใดก็ได้

1.3.4 วิธีของฮอยท์ (Hoyt's ANOVA procedure) การหาค่าความเชื่อมั่นโดยวิธีนี้เหมาะสำหรับเครื่องมือที่ใช้ในการเก็บข้อมูลประเภทตรวจให้คะแนนต่างกัน ในแต่ละข้อ เช่นเดียวกับการหาค่าความเชื่อมั่นแบบสัมประสิทธิ์แอลฟา แต่วิธีการคำนวณแบบฮอยท์นี้ใช้หลักสถิติของการวิเคราะห์ความแปรปรวน

1.4 ความเชื่อมั่นที่มีผู้ให้คะแนนมากกว่า 1 คน ซึ่งหาในรูปของสัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient)

1.5 ความเชื่อมั่นของคะแนนผลต่างของแบบทดสอบ คะแนนผลต่างของแบบทดสอบ (Difference Score : $D = X - Y$) มี 3 ลักษณะคือ

1.5.1 คะแนนผลต่างระหว่างนักเรียน 2 คนที่สอบแบบทดสอบฉบับเดียวกัน

1.5.2 คะแนนผลต่างระหว่างคะแนนของแบบทดสอบ 2 ฉบับ ที่สอบกับนักเรียนคนเดียว

1.5.3 คะแนนผลต่างของการทดสอบ 2 ครั้งในแบบทดสอบฉบับเดียวกัน
ของนักเรียนคนเดียวกัน (Pretest-Posttest)

1.6 ความเชื่อมั่นจากการแบ่งส่วนย่อยของแบบทดสอบ โดยแบ่งแบบทดสอบเป็นส่วนย่อย (Parts) และควรจะหาในรูปแบบของคอนเจนเนอริก (Congeneric form) คือมีข้อตกลงเบื้องต้น (Assumption) ว่าส่วนย่อยแต่ละส่วนนั้นมีคะแนนจริง (True score) สัมพันธ์กันในเชิงเส้นตรง การหาความเชื่อมั่นประเภทนี้ เป็นแบบความสอดคล้องภายใน (Internal Consistency) โดยเอาคะแนนจากการทดสอบเพียงครั้งเดียวมาคำนวณจากสูตรของราจู (Raju. 1970) , สูตรของคริสทอฟ (Kristof. 1971) , สูตรของราจู (Raju. 1977) และเฟลด์ท์ (Feldt. 1975)

2. ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์

2.1 ความเชื่อมั่นแบบหาความคงที่ของความรอบรู้ (Stability reliability)

2.2 ความเชื่อมั่นแบบความสอดคล้องในการตัดสินใจ (Decision consistency)

2.3 ความเชื่อมั่นของแบบทดสอบอิงเกณฑ์จากการทดสอบเพียงครั้งเดียว

2.4 ความเชื่อมั่นที่คำนึงจุดประสงค์เชิงพฤติกรรมของแบบทดสอบอิงเกณฑ์

3. ความยากง่าย (Difficulty) คือสัดส่วนที่แสดงว่าข้อทดสอบนั้นมีคนทำถูกมากหรือน้อย ถ้ามีคนทำถูกมากก็เป็นข้อทดสอบง่าย ถ้ามีคนทำถูกน้อยก็เป็นข้อทดสอบยาก การหาค่าความยากง่ายเป็นวิธีตรวจสอบคุณภาพด้านพุทธิพิสัย (Cognitive domain) และเป็นแบบทดสอบในระบบอิงกลุ่ม (Norm-referenced test) มีลักษณะเป็นการวิเคราะห์รายข้อ (Item analysis) ค่าความยากง่ายมีค่าอยู่ระหว่าง 0 ถึง 1 นิยมเขียนแทนด้วย P ค่าร้อยละหรือสัดส่วนที่คำนวณได้มีความหมายดังนี้

ตาราง 2 ค่าและความหมายความยากง่าย

ค่าความยากง่าย		ความหมายระดับความยาก	คุณภาพข้อทดสอบ
ร้อยละ	สัดส่วน		
80-100	0.8-1.0	ง่ายมาก	ไม่ดีต้องตัดทิ้งหรือปรับปรุงใหม่
60-79	0.6-0.79	ง่าย	พอใช้ได้
40-59	0.4-0.59	ปานกลาง	ดีมาก
20-39	0.2-0.39	ยาก	พอใช้ได้
0-19	0-0.19	ยากมาก	ไม่ดีต้องตัดทิ้งหรือปรับปรุงใหม่

ที่มา: ศูนย์พัฒนาทรัพยากรการศึกษา มหาวิทยาลัยมหาสารคาม. (2004). การตรวจสอบคุณภาพเครื่องมือที่ใช้ในการวิจัย. (ออนไลน์)

จากตารางสามารถสรุปได้ว่า ถ้าค่า P น้อยแปลว่าข้อทดสอบมีความยาก ถ้าค่า P มากแปลว่า ข้อทดสอบมีความง่าย ข้อทดสอบที่คัดเลือกมาใช้ในการเก็บรวบรวมข้อมูลควรเป็นข้อทดสอบที่มีความยากปานกลางคือ ประมาณ 0.5 แต่ในทางปฏิบัติมักกำหนดเกณฑ์ระดับความยากของข้อทดสอบที่จะเลือกไว้ใช้ในช่วง 0.2–0.8 ถ้าค่า P อยู่นอกขอบเขตนี้ถือว่าข้อทดสอบนั้นง่ายหรือยากเกินไป

4. อำนาจจำแนก (Discrimination) คือความสามารถของเครื่องมือในการจำแนกบุคคลออกเป็นสองกลุ่มที่ต่างกัน คือกลุ่มเก่ง-กลุ่มอ่อน หรือกลุ่มสูง-กลุ่มต่ำ การหาค่าอำนาจจำแนกใช้ในการตรวจสอบคุณภาพเครื่องมือ ในการวิจัยประเภทแบบทดสอบ แบบสอบถามและแบบวัดเจตคติ มีลักษณะเป็นการวิเคราะห์รายข้อ ค่าอำนาจจำแนกจะมีค่าอยู่ระหว่าง (-1.00) ถึง (+1.00) นิยมแทนด้วย r เกณฑ์การพิจารณาค่าอำนาจจำแนกแสดงในตารางได้ดังนี้

ตาราง 3 เกณฑ์การพิจารณาค่าอำนาจจำแนก

ค่าอำนาจจำแนก	ความหมายของคุณภาพข้อทดสอบ
0.40 ขึ้นไป	ดีมาก
0.30-0.39	ดีพอสมควร
0.20-0.29	พอใช้ได้แต่ควรปรับปรุง
0.19 ลงไป	ไม่ดีต้องตัดทิ้งหรือปรับปรุงใหม่

ที่มา: ศูนย์พัฒนาทรัพยากรการศึกษา มหาวิทยาลัยมหาสารคาม. (2004). *การตรวจสอบคุณภาพเครื่องมือที่ใช้ในการวิจัย*. (ออนไลน์)

บุญเชิด ภิญโญนนตพงษ์ (2548: 135) สรุปว่า ถ้าข้อทดสอบใดมีค่าอำนาจจำแนกเข้าใกล้ +1.00 แสดงว่าข้อทดสอบนั้นมีค่าอำนาจจำแนกมาก หรือมีประสิทธิภาพในการแยกผู้ตอบที่มีความสามารถและผู้ตอบที่ไม่มีความสามารถได้มาก ถ้าข้อทดสอบมีค่าอำนาจจำแนกเข้าใกล้ 0 แสดงว่าข้อทดสอบนั้นไม่มีอำนาจจำแนกหรือไม่มีประสิทธิภาพในการแยกผู้ตอบที่มีความสามารถออกจากผู้ตอบที่ไม่มีความสามารถ สำหรับข้อทดสอบที่มีค่าอำนาจจำแนกเป็นลบเกิดจากข้อทดสอบซึ่งผู้ตอบที่มีความสามารถส่วนใหญ่ตอบผิด แต่ผู้ตอบที่ไม่มีความสามารถส่วนใหญ่ตอบถูก ซึ่งถือว่าข้อทดสอบข้อนั้นเป็นข้อทดสอบที่ไม่ดี

5. ความเป็นปรนัย (Objectivity) เป็นคุณสมบัติที่จำเป็นสำหรับเครื่องมือวิจัยทุกชนิด ที่แสดงให้เห็นว่า คำถามที่ใช้มีความชัดเจน เมื่ออ่านแล้วมีความเข้าใจตรงกัน การตรวจให้คะแนนมีความแน่นอนตรงกันตลอดจนมีความชัดเจนในการแปลความหมายของคะแนนที่ตรงกันด้วย เครื่องมือที่มีความเป็นปรนัยจะมีความเที่ยงตรงและความเชื่อมั่นสูง การตรวจสอบความเป็นปรนัยนี้สามารถกระทำโดยผู้ทำวิจัยเองหรือให้ผู้เชี่ยวชาญทำการตรวจสอบอีกครั้งหนึ่งก็ได้

นอกจากนี้แฮริส (Harris. 1969: 21-23) ยังมีความเห็นเพิ่มเติมว่านอกจากแบบทดสอบที่มีคุณภาพด้าน ความเที่ยงตรง และความเชื่อมั่นแล้ว ควรมีการคำนึงถึงความสะดวกของแบบทดสอบในการนำไปใช้ด้วยเช่นกัน คุณสมบัติดังกล่าวประกอบด้วย ความประหยัด ความสะดวกในการบริหารจัดการและการให้คะแนน ตลอดจนความสะดวกและง่ายในการวิเคราะห์ตีความ

3.3 เกณฑ์ปกติ (Norm)

เกณฑ์ปกติหมายถึง ข้อเท็จจริงทางสถิติที่บรรยายการแจกแจงของคะแนนจากประชากรที่นิยามไว้อย่างดีแล้ว และเป็นคะแนนตัวที่บอกระดับความสามารถของผู้รับการทดสอบว่าอยู่ระดับใดของกลุ่มประชากร (ล้วน สายยศ; และอังคณา สายยศ. 2539: 313-314)

เยาวดี วิบูลย์ศรี (2539: 84) กล่าวว่า เกณฑ์ปกติของแบบทดสอบ (Norm) หมายถึง คะแนนมาตรฐานที่ได้จากการที่กลุ่มตัวอย่างประชากรขนาดใหญ่ทำแบบทดสอบฉบับนั้นๆ ในสถานะที่เป็นมาตรฐานเดียวกัน คะแนนมาตรฐานดังกล่าวได้มาโดยการแปลงคะแนนดิบจากการทดสอบให้เป็นมาตรฐานในรูปแบบใดรูปแบบหนึ่งเพื่อประโยชน์ในการเปรียบเทียบ ทั้งนี้เมื่อต้องการทราบสถานภาพของผู้รับการทดสอบอื่นที่ตอบแบบทดสอบฉบับนั้นภายหลัง ซึ่งสามารถกระทำได้โดยการเปรียบเทียบคะแนนของผู้รับการทดสอบนั้นกับเกณฑ์ปกติที่สร้างขึ้น

ชวาล แพรัตกุล (2548: 174-182) อธิบายเพิ่มเติมว่า แนวความคิดในการสร้างเกณฑ์ปกติเกิดจากความเชื่อที่ว่า ข้อทดสอบที่มีคุณภาพดีและเหมาะสมมีความยากง่ายเป็นมาตรฐานไม่ได้แสดงว่าข้อทดสอบนั้นมีคุณสมบัติที่ดีในด้านใด หรือความยากง่ายเป็นเท่าใด การนำตัวเลขคะแนนมาเป็นเครื่องชี้บอกปริมาณความรู้ของผู้เรียนว่าเก่งกว่าอีกวิชาหนึ่งเป็นการพิจารณาที่ไม่ถูกต้องเนื่องจากคะแนนแต่ละวิชามีหน่วยที่ไม่เท่ากัน ไม่สามารถนำคะแนนดิบของวิชาต่างๆ มาเทียบรวมกันได้โดยตรงเพราะมีความแตกต่างกันในธรรมชาติแต่ละเนื้อหาของแต่ละวิชา จึงทำให้คะแนนของแต่ละวิชาไม่เท่ากัน

สรุปได้ว่าเกณฑ์ปกติหมายถึง คะแนนมาตรฐานที่ใช้ในการเปรียบเทียบความสามารถของผู้รับการทดสอบว่าอยู่ในระดับใด โดยการนำคะแนนดิบที่ได้มาเปลี่ยนให้เป็นคะแนนมาตรฐานที่มีความหมาย สามารถนำไปใช้ประโยชน์ในการวัดผลต่อไป

ความสำคัญของการสร้างเกณฑ์ปกติอยู่ที่จะต้องแบ่งคะแนนออกจากจุดหลักให้เป็นช่วงๆ ที่เท่ากันอย่างมีความหมาย ซึ่งไม่สามารถใช้คะแนนดิบหรือเปอร์เซ็นต์มาแบ่งได้เนื่องจากคะแนนดิบมีหน่วยที่ไม่เท่ากันจำเป็นต้องใช้วิธีการทางสถิติมาช่วยแปลงคะแนนต่างๆ ให้มีหน่วยที่เทียบเท่ากันก่อน ผลที่ได้นำไปเทียบหาคะแนนมาตรฐานซึ่งเป็นคะแนนชี้ หรือคะแนนที่ปกติจากตารางสำเร็จรูปต่อไป

3.4 คะแนนมาตรฐาน (Standard score)

ภัทธา นิคมานนท์ (2543: 273-247) อธิบายว่า คะแนนมาตรฐานเป็นคะแนนที่มีการกระจายของคะแนนเป็นโค้งปกติ (Normal curve) ซึ่งโค้งปกติเป็นโค้งที่เกิดจากการแจกแจงของข้อมูลชุดหนึ่งๆเป็นรูประฆังคว่ำ (Bell shape) ที่มีค่าเฉลี่ย ($\bar{X}$) อยู่ตรงกลาง และแบ่งพื้นที่ด้านซ้ายและขวาออกเป็น 2 ส่วนที่เท่ากัน มีช่วงคะแนนที่เท่ากัน และสามารถนำคะแนนมา บวก ลบ คูณหาร หรือเปรียบเทียบกันได้ รัตนะ บัวสนธ์ (2545: 190-193) เพิ่มเติมว่า คะแนนมาตรฐาน

เป็นคะแนนที่ได้จากการแปลงคะแนนดิบให้มีค่าเฉลี่ย ($\bar{X}$) และส่วนเบี่ยงเบนมาตรฐาน (S) คงที่ หรือให้เป็นหน่วยเดียวกัน คะแนนมาตรฐานมีหลายประเภทได้แก่

1. คะแนนมาตรฐานซี (Standard Z-score) หาได้โดยการนำคะแนนที่ได้ ลบ คะแนนเฉลี่ยและหารด้วยส่วนเบี่ยงเบนมาตรฐานของคะแนนชุดนั้น คะแนนมาตรฐานซีจะมีคะแนนเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ซึ่งคะแนนมาตรฐานซีอาจมีค่าได้เป็นทั้งค่าบวกและค่าลบ ดังนั้นอาจทำให้เข้าใจได้ยากจึงทำให้มีการแปลงคะแนนดังกล่าวเป็นคะแนนมาตรฐานที่ เพื่อให้คะแนนที่ได้ไม่มีค่าติดลบ

2. คะแนนมาตรฐานที (Standard T-score) เป็นคะแนนที่ได้จากการแปลง คะแนนดิบในเชิงเส้นตรง (Linear transformation) เป็นคะแนนมาตรฐานที่แปลงมาจากคะแนนซี เพื่อแก้ปัญหาคะแนนที่มีค่าเป็นลบและเป็นทศนิยม คะแนนมาตรฐานทีจะมีค่าเฉลี่ยเท่ากับ 50 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10

3. คะแนนมาตรฐานทีปกติ (Standard normalized T-score) หรือคะแนนทีปกติ เป็นคะแนนที่ได้จากการแปลงคะแนนดิบให้เป็นการกระจายแบบโค้งปกติ (Normal curve) คำนวณ โดยอาศัยพื้นที่ใต้โค้งปกติเป็นหลัก โดยถือว่าพื้นที่ดังกล่าวใช้แทนจำนวนคนในกลุ่มที่เข้าสอบจะเป็นการบอกให้รู้ว่าจุดคะแนนนั้นๆมีจำนวนคนที่ได้คะแนนน้อยกว่าอยู่ร้อยละเท่าไร คะแนนทีปกติ จะมีช่วงคะแนนระหว่าง 0 ถึง 100 มีคะแนนเฉลี่ย ($\bar{X}$) เท่ากับ 50 ความเบี่ยงเบนมาตรฐาน (Standard deviation) เท่ากับ 10 (โชติ เพชรชื่น. 2548: 89) การแปลงคะแนนดิบให้เป็นคะแนนที่ ปกติทำได้โดยแปลงคะแนนดิบให้เป็นตำแหน่งเปอร์เซ็นต์ไทล์ก่อน แล้วเปรียบเทียบเปอร์เซ็นต์ไทล์ให้ เป็นคะแนนทีปกติจากตารางสำเร็จรูป ขั้นตอนดังกล่าวไม่จำเป็นต้องคำนวณค่าเฉลี่ย ($\bar{X}$) และส่วน เบี่ยงเบนมาตรฐาน (S) ของกลุ่ม (สมนึก ภัททิยธนี. 2546: 265–266)

วิธีคำนวณหาเปอร์เซ็นต์ไทล์ และหาคะแนนทีปกติมีลำดับขั้นตอนดังนี้

1. เรียงคะแนนจากมากไปน้อย
2. แจกแจงความถี่ของคะแนนดิบ (f) และความถี่สะสม (cf)
3. คำนวณความถี่สะสม $cf + \frac{1}{2}f$ (จะหาค่า $cf + \frac{1}{2}f$ ของขั้นใดต้องใช้ค่า cf ที่อยู่ก่อนถึงขั้นนั้นแต่ใช้ค่า f ของขั้นนั้น)
4. คำนวณเปอร์เซ็นต์ไทล์โดยเอา $\frac{100}{N}$ (N คือจำนวนผู้เข้าสอบทั้งหมด) ไปคูณกับค่า $cf + \frac{1}{2}f$ ของคะแนนแต่ละคะแนน
5. นำค่าเปอร์เซ็นต์ไทล์ที่ได้ไปเทียบหาค่า คะแนนทีปกติจากตาราง สำเร็จรูป

ตาราง 4 ตัวอย่างการแปลงคะแนนดิบให้เป็นเปอร์เซ็นต์ไทล์ และคะแนนที่ปกติ

X (คะแนนดิบ)	f (ความถี่)	cf (ความถี่สะสม)	$cf + \frac{1}{2}f$	% ile (เปอร์เซ็นต์ไทล์)	T (คะแนนที่ปกติ)
24	1	40	39.5	98.75	72
23	2	39	38.0	95.00	67
22	0	37	37.0	92.05	64
21	3	37	35.5	88.75	62
20	2	34	33.0	82.50	59
19	3	32	30.5	76.25	57
18	3	29	27.5	68.75	55
17	2	26	25.0	62.50	53
16	2	24	23.0	57.50	52
15	5	22	19.5	48.75	50
14	2	17	16.0	40.00	47
13	4	15	13.0	32.50	45
12	3	11	9.5	23.75	43
11	3	8	6.5	16.25	40
10	1	5	11.25	11.25	38
9	1	4	3.5	8.75	36
8	2	3	2.0	5.00	33
7	1	1	0.5	1.25	28

ที่มา: สมนึก ภัททิยธนี. (2541, กรกฎาคม). เกณฑ์ปกติ. วารสารการวัดผลการศึกษา.

4(1): 36.

ตาราง 5 ตารางสำหรับแปลงคะแนนที่ปกติให้เป็นจำนวนร้อยละที่อยู่เหนือกว่าผู้อื่น

คะแนนที่ ปกติ	% ที่อยู่เหนือ ผู้อื่น	คะแนนที่ ปกติ	% ที่อยู่เหนือ ผู้อื่น	คะแนนที่ ปกติ	% ที่อยู่เหนือ ผู้อื่น
10	.0032	37	9.68	64	91.92
11	.0048	38	11.51	65	93.32
12	.007	39	13.57	66	94.52
13	.001	40	15.87	67	95.54
14	.016	41	18.41	68	96.41
15	.023	42	21.19	69	97.13
16	.034	43	24.20	70	97.72
17	.048	44	27.43	71	98.21
18	.069	45	30.85	72	98.61
19	.097	46	34.46	73	99.93
20	.13	47	38.21	74	98.18
21	.19	48	42.07	75	99.38
22	.26	49	46.02	76	99.53
23	.35	50	50.00	77	99.65
24	.47	51	53.98	78	99.74
25	.62	52	57.93	79	99.81
26	.82	53	61.79	80	99.865
27	1.07	54	65.54	81	99.903
28	1.39	55	69.15	82	99.931
29	1.79	56	72.57	83	99.952
30	2.28	57	75.80	84	99.966
31	2.87	58	78.81	85	99.977
32	3.59	59	81.59	86	99.984
33	4.46	60	84.13	87	99.9890
34	5.48	61	86.43	88	99.9928
35	6.68	62	88.49	89	99.9952
36	8.08	63	90.32	90	99.9968

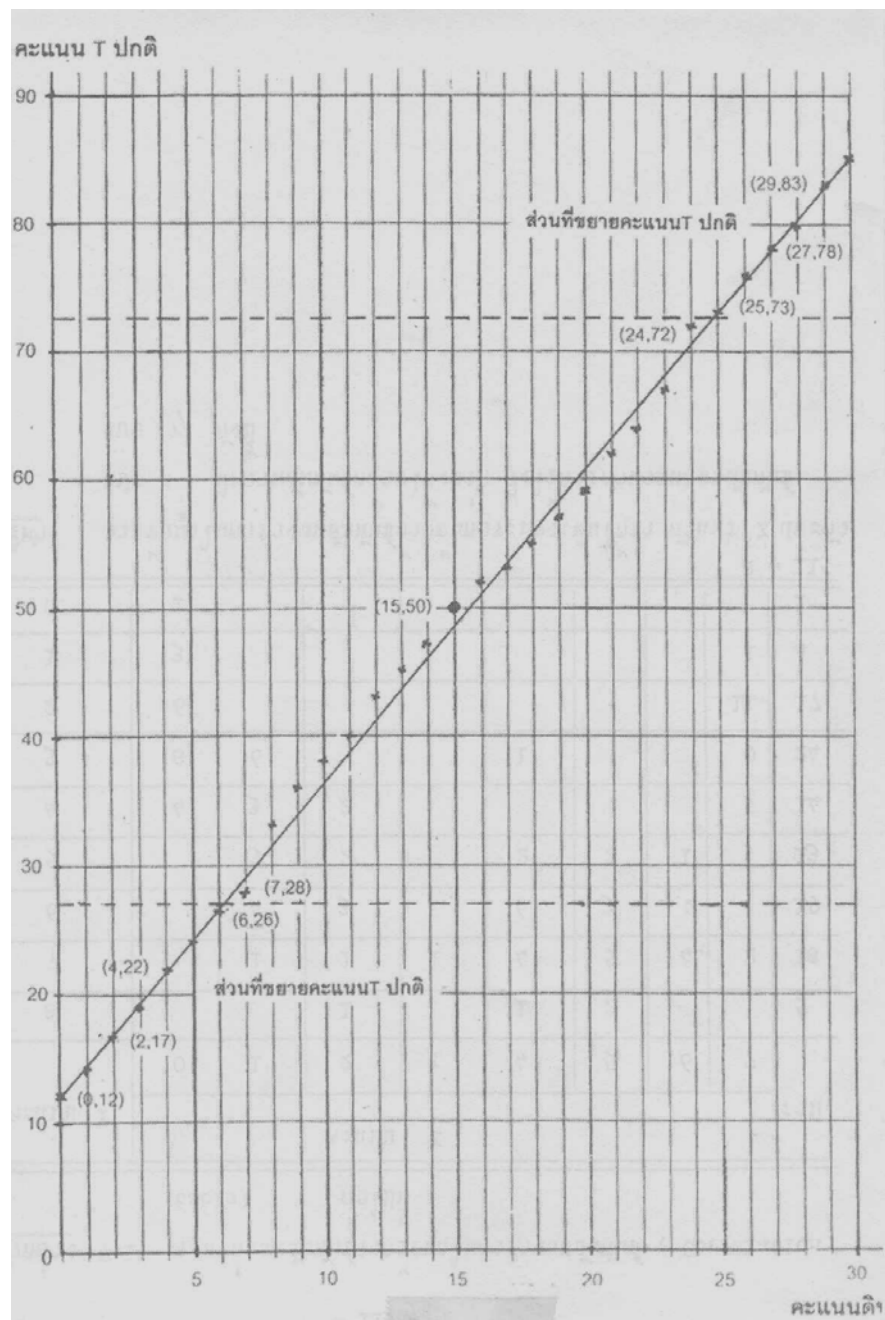
ที่มา: ชวาล แพริตกุล. (2548). เกณฑ์ปกติ. ใน สารานุกรมศึกษาศาสตร์. หน้า 178 – 179.

จากตารางและข้อมูลดังกล่าวจะเห็นว่า ผู้ที่มีคะแนนดิบที่ 20 คะแนนเมื่อนำมาคำนวณหา ค่าเปอร์เซ็นต์ไทล์จะมีค่าเท่ากับ 82.50 เมื่อนำมาเปรียบเทียบกับตารางคะแนนที่ปกติสำเร็จรูปจะมีค่า คะแนนที่เท่ากับ 59 แสดงว่าผู้รับการทดสอบจะอยู่ในตำแหน่งที่อยู่เหนือผู้อื่นมีจำนวน 83% จากนั้น ก็นำไปประเมินระดับความสามารถจากการเทียบค่าคะแนนที่ปกติจากตาราง (ตาราง 1) ซึ่งพบว่า ผู้รับการทดสอบจะมีระดับความสามารถอยู่ในระดับดี และจากตารางจะสังเกตเห็นว่ามีคะแนนที่บาง ตัวซ้ำกันที่ตรงหัวและตรงท้ายของช่วงคะแนน เช่น คะแนนที่ 35, 45, 55, และ 65 เป็นต้น ขวาล แพร์ตกุล (2548: 180) อธิบายว่า การที่เป็นเช่นนี้เนื่องจากคะแนนที่ดังกล่าวเป็นจุดแบ่งเขตระหว่าง กลุ่ม ซึ่งอาจทำให้เกิดความไม่แน่ใจในการจัดกลุ่มได้ การแก้ไขในกรณีนี้ทำได้โดยจัดผู้เรียนที่อยู่ใน จุดดังกล่าวขึ้นไปอยู่ในกลุ่มสูงที่อยู่ถัดไปเสมอ ทั้งนี้เพื่อผลทางด้านจิตวิทยาสำหรับผู้เรียน

3.5 วิธีสร้างเกณฑ์ปกติคะแนน

การเปลี่ยนคะแนนดิบเป็นคะแนนที่ปกติตามขั้นตอนข้างต้นทำให้คะแนนดิบทุก คะแนนที่ได้จากกลุ่มตัวอย่างถูกแปลงเป็นคะแนนที่ปกติ แต่คะแนนที่ได้ดังกล่าวอาจจะไม่ครอบคลุม คะแนนดิบทุกคะแนน (ตั้งแต่ 0–คะแนนเต็ม) ทั้งนี้เนื่องมาจากจำนวนผู้ที่เข้ารับการทดสอบอาจมีไม่ มากพอหรือข้อทดสอบมีความยากง่ายเกินไป ดังนั้นจึงมีความจำเป็นที่จะต้องสร้างเกณฑ์ปกติให้ ครอบคลุมคะแนนดิบทุกคะแนนด้วยการขยายคะแนนที่ปกติ เพื่อให้เป็นหลักเกณฑ์ในการสร้าง เกณฑ์ปกติและเกิดความสะดวกในการนำไปใช้ต่อไป

การขยายคะแนนที่ปกติทำได้โดยการเขียนกราฟจากคู่อันดับระหว่างคะแนนดิบกับ คะแนนที่ปกติ ลากเส้นตรงให้ผ่านจุดกราฟต่างๆที่มีอยู่ให้มากที่สุดโดยจะต้องพยายามลากเส้นตรง ให้ผ่านคะแนนที่ปกติที่ 50 (เนื่องจากคะแนนที่ปกติที่ 50 จะตรงกับค่าเฉลี่ยโดยประมาณของคะแนน ดิบ) จากนั้นอ่านกราฟโดยพิจารณาความสัมพันธ์ระหว่างคะแนนดิบและคะแนนที่ปกติ นำมา เรียงลำดับจากคะแนนเต็มไปยังคะแนนต่ำสุดให้ครอบคลุมคะแนนดิบทุกคะแนน (คะแนนที่ 0– คะแนนเต็ม) เพื่อนำไปใช้ในการสร้างเกณฑ์ปกติ (Norm) เรียงลำดับคะแนนดิบทุกคะแนนจาก คะแนนเต็ม และคะแนนที่ปกติที่ได้จากการอ่านค่ากราฟ มาสร้างเกณฑ์ปกติคะแนนที่ โดยเทียบจาก ตารางอ่านค่าที่ปกติให้ครบทุกค่าตั้งแต่คะแนน 0–คะแนนเต็ม แล้วจัดระดับกลุ่มผู้ทดสอบตาม ความสามารถที่ได้จากนั้นนำคะแนนดิบที่ได้จากการทดสอบครั้งต่อไปมาเทียบกับเกณฑ์ปกติที่สร้าง ขึ้น อ่านค่าเกณฑ์ปกติและสามารถจัดระดับความสามารถของผู้ทดสอบได้ทันทีโดยไม่ต้องเข้าสู่ กระบวนการหาเกณฑ์ปกติใหม่ และไม่ต้องเปรียบเทียบกับคะแนนของผู้อื่น



กราฟแสดงการสร้างเกณฑ์ปกติ และการขยายคะแนนที่ปกติ

ภาพประกอบ 3 กราฟแสดงการสร้างเกณฑ์ปกติ และการขยายคะแนนที่ปกติ

ที่มา: สมนึก ภัททิยธนี. (2541, กรกฎาคม). เกณฑ์ปกติ. วารสารการวัดผลการศึกษา.

ขั้นตอนนี้ทำให้สามารถแยกผู้เรียนออกเป็นกลุ่มตามความสามารถได้อย่างถูกต้อง และจากความรู้ดังกล่าว ผู้วิจัยจึงได้ศึกษาและสร้างเกณฑ์ปกติคะแนนที่ขึ้น ซึ่งนอกเหนือจากการศึกษาคุณภาพของแบบทดสอบแล้ว เกณฑ์ปกติคะแนนที่ยังช่วยจัดระดับความสามารถของพนักงานต้อนรับบนเครื่องบินให้เข้ารับการอบรมภาษาอังกฤษตามหลักสูตรที่จัดทำขึ้นได้อย่างถูกต้อง มีประสิทธิภาพ ทั้งยังสามารถนำไปทดลองกับผู้รับการทดสอบกลุ่มอื่นๆได้อีกด้วย

4. งานวิจัยที่เกี่ยวข้องกับคุณภาพของแบบทดสอบ

วรรณ ปฏิบัติศน์ (2533) ได้ศึกษาคุณภาพของแบบทดสอบที่ใช้ในการทดสอบคัดเลือกบุคคลเข้าศึกษาในมหาวิทยาลัยขอนแก่นโดยวิธีรับตรง โปรแกรมทั่วไป ปีการศึกษา 2531 โดยพิจารณาในเรื่อง ค่าความยากง่าย อำนาจจำแนก ความเชื่อมั่น และความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบ สำหรับวิชาภาษาไทย สังคมศึกษา ภาษาอังกฤษ คณิตศาสตร์ และวิทยาศาสตร์-กายภาพ-ชีวภาพ กลุ่มตัวอย่างมีจำนวน 8,005 คน ใช้วิธีสุ่มอย่างเป็นระบบจากทั้งหมด 19,446 คน ในการศึกษาความเที่ยงตรงเชิงพยากรณ์ ใช้กลุ่มตัวอย่าง 123 คน ซึ่งเป็นผู้ผ่านการสอบคัดเลือกเข้าศึกษาในคณะศึกษาศาสตร์ และคณะมนุษยศาสตร์ และสังคมศาสตร์ โดยใช้เกรดเฉลี่ยสะสมภาคเรียนที่ 1 เป็นตัวเกณฑ์ ผลจากการศึกษามีดังนี้

1. ความยากเฉลี่ยของแบบทดสอบ วิชาสามัญ 2 ตอนที่ 1 ภาษาไทย คณิตศาสตร์ ก ภาษาอังกฤษ กขค ภาษาไทย กข สังคมศึกษา กข ฝรั่งเศส พื้นฐานทางคณิตศาสตร์ พื้นฐานทางภาษาไทย พื้นฐานทางวิทยาศาสตร์ และพื้นฐานทางสังคมศึกษามีค่าเท่ากับ .43, .27, .34, .46, .36, และ .51 ส่วนค่าอำนาจจำแนกเฉลี่ยของแบบทดสอบมีค่าเท่ากับ .23, .20, .22, .21, .18, .28, .22, .25, .20 และ .24 ตามลำดับ สำหรับแบบทดสอบวิชาสามัญ 2 ตอนที่ 2 ซึ่งมีทั้งหมด 14 เรื่อง มีค่าความยากง่ายเฉลี่ยอยู่ในช่วง .26 ถึง .64 อำนาจจำแนกเฉลี่ยมีค่าอยู่ในช่วง .20 ถึง .33

2. ค่าสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบมีค่าอยู่ระหว่าง .07 ถึง .82 แบบทดสอบวิชาสามัญ 2 ตอนที่ 1 ภาษาไทย คณิตศาสตร์ ก ภาษาอังกฤษ กขค ภาษาไทย กข สังคมศึกษา กข ฝรั่งเศส พื้นฐานทางคณิตศาสตร์ พื้นฐานภาษาไทย พื้นฐานทางวิทยาศาสตร์ และพื้นฐานทางสังคมศึกษา มีค่าสัมประสิทธิ์ความเชื่อมั่นเท่ากับ .35, .09, .74, .62, .49, .82, .07, .53, .39 และ .54 ตามลำดับ สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบวิชาสามัญ 2 ตอนที่ 2 วิทยาศาสตร์กายภาพชีวภาพมีค่าอยู่ระหว่าง .00 ถึง .24

3. แบบทดสอบมีค่าสัมประสิทธิ์ความเที่ยงตรงเชิงพยากรณ์อยู่ระหว่าง .05 ถึง .36 และแบบทดสอบที่มีความเที่ยงตรงเชิงพยากรณ์แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ได้แก่ แบบทดสอบวิชาฝรั่งเศส และแบบทดสอบวิชาพื้นฐานทางคณิตศาสตร์ ซึ่งมีค่าความเที่ยงตรงเชิงพยากรณ์เท่ากับ .32 และ .36 ตามลำดับ

วีระพงษ์ ศรีชาติ (2535) ศึกษาการสร้างแบบทดสอบอิงเกณฑ์วิชาภาษาอังกฤษด้านทักษะการอ่าน เพื่อความเข้าใจ ชั้นมัธยมศึกษาปีที่ 3 หาคะแนนเกณฑ์ที่เหมาะสมของแบบทดสอบ

อิงเกณฑ์ที่สร้างขึ้น และหาคุณภาพของแบบทดสอบอิงเกณฑ์ที่สร้างขึ้น กลุ่มตัวอย่างได้แก่นักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2534 ของโรงเรียนมัธยมศึกษาในอำเภอสว่างแดนดิน จังหวัดสกลนคร ซึ่งสุ่มมาโดยวิธีการสุ่มแบบหลายขั้นตอน ได้นักเรียน 131 คน จาก 4 โรงเรียน แบ่งเป็น 2 กลุ่ม คือ กลุ่มแรก 68 คน เพื่อคัดเลือกและปรับปรุงข้อทดสอบ และกลุ่มหลังจำนวน 63 คน เพื่อหาคุณภาพของแบบทดสอบ แบบทดสอบที่สร้างขึ้นเป็นแบบทดสอบอิงเกณฑ์วิชาภาษาอังกฤษด้านทักษะการอ่านเพื่อความเข้าใจ ชั้นมัธยมศึกษาปีที่ 3 จำนวน 4 ฉบับ เป็นแบบทดสอบเลือกตอบชนิด 5 ตัวเลือก ซึ่งได้ดำเนินการทดสอบ 2 ครั้ง เพื่อคัดเลือกและปรับปรุงข้อทดสอบ วิเคราะห์คุณภาพ ผลการวิจัยฉบับที่ 1 แบบทดสอบเรื่องการจับใจความ จำนวน 21 ข้อ มีค่าอำนาจจำแนก (ค่าดัชนี S) อยู่ในช่วง .206-.698 ค่าความเที่ยงตรงตามโครงสร้างเท่ากับ .968 ค่าความเชื่อมั่นเท่ากับ .864 และคะแนนเกณฑ์เป็น 38 เปอร์เซนต์ของคะแนนเต็ม ฉบับที่ 2 แบบทดสอบเรื่อง การสรุปความ จำนวน 12 ข้อ มีค่าอำนาจจำแนก (ค่าดัชนี S) อยู่ในช่วง .270-.825 ค่าความเที่ยงตรงตามโครงสร้างเท่ากับ .968 ค่าความเชื่อมั่นเท่ากับ .721 และคะแนนเกณฑ์เป็น 58 เปอร์เซนต์ของคะแนนเต็ม ฉบับที่ 3 แบบทดสอบเรื่อง การตีความ จำนวน 15 ข้อ มีค่าอำนาจจำแนก (ค่าดัชนี S) อยู่ในช่วง .222-.619 ค่าความเที่ยงตรงตามโครงสร้างเท่ากับ .968 ค่าความเชื่อมั่นเท่ากับ .417 และคะแนนเกณฑ์เป็น 53 เปอร์เซนต์ของคะแนนเต็ม ฉบับที่ 4 แบบทดสอบเรื่อง การสรุปความเมื่อกำหนดเวลาในการอ่าน จำนวน 12 ข้อ มีค่าอำนาจจำแนก (ค่าดัชนี S) อยู่ในช่วง .26-.460 ค่าความเที่ยงตรงตามโครงสร้างเท่ากับ .873 ค่าความเชื่อมั่นเท่ากับ .570 และคะแนนเกณฑ์เป็น 58 เปอร์เซนต์ของคะแนนเต็ม

รัตนา เกิดผล (2536) ศึกษาความสัมพันธ์ของคะแนนที่ได้จากแบบทดสอบสองชนิด ได้แก่ แบบทดสอบการเขียนตามคำบอกแบบเดิมคำกับแบบทดสอบมาตรฐานวิชาภาษาอังกฤษ นอกจากนี้ยังได้ศึกษาความสามารถของนักเรียนในการทำแบบทดสอบการเขียนตามคำบอกที่มีลักษณะเนื้อหาต่างประเภทกัน ได้แก่ บทสนทนา จดหมาย และข้อความเกี่ยวกับสถานที่สำคัญในประเทศ พร้อมกันนี้มีการศึกษาเจตคติที่มีต่อแบบทดสอบการเขียนตามคำบอกแบบเดิมคำของนักเรียนชั้นมัธยมศึกษาปีที่ 4 เครื่องมือที่ใช้ในการวิจัยครั้งนี้คือ

1. แบบทดสอบการเขียนตามคำบอกแบบเดิมคำ ซึ่งสร้างจากเนื้อเรื่องสามประเภทโดยการตัดข้อความที่มีลักษณะเป็นกลุ่มคำ ตั้งแต่สองถึงเจ็ดคำ รวมทั้งหมด 45 คำ
2. แบบทดสอบมาตรฐานวิชาภาษาอังกฤษ Form A ของสถาบันภาษา จุฬาลงกรณ์ มหาวิทยาลัย
3. แบบทดสอบถามด้านเจตคติที่มีต่อแบบทดสอบการเขียนตามคำบอกแบบเดิมคำ ชนิดอัตราส่วนประมาณค่า 5 ระดับของ Likert จำนวน 20 ข้อ

นำแบบทดสอบมาตรฐานวัดความสามารถภาษาอังกฤษและแบบทดสอบการเขียนตามคำบอกแบบเดิมคำไปทดสอบกับกลุ่มตัวอย่าง ซึ่งได้นักเรียนชั้นมัธยมศึกษาปีที่ 4 ที่กำลังศึกษาอยู่ใน ภาคเรียนที่ 2 ปีการศึกษา 2535 ของโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดกรุงเทพมหานคร รวม 8 โรงเรียน จำนวน 260 คน ที่ได้จากการสุ่มแบบหลายขั้นตอน โดยทำ

ต่อเนื่องกันตามลำดับ จากนั้นแจกแบบทดสอบถามด้านเจตคติ รวบรวมข้อมูล และวิเคราะห์ความสัมพันธ์ของคะแนนที่ได้จากการวัดโดยแบบทดสอบทั้ง 2 ชนิด โดยใช้สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (The Pearson Product Moment Correlation Coefficient) รวมทั้งเปรียบเทียบความแตกต่างระหว่างประเภทของเนื้อเรื่องโดยใช้วิธีวิเคราะห์ความแปรปรวนทางเดียว (One Way Analysis of Variance) ในด้านที่เกี่ยวกับเจตคตินั้น หาค่าเฉลี่ยส่วนเบี่ยงเบนมาตรฐาน และร้อยละของน้ำหนักคะแนนทั้งฉบับและคิดเป็นรายข้อ ผลการวิจัยสรุปได้ดังต่อไปนี้คือ ในการศึกษาความสัมพันธ์ระหว่างคะแนนที่ได้จากแบบทดสอบการเขียนตามคำบอกแบบเดิมคำ และคะแนนที่ได้จากแบบทดสอบมาตรฐานนั้นพบว่าแบบทดสอบทั้งสองชนิดมีสหสัมพันธ์เท่ากับ .35 คะแนนที่ได้จากการวัดโดยแบบทดสอบการเขียนตามคำบอกที่มีเนื้อหาต่างประเภทกันมีความแตกต่างกันทั้ง 3 ประเภทคือ นักเรียนสามารถทำคะแนนประเภทบทสนทนาได้สูงที่สุด รองลงมาคือ ประเภทจดหมาย และต่ำที่สุดคือ ประเภทข้อความที่เกี่ยวกับสถานที่สำคัญในประเทศ ค่าความคิดเห็นที่ได้จากแบบทดสอบถามด้านเจตคติที่มีต่อแบบทดสอบการเขียนตามคำบอก แบบเดิมคำไม่ว่าจะเป็นค่าเฉลี่ยรายข้อหรือค่าเฉลี่ยโดยรวมมีค่าปานกลาง คือ 3.39 หรือ 67.86

เฉลิมเกียรติ เฟื่องแก้ว (2539) ศึกษาการสร้างแบบทดสอบวัดระดับความสามารถด้านการอ่านภาษาฝรั่งเศสของนักเรียนชั้นมัธยมศึกษาปีที่ 5 เขตกรุงเทพมหานคร กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นนักเรียนชั้น มัธยมศึกษาปีที่ 5 เขตกรุงเทพมหานครที่เรียนภาษาฝรั่งเศส ภาคการศึกษาที่ 2 ปีการศึกษา 2538-2539 จำนวน 302 คน เครื่องมือที่ใช้ในการวิจัยคือ แบบทดสอบวัดระดับความสามารถ ด้านการอ่านภาษาฝรั่งเศสแบบปรนัย 4 ตัวเลือก จำนวน 50 ข้อ ผลการวิจัยสรุปได้ว่า แบบทดสอบวัดระดับความสามารถด้านการอ่านภาษาฝรั่งเศสจำนวน 50 ข้อ ที่สร้างขึ้นเป็นแบบทดสอบที่ดีคือ มีค่าดัชนีความง่ายที่พอเหมาะระหว่าง .25-.79 ค่าดัชนีอำนาจจำแนกที่แยกความสามารถของกลุ่มนักเรียนได้ระหว่าง .20-.56 ค่าเบี่ยงเบนมาตรฐานที่กระจายไม่มากกว่า 7.01 ค่าความเที่ยงตรงที่สูงที่ค่า 0.80 และค่าความคลาดเคลื่อนมาตรฐานของการวัดมีค่าต่ำกว่า 3.13

กัลยาณี นันทะศรี (2542) เปรียบเทียบคุณภาพของแบบทดสอบโมดิฟายด์โคลซ 5 รูปแบบในด้าน ค่าความยากง่าย ค่าอำนาจจำแนก ค่าความเชื่อมั่น และค่าความเที่ยงตรงของแบบทดสอบ เปรียบเทียบคะแนนเฉลี่ยที่ได้จากแบบทดสอบโมดิฟายด์โคลซ 5 รูปแบบในกลุ่มนักเรียน ที่มีผลสัมฤทธิ์ทางการเรียนสูง กลุ่มนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนปานกลาง กลุ่มนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนต่ำ และศึกษาปฏิสัมพันธ์ระหว่างรูปแบบของแบบทดสอบกับระดับผลสัมฤทธิ์ทางการเรียนที่มีต่อคะแนนความเข้าใจในการอ่านภาษาอังกฤษ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 2 ปีการศึกษา 2541 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษาอำเภอเมือง จังหวัดอุดรธานี จำนวน 558 คน ได้มาโดยวิธีการสุ่มแบบหลายขั้นตอน (Multi-Stage Random Sampling) เครื่องมือที่ใช้ในการศึกษาค้นคว้าเป็นแบบทดสอบวัดความรู้พื้นฐานทางภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 และแบบทดสอบโมดิฟายด์โคลซ 5 รูปแบบ คือ แบบเลือกตอบ แบบจับคู่ แบบเอ็มซี แบบจี และแบบมีรูปภาพชี้แนะ สถิติที่ใช้

ในการวิเคราะห์ข้อมูลคือ การวิเคราะห์ความแปรปรวน (ANOVA) และไค-สแควร์ (Chi-Square) ผลการวิจัย พบว่า

1. แบบทดสอบโมติฟายด์โคลซ 5 รูปแบบ มีค่าความยากแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยแบบทดสอบโมติฟายด์โคลซแบบเลือกตอบมีค่าความยากมากกว่าแบบอื่นๆทุกรูปแบบ แบบมีรูปภาพชี้แนะมีค่าความยากง่ายมากกว่าแบบจับคู่ แบบเอ็มซี และแบบจี

2. แบบทดสอบโมติฟายด์โคลซ 5 รูปแบบ มีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยแบบทดสอบโมติฟายด์โคลซแบบเลือกตอบมีค่าความเชื่อมั่นสูงกว่าแบบอื่นๆทุกรูปแบบ แบบจีมีค่าความเชื่อมั่นสูงกว่าแบบมีรูปภาพชี้แนะ

3. แบบทดสอบโมติฟายด์โคลซ 5 รูปแบบ มีค่าอำนาจจำแนกและค่าความเที่ยงตรงไม่แตกต่างกัน

4. นักเรียนกลุ่มที่มีผลสัมฤทธิ์ทางการเรียนสูงมีคะแนนเฉลี่ยจากแบบทดสอบโมติฟายด์โคลซ 5 รูปแบบ ไม่แตกต่างกัน นักเรียนกลุ่มที่มีผลสัมฤทธิ์ทางการเรียนปานกลางมีคะแนนเฉลี่ยจากแบบทดสอบโมติฟายด์โคลซแบบจับคู่สูงกว่าแบบเลือกตอบและแบบมีรูปภาพชี้แนะอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 นักเรียนกลุ่มที่มีผลสัมฤทธิ์ทางการเรียนต่ำมีคะแนนเฉลี่ยจากแบบทดสอบโมติฟายด์โคลซแบบเอ็มซีสูงกว่าแบบมีรูปภาพชี้แนะอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

5. ไม่มีปฏิสัมพันธ์ระหว่างรูปแบบของแบบทดสอบกับระดับผลสัมฤทธิ์ทางการเรียนที่มีต่อคะแนนความเข้าใจในการอ่านภาษาอังกฤษ

ประโมกษ์ ดุลนีย์ (2542) ศึกษาการสร้างและหาคุณภาพแบบทดสอบวัดภาคปฏิบัติด้านทักษะการพูดวิชาภาษาอังกฤษสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 90 คน ที่กำลังเรียนอยู่ในภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษาในเขตอำเภอคำม่วง และกิ่งอำเภอสามชัย จังหวัดกาฬสินธุ์ ซึ่งได้มาโดยวิธีการสุ่มแบบหลายขั้นตอน เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลประกอบด้วย แบบสังเกตวัดภาคปฏิบัติด้านทักษะการพูดวิชาภาษาอังกฤษ ประกอบด้วย 2 ส่วน คือ ข้อคำถาม และตารางการให้คะแนนการวัดภาคปฏิบัติด้านทักษะการพูดวิชาภาษาอังกฤษ ซึ่งแบ่งเนื้อหาเป็นส่วนย่อย 5 พฤติกรรม ได้แก่ การออกเสียงคำศัพท์ ไวยากรณ์ ความคล่องแคล่วในการใช้คำ และความสามารถพูดให้ผู้อื่นเข้าใจ ผู้ดำเนินการจะถามข้อคำถามนักเรียนแล้วให้นักเรียนตอบ ตามสถานการณ์หรือตามคำถามนั้น แล้วผู้ดำเนินการทดสอบสังเกตความถูกต้องในด้านการพูดและบันทึกเสียงไว้ในเทปบันทึกเสียง เพื่อให้อาสาสมัครหรืออาจารย์ภาษาอังกฤษผู้ที่เป็นเจ้าของภาษาตรวจให้คะแนนอีกครั้ง การวิเคราะห์หาคุณภาพของแบบทดสอบวัดภาคปฏิบัติด้านทักษะการพูดวิชาภาษาอังกฤษพบว่า แบบทดสอบวัดภาคปฏิบัติด้านทักษะการพูด วิชาภาษาอังกฤษมีค่าอำนาจจำแนกระหว่าง 0.09 ถึง 0.81 มีค่าความเชื่อมั่นอยู่ระหว่าง 0.54 ถึง 0.79 และค่าความเชื่อมั่นของผู้สังเกตโดยหาค่าตรรกษณ์ความสอดคล้องของผู้สังเกต 2 คน มีค่าระหว่าง 0.88 ถึง 0.97

อมรพรรณ จำปา (2546) ศึกษาความสามารถในการใช้ภาษาอังกฤษธุรกิจของนักศึกษา คณะบริหารธุรกิจ สถาบันเทคโนโลยีราชมงคล และเพื่อพัฒนาข้อทดสอบมาตรฐาน ซึ่งใช้วัดความสามารถในการใช้ภาษาอังกฤษของนักศึกษา คณะบริหารธุรกิจ ตามเนื้อหาและทักษะที่กำหนดไว้ในหลักสูตร กลุ่มประชากรเป็นนักศึกษาชั้นปีที่ 4 และนักศึกษาชั้นปีที่ 2 วิชาเอกของภาษาอังกฤษธุรกิจ จำนวนทั้งหมด 150 คน แบบทดสอบที่สร้างขึ้นประกอบด้วย ข้อทดสอบแบบเลือกตอบ จำนวน 75 ข้อ และข้อทดสอบอัตนัย 1 ข้อ โดยแบ่งออกเป็น 4 ส่วนคือ การฟัง การพูด การอ่าน และ การเขียน แบบทดสอบที่สร้างขึ้นมีค่าความยากง่ายอยู่ระหว่าง 0.20-0.75 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.20-0.77 และมีค่าสัมประสิทธิ์ความเที่ยงตรงเท่ากับ 0.86

อาภา มณีสถิตย์ (2545) ศึกษาการสร้างแบบทดสอบบุคลิกภาพผู้จัดการสาขาของธนาคารตามแนวบุคลิกภาพของแคทเทล และหาคุณภาพของแบบทดสอบด้านค่าอำนาจจำแนก ความเที่ยงตรงเชิงโครงสร้าง ความเชื่อมั่น และสร้างเกณฑ์ปกติในแต่ละองค์ประกอบในรูปคะแนนที่ปกติ มีผลการวิจัยสรุปได้ว่า ค่าอำนาจจำแนกของแบบทดสอบวัดบุคลิกภาพผู้จัดการสาขาของธนาคารในองค์ประกอบด้านความวิตกกังวล ด้านความกล้าเผชิญ ด้านการควบคุมอารมณ์ ด้านการปรับตัว ด้านการเข้าสังคม ด้านสติปัญญา ด้านมโนธรรม และด้านจิตใจอ่อนโยนมีค่าอำนาจจำแนกอยู่ระหว่าง 2.053-7.087, 1.957-7.419, 2.484-7.783, 2.461-5.958, 2.137-6.012, 2.239-6.005, 2.250-5.283 และ 2.430-5.851 ตามลำดับ และเมื่อรวมทั้งฉบับพบว่า มีค่าอำนาจจำแนกรายข้ออยู่ระหว่าง 1.957-7.783 อย่างมีนัยสำคัญทางสถิติอย่างน้อยที่ระดับ .05 ค่าความเที่ยงตรงเชิงโครงสร้างในแต่ละองค์ประกอบด้านความวิตกกังวล ด้านความกล้าเผชิญ ด้านการควบคุมอารมณ์ ด้านการปรับตัว ด้านการเข้าสังคม ด้านสติปัญญา ด้านมโนธรรม และด้านจิตใจอ่อนโยน มีค่าความเที่ยงตรงที่เป็นน้ำหนักองค์ประกอบอยู่ระหว่าง 0.304-0.612, 0.306-0.568, 0.418-0.736, 0.347-0.641, 0.309-0.562, 0.301-0.736, 0.339-0.631 และ 0.353-0.421 ตามลำดับ และเมื่อรวมทั้งฉบับแล้วพบว่า มีค่าน้ำหนักองค์ประกอบอยู่ระหว่าง 0.301-0.736 ค่าความเชื่อมั่นของแบบทดสอบวัดบุคลิกภาพผู้จัดการสาขาของธนาคารในองค์ประกอบด้านความวิตกกังวล ด้านความกล้าเผชิญ ด้านการควบคุมอารมณ์ ด้านการปรับตัว ด้านการเข้าสังคม ด้านสติปัญญา ด้านมโนธรรม และด้านจิตใจอ่อนโยน มีค่าความเชื่อมั่นเท่ากับ 0.8741, 0.6950, 0.8676, 0.6896, 0.5913, 0.5890, 0.3711 และ 0.8226 ตามลำดับ และเมื่อรวมทั้งฉบับมีค่าความเชื่อมั่นเท่ากับ 0.8847 เกณฑ์ปกติของแบบทดสอบวัดบุคลิกภาพผู้จัดการสาขาของธนาคารในองค์ประกอบด้านความวิตกกังวล ด้านความกล้าเผชิญ ด้านการควบคุมอารมณ์ ด้านการปรับตัว ด้านการเข้าสังคม ด้านสติปัญญา ด้านมโนธรรม และด้านจิตใจอ่อนโยน มีค่าความเชื่อมั่น มีเกณฑ์คะแนนที่ปกติอยู่ระหว่าง T20 ถึง T81, T21 ถึง T78, T20 ถึง T74, T22 ถึง T76, T21 ถึง T77, T13 ถึง T76, T20 ถึง T76 และ T20 ถึง T70 ตามลำดับ และเมื่อรวมทั้งองค์ประกอบ มีเกณฑ์คะแนนที่ปกติอยู่ระหว่าง T13 ถึง T81

จากงานวิจัยดังกล่าวสรุปได้ว่า การสร้างแบบทดสอบให้มีคุณภาพที่ดีจะต้องพิจารณาถึงองค์ประกอบต่าง ๆ เช่น ค่าความเที่ยงตรง ความเชื่อมั่น ความยากง่าย และค่าอำนาจจำแนก ตลอดจนมีความสอดคล้องกับแบบทดสอบอื่นที่มีความเป็นมาตรฐานด้วย

ด้วยเหตุนี้ผู้วิจัยจึงได้ศึกษาแบบทดสอบจัดระดับความสามารถทางภาษาของพนักงานต้อนรับฯ บริษัทการบินไทยจำกัด (มหาชน) ที่ได้จัดทำขึ้นว่ามีคุณภาพตามหลักทฤษฎีการสร้างแบบทดสอบและคุณภาพของแบบทดสอบหรือไม่ ทั้งนี้ผลที่ได้จะได้นำไปสร้างและปรับปรุงแบบทดสอบในส่วนอื่นๆให้มีคุณภาพได้มาตรฐานต่อไป

บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง
2. เครื่องมือที่ใช้ในการวิจัย
3. การเก็บรวบรวมข้อมูล
4. การจัดกระทำและการวิเคราะห์ข้อมูล

การกำหนดประชากรและการเลือกกลุ่มตัวอย่าง

ประชากร

ประชากรและกลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้เป็นพนักงานต้อนรับบนเครื่องบินรุ่นใหม่ของบริษัท การบินไทย จำกัด (มหาชน)

การเลือกกลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นกลุ่มตัวอย่างที่ถูกเลือกแบบเจาะจงจากพนักงานต้อนรับบนเครื่องบินรุ่นใหม่ของบริษัท การบินไทย จำกัด (มหาชน) ที่เริ่มปฏิบัติงานในปี พ.ศ. 2546 ทั้งหญิงและชาย ที่เข้ารับการทดสอบความสามารถทางภาษาอังกฤษ (Placement test) จากกองฝึกอบรมภาษา บริษัทการบินไทย จำกัด (มหาชน) จำนวน 489 คน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ (Placement test) ของพนักงานต้อนรับบนเครื่องบินที่เริ่มปฏิบัติงานกับบริษัทการบินไทย จำกัด (มหาชน) ที่เริ่มปฏิบัติงานในปี พ.ศ. 2546 แบบทดสอบที่นำมาศึกษาเป็นแบบทดสอบด้านการอ่านและเขียนจำนวน 90 ข้อ เป็นคำถามแบบเลือกตอบ 4 ตัวเลือก (Multiple choices) ตามรายละเอียดดังนี้

1. แบบทดสอบการใช้ภาษาเพื่อการสื่อสาร (Service language) ประเภทเลือกตอบ จำนวน 20 ข้อ
2. แบบทดสอบความเข้าใจในการอ่าน และคำศัพท์ (Reading comprehension and vocabulary) จำนวน 40 ข้อ
3. แบบทดสอบโครงสร้างทางไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์ (Error recognition) และประเภทเลือกเติมคำ จำนวน 30 ข้อ

ในการสร้างข้อทดสอบเพื่อใช้เป็นแบบทดสอบดังกล่าวมีการดำเนินการดังต่อไปนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบคือเพื่อสร้างแบบทดสอบจัดระดับความสามารถทางภาษาอังกฤษ (Placement test) สำหรับพนักงานต้อนรับบนเครื่องบิน บริษัทการบินไทย จำกัด (มหาชน)
 2. ศึกษาเนื้อหาที่ใช้ในการสร้างแบบทดสอบเพื่อให้ครอบคลุมการใช้ภาษาเพื่อการสื่อสารในระหว่างปฏิบัติการบิน รวมทั้งภาษาที่ใช้ในการติดต่อสื่อสารกับบุคคลต่างๆ ที่เกี่ยวข้องในชีวิตประจำวัน
 3. สร้างแบบทดสอบจัดระดับความสามารถทางการใช้ภาษาอังกฤษ โดยมีเนื้อหาครอบคลุมการใช้ภาษาเพื่อการสื่อสาร ไวยากรณ์ ความเข้าใจในการอ่าน ในการปฏิบัติงานของพนักงานต้อนรับบนเครื่องบิน
 4. พิจารณาความเที่ยงตรงของแบบทดสอบรายข้อ และปรับปรุงแก้ไขโดยกลุ่มผู้เชี่ยวชาญ จำนวน 5 ท่านซึ่งมีคุณสมบัติดังนี้
 - 4.1 ผู้เชี่ยวชาญที่สอนภาษาอังกฤษให้กับพนักงานต้อนรับบนเครื่องบินที่ปฏิบัติงานในกองฝึกอบรมภาษา บริษัทการบินไทย จำกัด (มหาชน) ไม่น้อยกว่า 5 ปี และเคยเป็นผู้ที่มีประสบการณ์ในการปฏิบัติงานเป็นพนักงานต้อนรับบนเครื่องบินไม่น้อยกว่า 15 ปี จำนวน 2 ท่าน
 - 4.2 ผู้เชี่ยวชาญที่สอนภาษาอังกฤษให้กับพนักงานต้อนรับบนเครื่องบินที่ปฏิบัติงานในกองฝึกอบรมภาษา บริษัทการบินไทย จำกัด (มหาชน) ที่สำเร็จการศึกษาปริญญาโท ด้านการสอนภาษาอังกฤษ และเป็นพนักงานต้อนรับบนเครื่องบิน จำนวน 1 ท่าน
 - 4.3 ผู้เชี่ยวชาญที่สอนภาษาอังกฤษให้กับพนักงานต้อนรับบนเครื่องบินที่ดำรงตำแหน่งผู้จัดการแผนกฝึกอบรมภาษาอังกฤษขั้นพื้นฐาน กองฝึกอบรมภาษา บริษัทการบินไทย จำกัด (มหาชน) จำนวน 1 ท่าน
 - 4.4 ผู้จัดการกองฝึกอบรมภาษา บริษัทการบินไทย จำกัด (มหาชน) จำนวน 1 ท่าน
- ผู้เชี่ยวชาญร่วมพิจารณาคัดเลือกข้อทดสอบที่เหมาะสม จำนวน 90 ข้อ จากจำนวนข้อทดสอบ 120 ข้อ จัดทำเป็นแบบทดสอบ

การเก็บรวบรวมข้อมูล

ผู้วิจัยทำการเก็บรวบรวมข้อมูลตามขั้นตอนต่อไปนี้

1. ติดต่อขออนุญาตผู้จัดการกองฝึกอบรมภาษา บริษัท การบินไทย จำกัด (มหาชน) นำข้อสอบที่ใช้สอบกับพนักงานต้อนรับบนเครื่องบินมาทำการวิจัยหาคุณภาพของแบบทดสอบ
2. นำผลการทดสอบมาวิเคราะห์หาค่าสถิติ และหาคุณภาพของแบบทดสอบรายข้อ ด้านความเที่ยงตรง ความยากง่าย และค่าอำนาจจำแนก
3. หาคุณภาพของแบบทดสอบทั้งฉบับโดยใช้ค่าความเชื่อมั่น

4. ศึกษาความสัมพันธ์ของผลการทดสอบ (Correlation) กับแบบทดสอบมาตรฐาน TOEIC

5. สร้างเกณฑ์ปกติคะแนนที่ (Normalized T-score)

การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล

1. การตรวจสอบคุณภาพด้านความเที่ยงตรง (Validity)

1.1 หาความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบ โดยหาค่าสัมประสิทธิ์ความสอดคล้องด้วยการหาความสัมพันธ์ระหว่างคะแนนรวมทั้งฉบับกับคะแนนรายข้อ โดยใช้สูตรการหาสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple correlation) ของเพียร์สัน (Pearson product moment correlation)

1.2 นำผลการวัดที่ได้ไปวิเคราะห์ข้อคำถามเป็นรายข้อ (Item analysis) เพื่อหาค่าอำนาจจำแนก และความยากง่าย โดยใช้เทคนิค 27% ของ จุง เตห์ ฟาน (Chung The Fan)

2. การตรวจสอบคุณภาพของแบบทดสอบทั้งฉบับด้านความเชื่อมั่น พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ของความเชื่อมั่น (Reliability coefficient) โดยวิเคราะห์หาค่าสัมประสิทธิ์ของความเชื่อมั่นโดยใช้สูตร K-R 20 ของคูเดอร์และริชาร์ดสัน (Kuder Richardson)

3. นำข้อทดสอบจัดระดับความสามารถทางการใช้ภาษาอังกฤษที่ผ่านการวิเคราะห์คุณภาพมาพิจารณาความสัมพันธ์ (Correlation) กับแบบทดสอบ TOEIC โดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson product moment correlation)

4. คัดเลือกข้อทดสอบเพื่อเก็บไว้ในคลังข้อทดสอบ

5. สร้างเกณฑ์ปกติคะแนนที่ (Normalized T-score) ของแบบทดสอบที่ผ่านการวิเคราะห์คุณภาพ โดยการหาตำแหน่งเปอร์เซ็นต์ไทล์แล้วนำไปเทียบเป็นค่าคะแนนที่ปกติจากตารางสำเร็จรูป จากนั้นนำคู่อันดับระหว่างคะแนนผลการทดสอบ (คะแนนดิบ) กับคะแนนที่ปกติ มาลงจุดกราฟ ขยายคะแนนที่ปกติและคะแนนดิบให้ครบทุกคะแนน (0-90 คะแนน) โดยการใช้เส้นตรงที่เป็นตัวแทนของจุดกราฟ แล้วอ่านค่าความสัมพันธ์ระหว่างคะแนนดิบกับคะแนนที่ นำมาสร้างเป็นตาราง เกณฑ์ปกติของแบบทดสอบ เพื่อให้สามารถนำคะแนนที่ได้จากการทดสอบครั้งต่อไปมาเทียบกับ เกณฑ์ปกติคะแนนที่สร้างขึ้น นำมาใช้ในการจัดระดับความสามารถของผู้รับการทดสอบ

6. สถิติที่ใช้ในการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลครั้งนี้ผู้วิจัยได้ใช้สถิติต่าง ๆ ดังนี้

6.1 สถิติพื้นฐานคือค่าเฉลี่ย ($\bar{X}$) และค่าส่วนเบี่ยงเบนมาตรฐาน (S)

6.2 ค่าความยากง่าย และอำนาจจำแนกโดยใช้เทคนิค 27% ของ จุง เตห์ ฟาน

6.3 ค่าสัมประสิทธิ์ความสอดคล้องของคะแนนรวมทั้งฉบับ และคะแนนรายข้อ โดยใช้สูตรสหสัมพันธ์ของเพียร์สัน (Pearson Product Moment Correlation Coefficient)

6.4 ค่าความเชื่อมั่นของแบบทดสอบ โดยใช้สูตรของคูเดอร์และริชาร์ดสัน (Kuder Richardson) ที่ 20

6.5 เกณฑ์ปกติคะแนนที่ (Normalized T-score)

บทที่ 4

ผลการวิเคราะห์ข้อมูล

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

ในการวิเคราะห์ข้อมูลและแปลความหมายของการวิเคราะห์ข้อมูล เพื่อให้เกิดความเข้าใจตรงกัน ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล ดังต่อไปนี้

N	แทน	จำนวนกลุ่มตัวอย่าง
n	แทน	จำนวนข้อของแบบทดสอบแต่ละองค์ประกอบ
p	แทน	ค่าความยากง่าย
r	แทน	ค่าอำนาจจำแนก
r_{tt}	แทน	สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบ
r_{xy}	แทน	สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับ
S	แทน	ความเบี่ยงเบนมาตรฐาน
T	แทน	คะแนนที่-ปกติ (Normalized T-score) ในเกณฑ์ปกติ (Norm)
X	แทน	คะแนนดิบ
$\bar{X}$	แทน	ค่าเฉลี่ยรายข้อของแบบทดสอบ
*	แทน	มีนัยสำคัญทางสถิติที่ระดับ .05
**	แทน	มีนัยสำคัญทางสถิติที่ระดับ .01

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามจุดมุ่งหมายของการศึกษาค้นคว้าดังนี้

1. การหาความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบ โดยหาค่าสัมประสิทธิ์ความสอดคล้องของความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับ โดยใช้สูตรการหาสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple correlation) ของเพียร์สัน (Pearson product moment correlation)
2. การหาคุณภาพรายข้อด้านค่าอำนาจจำแนกและความยากง่าย โดยใช้เทคนิค 27% ของ จุง เตห์ ฟาน (Chung The Fan)
3. การหาค่าความเชื่อมั่นของแบบทดสอบ จากค่าสัมประสิทธิ์สหสัมพันธ์ของความเชื่อมั่น (Reliability coefficient) โดยวิเคราะห์หาค่าสัมประสิทธิ์ของความเชื่อมั่นโดยใช้สูตร K-R 20 ของคูเดอร์และริชาร์ดสัน (Kuder-Richardson)

4. การพิจารณาความสอดคล้อง (Correlation) ของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษที่สร้างขึ้นกับแบบทดสอบ TOEIC โดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson product moment correlation)
5. การสร้างเกณฑ์ปกติคะแนนที่ (Normalized T-score) ของแบบทดสอบ

ผลการวิเคราะห์ข้อมูล

ข้อมูลที่ใช้ในการวิเคราะห์เป็นแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ (Placement test) จำนวน 90 ข้อ จากกองฝึกอบรมภาษา บริษัท การบินไทย จำกัด (มหาชน) ข้อมูลที่ใช้นำมาจากการทดสอบกับกลุ่มตัวอย่างซึ่งเป็นพนักงานต้อนรับบนเครื่องบินรุ่นใหม่ของบริษัท การบินไทย จำกัด (มหาชน) ที่เริ่มปฏิบัติงานในปี พ.ศ. 2546 ทั้งหญิงและชาย จำนวน 489 คน ผลการวิเคราะห์ข้อมูลปรากฏตามรายละเอียดดังต่อไปนี้

1. คุณภาพด้านความเที่ยงตรง (Validity)

1.1 ความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบ

ผู้วิจัยได้ทำการหาความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบโดยการหาค่าสัมประสิทธิ์ความสอดคล้องของความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับ โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple correlation) ของเพียร์สัน (Pearson product moment correlation) ได้ผลดังปรากฏในตาราง 6

ตาราง 6 ผลการวิเคราะห์จำนวนข้อของแบบทดสอบที่มีความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับ

แบบทดสอบ	n	ความสัมพันธ์ที่ไม่มีนัยสำคัญ			ความสัมพันธ์ที่มีนัยสำคัญ*		
	90	ทางลบ	ทางบวก	รวม	ทางลบ	ทางบวก	รวม
การใช้ภาษาเพื่อการสื่อสาร	20	1	17	18	-	2	2
ความเข้าใจในการอ่าน และคำศัพท์	40	2	13	15	1	24	25

ตาราง 6 (ต่อ)

แบบทดสอบ	n	ความสัมพันธ์ที่ไม่มีนัยสำคัญ			ความสัมพันธ์ที่มีนัยสำคัญ*		
		ทางลบ	ทางบวก	รวม	ทางลบ	ทางบวก	รวม
โครงสร้างทางไวยากรณ์ทางภาษา ประเภทการหาข้อบกพร่องทางไวยากรณ์	30	-	4	4	-	26	26
ความสามารถในการใช้ภาษาทั้งฉบับ	90	3	34	37	1	52	53
ร้อยละของแบบทดสอบทั้งฉบับ		3.3	37.7	41.1	1.1	57.7	58.9

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 6 ผลการวิเคราะห์จำนวนข้อของแบบทดสอบที่มีความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ มีรายละเอียดดังนี้

1. แบบทดสอบรายข้อด้านการใช้ภาษาเพื่อการสื่อสารประเภทเลือกตอบ ที่มีความสัมพันธ์กับคะแนนรวมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวน 2 ข้อ แบบทดสอบข้อที่ไม่มีความสัมพันธ์กับคะแนนรวมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวน 18 ข้อ โดยเป็นความสัมพันธ์ทางบวกจำนวน 17 ข้อ และทางลบจำนวน 1 ข้อ

2. แบบทดสอบรายข้อด้านความเข้าใจในการอ่าน และคำศัพท์ มีความสัมพันธ์กับคะแนนรวมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวน 25 ข้อ โดยเป็นความสัมพันธ์ทางบวกจำนวน 24 ข้อและทางลบจำนวน 1 ข้อ นอกจากนี้ข้อสอบข้อที่ไม่มีความสัมพันธ์กับคะแนนรวมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวน 15 ข้อ โดยเป็นความสัมพันธ์ทางบวกจำนวน 13 ข้อ และทางลบจำนวน 2 ข้อ

3. แบบทดสอบรายข้อด้านโครงสร้างไวยากรณ์ทางภาษา ประเภทการหาข้อบกพร่องทางไวยากรณ์และประเภทเลือกเติมคำที่มีความสัมพันธ์กับคะแนนรวมอย่างมีนัยสำคัญ

ทางสถิติที่ระดับ .05 จำนวน 26 ข้อ แบบทดสอบข้อที่ไม่มีความสัมพันธ์กับคะแนนรวมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวน 4 ข้อ โดยเป็นความสัมพันธ์ในทางบวก

สรุปผลการวิเคราะห์ความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษพบว่า แบบทดสอบรายข้อมีความสัมพันธ์กับคะแนนรวมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวน 53 ข้อ หรือเท่ากับร้อยละ 58.9 โดยเป็นความสัมพันธ์ทางบวกจำนวน 52 ข้อ คิดเป็นร้อยละ 57.7 ทางลบจำนวน 1 ข้อ คิดเป็นร้อยละ 1.1 และแบบทดสอบที่ไม่มีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จำนวนทั้งสิ้น 37 ข้อ หรือเท่ากับร้อยละ 41.1 โดยเป็นความสัมพันธ์ที่ไม่มีนัยสำคัญทางสถิติในทางบวกจำนวน 34 ข้อ หรือร้อยละ 37.7 และเป็นความสัมพันธ์ในทางลบจำนวน 3 ข้อหรือเท่ากับร้อยละ 3.3 ตามลำดับ (ข้อมูลของความสัมพันธ์รายข้อของแต่ละด้านดูจากตาราง 8)

1.2 คุณภาพรายข้อด้านค่าความยากง่าย และค่าอำนาจจำแนก

ผู้วิจัยได้ทำการหาคุณภาพรายข้อด้านค่าความยากง่ายและค่าอำนาจจำแนก โดยใช้เทคนิค 27% ของ จุง เตห์ ฟาน (Chung The Fan) ซึ่งมีกลุ่มสูงจำนวน 135 คน และกลุ่มต่ำจำนวน 137 คน ปรากฏผลในตาราง 7 และ 8 ดังนี้

ตาราง 7 ค่าความยากง่ายและค่าอำนาจจำแนกของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับ บริษัทการบินไทย จำกัด (มหาชน)

แบบทดสอบ	n	p	r
การใช้ภาษาเพื่อการสื่อสาร (ข้อ 1-20)	20	0.23 – 0.99	0.03 – 0.36
ความเข้าใจในการอ่าน และคำศัพท์ (ข้อ 21-60)	40	0.11 – 0.91	-0.26 – 0.66
โครงสร้างไวยากรณ์ทางภาษา ประเภทการหาข้อบกพร่องทาง ไวยากรณ์ (ข้อ 61-90)	30	0.28 – 0.99	0.02 – 0.58
แบบทดสอบจัดระดับความสามารถ ในการใช้ภาษารวมทั้งฉบับ	90	0.11 – 0.99	-0.26 – 0.66

จากตาราง พบว่า แบบทดสอบจัดระดับความสามารถการใช้ภาษาของพนักงานต้อนรับบนเครื่องบินทั้งฉบับมีค่าความยากง่ายอยู่ระหว่าง 0.11 ถึง 0.99 และค่าอำนาจจำแนกอยู่ระหว่าง -0.26-0.66 โดยมีรายละเอียดสรุปได้ดังนี้

1. แบบทดสอบการใช้ภาษาเพื่อการสื่อสาร ประเภทเลือกตอบ มีค่าความยากง่ายอยู่ระหว่าง 0.23 ถึง 0.99 และค่าอำนาจจำแนกอยู่ระหว่าง 0.03 ถึง 0.36

2. แบบทดสอบความเข้าใจในการอ่านและคำศัพท์ มีค่าความยากง่ายอยู่ระหว่าง 0.11 ถึง 0.91 และค่าอำนาจจำแนกอยู่ระหว่าง -0.26 ถึง 0.66

3. แบบทดสอบโครงสร้างไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์ และประเภทเลือกเติมคำ มีค่าความยากง่ายอยู่ระหว่าง 0.28 ถึง 0.99 และค่าอำนาจจำแนกอยู่ระหว่าง 0.02 ถึง 0.58

ทั้งนี้รายละเอียดความยากง่าย และค่าอำนาจจำแนกรายข้อปรากฏในตาราง 8

ผู้วิจัยได้ทำการหาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ จากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรวมกับคะแนนรายข้อ ด้านค่าความยากง่าย และค่าอำนาจจำแนกเป็นรายข้อ ได้ผลดังปรากฏในตาราง 8

ตาราง 8 สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับ ค่าความยากง่ายและ
ค่าอำนาจจำแนกเป็นรายข้อ

ข้อ	r_{xy}	p	r	ข้อ	r_{xy}	p	r
1	-.025	.95	.08	46	.147	.34	.23
2	.051	.99	.04	47	.369*	.61	.60
3	.043	.96	.03	48	.130	.53	.21
4	.030	.67	.08	49	.206*	.84	.42
5	.100	.43	.21	50	.311*	.64	.48
6	.099	.40	.16	51	.341*	.74	.47
7	.023	.89	.09	52	.159	.59	.25
8	.139	.96	.08	53	.056	.45	.10
9	.109	.97	.07	54	.244*	.81	.42
10	.191*	.23	.32	55	.184*	.29	.32
11	.152	.86	.19	56	.179*	.83	.29
12	.139	.97	.07	57	-.145	.11	-.24
13	.152	.89	.25	58	.275*	.80	.43
14	.001	.38	.11	59	.187*	.56	.31
15	.137	.58	.21	60	-.230*	.19	-.26
16	.097	.57	.16	61	.382*	.58	.53
17	.093	.79	.24	62	.183*	.80	.35
18	.251*	.85	.36	63	.024	.89	.02
19	.028	.79	.13	64	.231*	.81	.41
20	.130	.96	.08	65	.176*	.28	.33
21	.053	.49	.11	66	.170*	.59	.27
22	.071	.69	.15	67	.309*	.84	.53
23	.354*	.42	.55	68	.130	.93	.30
24	.091	.63	.16	69	.187*	.83	.37
25	.176*	.64	.36	70	.076	.99	.05
26	-.013	.48	.06	71	.395*	.52	.56
27	.298*	.68	.48	72	.376*	.69	.58
28	.097	.13	.24	73	.323*	.82	.52

ตาราง 8 (ต่อ)

ข้อ	r_{xy}	p	r	ข้อ	r_{xy}	p	r
29	.364*	.85	.66	74	.235*	.93	.36
30	.336*	.63	.49	75	.239*	.76	.39
31	.188*	.91	.28	76	.302*	.88	.26
32	.285*	.89	.49	77	.244*	.92	.53
33	.184*	.84	.34	78	.282*	.56	.41
34	.124	.73	.26	79	.223*	.95	.41
35	.206*	.86	.43	80	.145	.78	.27
36	.225*	.84	.34	81	.243*	.63	.42
37	.049	.40	.09	82	.295*	.76	.50
38	.270*	.87	.42	83	.265*	.55	.42
39	.155	.81	.23	84	.351*	.59	.56
40	.166*	.84	.32	85	.374*	.64	.55
41	.206*	.77	.39	86	.261*	.77	.44
42	.348*	.52	.49	87	.240*	.65	.38
43	.108	.57	.13	88	.235*	.55	.40
44	.193*	.86	.33	89	.394*	.35	.52
45	.084	.23	.16	90	.342*	.60	.48

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง พบว่า แบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ ของ พนักงานต้อนรับบนเครื่องบิน จำนวน 90 ข้อ มีค่าสัมประสิทธิ์ความสอดคล้องระหว่างคะแนนรายข้อ กับคะแนนรวมทั้งฉบับ ค่าความยากง่าย และค่าอำนาจจำแนกสรุปได้ดังนี้

1. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับมีค่า ตั้งแต่ -0.230 ถึง 0.395 โดยมีรายละเอียดดังนี้

1.1 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับ ของแบบทดสอบทั้งหมดจะมีค่าเป็นบวก (ยกเว้นข้อ 1, 26, 57, และ 60 ที่มีค่าเป็นลบ) แสดงว่า แบบทดสอบส่วนใหญ่วัดมิติเดียวกัน

1.2 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับที่มีค่าเป็นลบมีจำนวน 4 ข้อ คือ ข้อที่ 1, 26, 57 และ 60 แสดงว่าแบบทดสอบวัดในมิติที่แตกต่างกัน ไม่สอดคล้องกัน

1.3 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับที่มีค่าเกือบเท่ากับ 0 (0.001) มีจำนวน 1 ข้อ คือข้อที่ 14 แสดงถึงความไม่สัมพันธ์กันของข้อมูล ไม่สามารถเป็นคะแนนที่แทนโครงสร้างรวมของแบบทดสอบได้

2. ค่าความยากง่ายของข้อสอบอยู่ระหว่าง 0.11 ถึง 0.99 โดยข้อสอบส่วนใหญ่มีความยากง่ายปานกลางถึงค่อนข้างง่าย มีรายละเอียดดังนี้

2.1 ข้อสอบที่มีค่าความยากง่ายอยู่ระหว่าง 0.2 ถึง 0.8 ซึ่งเป็นข้อสอบที่มีความยากง่ายพอเหมาะมีจำนวน 63 ข้อ ได้แก่ข้อที่ 4, 5, 6, 10, 14-17, 19, 21-27, 30, 33, 34, 36, 37, 39-43, 45-56, 58, 59, 61, 62, 64-67, 69, 71, 72, 73, 78, และ 80-90

2.2 ข้อสอบที่มีค่าความยากง่ายต่ำกว่า 0.2 ซึ่งเป็นข้อสอบที่ยากเกินไปไม่เหมาะสมกับการนำไปใช้ควรปรับปรุงให้ง่ายขึ้นมีจำนวน 3 ข้อ ได้แก่ข้อที่ 28, 57, และ 60

2.3 ข้อสอบที่มีค่าความยากง่ายสูงกว่า 0.8 ซึ่งเป็นข้อสอบที่ง่ายเกินไปไม่เหมาะสมกับการนำไปใช้ควรปรับปรุงให้ยากขึ้นมีจำนวน 24 ข้อ ได้แก่ข้อที่ 1, 2, 3, 7, 8, 9, 11, 12, 13, 18, 20, 29, 31, 32, 35, 38, 44, 63, 68, 70, 74, 76, 77, และ 79

3. ค่าอำนาจจำแนกของแบบทดสอบอยู่ระหว่าง -0.26 ถึง 0.66 โดยส่วนใหญ่มีค่าเป็นบวก ตามรายละเอียดดังนี้

3.1 ค่าอำนาจจำแนกของข้อสอบ ที่มีค่าตั้งแต่ 0.2 ขึ้นไปซึ่งสามารถนำไปใช้ได้มีจำนวนทั้งสิ้น 64 ข้อ คือข้อที่ 5, 10, 13, 15, 17, 23, 25, 27-36, 38-42, 44, 46-52, 54, 55, 56, 58, 59, 61, 62, 64-69, และ 71-90

3.2 ค่าอำนาจจำแนกของข้อสอบที่มีค่าน้อยกว่า 0.2 มีค่าอยู่ระหว่าง 0.02 ถึง 0.19 ซึ่งต้องทำการปรับปรุงแก้ไขก่อนนำไปใช้ มีจำนวน 24 ข้อ ได้แก่ ข้อที่ 1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 16, 19, 20, 21, 22, 24, 26, 37, 43, 45, 53, 63 และ 70

3.3 ค่าอำนาจจำแนกที่มีค่าเป็นลบมีจำนวน 2 ข้อ คือข้อที่ 57 และ 60 แสดงว่าข้อสอบข้อนั้นจำแนกผู้รับการทดสอบตรงข้ามกับความเป็นจริง คือคนเก่งตอบคำถามข้อนั้นผิด แต่คนอ่อนตอบถูก จึงไม่ควรนำข้อสอบดังกล่าวไปใช้ หรือต้องทำการปรับปรุงก่อนนำไปใช้

2. ค่าความเชื่อมั่นของแบบทดสอบ (Reliability)

ผู้วิจัยได้ดำเนินการหาค่าความเชื่อมั่นจากค่าสัมประสิทธิ์สหสัมพันธ์ของความเชื่อมั่น (Reliability coefficient) โดยวิเคราะห์หาค่าสัมประสิทธิ์ของความเชื่อมั่นโดยใช้สูตร K-R 20 ของคูเดอร์และริชาร์ดสัน (Kuder Richardson) เพื่อหาคุณภาพของแบบทดสอบทั้งฉบับด้านความเชื่อมั่นได้ผลดังปรากฏในตาราง 9

ตาราง 9 ค่าสถิติพื้นฐานและค่าความเชื่อมั่นของแบบทดสอบรายด้าน และแบบทดสอบรวมทั้งฉบับ
ของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ (Placement test)

แบบทดสอบ	n	$\bar{x}$	S	r_{tt}
การใช้ภาษาเพื่อการสื่อสาร	20	15.36	1.81	.26
ความเข้าใจในการอ่าน และคำศัพท์	40	24.67	4.54	.65
โครงสร้างไวยากรณ์ทางภาษา ประเภทการหา ข้อบกพร่องทางไวยากรณ์	30	21.25	4.60	.78
แบบทดสอบจัดระดับความสามารถในการใช้ ภาษารวมทั้งฉบับ	90	61.29	8.53	.80

จากตารางพบว่า แบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ มีค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับเท่ากับ .80 และเมื่อพิจารณาค่าความเชื่อมั่นของแบบทดสอบรายด้านพบว่า แบบทดสอบโครงสร้างไวยากรณ์ทางภาษา ประเภทการหาข้อบกพร่องทางไวยากรณ์ และประเภทเลือกเติมคำ มีค่าความเชื่อมั่นมากเป็นอันดับแรกเท่ากับ .78 รองลงมาคือ แบบทดสอบความเข้าใจในการอ่าน และคำศัพท์ มีค่าความเชื่อมั่นเท่ากับ .65 และในแบบทดสอบการใช้ภาษาเพื่อการสื่อสาร ประเภทเลือกตอบ มีค่าความเชื่อมั่นน้อยและเป็นอันดับสุดท้ายเท่ากับ .26

3. ค่าความสอดคล้อง (Correlation) ของคะแนนจากแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษที่สร้างขึ้นกับคะแนนจากแบบทดสอบ TOEIC

ผู้วิจัยได้นำคะแนนจากแบบทดสอบจัดระดับการใช้ภาษาอังกฤษไปหาค่าความสอดคล้องกับคะแนนจากแบบทดสอบ TOEIC ซึ่งพนักงานต้อนรับบนเครื่องบินนำมาใช้ประกอบในการสมัครตามข้อกำหนดของบริษัท โดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson product moment correlation) ปรากฏผลในตาราง 10 ดังนี้

ตาราง 10 ค่าสัมประสิทธิ์สหสัมพันธ์ของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ
กับแบบทดสอบ TOEIC

แบบทดสอบ	TOEIC	การใช้ภาษา เพื่อการ สื่อสาร	ความเข้าใจใน การอ่าน และ คำศัพท์	โครงสร้าง ไวยากรณ์ทาง ภาษา	คะแนนรวม ทั้งฉบับ
TOEIC	1	.261**	.438**	.548**	.585**
การใช้ภาษาเพื่อการ สื่อสาร		1	.311**	.220**	.494**
ความเข้าใจในการ อ่าน และคำศัพท์			1	.452**	.841**
โครงสร้างไวยากรณ์ ทางภาษาประเภท การหาข้อบกพร่อง ทางไวยากรณ์				1	.829**
คะแนนรวมทั้งฉบับ					1

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง พบว่า คะแนนรวมของแบบทดสอบทั้งฉบับมีความสัมพันธ์ทางบวกกับคะแนนแบบทดสอบ TOEIC อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .585

เมื่อพิจารณาแบบทดสอบรายด้านพบว่า แบบทดสอบโครงสร้างไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์และประเภทเลือกเติมคำ มีความสัมพันธ์กับแบบทดสอบ TOEIC มากเป็นอันดับแรก โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .548 รองลงมาคือ แบบทดสอบความเข้าใจในการอ่านและคำศัพท์ มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .438 และแบบทดสอบการใช้ภาษาเพื่อการสื่อสารประเภทเลือกตอบ มีความสัมพันธ์กับแบบทดสอบ TOEIC เป็นอันดับสุดท้าย โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .261

ในส่วน of ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนจากแบบทดสอบรายด้านกับคะแนนรวมทั้งฉบับพบว่า แบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษด้านความเข้าใจในการ

อ่านและคำศัพท์ มีความสัมพันธ์กับคะแนนรวมทั้งฉบับมากที่สุดโดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .841 รองลงมาคือแบบทดสอบความสามารถในการใช้ภาษาด้านโครงสร้างไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์ และประเภทเลือกเติมคำมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .829 และแบบทดสอบความสามารถในการใช้ภาษาเพื่อการสื่อสารมีความสัมพันธ์กับคะแนนทั้งฉบับน้อยที่สุด โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .494 ตามลำดับ

นอกจากนี้ยังพบว่าความสัมพันธ์ของคะแนนของแบบทดสอบด้านความเข้าใจในการอ่านและคำศัพท์ มีความสัมพันธ์กับคะแนนของแบบทดสอบด้านโครงสร้างไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์และประเภทเลือกเติมคำมากที่สุดเท่ากับ .452 รองลงมาคือแบบทดสอบความสามารถในการใช้ภาษาเพื่อการสื่อสารมีความสัมพันธ์กับคะแนนของแบบทดสอบด้านความเข้าใจในการอ่านและคำศัพท์ มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .311 และแบบทดสอบความสามารถในการใช้ภาษาเพื่อการสื่อสาร มีความสัมพันธ์กับแบบทดสอบความสามารถในการใช้ภาษาด้านโครงสร้างไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์และประเภทเลือกเติมคำน้อยที่สุดโดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .220 ตามลำดับ

5. เกณฑ์ปกติคะแนนที่ (Normalized T-score) ของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของพนักงานต้อนรับบนเครื่องบินบริษัท การบินไทย จำกัด (มหาชน)

ผู้วิจัยนำคะแนนที่ได้จากการทดสอบกลุ่มตัวอย่างมาสร้างเกณฑ์ปกติ (Norm) ในรูปของคะแนนมาตรฐานที่ปกติ (Normalized T-score) โดยการนำค่าคะแนนดิบและคะแนนที่-ปกติที่ได้จากการวิเคราะห์ข้อมูลมาสร้างกราฟและปรับเส้นกราฟ จากนั้นผู้วิจัยได้อ่านค่าคะแนนดิบและคะแนนที่ปกติจากกราฟที่ปรับแล้ว ดังแสดงในตาราง 11, 12, 13, และ 14

ตาราง 11 เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของพนักงานต้อนรับบนเครื่องบิน ในรูปของคะแนนมาตรฐานที่-ปกติ

X	T	X	T	X	T	X	T	X	T
90	82	70	61	50	36	30	13	10	10
89	81	69	60	49	35	29	12	9	10
88	80	68	59	48	34	28	11	8	10
87	79	67	57	47	33	27	10	7	10
86	78	66	55	46	32	26	10	6	10
85	77	65	54	45	31	25	10	5	10
84	76	64	53	44	30	24	10	4	10
83	74	63	52	43	29	23	10	3	10
82	73	62	50	42	28	22	10	2	10
81	72	61	49	41	27	21	10	1	10
80	71	60	48	40	25	20	10	0	10
79	70	59	47	39	24	19	10		
78	69	58	46	38	23	18	10		
77	68	57	45	37	22	17	10		
76	67	56	44	36	21	16	10		
75	66	55	42	35	20	15	10		
74	65	54	41	34	19	14	10		
73	64	53	40	33	18	13	10		
72	63	52	39	32	16	12	10		
71	62	51	38	31	15	11	10		

จากตาราง 11 เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของพนักงานต้อนรับบนเครื่องบิน ในรูปของคะแนนมาตรฐานที่-ปกติ มีคะแนนดิบตั้งแต่ 0-90 คะแนน และมีคะแนนที่-ปกติอยู่ระหว่าง 10-82

ตาราง 12 เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของพนักงานต้อนรับบนเครื่องบินด้านการใช้ภาษาเพื่อการสื่อสารประเภทเลือกตอบในรูปของคะแนนมาตรฐานที่-ปกติ

X	T
20	74
19	69
18	64
17	59
16	53
15	48
14	42
13	37
12	31
11	27
10	22

X	T
9	16
8	11
7	10
6	10
5	10
4	10
3	10
2	10
1	10
0	10

จากตาราง 12 เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของพนักงานต้อนรับบนเครื่องบิน ด้านการใช้ภาษาเพื่อการสื่อสารประเภทเลือกตอบ ในรูปของคะแนนมาตรฐานที่-ปกติ มีคะแนนดิบตั้งแต่ 0-20 คะแนน และมีคะแนนที่-ปกติอยู่ระหว่าง 10-74

ตาราง 13 เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของพนักงานต้อนรับบนเครื่องบิน ด้านความเข้าใจในการอ่าน และคำศัพท์ ในรูปของคะแนนมาตรฐานที่-ปกติ

X	T	X	T	X	T	X	T
40	85	29	60	18	35	7	10
39	83	28	58	17	33	6	10
38	81	27	55	16	30	5	10
37	78	26	53	15	28	4	10
36	76	25	51	14	26	3	10
35	74	24	49	13	23	2	10
34	72	23	46	12	22	1	10
33	69	22	44	11	19	0	10
32	67	21	42	10	17		
31	65	20	39	9	14		
30	62	19	37	8	11		

จากตาราง 13 เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของพนักงานต้อนรับบนเครื่องบิน ด้านความเข้าใจในการอ่าน และคำศัพท์ ในรูปของคะแนนมาตรฐานที่-ปกติ มีคะแนนดิบตั้งแต่ 0-40 คะแนน และมีคะแนนที่-ปกติอยู่ระหว่าง 10-85

ตาราง 14 เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของพนักงานต้อนรับบนเครื่องบิน ด้านโครงสร้างไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์และประเภทเลือกเติมคำในรูปของคะแนนมาตรฐานที่-ปกติ

X	T	X	T	X	T
30	67	19	45	8	23
29	65	18	43	7	21
28	63	17	41	6	19
27	61	16	39	5	17
26	59	15	37	4	15
25	57	14	35	3	13
24	55	13	33	2	11
23	53	12	31	1	10
22	51	11	29	0	10
21	49	10	27		
20	47	9	25		

จากตาราง 14 เกณฑ์ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาของพนักงานต้อนรับบนเครื่องบินด้านโครงสร้างไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์และประเภทเลือกเติมคำในรูปของคะแนนมาตรฐานที่-ปกติ มีคะแนนดิบตั้งแต่ 0-30 คะแนน และมีคะแนนที่-ปกติอยู่ระหว่าง 10-67

การนำเกณฑ์ปกติที่ไปใช้ในการประเมินผล

ในการประเมินผล ถ้าต้องการทราบระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบิน ให้นำคะแนนดิบมาเทียบกับคะแนนที่ปกติโดยพิจารณาจากเกณฑ์ต่อไปนี้ (ชวาล แพรัตกุล. 2548: 180)

เกณฑ์ในการแปลความหมายคะแนน

ตั้งแต่ T65 - ขึ้นไป	แปลว่า มีความสามารถในการใช้ภาษาอังกฤษอยู่ในระดับสูง
ตั้งแต่ T55 - T65	แปลว่า มีความสามารถในการใช้ภาษาอังกฤษอยู่ในระดับ ค่อนข้างสูง
ตั้งแต่ T45 - T55	แปลว่า มีความสามารถในการใช้ภาษาอังกฤษอยู่ในระดับ ปานกลาง
ตั้งแต่ T35 - T45	แปลว่า มีความสามารถในการใช้ภาษาอังกฤษอยู่ในระดับที่ต้อง ปรับปรุง
ต่ำกว่า T35	แปลว่า มีความสามารถในการใช้ภาษาอังกฤษอยู่ในระดับที่ต้อง ปรับปรุงอย่างเร่งด่วน

ถ้าผู้เข้ารับการทดสอบได้คะแนนตรงจุดแบ่งพอดี คือ T35 T45 T55 และ T65 ให้เลื่อนไปอยู่ในกลุ่มสูงที่ถัดไปเสมอ

ดังนั้นในการใช้แบบทดสอบในการจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบิน จะต้องนำคะแนนดิบที่ได้จากผู้เข้ารับการอบรมมาเทียบกับเกณฑ์ปกติเพื่อตัดสินว่าผู้นั้นมีความสามารถในการใช้ภาษาอังกฤษอยู่ในระดับใด

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

ความมุ่งหมายของการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยได้ตั้งจุดมุ่งหมายไว้ดังนี้

1. เพื่อคัดเลือกข้อทดสอบจัดระดับความสามารถทางภาษาอังกฤษ (Placement test) สำหรับพนักงานต้อนรับบนเครื่องบิน บริษัท การบินไทย จำกัด (มหาชน) ที่มีคุณภาพจัดเก็บไว้ในคลังข้อทดสอบ (Test bank)
2. เพื่อสร้างเกณฑ์ปกติ (Norm) ของแบบทดสอบสำหรับเป็นเกณฑ์ในการกำหนดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบินว่ามีความสามารถอยู่ในระดับใด อันจะนำไปเป็นแนวทางในการจัดกลุ่มของพนักงานต้อนรับบนเครื่องบินเข้ารับการอบรมในหลักสูตรภาษาอังกฤษได้อย่างเหมาะสมต่อไป

การกำหนดประชากรและการเลือกกลุ่มตัวอย่าง

ประชากร

ประชากรและกลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้เป็นพนักงานต้อนรับบนเครื่องบินที่เริ่มปฏิบัติงานในปี พ.ศ. 2546 ทั้งหญิงและชาย ของบริษัท การบินไทย จำกัด (มหาชน)

การเลือกกลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นกลุ่มตัวอย่างที่ถูกเลือกแบบเจาะจงจากพนักงานต้อนรับบนเครื่องบินรุ่นใหม่ของบริษัท การบินไทย จำกัด (มหาชน) ที่เริ่มปฏิบัติงานในปี พ.ศ. 2546 ทั้งหญิงและชาย ที่เข้ารับการทดสอบความสามารถทางภาษาอังกฤษ (Placement test) จากกองฝึกอบรมภาษา บริษัท การบินไทย จำกัด (มหาชน) จำนวน 489 คน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ (Placement test) ของพนักงานต้อนรับบนเครื่องบินที่เริ่มปฏิบัติงานกับบริษัท การบินไทย จำกัด (มหาชน) ในปี พ.ศ. 2546 แบบทดสอบที่นำมาศึกษาดังกล่าวประกอบด้วยข้อคำถามแบบเลือกตอบ (Multiple choices) 4 ตัวเลือก จำนวน 90 ข้อ โดยมีโครงสร้างของข้อคำถามดังนี้

1. เป็นแบบทดสอบการใช้ภาษาเพื่อการสื่อสาร (Service Language Part) ประเภทเลือกตอบ จำนวน 20 ข้อ

2. เป็นแบบทดสอบความเข้าใจในการอ่านและคำศัพท์ (Reading comprehension and vocabulary) จำนวน 40 ข้อ

3. เป็นแบบทดสอบโครงสร้างทางไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์ (Error recognition) และประเภทเลือกเติมคำ จำนวน 30 ข้อ

การเก็บรวบรวมข้อมูล

ผู้วิจัยทำการเก็บรวบรวมข้อมูลตามขั้นตอนต่อไปนี้

1. ติดต่อขออนุญาตผู้จัดการกองฝึกอบรมภาษา บริษัท การบินไทย จำกัด (มหาชน) นำข้อสอบที่ใช้สอบกับพนักงานต้อนรับบนเครื่องบินมาทำการวิจัยหาคุณภาพของแบบทดสอบ
2. นำผลการทดสอบมาวิเคราะห์หาค่าสถิติ และหาคุณภาพของแบบทดสอบรายข้อ ด้านความเที่ยงตรงความยากง่ายและอำนาจจำแนก
3. หาคุณภาพของแบบทดสอบทั้งฉบับโดยใช้ค่าความเชื่อมั่น และความสอดคล้องกับผลการทดสอบด้วยแบบทดสอบมาตรฐาน TOEIC
4. สร้างเกณฑ์ปกติคะแนนที่ (Normalized T-score)

การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล

1. การตรวจสอบคุณภาพด้านความเที่ยงตรง (Validity)
 - 1.1 หาความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบ โดยหาค่าสัมประสิทธิ์ความสอดคล้องของความสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับโดยใช้สูตรการหาสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple correlation) ของเพียร์สัน (Pearson product moment correlation)
 - 1.2 นำผลการวัดที่ได้ไปวิเคราะห์ข้อคำถามเป็นรายข้อ (Item analysis) เพื่อหาความยากง่าย และอำนาจจำแนก โดยใช้เทคนิค 27% ของ จุง เตห์ ฟาน (Chung The Fan)
2. การตรวจสอบคุณภาพของแบบทดสอบทั้งฉบับด้านความเชื่อมั่น พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ของความเชื่อมั่น (Reliability coefficient) โดยวิเคราะห์หาค่าสัมประสิทธิ์ของความเชื่อมั่นโดยใช้สูตร K-R 20 ของคูเดอร์และริชาร์ดสัน (Kuder Richardson)
3. นำข้อทดสอบที่ผ่านการวิเคราะห์คุณภาพมาพิจารณาความสอดคล้อง (Correlation) ของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษที่สร้างขึ้นกับแบบทดสอบ TOEIC โดยใช้สัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson product moment correlation)
4. คัดเลือกข้อทดสอบเพื่อเก็บไว้ในคลังข้อทดสอบ
5. สร้างเกณฑ์ปกติคะแนนที่ (Normalized T-score) ของแบบทดสอบที่ผ่านการวิเคราะห์คุณภาพ โดยการหาตำแหน่งเปอร์เซ็นต์ไทล์แล้วนำไปเทียบเป็นค่าคะแนนที่ปกติจากตารางสำเร็จรูป จากนั้นนำคู่อันดับระหว่างคะแนนผลการทดสอบ (คะแนนดิบ) กับคะแนนที่ปกติ มาลงจุดกราฟ

ขยายคะแนนที่ปกติและคะแนนดิบให้ครบทุกคะแนน (0 ถึง 90 คะแนน) โดยการใช้เส้นตรงที่เป็นตัวแทนของจุดกราฟ แล้วอ่านค่าความสัมพันธ์ระหว่างคะแนนดิบกับคะแนนที่นำมาสร้างเป็นตารางเกณฑ์ปกติของแบบทดสอบ เพื่อให้สามารถนำคะแนนที่ได้จากการทดสอบครั้งต่อไปมาเทียบกับเกณฑ์ปกติคะแนนที่สร้างขึ้น นำมาใช้ในการจัดระดับความสามารถของผู้รับการทดสอบ

สรุปผลการวิจัย

จากการวิเคราะห์ข้อมูลเพื่อศึกษาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบิน บริษัท การบินไทย จำกัด (มหาชน) จำนวน 90 ข้อ เพื่อคัดเลือกข้อทดสอบที่มีคุณภาพจัดเก็บไว้ในคลังข้อสอบ (Test bank) พบว่า

1. แบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษที่มีคุณภาพด้านความเที่ยงตรงสามารถจัดเก็บไว้ในคลังข้อสอบมีจำนวนทั้งสิ้น 42 ข้อ หรือเท่ากับ 46.6 % ได้แก่ แบบทดสอบข้อที่ 10, 23, 25, 27, 30, 33, 36, 40, 41, 42, 47, 49, 50, 51, 52, 54, 55, 56, 58, 59, 61, 62, 64, 65, 66, 67, 69, 71, 72, 73, 75, 78, และข้อที่ 81-90 โดยมีคุณสมบัติดังนี้

1.1 เป็นข้อทดสอบที่มีความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบที่มีสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับที่มีนัยสำคัญทางสถิติที่ระดับ .05

1.2 เป็นข้อทดสอบที่มีค่าความยากง่ายอยู่ในระดับที่กำหนดคือมีค่าระดับความยากง่ายตั้งแต่ระดับ 0.2 ถึง 0.8

1.3 เป็นข้อทดสอบที่มีค่าอำนาจจำแนกตั้งแต่ 0.2 ขึ้นไป คือเป็นข้อทดสอบที่สามารถจำแนกคนเก่งกับคนอ่อนได้ดี

2. แบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบินทั้งฉบับมีค่าความเชื่อมั่นเท่ากับ 0.8 โดยมีค่าความเชื่อมั่นรายด้านของแบบทดสอบโครงสร้างทางไวยากรณ์ทางภาษาประเภทการหาข้อบกพร่องทางไวยากรณ์และประเภทเลือกเติมคำ และด้านความเข้าใจในการอ่านและคำศัพท์อยู่ในระดับสูง และค่าความเชื่อมั่นด้านการใช้ภาษาเพื่อการสื่อสารประเภทเลือกตอบอยู่ในระดับต่ำซึ่งจะต้องทำการปรับปรุงแก้ไขให้มีคุณภาพต่อไป

3. แบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษมีค่าความสอดคล้องกับแบบทดสอบ TOEIC ในระดับปานกลางโดยมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.585 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงถึงการยืนยันคุณภาพของแบบทดสอบที่มีความสอดคล้องกับแบบทดสอบมาตรฐาน TOEIC ว่าเป็นไปในทิศทางเดียวกัน

4. เกณฑ์คะแนนที่ปกติของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบินทั้งฉบับอยู่ระหว่าง T10 ถึง T82 โดยมีเกณฑ์คะแนน

ที่ปกติด้านการใช้ภาษาเพื่อการสื่อสาร ด้านความเข้าใจในการอ่าน และคำศัพท์ และด้านโครงสร้างทางไวยากรณ์ประเภทการหาข้อบกพร่องทางไวยากรณ์ อยู่ระหว่าง T10 ถึง T76, T10 ถึง T87, และ T10 ถึง T67 ตามลำดับ

อภิปรายผลที่ได้จากการวิจัย

ผลที่ได้จากการวิจัยเรื่องคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบิน บริษัทการบินไทยจำกัด (มหาชน) สามารถอภิปรายตามประเด็นต่างๆดังนี้

1. ความเที่ยงตรงของแบบทดสอบ

1.1 ความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบได้จากการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับเพื่อยืนยันสภาพของความเที่ยงตรงเชิงโครงสร้าง โดยถือว่าคะแนนรวมทั้งฉบับเป็นคะแนนที่แทนโครงสร้างรวมของคุณลักษณะที่ต้องการวัด ถ้ามีค่าสหสัมพันธ์กับคะแนนรวมสูง ข้อคำถามนั้นก็มีความเที่ยงตรงตามโครงสร้างสูงด้วยเช่นกัน ในทางตรงข้ามถ้าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับต่ำ ข้อคำถามนั้นก็มีความเที่ยงตรงเชิงโครงสร้างต่ำด้วย (พวงรัตน์ ทวีรัตน์. 2540: 118) จากผลการวิเคราะห์คะแนนรายข้อและคะแนนรวมทั้งฉบับพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ทางบวกเท่ากับร้อยละ 57.7 โดยมีข้อสอบจำนวน 38 ข้อ ที่ยังต้องปรับปรุงคุณภาพซึ่งส่วนใหญ่เป็นข้อสอบด้านการใช้ภาษาเพื่อการสื่อสารจำนวน 18 ข้อ สาเหตุสำคัญประการหนึ่งที่ทำให้ค่าความเที่ยงตรงไม่สูงมากโดยเฉพาะในด้านดังกล่าวน่าจะมาจาก การที่ข้อสอบมีความยากหรือง่ายเกินไปดังตัวอย่างที่ได้จากผลการวิเคราะห์ข้อมูลในตาราง 8 ซึ่งได้แก่ข้อที่ 1 ($p=.95$, $r_{xy} = -.025$), ข้อที่ 2 ($p=.99$, $r_{xy} = -.051$), และ ข้อที่ 3 ($p=.96$, $r_{xy}=.043$) เป็นต้น ที่พบว่ามีความง่ายเกินไป และข้อที่ 28 ($p=.13$, $r_{xy}=.097$), ข้อที่ 57 ($p=.11$, $r_{xy} = -.145$), และข้อที่ 60 ($p=.19$, $r_{xy} = -.230$) ที่มีความยากเกินไปไม่เหมาะสมกับการนำไปใช้ ตรงกับฉันทะญูวรรณ หลาวทอง (2548: 10-11) ที่กล่าวว่า ข้อสอบที่ง่ายหรือยากมากจะมีผลให้ค่าความเที่ยงตรงที่คำนวณจากสูตรความสัมพันธ์มีค่าน้อยกว่าข้อสอบที่มีความยากง่ายปานกลางนอกจากนี้ปัจจัยที่สำคัญอีกประการหนึ่งที่มีผลต่อความเที่ยงตรงคือ ค่าอำนาจจำแนกกล่าวคือ ข้อทดสอบที่มีค่าอำนาจจำแนกต่ำก็จะมีค่าความเที่ยงตรงต่ำด้วยเช่นกัน (พวงรัตน์ ทวีรัตน์. 2540: 120) ผลจากการวิเคราะห์แบบทดสอบพบว่าข้อสอบที่มีค่าอำนาจจำแนกต่ำกว่ามาตรฐานได้แก่แบบทดสอบข้อที่ 1 ($r=.08$, $r_{xy} = -.025$), ข้อที่ 2 ($r=.04$, $r_{xy}=.051$), และข้อที่ 3 ($r=.03$, $r_{xy}=.043$) เป็นต้น

สรุปว่าความเที่ยงตรงของแบบทดสอบอยู่ในระดับปานกลางโดยมีผลจากการที่มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมทั้งฉบับเท่ากับร้อยละ 58.9 ของจำนวนข้อสอบทั้งฉบับ มีข้อสอบที่มีความยากง่ายพอเหมาะที่สามารถนำไปใช้ได้คิดเป็นจำนวน 63 ข้อ หรือเท่ากับร้อยละ 70 และมีข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์ที่กำหนดจำนวน 64 ข้อ

หรือเท่ากับร้อยละ 71 ทั้งนี้หากได้มีการปรับปรุงข้อสอบที่มีคุณภาพยังไม่เป็นไปตามเกณฑ์ดังกล่าว ก็จะทำให้ความเที่ยงตรงของแบบทดสอบมีค่าสูงขึ้น

1.2 ความยากง่ายของแบบทดสอบมีค่าอยู่ระหว่าง 0.11 ถึง 0.99 จากผลการวิเคราะห์ข้อมูลแสดงว่าแบบทดสอบที่ใช้ในการจัดระดับค่อนข้างง่าย ทั้งนี้อาจเป็นเพราะวัตถุประสงค์ในการจัดทำแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบินมิได้มีไว้เพื่อเป็นเกณฑ์ในการคัดเลือกหรือแบ่งระดับความสามารถของพนักงานออกจากกันอย่างเด็ดขาด การจัดระดับความสามารถของพนักงานเป็นเพียงเครื่องมือที่ช่วยอำนวยความสะดวกในการจัดการเรียนการสอนและการแบ่งกลุ่มของพนักงานเพื่อเข้ารับการอบรมเพื่อเพิ่มศักยภาพในการใช้ภาษาอังกฤษ ทั้งนี้เนื่องจากการเรียนการสอนสำหรับพนักงานระดับนี้เน้นการใช้ภาษาอังกฤษเพื่องานอาชีพมากกว่าการใช้ภาษาอังกฤษเพื่อจุดประสงค์ทางวิชาการ ดังนั้นแบบทดสอบที่จัดทำขึ้นจึงมีความง่ายกว่าแบบทดสอบวัดความสามารถทางภาษาทั่วไป ซึ่งภัทธา นิคมานนท์ (2543: 155) ได้กล่าวว่า การพิจารณาคุณภาพของแบบทดสอบจากค่าความยากง่ายอาจแตกต่างกันไปตามจุดมุ่งหมายในการสอบ เช่น แบบทดสอบสำหรับการประเมินผลการเรียนการสอนโดยทั่วไป อาจใช้แบบทดสอบที่มีค่าความยากง่ายระหว่าง 0.2 ถึง 0.8 แบบทดสอบสำหรับการคัดเลือกอาจใช้แบบทดสอบที่มีค่าความยากง่ายอยู่ระหว่าง .15 ถึง .75 และแบบทดสอบสำหรับการประเมินผลแบบอิงเกณฑ์อาจใช้แบบทดสอบที่ง่ายขึ้นไปอีกก็ได้

สาเหตุประการหนึ่งที่ทำให้แบบทดสอบค่อนข้างง่ายโดยเฉพาะแบบทดสอบด้านการใช้ภาษาเพื่อการสื่อสารซึ่งมีความง่ายกว่าแบบทดสอบในด้านอื่นๆ คือในจำนวนแบบทดสอบ 20 ข้อ จากผลการวิเคราะห์พบว่ามีข้อสอบที่พบว่ามีความง่ายเกินไปถึง 11 ข้อ ทั้งนี้อาจมีสาเหตุเนื่องมาจากผู้เข้ารับการทดสอบได้ผ่านการทดสอบ TOEIC ที่ระดับ 600 คะแนนมาแล้ว ตลอดจนได้รับการฝึกอบรมความรู้ทั่วไปในการปฏิบัติงานจากหน่วยงานต้นสังกัดและได้เริ่มปฏิบัติงานจริงไปบ้างแล้วทำให้ผู้เข้ารับการทดสอบรู้สึกคุ้นเคยกับบริบท (Context) ที่นำมาใช้ในการสร้างแบบทดสอบทำผู้เข้ารับการทดสอบสามารถทำข้อสอบดังกล่าวได้มากขึ้น

1.3 ค่าอำนาจจำแนกของแบบทดสอบมีค่าอยู่ระหว่าง -0.26 ถึง 0.66 มีแบบทดสอบจำนวน 64 ข้อ หรือคิดเป็นร้อยละ 71 ที่มีค่าอำนาจจำแนกตามเกณฑ์ที่กำหนด (0.2 ขึ้นไป) แสดงว่าแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบินมีค่าอำนาจจำแนกอยู่ในระดับค่อนข้างสูง สามารถจำแนกผู้เรียนที่เก่งออกจากผู้เรียนที่อ่อนได้ อย่างไรก็ตามพบว่ามีแบบทดสอบจำนวน 26 ข้อที่ขาดคุณภาพด้านค่าอำนาจจำแนกซึ่งส่วนใหญ่เป็นข้อสอบด้านการใช้ภาษาเพื่อการสื่อสารถึง 14 ข้อ สาเหตุดังกล่าวมีผลทำให้แบบทดสอบมีความเที่ยงตรงที่ลดลงด้วยตรงตามความเห็นของ พวงรัตน์ ทวีรัตน์ (2540: 120) ที่กล่าวว่าค่าอำนาจจำแนกจะมีความสัมพันธ์อย่างมากกับความเที่ยงตรง เครื่องมือที่มีค่าอำนาจจำแนกสูงก็จะมี ความเที่ยงตรงสูงด้วยเช่นกัน จากตัวอย่างที่ได้จากการวิเคราะห์ (ตาราง 8) แสดงให้เห็นว่าข้อสอบข้อที่มีค่าอำนาจจำแนก (r) ไม่สูงมากก็จะมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับ

คะแนนรวมทั้งฉบับ (r_{xy}) ที่ไม่สูงเช่นกัน เช่น แบบทดสอบข้อที่ 1 ($r=.08$, $r_{xy}= -.025$), ข้อที่ 2 ($r=.04$, $r_{xy}= -.051$), ข้อที่ 7 ($r=.09$, $r_{xy}=.023$), และ ข้อที่ 14 ($r=.11$, $r_{xy}= .001$) เป็นต้น ทั้งนี้หากได้ทำการปรับปรุงแก้ไขแบบทดสอบให้มีข้อคำถามที่ดี เข้าใจง่าย ไม่วกวน คลุมเครือ ตลอดจนการปรับปรุงคุณภาพของตัวลวงให้ดีขึ้นก็เชื่อว่าจะทำให้แบบทดสอบนี้มีคุณภาพด้านค่าอำนาจจำแนกที่ดีขึ้นเช่นเดียวกัน

2. ด้านความเชื่อมั่นของแบบทดสอบพบว่า แบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบินมีค่าความเชื่อมั่นเท่ากับ .80 ซึ่งอยู่ในระดับสูง ทั้งนี้อาจเป็นผลมาจากการที่มีจำนวนข้อของแบบทดสอบทั้งหมดถึง 90 ข้อ ซึ่งณัฐภรณ์ หลาวทอง (2548: 10-17) ได้กล่าวว่าแบบทดสอบที่มีจำนวนข้อสอบมาก ค่าความเชื่อมั่นจะสูงกว่าผลสอบที่เกิดจากแบบทดสอบที่มีจำนวนน้อย อย่างไรก็ตามเมื่อพิจารณาค่าความเชื่อมั่นรายด้านผลการวิเคราะห์พบว่าค่าความเชื่อมั่นด้านแบบทดสอบการใช้ภาษาเพื่อการสื่อสารมีค่าเท่ากับ 0.26 อยู่ในระดับค่อนข้างต่ำซึ่งจะต้องทำการปรับปรุงเพื่อให้มีความเชื่อมั่นสูงขึ้นต่อไป ทั้งนี้สาเหตุที่ทำให้ค่าความเชื่อมั่นของแบบทดสอบด้านดังกล่าวอยู่ในระดับต่ำอาจเป็นเพราะจำนวนข้อของแบบทดสอบในด้านนี้มีอยู่เพียง 20 ข้อ ในขณะที่แบบทดสอบด้านความเข้าใจในการอ่าน และคำศัพท์ มีจำนวนถึง 40 ข้อ และแบบทดสอบด้านโครงสร้างทางไวยากรณ์ประเภทการหาข้อบกพร่องทางไวยากรณ์ และประเภทเลือกเติมคำมีจำนวน 30 ข้อ ซึ่งสอดคล้องกับความเห็นของ อรพินท์ ชูชม (2545: 323) ที่กล่าวว่าข้อคำถามที่มีจำนวนข้อน้อยเกินไปจะทำให้เกิดความคลาดเคลื่อนในการวัดเป็นปัจจัยที่ส่งผลต่อค่าความเชื่อมั่น

นอกจากนี้ปัจจัยที่มีผลต่อค่าความเชื่อมั่นของแบบทดสอบได้แก่คุณภาพของแบบทดสอบรายข้อด้านความยากง่ายและค่าอำนาจจำแนกซึ่งพบว่าแบบทดสอบด้านการใช้ภาษาเพื่อการสื่อสารมีจำนวนข้อสอบที่มีคุณภาพด้านความยากง่ายไม่เป็นไปตามเกณฑ์ที่กำหนดคิดเป็นร้อยละ 55 และค่าอำนาจจำแนกคิดเป็นร้อยละ 70 ตรงตามแนวความเห็นของ ณัฐภรณ์ หลาวทอง (2548: 10-17) และอรพินท์ ชูชม (2545: 323) ที่ได้กล่าวไว้ว่า แบบทดสอบที่ยากหรือง่ายเกินไปและมีค่าอำนาจจำแนกต่ำจะทำให้แบบทดสอบดังกล่าวมีค่าความเชื่อมั่นต่ำด้วยเช่นกัน

3. ความสอดคล้อง (Correlation) ของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษกับแบบทดสอบ TOEIC พบว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ที่ .585 แสดงว่าแบบทดสอบที่สร้างขึ้นมีความสัมพันธ์ทางบวกกับแบบทดสอบมาตรฐาน TOEIC กล่าวคือผู้ที่ทำแบบทดสอบ TOEIC ได้คะแนนสูงก็มีแนวโน้มที่จะสามารถทำแบบทดสอบจัดระดับความสามารถได้สูงเช่นกัน ทั้งนี้น่าจะมีสาเหตุมาจากแบบทดสอบของหน่วยงานที่จัดทำขึ้นมีรูปแบบและโครงสร้างใกล้เคียงกับแบบทดสอบ TOEIC แม้ว่าเนื้อหา (Content) และบริบท (Context) ที่ใช้จะเกี่ยวข้องกับลักษณะงานของพนักงานมากกว่าแบบทดสอบ TOEIC และด้วยรูปแบบและโครงสร้างที่ใกล้เคียงกับ

แบบทดสอบ TOEIC นี้ จะทำให้ผู้เข้ารับการทดสอบรู้สึกคุ้นเคย และสามารถทำความเข้าใจกับวิธีการทำแบบทดสอบได้อย่างรวดเร็ว

จากการค้นคว้าเพิ่มเติมได้พบงานวิจัยที่ใกล้เคียงกับความสอดคล้องของแบบทดสอบข้างต้นจัดทำโดย ฮิราอิ (Hirai, 2002: 1-7) เรื่องความสอดคล้องระหว่างคะแนนของแบบทดสอบทักษะเชิงรุกและแบบทดสอบทักษะเชิงรับ (Correlations between active skill and passive skill test scores) ซึ่งได้ทำการศึกษาความสอดคล้อง (Correlation) ระหว่างแบบทดสอบสัมภาษณ์ภาษาอังกฤษ (Company-internal interview) ของพนักงานบริษัท Hitachi ที่เข้ารับการอบรมภาษาอังกฤษตามหลักสูตรที่กำหนดขึ้นกับแบบทดสอบ TOEIC พบว่าสัมประสิทธิ์ความสอดคล้องของแบบทดสอบทั้งสองมีค่าเท่ากับ 0.78 ซึ่งมีความสอดคล้องอยู่ในระดับค่อนข้างสูง นอกจากนี้ ฮิราอิยังได้ทำการศึกษาความสอดคล้องของแบบทดสอบ BULATS (Business Language Testing Service) ซึ่งเป็นแบบทดสอบด้านการเขียน (Writing test) ที่พัฒนาโดยมหาวิทยาลัย Cambridge กับแบบทดสอบ TOEIC อีกด้วย ผลการศึกษาพบว่าสัมประสิทธิ์ความสอดคล้องของแบบทดสอบดังกล่าวมีค่าเท่ากับ .66 ซึ่งต่ำกว่าสัมประสิทธิ์ความสอดคล้องระหว่างแบบทดสอบสัมภาษณ์ภาษาอังกฤษกับแบบทดสอบ TOEIC อยู่เล็กน้อย อย่างไรก็ตามจะเห็นว่าทั้งแบบทดสอบการสัมภาษณ์ภาษาอังกฤษ และแบบทดสอบ BULATS ซึ่งเป็นแบบทดสอบการใช้ทักษะเชิงรุก (Active skill) กับแบบทดสอบ TOEIC ซึ่งเป็นแบบทดสอบทักษะเชิงรับ (Passive skill) มีความสอดคล้องกัน

4. เกณฑ์ปกติ (Norm) ของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบิน มีคะแนนดิบตั้งแต่ 27 ถึง 83 คะแนน มีค่าคะแนนที่ปกติอยู่ที่ T 10 ถึง T 82 คะแนนที่ปกติของแบบทดสอบดังกล่าวได้ถูกนำมาสร้างและปรับขยายให้ครอบคลุมคะแนนที่ได้จากการสอบหรือคะแนนดิบต่ำสุด (0 คะแนน) ถึงคะแนนเต็ม (90 คะแนน และคะแนนเต็มในแต่ละด้าน) ซึ่งทำให้แบบทดสอบด้านการใช้ภาษาเพื่อการสื่อสารประเภทเลือกตอบมีคะแนนที่ปกติอยู่ระหว่าง T10 ถึง T74 แบบทดสอบความเข้าใจในการอ่าน และคำศัพท์มีคะแนนที่ปกติอยู่ระหว่าง T10 ถึง T85 และแบบทดสอบโครงสร้างทางไวยากรณ์ประเภทการหาข้อบกพร่องทางไวยากรณ์และประเภทเลือกเติมคำมีคะแนนที่ปกติอยู่ระหว่าง T10 ถึง T67 ตามลำดับ

คะแนนที่ปกติที่สร้างขึ้นนี้สามารถนำไปใช้ในการแปลความหมายคะแนนตามเกณฑ์กับกลุ่มพนักงานต้อนรับบนเครื่องบิน บริษัท การบินไทย จำกัด (มหาชน) ที่ได้นำแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษที่ผู้วิจัยได้ทำการศึกษาไปใช้ และนอกจากนี้เมื่อนำคะแนนที่ปกติในแต่ละด้านมาทำการลงจุดกราฟก็จะทำให้ทราบผลการทดสอบหรือระดับความสามารถในแต่ละด้านของผู้รับการทดสอบได้อีกด้วย ผลที่ได้รับดังกล่าวจะเป็นประโยชน์ให้ผู้สอนและหน่วยงานที่เกี่ยวข้องสามารถทำการวางแผน ปรับปรุง พัฒนาแบบทดสอบที่จะสร้าง

ขึ้นในครั้งต่อไปให้มีความสมบูรณ์มากยิ่งขึ้น อีกทั้งยังสามารถใช้เป็นแนวทางในการช่วยเหลือผู้เรียนที่ต้องการพัฒนาความสามารถในการใช้ภาษาอังกฤษในแต่ละด้านให้ตรงตามเป้าหมายและวัตถุประสงค์มากที่สุดต่อไป

ข้อเสนอแนะทั่วไป

1. แม้แบบทดสอบที่นำมาใช้ในการวิจัยจะค่อนข้างมีคุณภาพตามเกณฑ์ แต่ก่อนที่จะมีการนำไปเก็บไว้ในคลังข้อสอบ คณะกรรมการจัดทำแบบทดสอบควรรับทราบผลการวิจัยเพื่อทำการศึกษาข้อทดสอบที่ดีและไม่ดีร่วมกันและทำการปรับปรุงข้อสอบบางข้อให้มีคุณภาพด้านความเที่ยงตรง ความยากง่าย และคำอ่านจำแนกเพิ่มมากขึ้น
2. หลังจากได้ทำการปรับปรุงคุณภาพเพิ่มเติมแล้ว เมื่อมีกลุ่มตัวอย่างใหม่หรือเมื่อมีพนักงานต้อนรับชุดใหม่ควรได้มีการนำแบบทดสอบที่จัดเก็บไว้มาทำการทดลองซ้ำหลายๆ ครั้ง (Try out) เพื่อนำข้อมูลจากการทดลองมาใช้ในการปรับปรุงให้แบบทดสอบของหน่วยงานมีคุณภาพมากขึ้น
3. ฝ่ายบริหารหรือผู้ที่เกี่ยวข้องอาจใช้ผลการวิจัยครั้งนี้เป็นข้อมูลประกอบการกำหนดนโยบายในการวางแผนพัฒนาบุคลากรของบริษัทให้มีคุณสมบัติและความสามารถในการใช้ภาษาอังกฤษตามเกณฑ์ที่บริษัทกำหนด (Core competency) ต่อไป

ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

1. ควรมีการหาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษของพนักงานต้อนรับบนเครื่องบินในทักษะอื่นๆที่ยังไม่ได้ทำการหาคุณภาพ เช่น แบบทดสอบความสามารถในการใช้ภาษาอังกฤษด้านทักษะการฟัง (Listening skill) ทักษะด้านการพูด (Speaking skill) และทักษะการเขียน (Writing skill)
2. สร้างเกณฑ์ปกติของแบบทดสอบความสามารถในการใช้ภาษาอังกฤษในแต่ละทักษะเพื่อให้สามารถนำผลการทดสอบมาใช้เป็นแนวทางในการพัฒนาศักยภาพการใช้ภาษาอังกฤษของบุคลากรในหน่วยงานว่ามีความต้องการเพิ่มเติมหรือซ่อมเสริมในด้านใด
3. ควรมีการหาคุณภาพของแบบทดสอบจัดระดับความสามารถในการใช้ภาษาต่างประเทศอื่นๆนอกเหนือจากภาษาอังกฤษ ตลอดจนสร้างเกณฑ์ปกติของแบบทดสอบเพื่อเป็นประโยชน์แก่หน่วยงานต่อไป

บรรณานุกรม

บรรณานุกรม

- กัลยาณี นันทะศรี. (2542). การเปรียบเทียบคุณภาพของแบบทดสอบโมดิฟายด์โคลซ 5 รูปแบบ ในการวัดความเข้าใจในการอ่านภาษาอังกฤษ ของนักเรียนชั้นประถมศึกษาปีที่ 6. วิทยานิพนธ์ กศ.ม. (การบริหารการศึกษา). มหาสารคาม: บัณฑิตวิทยาลัย มหาวิทยาลัยมหาสารคาม. ถ่ายเอกสาร.
- เฉลิมเกียรติ เฟื่องแก้ว. (2539). การสร้างแบบทดสอบวัดระดับความสามารถด้านการอ่าน ภาษา ฝรั่งเศสของนักเรียนชั้นมัธยมศึกษาปีที่ 5 เขตกรุงเทพมหานคร. วิทยานิพนธ์ ศศ.ม. (ศึกษาศาสตร์-การสอนภาษาฝรั่งเศส). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์. ถ่ายเอกสาร.
- ชวาล แพรัตกุล. (2518). เทคนิคการวัดผล. พิมพ์ครั้งที่ 6. กรุงเทพฯ: เจริญผล.
- ชวาล แพรัตกุล. (22548). เกณฑ์ปรกติ. ใน สารานุกรมศึกษาศาสตร์ ฉบับรวมเรื่องเฉพาะเล่ม อันดับ 3 (การวัดและประเมินผลการศึกษา). หน้า 172-182. กรุงเทพฯ: ก๊อปปี้ แอนด์ พรินท์.
- ชาติเชื้อ สุวรรณมุสิก. (2547). การพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาการปฐม พยาบาล มหาวิทยาลัยราชภัฏยะลา. ยะลา: คณะวิทยาศาสตร์และเทคโนโลยี.
- โชติ เพชรชื่น. (2548). คะแนน. ใน สารานุกรมศึกษาศาสตร์ ฉบับรวมเรื่องเฉพาะเล่ม อันดับ 3 (การวัดและประเมินผลการศึกษา). หน้า 87-90. กรุงเทพฯ: ก๊อปปี้ แอนด์ พรินท์.
- ณัฐฐารณ์ หลาวทอง. (2548). เอกสารประกอบการสอนวิชา 2702303 การวัดและการประเมินผล ทางการศึกษา. กรุงเทพฯ: คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ดิลก ดิลกานนท์. (2547). การเขียนข้อสอบแบบเลือกตอบ. ใน เอกสารประกอบการอบรม หลักสูตร การวัดและประเมินผลทางการศึกษา. หน้า 77-88. กรุงเทพฯ: สำนักทดสอบทาง การศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ.
- บุญชม ศรีสะอาด. (2520). การวัดผลและการประเมินผลทางการศึกษา. มหาสารคาม: คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- บุญเชิด ภิญโญอนันตพงษ์. (2524). การวัดและการประเมินผลการศึกษา : ทฤษฎีและการประยุกต์. กรุงเทพฯ: อักษรเจริญทัศน์.
- ประโมกษ์ ดุลนัย. (2542). การสร้างแบบทดสอบวัดภาคปฏิบัติด้านทักษะการพูดวิชาภาษาอังกฤษ สำหรับนักเรียน ชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ กศ.ม. (การบริหารการศึกษา). มหาสารคาม: บัณฑิตวิทยาลัย มหาวิทยาลัยมหาสารคาม. ถ่ายเอกสาร.
- พวงรัตน์ ทวีรัตน์. (2540). วิธีการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์ (ฉบับปรับปรุงใหม่ล่าสุด). พิมพ์ครั้งที่ 7. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.

- พิชิต ฤทธิ์จัญญ. (2544). *หลักการวัดและประเมินผลการศึกษา*. กรุงเทพฯ: คณะครุศาสตร์ สถาบันราชภัฏพระนคร.
- เพลินพิศ ธรรมรัตน์. (2542). *การประเมินผลการเรียน*. สกลนคร: คณะครุศาสตร์สถาบันราชภัฏสกลนคร.
- เพิ่มสุข วิชญวงค์. (2547, 20 ธันวาคม). สัมภาษณ์โดย วัชรพงษ์ กริบุตร ที่บริษัทการบินไทย จำกัด (มหาชน). เมื่อ 20 ธันวาคม 2547.
- ไพศาล หวังพานิช. (2547). ลักษณะของเครื่องมือวัดที่ดี. ใน *เอกสารประกอบการอบรมหลักสูตรการวัดและประเมินผลทางการศึกษา*. หน้า 57-76. กรุงเทพฯ: สำนักทดสอบทางการศึกษา และจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ภัทรา นิคมานนท์. (2522). *การประเมินผลและการสร้างแบบทดสอบ*. กรุงเทพฯ: อักษราพัฒนา.
- เยาวดี วิบูลย์ศรี. (2540). *การวัดผลและการสร้างแบบทดสอบผลสัมฤทธิ์*. พิมพ์ครั้งที่ 2. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- รัตนะ บัวสนธ์. (2541). คะแนนและการใช้ผลคะแนน. ใน *เอกสารการสอน การวัดและการประเมินผลการศึกษา*. บุญชม ศรีสะอาด และ สมนึก ภัทธีธรณี. หน้า 185-219. กรุงเทพฯ: สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี.
- รัตนา เกิดผล. (2536). *การศึกษาความสัมพันธ์ระหว่างคะแนนที่ได้จากแบบทดสอบการเขียนตามคำบอกแบบเติมคำ กับคะแนนที่ได้จากแบบทดสอบมาตรฐานวิชาภาษาอังกฤษ และเจตคติต่อแบบทดสอบการเขียนตามคำบอกแบบเติมคำ ของนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนในสังกัดกรมสามัญศึกษา กรุงเทพมหานคร*. วิทยานิพนธ์ ศษ.ม. (การสอนภาษาอังกฤษในฐานะภาษาต่างประเทศ). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร. ถ่ายเอกสาร.
- ล้วน สายยศ; และ อังคณา สายยศ. (2540). *สถิติวิทยาทางการวิจัย*. พิมพ์ครั้งที่ 3. กรุงเทพฯ: สุวีริยาสาส์น.
- วรรณ ภูมิทัศน์. (2533). *คุณภาพของแบบทดสอบคัดเลือกบุคคลเข้าศึกษาในมหาวิทยาลัยขอนแก่นโดยวิธีรับตรง : โปรแกรมทั่วไป*. วิทยานิพนธ์ ศษ.ม. (การวัดและประเมินผลการศึกษา). ขอนแก่น: บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น. ถ่ายเอกสาร.
- วรัญญา วิศาลาภรณ์. (2540). *การวิจัยทางการศึกษาหลักการและแนวทางในการปฏิบัติ*. กรุงเทพฯ: คอมแพคพริ้นท์.
- วิรัช วรรณรัตน์. (2547). จุดประสงค์การสอนและการสอบ. ใน *เอกสารประกอบการอบรมหลักสูตรการวัดและประเมินผลทางการศึกษา*. หน้า 21-32. กรุงเทพฯ: สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ.

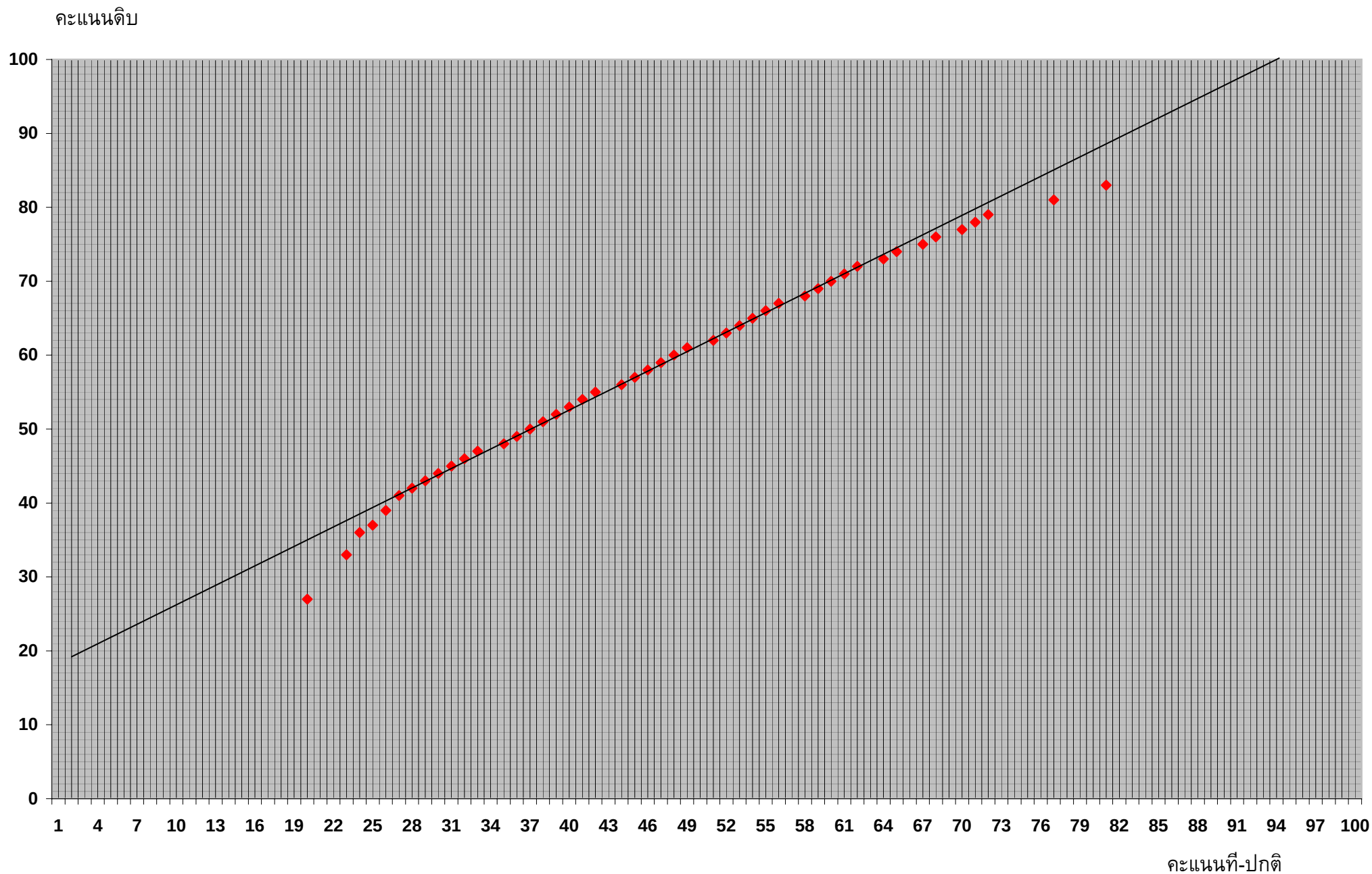
- วีระพงษ์ ศรีชาติ. (2535). การสร้างแบบทดสอบอิงเกณฑ์วิชาภาษาอังกฤษด้านทักษะการอ่าน เพื่อความเข้าใจ ชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา).
มหาสารคาม: บัณฑิตวิทยาลัย มหาวิทยาลัยมหาสารคาม. ถ่ายเอกสาร.
- สมนึก ภัททิยธนี. (2546). การวัดผลการศึกษา. พิมพ์ครั้งที่ 4. กาลสินธุ์: ประสานการพิมพ์.
- สุทิน พูลสวัสดิ์. (2546). เตรียมสอบ TOEIC. พิมพ์ครั้งที่ 3. กรุงเทพฯ: TGRE.
- สุมิตรา อังวัฒนกุล. (2539). การทดสอบความสามารถทางภาษา (Language Testing). กรุงเทพฯ : คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- สำเร็จ บุญเรืองรัตน์. (2548). ความเที่ยงตรง. ใน สารานุกรมศึกษาศาสตร์ ฉบับรวมเรื่องเฉพาะ เล่ม อันดับ 3 (การวัดและประเมินผลการศึกษา). หน้า 172-178. กรุงเทพฯ: ก๊อปปี้ แอนด์ พรินท์
- อมรพรรณ จำปา. (2546). ความสามารถในการใช้ภาษาอังกฤษธุรกิจของนักศึกษาคณะ บริหารธุรกิจ สถาบันเทคโนโลยีราชมงคล. วิทยานิพนธ์ศึกษาศาสตร์บัณฑิต (ศึกษาศาสตร์-การสอนภาษาอังกฤษ). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยเกษตรศาสตร์. ถ่ายเอกสาร.
- อรพิน ชูชม. (2545). การสร้างและพัฒนาเครื่องมือวัดทางพฤติกรรมศาสตร์. ใน เอกสารการสอน วิชา วป/502. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- อาภา มณีสถิตย์. (2545). การสร้างแบบทดสอบวัดบุคลิกภาพผู้จัดการสาขาของธนาคาร ตามแนว ทฤษฎีบุคลิกภาพของแคทเทิล. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- อนันต์ ศรีโสภา. (2525). การวัดผลการศึกษา. กรุงเทพฯ: ไทยวัฒนาพานิช.
- อัจฉรา วงศ์โสธร. (2538). แนวทางการสร้างข้อสอบภาษา. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- อัจฉรา วงศ์โสธร. (2544). การทดสอบและประเมินผลการเรียนการสอนภาษาอังกฤษ. พิมพ์ครั้งที่ 2. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- อัจฉรา วงศ์โสธร. (2544). การพัฒนาข้อทดสอบอิงวัตถุประสงค์แบบอุดมวัดความสามารถและการ กระทำปกติทางภาษา. กรุงเทพฯ: เท็กซ์ แอนด์ เจอร์นัล.
- อนุวัติ คุณแก้ว. (2547). ข้อสอบแบบอัตนัย. สืบค้นเมื่อ 17 พฤศจิกายน 2547, จาก <http://www.ripb.ac.th/Dranuwat/chapter6/f14html>
- Anastasi. (1982). *Psychological testing*. 6th Ed. New York: Macmillan.
- Bailey, Kathleen M. (1998). *Learning About Language Assessment: Dilemmas, Decisions, and Directions*. Albany: Heinle & Heinle.
- Brown, F.G. (1976). *Principles of Education and Psychological Testing*. New York: Holt, Rinehart and Winston.
- Coombe, Christine. (2005). *Testing*. (Thailand TESOL Handouts). Bangkok. Photocopied.

- Coombe, Christine; & Hubley, Nancy. (2005). The Cornerstones of Languages Testing. In *The Fundamentals of Language Assessment: A Practical Guide for Teachers in the Gulf*. pp. 3-16. Dubai: Midex.
- Davies, Roger; & Houghton, Peter. (1995). *Mastering Psychology*. 2nd ed. London: Macmillan.
- Dwaretzky, John P. (1997). *Psychology*. 6th. California: Brooks/ Gole.
- Ebel, Robert L. ; & Frisbie, David A. (1996). *Essentials of Education Measurement*. 4th ed. New Jersey: Prentice-Hall.
- Eysenck. J. Hans; & Eysenck Sybil B.G. (1969). *Personality Structure and Measurement*. U.S.A.: Robert R. Knapp.
- Gronlund, Norman E. (1976). *Measurement and Evaluation in Teaching*. 3rd ed. New York: Macmillan.
- Harris, David P. (1969). *Testing English as a Second Language*. New York: McGraw-Hill.
- Harrison, Andrew. (1986). *A Language Testing Handbook*. 2nd ed. Hong Kong: Macmillian.
- Heaton, J.B. (1988). *Writing English Language Tests (New Edition)*. London: Longman.
- Hirai, Michihiro. (2002). *Correlations between active skill and passive skill test scores*. Retrived March, 2006, from http://www.jalt.org/test/hir_1.htm
- Huges, Arthur. (2003). *Testing for Language Teachers*. 2nd ed. United Kingdom: Cambridge.
- Kubiszyn, Tom ; & Gary, Borich. (1996). *Educational Testing and Measurement, Classroom Application and Practice*. 5th ed. United States of America: HarperCollins.
- Oller, J.W. (1979). *Language tests at school*. London: Longman
- Pervin, Lawrence A. (1993). *Sixth Edition Personality and Research*. 6th ed. Canada: TC Systems.
- Puckett, Magaret B. ; & Black, Janet K. (2000). *Authentic Assessment of the Young Child : Celebrating Developmentr and Learning*. New Fersey: Prentice-Hall.
- Raeaz, Ulrich; & Klien-Braley, Christine. (2004). *Introduction to Language testing and C-tests*. Retrived November 5, 2004, from <http://www.uniduisburg.de/FB3/ANGLING/FORSCHUNG/HOWTO.HTM>
- Takeuchi. (2003). Dictation: *Is it really effective for language teaching?*. Retrived February 14, 2003, from <http://www.2.ipcku.kansaiu.ac.jp/~takeuchi/papers/av20.html>

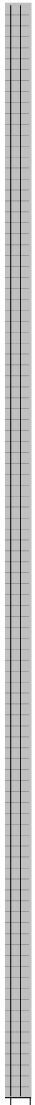
Thai Airways International PCL. (2003). *Career* . Retrieved September 4, 2003, from <http://www.thaiairways.com>

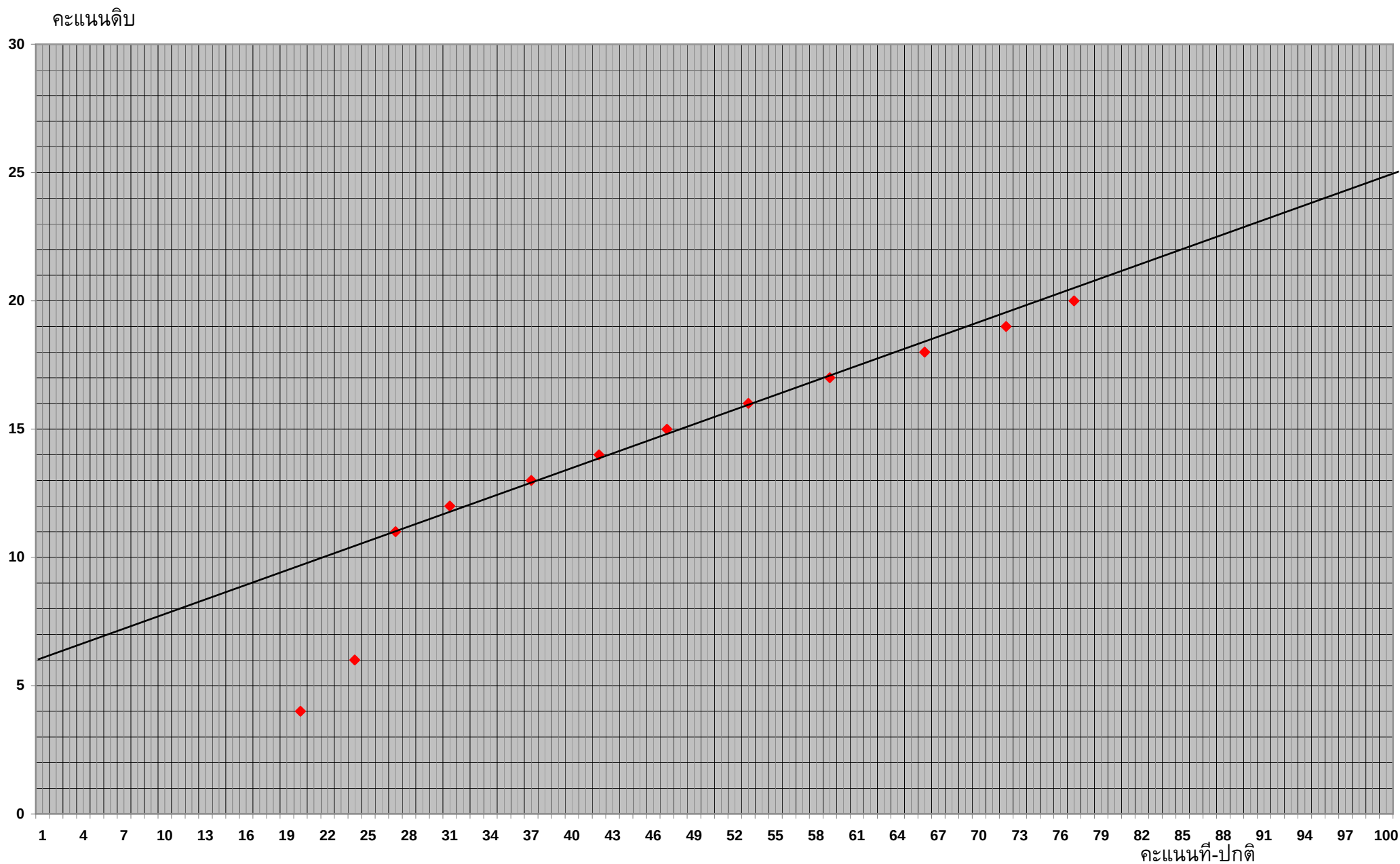
Vallette, Rebecca M. (1977). *Modern language Testing*. 2nd ed. New York: Harcourt Brace Jovanovich.

ภาคผนวก

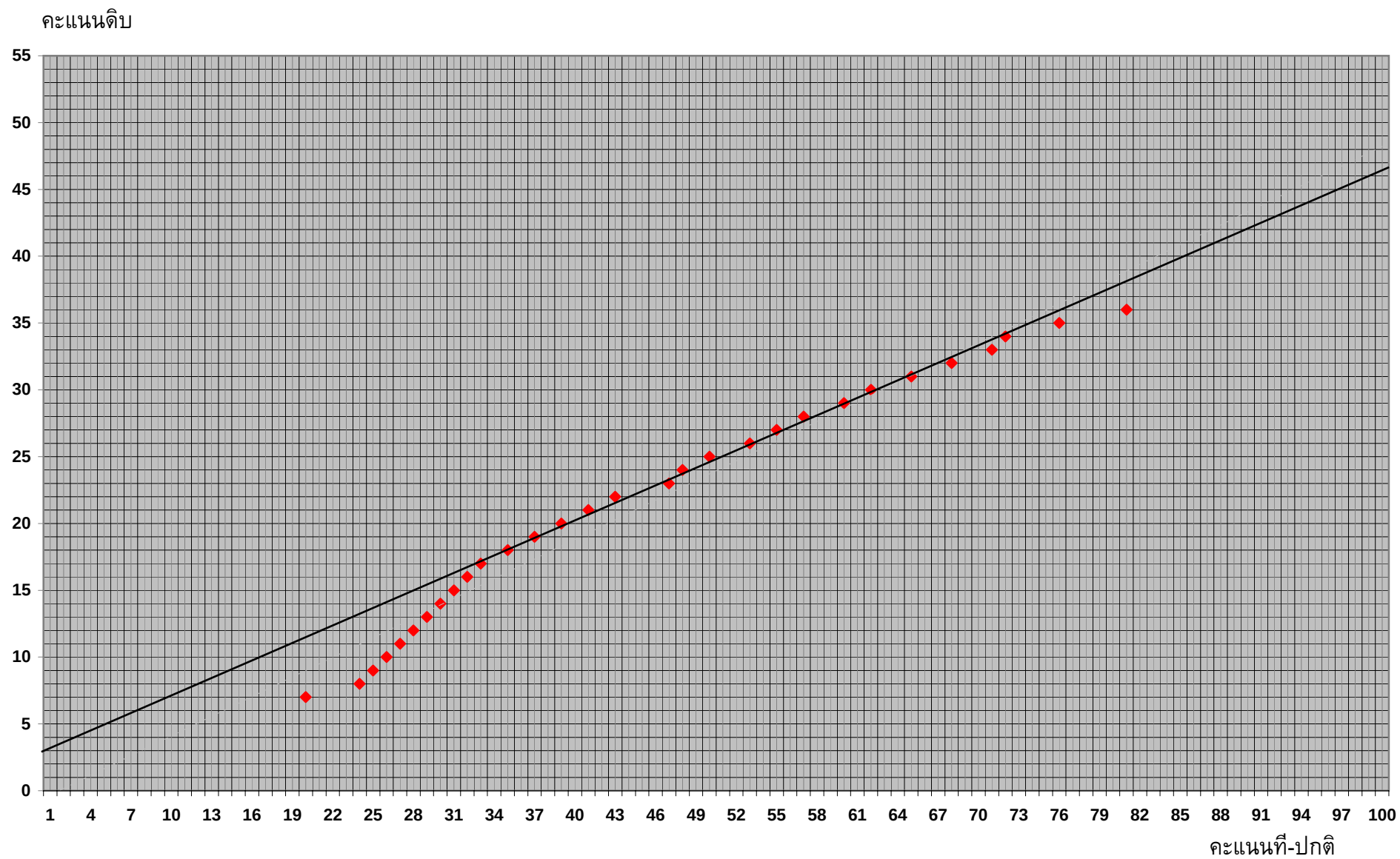


ภาพประกอบ 4 กราฟแสดงความสัมพันธ์ระหว่างคะแนนดิบและคะแนนมาตรฐานที่-ปกติ แบบทดสอบจัดระดับความสามารถในการใช้ภาษาอังกฤษ



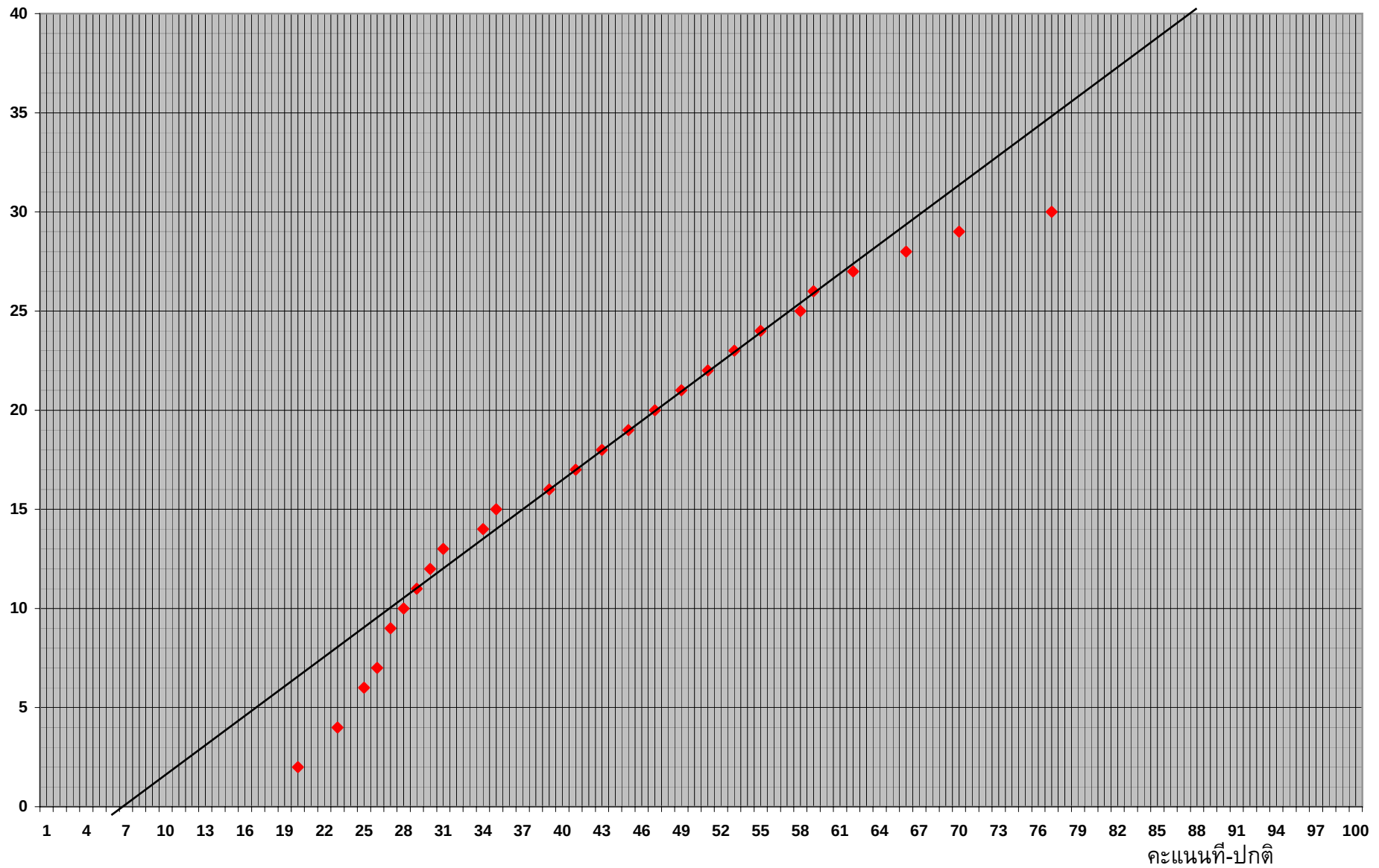


ภาพประกอบ 5 กราฟแสดงความสัมพันธ์ระหว่างคะแนนดิบและคะแนนมาตรฐานที่-ปกติ แบบการทดสอบการใช้ภาษาเพื่อการสื่อสารประเภทเลือกตอบ



ภาพประกอบ 6 กราฟแสดงความสัมพันธ์ระหว่างคะแนนดิบและคะแนนมาตรฐานที่-ปกติ แบบทดสอบความเข้าใจในการอ่าน และคำศัพท์

คะแนนดิบ



ภาพประกอบ 7 กราฟแสดงความสัมพันธ์ระหว่างคะแนนดิบและคะแนนมาตรฐานที่-ปกติ แบบทดสอบด้านโครงสร้างทางไวยากรณ์ทางภาษาประเภท

การหาข้อบกพร่องทางไวยากรณ์ และประเภทเลือกเติมคำ

ประวัติย่อผู้ทำสารนิพนธ์

ประวัติย่อผู้ทำสารนิพนธ์

ชื่อ ชื่อสกุล	นายวัชรพงษ์ กรีบุตร
วันเดือนปีเกิด	11 ตุลาคม 2503
สถานที่เกิด	อำเภอ บางเขน กรุงเทพมหานคร
สถานที่อยู่ปัจจุบัน	22/33 หมู่บ้านเปรมประชา ถนนวิภาวดีรังสิต แขวงสีกัน เขตดอนเมือง กรุงเทพมหานคร 10210
สถานที่ทำงานปัจจุบัน	บริษัทการบินไทย จำกัด (มหาชน)
ประวัติการศึกษา	
พ.ศ. 2514	โรงเรียนสาริต
พ.ศ. 2522	จากมหาวิทยาลัยเกษตรศาสตร์ มนุษยศาสตร์ วิชาเอกภาษาอังกฤษ จากมหาวิทยาลัยเชียงใหม่
พ.ศ. 2545	ศิลปศาสตรมหาบัณฑิต วิชาเอกการสอนภาษาอังกฤษใน ฐานะภาษาต่างประเทศ จากมหาวิทยาลัยศรีนครินทรวิโรฒ