



บทที่ ๑

บทนำ

ความเป็นมาและแนวเหตุผลเชิงทฤษฎีที่สำคัญของปัญหา

คุณสมบัติที่สำคัญ ๒ ประการของแบบสอบคือ ความตรง (Validity) และ ความเที่ยง (Reliability)

สำหรับความตรงของแบบสอบนั้น โดยทั่วไปหมายถึง ความถูกต้องของสิ่งที่วัดที่แบบสอบสามารถวัดได้ (Ebel, ๑๙๗๐:๔๐๕) แต่ความหมายเฉพาะของความตรงนั้น แตกต่างกันไปตามวิธีการคำนวณ และจุดมุ่งหมายของการวัด หรือสิ่งที่ต้องการอ้างอิง ดังนั้น จึงมีความตรงของแบบสอบหลายอย่าง เช่น ความตรงเชิงพยากรณ์ (Predictive Validity) ความตรงร่วมสมัย (Concurrent Validity) ความตรงเชิงภาวะสันนิฐาน (Construct Validity) ความตรงเชิงประจักษ์ (Face Validity) และความตรงเชิงเนื้อหา (Content Validity) เป็นต้น ในที่นี้ผู้วิจัยสนใจที่จะศึกษาเฉพาะเรื่อง ความตรงเชิงพยากรณ์เท่านั้น ซึ่งได้แก่ ความถูกต้องของผลการวัด ในการพยากรณ์ พฤติกรรม หรือสิ่งที่ต้องการวัดที่จะเกิดขึ้นในอนาคต ซึ่งหาได้จากค่าสหสัมพันธ์ระหว่าง คะแนนของแบบสอบนั้น กับผลการสอบของแบบสอบที่ใช้เป็นเกณฑ์ (Ebel, ๑๙๗๐:๔๓๖; Cronbach, ๑๙๔๕:๑๐๓) ตามทฤษฎีของการวัด ความเที่ยงและความตรงเชิงพยากรณ์ มีความสัมพันธ์กัน กล่าวคือ ในกรณีที่เกณฑ์ของการพยากรณ์มีความเที่ยงสูงสุด ค่าความตรง จะมีค่าสูงสุดดังนี้ (Magnuson, ๑๙๖๖:๑๔๔)

$$\begin{aligned} r_{tg(\max)} &= \sqrt{r_{tt}} \\ &= \sqrt{\sigma_T^2 / \sigma_t^2} \\ &= \sqrt{(1 - \sigma_e^2) / \sigma_t^2} \end{aligned}$$

ดังนั้น จากสมการข้างต้นจะเห็นได้ว่า ค่าความตรงเชิงพยากรณ์จะมีค่าได้สูงสุด ไม่เกินค่ารากกำลังสองของค่าความเที่ยง และค่าความตรงนี้มีความสัมพันธ์กับการกระจายของคะแนนการสอบของแบบสอบ และความยาวของแบบสอบเช่นเดียวกับค่าความเที่ยง ดังจะได้กล่าวต่อไป ทั้งนี้ภายใต้เงื่อนไขที่ว่า เกณฑ์ของการพยากรณ์นั้นจะต้องมีค่าความเที่ยงสูงสุด (ค่าความเที่ยงสูงสุด = ๑.๐) แต่หากว่าเกณฑ์ดังกล่าวมีความคลาดเคลื่อน ค่าสูงสุดของค่าความตรงก็จะเปลี่ยนแปลงไปได้ โดยจะมีค่าน้อยลงจากค่าสูงสุดดังกล่าวแล้ว

ส่วนความเที่ยงของแบบสอบอิงกลุ่ม (Norm-referenced Test) มีความหมายได้หลายอย่าง แล้วแต่การตีความและที่มาของสูตรในการคำนวณ เช่น อาจหมายถึง "ดัชนีของการทำนายค่าคะแนนของแบบสอบฉบับหนึ่งจากแบบสอบคู่ขนานอีกชุดหนึ่ง" (Ghiselli, ๑๙๖๔:๒๑๘) หรือ "ความคงที่ของผลการสอบของแบบสอบฉบับใดฉบับหนึ่งที่ต้องการวัดสิ่งใดสิ่งหนึ่ง" (Ebel, ๑๙๗๒:๔๐๙) หรือ "ร้อยละของความแปรปรวนของสิ่งที่แบบสอบนั้นสามารถวัดได้" (Stanley, ๑๙๗๑:๓๗๔) เป็นต้น ปัจจัยที่มีผลต่อค่าความเที่ยงที่สำคัญมี ๒ อย่าง (Magnuson, ๑๙๖๖:๕๙) คือ ๑) ความยาวของแบบสอบ ถ้าแบบสอบมีความยาวเพิ่มขึ้น ค่าความเที่ยงจะเพิ่มขึ้น และ ๒) การกระจายของคะแนน ถ้าคะแนนของแบบสอบยิ่งกระจายมาก แบบสอบก็ยิ่งมีค่าความเที่ยงมาก ทั้งนี้จะสังเกตได้จากสูตรทั่วไปในการคำนวณหาความเที่ยง (Magnuson, ๑๙๖๖:๖๖-๗๒)

$$r_{tt} = 1 - \sigma_e^2 / \sigma_t^2 \dots\dots\dots[1]$$

และ $\sigma_T^2 = (1 - \sigma_e^2) \dots\dots\dots[2]$

$$\sigma_{nT}^2 = n^2 \sigma_T^2 \dots\dots\dots[3]$$

$$\sigma_{ne}^2 = n\sigma_e^2 \dots\dots\dots[4]$$

จากสูตรดังกล่าวแล้วจะสังเกตได้ว่า ค่าความเที่ยง คือ อัตราส่วนระหว่างค่าความแปรปรวนของคะแนนจริง (True Variance : $1-\sigma_e^2$ หรือ σ_T^2) กับค่าความแปรปรวนของคะแนนที่ปรากฏของแบบสอบทั้งฉบับ (σ_t^2) ถ้าข้อสอบยาวขึ้นกว่าเดิม n เท่า ค่าความแปรปรวนของคะแนนจริงเพิ่มขึ้น n^2 เท่า แต่ค่าความแปรปรวนของความคลาดเคลื่อน (σ_e^2) เพิ่มขึ้นเพียง n เท่า นี้แสดงว่า เมื่อแบบสอบมีความยาวเพิ่มขึ้น ความแปรปรวนของคะแนนจริงจะเพิ่มขึ้นมากกว่าค่าความแปรปรวนของความคลาดเคลื่อน จึงมีผลให้แบบสอบมีค่าความเที่ยงเพิ่มขึ้น เช่น แบบสอบชุดหนึ่งมีค่าความเที่ยง (r_{tt}) = ๐.๕๐ แสดงว่า แบบสอบนั้นมีความแปรปรวนของคะแนนจริง และความแปรปรวนของความคลาดเคลื่อนต่างมีค่า = ๐.๕๐ ถ้าหากเราเพิ่มความยาวของแบบสอบนี้เป็น ๒ เท่า ค่าความแปรปรวนของคะแนนจริงจะเป็น ๒.๐๐ ค่าความแปรปรวนของความคลาดเคลื่อนจะเป็น ๑.๐๐ และค่าความแปรปรวนของคะแนนที่ปรากฏทั้งหมดจะเป็น ๒.๐๐ + ๑.๐๐ = ๓.๐๐ ดังนั้น ความเที่ยงของแบบสอบที่มีความยาวเพิ่ม ๒ เท่า จะมีค่า = ๒.๐๐/๓.๐๐ = ๐.๖๗ เป็นต้น (Magnuson, ๑๙๖๖:๗๒)

ดังนั้น จะเห็นได้ว่าความยาวและการกระจายของคะแนนมีผลโดยตรงต่อค่าความเที่ยงของแบบสอบ ถ้าค่าดังกล่าวแล้วเพิ่มขึ้นจะทำให้ค่าความเที่ยงเพิ่มขึ้น

อนึ่ง สำหรับการสอบแบบอิงเกณฑ์ (Criterion-referenced Test) นั้น ความหมายของความเที่ยงแตกต่างจากแนวความคิดของความเที่ยงของแบบสอบอิงกลุ่ม ดังกล่าวมาแล้ว และมีแนวคิดที่ต่างกัน ๓ แนวคิด ดังนี้ (Berk, ๑๙๘๐:๓๒๓-๓๔๖) คือ

๑. ความเที่ยงตามแนวคิดของฟังก์ชันการสูญเสียที่น้อยที่สุด (Threshold Loss Function) หมายถึงความคงที่ของการตัดสินใจในการแบ่งผู้สอบออกเป็น ๒ กลุ่ม คือ กลุ่มผู้รู้แล้ว (mastery) และกลุ่มผู้ยังไม่รู้ (non-mastery) ว่าถูกต้องมากน้อยเพียงใด

๒. ความเที่ยงตามแนวคิดของฟังก์ชันการสูญเสียเกิดจากความคลาดเคลื่อนกำลังสอง (Squared-error Loss Function) หมายถึงความคงที่ของคะแนนสอบของผู้สอบแต่ละบุคคลว่าแตกต่างจากจุดตัด ซึ่งใช้เป็นเกณฑ์ในการสอบว่ามีมากน้อยเพียงใด

สำหรับความเที่ยงของแบบสอบตามแนวคิดที่สองนี้ อาจแบ่งได้ ๒ ประเภท ตามทฤษฎีของการวัด กล่าวคือ ความเที่ยงตามทฤษฎีการวัดแบบเดิม (Classical Testing Theory) และความเที่ยงตามทฤษฎีการอ้างอิงสรูป (Generalizability Theory) ซึ่งเรียกค่าความเที่ยงว่า ดัชนีความเชื่อถือ (Dependability Index) ความเที่ยงอย่างแรกอาศัยแนวคิดแบบสอบคู่ขนานแบบเดิม (Classically Parallel Test Forms) แต่ความเที่ยงแบบที่สอง หรือดัชนีความเชื่อถือนั้น อาศัยแนวคิดแบบสอบคู่ขนานแบบสุ่ม (Randomly Parallel Test Forms) ความเชื่อถือของแบบสอบอิงเกณฑ์วิธีนี้มีค่าเท่ากับค่าความเที่ยงของแบบสอบอิงกลุ่มที่คิดโดยวิธี KR₂₁ (Kuder-Richardson Formula 21) ในกรณีที่จุดตัดของการสอบเท่ากับค่ามัชฌิมเลขคณิตของการสอบ และจะเป็นค่าที่น้อยที่สุด แต่เมื่อความแตกต่างของจุดตัดกับค่าเฉลี่ยของการสอบมีขนาดมากขึ้น หรือน้อยลง ค่าดัชนีความเชื่อถือจะเพิ่มขึ้น (Brennan, ๑๙๘๑:๖๘)

๓. ความเที่ยงตามแนวคิดของการประมาณคะแนนปริเขต (Domain Score) หมายถึงความคงที่ของคะแนนสอบของผู้สอบแต่ละบุคคลที่แตกต่างจากคะแนนปริเขตของสิ่งที่ต้องการสอบว่ามีมากน้อยเพียงใด

สำหรับการวัดและประเมินผลนั้น เป็นที่ยอมรับกันทั่วไปแล้วว่า ความตรงและความเที่ยงเป็นคุณสมบัติที่จำเป็นของแบบสอบ เพราะนักทดสอบต้องการวัดสิ่งที่มุ่งวัดให้ได้มากที่สุด และต้องการให้ผลการวัดมีความคงที่มากที่สุดเท่าที่จะสามารถทำได้ (Ebel, ๑๙๗๒:๔๐๘; Gronlund, ๑๙๗๖:๗๔; Lord, ๑๙๘๐:๕) ดังนั้น ผู้ที่ทำการสอบจึงต้องการให้แบบสอบที่ตนสร้างขึ้น มีค่าทั้งสองอย่างนี้สูงที่สุดเท่าที่จะทำได้ แต่ในขณะเดียวกันก็พยายามให้แบบสอบมีข้อสอบจำนวนเหมาะสม เพราะหากว่าแบบสอบมีความยาวมากเกินไป และใช้เวลาสอบนานเกินไป จะทำให้ผลเสียทางจิตวิทยา เช่น ความเมื่อยล้า เบื่อหน่าย และหงุดหงิด เป็นต้น และอาจทำให้ผลการสอบคลาดเคลื่อนได้ง่าย ดังนั้น ในปัจจุบันนี้มีแนวโน้มว่า นักวัดผลพยายามสร้างแบบสอบให้มีความยาวสั้นลง (Ebel, ๑๙๗๒:๑๐๗) เช่น มีการศึกษาเกี่ยวกับการสอบที่ใช้เวลา และข้อสอบน้อยที่สุด (Stecklein and Others, ๑๙๕๑:๒๐๘) การกำหนดความยาวที่เหมาะสมของแบบสอบ (Lord, ๑๙๘๐:๑๗๓)

การกำหนดจำนวนตัวเลือกที่เหมาะสมของแบบสอบแบบเลือกตอบ (Lord, ๑๙๗๖:๓๓) และการสอบที่เหมาะสมกับรายบุคคล (Tailored Testing : Urry, ๑๙๗๗:๑๘๑ ; Lord, ๑๙๘๐:๑๕๐) เป็นต้น แต่วิธีการต่าง ๆ ดังกล่าวนี ก็ยังต้องการให้แบบสอบมีความตรงและความเที่ยงมากที่สุดเท่าที่จะสามารถทำได้ แต่ให้มีข้อสอบจำนวนน้อย

นอกจากนี้ การให้คะแนนที่เหมาะสมแก่ข้อสอบแต่ละข้อ หรือแต่ละตัวเลือก ก็เป็นอีกแนวทางหนึ่งที่นักทดสอบกำลังศึกษาในปัจจุบันนี้ เพราะตามทฤษฎีการวัดผลนั้น ปัจจัยที่สำคัญอย่างหนึ่งที่ทำให้ค่าความเที่ยงของแบบสอบเพิ่มขึ้น คือ การกระจายของคะแนนการสอบ (Gronlund, ๑๙๗๖:๑๑๘) ในกรณีที่แบบสอบมีจำนวนข้อสอบเท่ากัน แบบสอบที่มีการกระจายของคะแนนมากกว่าย่อมมีค่าความเที่ยงมากกว่าแบบสอบที่มีการกระจายของคะแนนน้อยกว่า ทั้งนี้เพราะว่า โดยทั่วไปแล้ว ค่าความเที่ยง (r_{tt}) หาได้จากสูตร $r_{tt} = 1 - \sigma_e^2 / \sigma_t^2$ (Magnuson, ๑๙๖๖:๗๕) ดังนั้น นักทดสอบเชื่อกันว่า คะแนนที่เหมาะสมสามารถทำให้แบบสอบมีค่าความเที่ยงเพิ่มขึ้นจากค่าดังกล่าวที่คิดจากคะแนนที่คิดวิธีประเพณีนิยม (Conventional Scoring Method) เพราะเชื่อว่าคะแนนดังกล่าวทำให้แบบสอบมีการกระจายของคะแนนการสอบได้มากกว่า และทำให้แบบสอบมีอำนาจจำแนกได้ดีกว่า การให้คะแนนที่คิดวิธีประเพณีนิยม ดังนั้น ในกรณีที่อำนาจจำแนกของแบบสอบบ่งบอกถึงระดับความแตกต่างของความสามารถของผู้สอบ และความสามารถนั้นสัมพันธ์กับเกณฑ์ที่ประสงค์จะทำการพยากรณ์แล้ว แบบสอบที่ให้คะแนนที่เหมาะสมอาจทำให้ค่าความตรงเชิงพยากรณ์เพิ่มขึ้น และสูงกว่าค่าดังกล่าวที่คิดจากคะแนนที่คิดวิธีประเพณีนิยม (Wang and Stanley, ๑๙๗๐:๖๖๓ ; Patnaik and Traub, ๑๙๗๓:๒๘๑)

ดังนั้น ด้วยความมุ่งหมายที่จะทำให้แบบสอบมีค่าความตรงและค่าความเที่ยงเพิ่มขึ้น แต่ความยาวคงที่ จึงทำให้นักทดสอบได้พยายามศึกษาหาวิธีต่าง ๆ ในการให้คะแนนของข้อสอบ เช่น ๑) ให้คะแนนโดยใช้ความเห็นของผู้เชี่ยวชาญ ๒) ใช้ค่ามาตรวัดจากวิธีของ Guttman ๓) ใช้ค่ามาตรวัดจากวิธีการเทียบรายคู่ของ Thurstone (Thurstone's Paired Comparison) ๔) ใช้คะแนนเมื่อหักค่าการเดาแล้ว ๕) ใช้คะแนนตามน้ำหนักความเชื่อมั่นในการตอบ (Confidence Weighting) และวิธีอื่น ๆ อีกมาก แต่วิธีการต่าง ๆ เหล่านี้

แม้ว่าจะมีผู้ศึกษาผลที่มีต่อค่าความเที่ยง และค่าความตรงติดต่อกันมาเรื่อย ๆ ประมาณ ๓๐ ปีมาแล้ว (อ้างจาก Wang and Stanley, ๑๙๗๐:๖๔๑) แต่ผลของการศึกษามักได้ผลไม่คงที่แน่นอน กล่าวคือ การให้คะแนนวิธีเดียวกัน บางครั้งทำให้ค่าความเที่ยงเพิ่มขึ้นอย่างมีนัยสำคัญ แต่บางครั้งก็ได้ผลไม่แตกต่างจากการให้คะแนนวิธีประเพณีนิยม หรือให้ค่าดังกล่าวน้อยกว่าค่าที่ได้จากการให้คะแนนวิธีประเพณีนิยม เป็นต้น นอกจากนี้ การให้คะแนนวิธีต่าง ๆ เหล่านี้ บางครั้งทำให้แบบสอบมีค่าความเที่ยงสูงกว่าค่าความเที่ยงที่ได้จากการให้คะแนนวิธีประเพณีนิยม แต่ให้ค่าความตรงต่ำกว่าค่าดังกล่าวที่ได้จากการให้คะแนนวิธีประเพณีนิยม ดังนั้น การให้คะแนนรายชื่อของแบบสอบโดยวิธีต่าง ๆ ยังให้ผลได้ไม่ดีเท่าที่ควร จนทำให้หนักทดสอบบางท่าน เช่น Guilford, Lovell และ Williams สรุปว่า เมื่อเทียบกับการให้คะแนนวิธีประเพณีนิยมแล้ว การให้คะแนนรายชื่อของแบบสอบวิธีต่าง ๆ ได้ผลไม่คุ้มค่าความยุ่งยากในการคำนวณ (อ้างจาก Wang and Stanley, ๑๙๗๐:๖๔๔)

แต่อย่างไรก็ตาม ปัจจุบันนี้นักทดสอบยังเหลือความหวังอย่างน้อยอีก ๒ วิธีที่จะกำหนดคะแนนของข้อสอบให้ต่างกัน เพื่อที่จะทำให้แบบสอบมีค่าความตรงและความเที่ยงเพิ่มขึ้น (Wang and Stanley, ๑๙๗๐:๖๗๐) คือ ๑) การให้คะแนนโดยอาศัยแนวคิดจากรูปแบบของความสามารถแฝง (Latent Trait Models) และ ๒) การให้คะแนนโดยอาศัยแนวคิดจากการถดถอยพหุคูณ (Multiple Regression)

ดังนั้น ตั้งแต่ปี ๑๙๗๐ เป็นต้นมา การศึกษาเกี่ยวกับการให้คะแนนว่ามีผลต่อค่าความตรงและค่าความเที่ยงมากน้อยเพียงใด มักจะใช้วิธีการต่าง ๆ ที่เกี่ยวข้องกันวิธีทั้งสองดังกล่าวแล้ว แต่ผลการศึกษาก็ยังได้ผลไม่แน่นอนพอที่จะสรุปรวมได้อย่างมั่นใจ แต่ทว่า Lord (Lord, ๑๙๗๕:๑๐) ผู้ซึ่งเป็นนักทดสอบที่มีชื่อเสียงมากในปัจจุบันนี้ได้ให้ข้อคิดเห็นว่า การให้คะแนนที่ต่างจากรีวิธีประเพณีนิยม แม้ว่าค่าความเที่ยงหรือค่าความตรงจะเพิ่มขึ้นเพียงเล็กน้อย เช่น ค่าความเที่ยงเพิ่มจาก .๕๐ เป็น .๕๑ หรือ .๖๐ เป็น .๖๐๓๓ ก็ถือว่าเป็นการพัฒนาแบบสอบที่คุ้มค่า เพราะสามารถลดจำนวนข้อสอบลงได้อย่างน้อยประมาณ ๑ ของจำนวนข้อสอบทั้งหมดเมื่อให้คะแนนวิธีประเพณีนิยม

ด้วยเหตุนี้ ผู้วิจัยจึงมีความสนใจที่จะเปรียบเทียบว่า วิธีการให้คะแนนรายข้อต่างกันจะมีผลต่อค่าความตรง ความเที่ยง และความคงที่ของอันดับที่ของคะแนนการสอบ การอ่านเข้าใจความภาษาอังกฤษมากขึ้นเพียงใด และเปรียบเทียบกับค่าต่าง ๆ ที่ได้จากการให้คะแนนวิธีประเพณีนิยม ดังนั้น วิธีให้คะแนนสำหรับการวิจัยครั้งนี้ จึงมี ๕ วิธี ได้แก่

๑. วิธีใช้ค่าเดลต้า (delta: Δ) ซึ่งเป็นค่าความยากง่ายรายข้อของแบบสอบวิธีหนึ่งที่ได้จากการวิเคราะห์แบบสอบรายข้อวิธีประเพณีนิยม (Classical Item Analysis)

๒. วิธีใช้ค่าเบต้าแปลง (bata: β_c) ซึ่งเป็นค่าความยากง่ายรายข้อของแบบสอบวิธีหนึ่งที่ได้จากการวิเคราะห์แบบสอบรายข้อตามรูปแบบราสช์ (Rasch Model) และแปลงแล้ว

๓. วิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (Optimal Scoring Weight: w) ซึ่งเป็นคะแนนรายข้อที่คำนวณมาจากค่าอำนาจจำแนก (a) ค่าความยากง่าย (p) และค่าการเดา (c) ซึ่งได้จากการวิเคราะห์แบบสอบรายข้อตามรูปแบบโลจิสติก ใช้พารามิเตอร์ ๓ ตัว (Three-parameter Logistic Model)

๔. วิธีใช้ค่าผลบวกของความแปรปรวนร่วม (Communality: h^2) ต่อสิ่งที่มุ่งสอบ ซึ่งได้จากการวิเคราะห์ตัวประกอบ

๕. วิธีประเพณีนิยม ซึ่งเป็นวิธีที่นิยมใช้กันทั่วไป ได้แก่ การให้คะแนนข้อสอบข้อที่ตอบถูก = ๑ คะแนน และให้คะแนนข้อที่ตอบผิด = ๐ คะแนน

จากวิธีการให้คะแนนทั้ง ๕ วิธีดังกล่าวแล้ว จะเห็นได้ว่า ผู้วิจัยเลือกใช้วิธีการให้คะแนนตามแนวคิดจากรูปแบบของความสามารถแฝง (คือวิธีที่ ๒ และ ๓) และตามแนวคิดจากการถดถอยพหุคูณ (คือวิธีที่ ๔) ซึ่งเป็นวิธีที่นักทดสอบคิดว่า จะทำให้แบบสอบที่ให้คะแนนตามแนวคิดดังกล่าวมีค่าความตรงและความเที่ยงเพิ่มขึ้นได้ดังได้กล่าวมาแล้ว ผู้วิจัยได้เลือกวิธีการให้คะแนนโดยอาศัยค่าความยากง่ายแบบ Δ มาศึกษาเปรียบเทียบกับ เพื่อ

ศึกษาความเป็นไปได้ในการนำผลการวิจัยไปประยุกต์ใช้ในภายหลัง และเลือกการให้คะแนนวิธีประเพณีนิยม เพื่อใช้เป็นมาตรฐานในการเปรียบเทียบค่าต่าง ๆ ดังกล่าว ส่วนเหตุผลเชิงทฤษฎีที่ผู้วิจัยเลือกวิธีการให้คะแนนรายข้อต่างกันทั้ง ๔ วิธี เพื่อการวิจัยครั้งนี้ มีดังนี้

๑. ค่า Δ เป็นค่าความยากง่ายของข้อสอบที่ได้จากการวิเคราะห์แบบสอบรายข้อวิธีประเพณีนิยม ค่า Δ นี้เป็นการแปลงค่าดัชนีความยากง่าย (p) ให้เป็นคะแนนมาตรฐาน โดยอาศัยพื้นที่ใต้โค้งปกติช่วยในการกำหนดค่า (Henrysson, ๑๙๗๑:๑๓๔) คือ

$$\Delta_i = 13 + 4z_i$$

ในเมื่อ z_i = ค่าความเบี่ยงเบนมาตรฐาน ณ จุดที่บอกอัตราส่วนของผู้ที่ตอบถูกและผิดของข้อสอบข้อที่ i

ข้อสอบยิ่งยากค่า Δ ยิ่งมีค่ามาก ดังนั้น ผู้วิจัยจึงเห็นว่าเมื่อความสามารถในการอ่านเข้าใจความมีลำดับชั้นจากง่ายที่สุดไปยังยากที่สุด ค่า Δ น่าจะใช้เป็นค่าน้ำหนักคะแนนของข้อสอบได้ด้วย นอกจากนี้ ปกติแล้ว z มีค่าระหว่าง ๓ ถึง -๓ ดังนั้น Δ มีค่าระหว่าง ๑ ถึง ๒๕ ซึ่งเป็นพิสัยที่ค่อนข้างกว้าง ด้วยเหตุนี้ ผู้วิจัยคาดว่าน้ำหนักคะแนนรายข้อที่ใช้ Δ นี้ คงจะทำให้คะแนนการสอบมีการกระจายมากกว่าคะแนนที่ได้จากการให้วิธีประเพณีนิยม ซึ่งมีค่าน้ำหนักคะแนน ๐ และ ๑ เท่านั้น จึงคาดว่า การให้คะแนนโดยวิธีใช้ค่า Δ จะทำให้แบบสอบมีค่าความตรงและความเที่ยงมากกว่าค่าดังกล่าวที่คิดจากคะแนนวิธีประเพณีนิยม ทั้งนี้เพราะว่า การกระจายของคะแนนเป็นปัจจัยที่สำคัญที่มีผลต่อค่าความตรงและความเที่ยงดังได้กล่าวมาแล้วข้างต้น

๒. แนวคิดของรูปแบบบราสซ์เชื่อว่า โอกาสที่ผู้สอบจะทำข้อสอบใด ๆ ได้ถูกต้องหรือไม่ ขึ้นอยู่กับความแตกต่างของความสามารถของผู้สอบ (θ) และความยากง่ายของข้อสอบ (β) เท่านั้น ไม่มีค่าอำนาจจำแนกหรือค่าการเดาเกี่ยวข้องกับ (Wright and Stone, ๑๙๗๔:๑๕) ดังนั้น ผู้วิจัยคาดว่า ข้อสอบที่มีความยากมากน่าจะมีความแม่นยำด้วย

เมื่อคำนึงว่าความเข้าใจในการอ่านมีลำดับชั้นจากง่ายที่สุดไปหายากที่สุดเช่นเดียวกัน ค่า β คำนวณจากสูตรดังนี้ (Wright and Stone, ๑๙๗๕:๒๕)

$$\beta_i = \bar{X} + \sqrt{(1+\sigma^2/1.7^2)} \ln[(1-p_i)/p_i]$$

ในเมื่อ \bar{X} = ค่ามัชฌิมเลขคณิตของคะแนนการสอบ

1.7 = ค่าคงที่เพื่อการปรับส่วนโค้งให้เป็น Logistic Model

σ^2 = ความแปรปรวนของคะแนนการสอบ

p_i = ความยากง่ายของข้อสอบข้อที่ i

แต่อย่างไรก็ตาม เนื่องจากค่า β ปกติแล้วมีค่าระหว่าง -๒.๕ ถึง + ๒.๕ (Ree, ๑๙๗๕:๓๗๒; Warm, ๑๙๗๕:๕๒) เพื่อจำกัดค่าติดลบ และเพื่อให้สอดคล้องกับแนวคิดของค่า Δ ผู้วิจัยจึงแปลงค่า β เป็นคะแนนมาตรฐาน โดยใช้สูตร

$$\beta_{t_i} = 13 + 4\beta_i$$

ดังนั้น โดยทั่ว ๆ ไปแล้ว ค่า β_{t_i} จะมีค่าระหว่าง ๓ ถึง ๒๓ ซึ่งเป็นพิสัยค่อนข้างกว้าง แต่ก็แคบกว่าพิสัยของค่า Δ ดังกล่าวแล้ว และมีพิสัยกว้างกว่าค่าพิสัยของคะแนนที่ให้วิธีประเมินมาก จึงคาดว่าจะทำให้แบบสอบมีค่าความตรงและความเที่ยงมากกว่าคะแนนที่ให้วิธีประเมินได้ด้วย

๓. จากการศึกษาของ Urry (Urry, ๑๙๗๗:๑๘๔) พบว่า รูปแบบวิธีโลจิสติกใช้พารามิเตอร์ ๓ ตัว มีความเหมาะสมกับแบบสอบแบบเลือกตอบมากที่สุด แม้ว่าแบบสอบจะมีค่าอำนาจจำแนกคงที่ทุกข้อ หรือค่าการเดาจะเป็นศูนย์ก็ตาม และพบว่าค่าดังกล่าวทั้งสองนี้ และค่าความยากง่ายเกี่ยวข้องกับความสามารถของผู้ตอบแบบสอบแบบเลือกตอบ ค่าความสัมพันธ์กันของค่าพารามิเตอร์ทั้งสาม เรียกว่าค่านำหนักคะแนนที่เหมาะสม (w) ซึ่งคำนวณจากสูตรดังนี้ (Lord, ๑๙๘๐:๗๖)

$$w_i = \frac{a_i \cdot p_i - c_i}{1 - c_i \cdot p_i}$$

ในเมื่อ a_i = ค่าอำนาจจำแนกของข้อสอบข้อที่ i จากการวิเคราะห์ข้อสอบ
ตามรูปแบบโลจิสติกใช้พารามิเตอร์ ๓ ตัว

c_i = ค่าการเดาของข้อสอบข้อที่ i

p_i = ความยากง่ายของข้อสอบข้อที่ i

ผู้วิจัยมีความเห็นว่า ค่า w นี้เป็นแนวคิดหนึ่งที่เหมาะสมสำหรับนำมาใช้เป็นค่า
น้ำหนักคะแนนรายข้อของข้อสอบได้ด้วยวิธีหนึ่ง เพราะว่าเป็นค่าที่คำนึงถึงความสัมพันธ์ของ
ค่าพารามิเตอร์ต่าง ๆ ที่สำคัญทั้งหมดเป็นรายข้อ ซึ่งถือได้ว่าเป็นวิธีที่ละเอียดมาก และที่
สำคัญยิ่งก็คือ เป็นวิธีที่คำนึงถึงค่าการเดา ซึ่งเป็นแหล่งของความคลาดเคลื่อนของการวัดที่
สำคัญมากแหล่งหนึ่ง (Magnuson, ๑๙๖๖:๑๐๑) ปกติแล้ว w มีค่าตั้งแต่ .๐๐ ถึง ๑.๕
ดังนั้น พิสัยของค่า w จะกว้างมากกว่าพิสัยของการให้คะแนนวิธีใช้ค่า Δ และวิธีใช้ค่า β_t
ดังนั้น เมื่อคำนึงถึงการคิดค่าการเดา และค่าพารามิเตอร์อื่นรวมทั้งพิสัยของค่าน้ำหนักคะแนน
ซึ่งคาดว่าจะทำให้คะแนนของการสอบมีการกระจายมากกว่าวิธีอื่น ๆ ที่กล่าวมาแล้ว จึงเป็น
ที่คาดว่าจะการกำหนดค่าน้ำหนักคะแนนวิธีนี้จะทำให้แบบสอบมีค่าความตรงและความเที่ยงสูงที่
สุดด้วย

๔. จากการศึกษาของ Wang and Stanley (Wang and Stanley,
๑๙๗๐:๖๕๐) เกี่ยวกับบรรณคดีที่เกี่ยวข้องกับการให้คะแนนข้อสอบหลายรูปแบบ พบว่านักทดสอบ
มีความหวังว่า การให้คะแนนข้อสอบโดยอาศัยแนวคิดจากทฤษฎีความสามารถแฝง (Latent
Trait Theory) และการถดถอยพหุคูณ เป็นวิธีที่อาจทำให้คุณภาพของแบบทดสอบดีขึ้น ผู้วิจัย
จึงเลือกใช้ค่าผลบวกของความแปรปรวนร่วม (h^2) จากวิธีวิเคราะห์ตัวประกอบ ซึ่งมีความ
หมายและค่าเท่ากับค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณยกกำลังสอง (Kerlinger and Pedhazur,
๑๙๗๓:๓๖๓) มาเป็นคะแนนของข้อสอบ ทั้งนี้เพราะว่าในกรณีของการวิเคราะห์ข้อสอบรายข้อ
นั้น วิธีการถดถอยพหุคูณ หรือวิธีอื่นที่มีวิธีการคำนวณในทำนองคล้ายคลึงกัน เช่น การหาค่า

สหสัมพันธ์แบบพอยท์ไบซีเรียล และแบบไบซีเรียล (Point Biserial และ Biserial Correlations) เป็นวิธีที่ให้ค่าไม่ถูกต้องตามความเป็นจริง เนื่องจากความสัมพันธ์ของคะแนนรายข้อกับคะแนนรวมมีคะแนนของรายข้อนั้นรวมอยู่ด้วย จึงเป็นตัวแปรตามที่ไม่ถูกต้อง กลุ่มตัวอย่างยิ่งมีขนาดเล็ก ค่าดังกล่าวยิ่งมีความผิดพลาดมาก (Henrysson, ๑๙๗๑: ๑๔๙) ค่า h^2 คำนวณมาจากสูตรต่อไปนี้ (Kerlinger and Pedhazur, ๑๙๗๓: ๓๖๓)

$$h_i^2 = \sum a_j^2$$

ในเมื่อ a_j = ค่าน้ำหนักของตัวประกอบ (factor loading) แต่ละตัวประกอบ หรือคือค่าของความแปรปรวนที่ข้อสอบแต่ละข้อมีส่วนรวม (หรือความสัมพันธ์) กับตัวประกอบตัวที่ j

นอกจากนี้ ค่า h^2 เป็นค่าผลบวกของความแปรปรวนของข้อสอบแต่ละข้อที่มีส่วนรวม หรือความสัมพันธ์กับสิ่งที่มีงวัด (ตัวประกอบ) ข้อสอบข้อใดที่มีความสัมพันธ์กับสิ่งที่มีงวัดมาก จะมีค่า h^2 มากกว่าข้อที่มีความสัมพันธ์น้อยกว่า ด้วยเหตุนี้ ผู้วิจัยจึงเห็นว่าข้อสอบข้อที่สัมพันธ์กับสิ่งที่มีงวัดมากก็ควรมีคะแนนมากกว่าข้อที่มีความสัมพันธ์กับสิ่งที่มีงวัดน้อยกว่า โดยทฤษฎีแล้ว h^2 มีค่าระหว่าง .๐๐ ถึง ๑.๐๐ ดังนั้น พิสัยของค่า h^2 จึงกว้างมาก เป็นอันดับที่ ๒ รองจากพิสัยของค่า w และกว้างกว่าพิสัยของคะแนนที่ใช้วิธีประเพณี ดังนั้น จึงคาดว่าทำให้คะแนนวิธีใช้ค่า h^2 นี้ จะทำให้ค่าความตรงและความเที่ยงของแบบสอบสูงกว่าค่าดังกล่าวที่ได้จากการให้คะแนนวิธีประเพณี และคาดว่าค่าความตรงและความเที่ยงที่ได้จะน้อยกว่าค่าดังกล่าวที่ได้จากการให้คะแนนวิธีใช้ค่า w เพราะค่าพิสัยของคะแนนแคบกว่า

นอกจากนี้ ผู้วิจัยยังทำการศึกษาเกี่ยวกับโครงสร้างความรู้ (Schema) ด้วยว่ามีอิทธิพลต่อการอ่านเข้าใจความภาษาอังกฤษหรือไม่ และมาน้อยเพียงใด ทั้งนี้เพื่อประโยชน์ในด้านการเตรียมการเรียนการสอนภาษาต่อไป

ในปัจจุบันนี้มีความเชื่อกันว่า การอ่านเป็นกระบวนการทางจิตสังคมภาษาศาสตร์ ที่มีบูรณาการผสมกลมกลืนกัน และเป็นที่ยอมรับกันทั่วไปว่า ความสามารถในการอ่านเข้าใจ ความ (Reading Comprehension) นั้น ไม่ได้ขึ้นอยู่กับความสามารถของผู้อ่านด้าน ภาษาศาสตร์เท่านั้น แต่ยังขึ้นอยู่กับความสามารถทั่ว ๆ ไปในด้านโลกทัศน์ และความสามารถ ของสมองในการจัดกระทำกับความรู้และความสามารถดังกล่าวด้วย (Carrell, ๑๙๘๓: ๑๙๓) และมีนักวิจัยจำนวนมาก เช่น Rumelhart (๑๙๗๕) Thorndyke (๑๙๗๗) Kintsch และ Van Dijk (๑๙๗๘) และ Carrell (๑๙๘๓, ๑๙๘๔b) เป็นต้น ที่ศึกษา พบว่า ผู้อ่านยิ่งมีพื้นฐานความรู้เกี่ยวกับโครงสร้างความรู้ของเรื่องที่อ่านมากเท่าใด ก็ยิ่ง สามารถอ่านเข้าใจเรื่องที่อ่านได้มากยิ่งขึ้น ทั้งนี้เพราะเชื่อว่า โครงสร้างความรู้ของ เรื่องสะท้อนให้เห็นโครงสร้างความรู้ของผู้เขียน และโครงสร้างความรู้ของเรื่อง มีลักษณะ แทนโครงสร้างของความจำของผู้อ่าน ซึ่งเป็นอิสระจากความซับซ้อนของโครงสร้างผิว (Surface Structure) ของเรื่องที่อ่าน (อ้างอิงจาก C. Daungploy, ๑๙๘๖:๒๔) ดังนั้น ถ้าผู้อ่านมีโครงสร้างความรู้สอดคล้องกับโครงสร้างความรู้ของผู้เขียน หรือของ เรื่องที่อ่านแล้ว จะทำให้เข้าใจเรื่องที่อ่านได้ง่ายขึ้น แต่หากว่าโครงสร้างทั้งสองแตกต่างกัน การอ่านก็จะมีปัญหา

โดยทั่ว ๆ ไปการเรียนการสอนภาษานั้น สิ่งที่มีักทำให้เกิดโครงสร้างความรู้ เกี่ยวกับเรื่องที่อ่าน ได้แก่ ชื่อเรื่อง ภาพประกอบเรื่อง คำศัพท์ที่มีความหมายสำคัญใน เรื่อง และความคุ้นเคยกับเนื้อหาสาระของเรื่องที่อ่าน Rumelhart (Rumelhart, ๑๙๘๐: ๔๑) กล่าวว่า ชื่อเรื่อง ภาพประกอบเรื่อง หรือความคุ้นเคยกับเนื้อหาสาระของเรื่องที่อ่าน กระตุ้นให้ผู้อ่านเกิดมีโครงสร้างความรู้ย่อย (Subschema) ที่จะทำให้ผู้อ่านเข้าใจเรื่องได้ ในลักษณะ "เบื้องสูง-สู่-เบื้องต่ำ" (top-down processing) กล่าวคือ ชื่อเรื่อง หรือ ภาพประกอบเรื่อง กระตุ้นใจผู้อ่านเกิดมีมโนทัศน์เกี่ยวกับเรื่องที่อ่านอย่างหยาบ ๆ แล้วจะ ใช้สารสนเทศที่ได้จากการอ่านตรวจสอบความถูกต้องของมโนทัศน์นั้นในระหว่างการอ่าน จนกว่าจะจบเรื่องที่อ่านนั้น ทำให้อาจเรียกการอ่านเข้าใจความลักษณะนี้ว่า เป็นกระบวนการ กระตุ้นเชิงมโนทัศน์ (Conceptually-driven Processing) ส่วนคำศัพท์ที่มีความ

หมายสำคัญต่อเรื่องที่ทำนั้น กระตุ้นให้ผู้อ่านเข้าใจเรื่องที่ทำได้ในลักษณะ "เบื้องต่ำ-สู่-เบื้องสูง" (Bottom-up Processing) กล่าวคือ คำศัพท์ที่สำคัญแต่ละคำจะกระตุ้นให้ผู้อ่านเกิดมีโครงสร้างความรู้ย่อยในระหว่างการอ่าน ผู้อ่านจะสังเคราะห์โครงสร้างย่อยต่าง ๆ ที่เกิดขึ้นในระหว่างการอ่าน เพื่อสรุปเป็นความเข้าใจในการอ่านของตนได้ในที่สุด จึงทำให้อาจเรียกการอ่านเข้าใจความลักษณะนี้ว่า เป็นกระบวนการกระตุ้นเชิงข้อมูล (Data-driven Processing)

จากการศึกษาของ Carrell (Carrell, ๑๙๘๓:๑๘๓-๒๐๖; Carrell, ๑๙๘๖:๘๗-๑๐๙) พบว่า ชื่อเรื่อง ภาพประกอบเรื่อง ความคุ้นเคยกับเรื่องที่ทำ และคำศัพท์ที่สำคัญในเรื่อง มีผลต่อการอ่านเข้าใจความภาษาอังกฤษของผู้ที่เป็นเจ้าของภาษาอย่างมีนัยสำคัญ คำศัพท์ที่มีความหมายสำคัญในเรื่อง มีผลต่อการอ่านเข้าใจความของผู้ที่ไม่ใช่เจ้าของภาษาอังกฤษ ที่มีความสามารถในการอ่านในระดับสูงกว่าปานกลางเล็กน้อยอย่างมีนัยสำคัญ และความคุ้นเคยกับเรื่องที่ทำมีผลต่อการอ่านเข้าใจความของผู้ที่ไม่ใช่เจ้าของภาษาอังกฤษที่มีความสามารถในการอ่านในระดับสูงอย่างมีนัยสำคัญ ส่วนชื่อเรื่อง และภาพประกอบเรื่อง ไม่มีผลต่อการอ่านเข้าใจความของผู้ที่ไม่ใช่เจ้าของภาษาอังกฤษที่มีความสามารถในการอ่านในระดับสูงกว่าปานกลางเล็กน้อย และในระดับสูงอย่างมีนัยสำคัญ

อนึ่ง จากการศึกษาของนักทดสอบทางภาษาจำนวนมาก ด้วยการตรวจสอบหลาย ๆ วิธี เช่น การวิเคราะห์กระบวนการตอบ (Scalogram Analysis) การวิเคราะห์วิถีสัมพันธ์ (Path Analysis) ทฤษฎีลำดับชั้น (Ordering Theory) และวิธีอื่น ๆ อีก เป็นที่สรุปค่อนข้างแน่ว่า การอ่านเข้าใจความมีลำดับชั้น (Hierarchies) ของความเข้าใจ (McKillop, ๑๙๕๓; Fry, ๑๙๖๓; Gagne, ๑๙๗๐:๒๖๔; Airasion, ๑๙๗๕:๑๖๓; Bergan, ๑๙๘๐:๖๒๔) ซึ่งหมายความว่า การอ่านมีระดับของการเข้าใจความหลายระดับ และการที่จะเข้าใจเรื่องลึกซึ้งยิ่งขึ้น ผู้อ่านจะต้องเข้าใจส่วนที่เป็นพื้นฐานง่าย ๆ ของเรื่องก่อน ดังนั้น นักทดสอบทางภาษาที่มีชื่อเสียงบางท่าน เช่น Valette, Disick และ Barrette (Valette and Disick, ๑๙๗๒:๑๖๑; Melnik and Merritt, ๑๙๗๔:๕๖) จึงได้จัดทำสสารระบบ (Taxonomy) เกี่ยวกับการอ่านเข้าใจความขึ้น

สำหรับการศึกษาค้างนี้ ปัญหาที่ผู้วิจัยสนใจศึกษาก็คือ การเปรียบเทียบผลของการให้คะแนนต่างกันหาวิธีต่อค่าความตรงเชิงพยากรณ์ ความเที่ยง และความคงที่ของอันดับที่ของคะแนนการสอบของแบบสอบการอ่านเข้าใจความภาษาอังกฤษ ซึ่งมีโครงสร้างความรู้แตกต่างกัน ในกรณีที่มีการสอบเป็นแบบอิงกลุ่มและแบบอิงเกณฑ์

วัตถุประสงค์ของการวิจัย

การวิจัยครั้งนี้ มีวัตถุประสงค์ดังต่อไปนี้

๑. เพื่อเปรียบเทียบว่า วิธีการให้คะแนนข้อสอบแต่ละข้อ ๕ วิธีดังกล่าวแล้ว คือ ๑) วิธีใช้ค่าเดลต้า (Δ) ๒) วิธีใช้ค่าเบต้าแปลง (β_t) ๓) วิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) ๔) วิธีใช้ค่าผลบวกของความแปรปรวนร่วม (h^2) และ ๕) วิธีประเพณีนิยม ทำให้แบบสอบการอ่านเข้าใจความภาษาอังกฤษมีค่าความตรงเชิงพยากรณ์แตกต่างกันอย่างมีนัยสำคัญหรือไม่ ในกรณีที่เรื่องทีอ่านมีโครงสร้างความรู้ต่างกัน และเป็นแบบสอบแบบอิงกลุ่ม

๒. เพื่อเปรียบเทียบว่า วิธีการให้คะแนนข้อสอบแต่ละข้อ ๕ วิธีดังกล่าวแล้ว ทำให้แบบสอบการอ่านเข้าใจความภาษาอังกฤษมีค่าความเที่ยงแตกต่างกันอย่างมีนัยสำคัญหรือไม่ ในกรณีที่เรื่องในการอ่านมีโครงสร้างความรู้ต่างกัน และเป็นแบบสอบอิงกลุ่ม

๓. เพื่อเปรียบเทียบว่า วิธีการให้คะแนนทั้ง ๕ วิธีนั้น ทำให้ค่าความเที่ยงของแบบสอบแบบอิงเกณฑ์ (ค่าดัชนีความเชื่อถือ) แตกต่างกันหรือไม่ ในกรณีที่เรื่องทีอ่านมีโครงสร้างความรู้ต่างกัน

๔. เพื่อศึกษาว่า วิธีการให้คะแนนข้อสอบแต่ละข้อต่างกัน ๕ วิธีนั้น ทำให้อันดับที่ของคะแนนการสอบแบบสอบการอ่านเข้าใจความภาษาอังกฤษคงที่หรือไม่ และมากน้อยเพียงใด เมื่อเปรียบเทียบกับอันดับที่ของคะแนนจากการให้คะแนนวิธีประเพณีนิยม ในกรณีที่เรื่องทีอ่านมีโครงสร้างความรู้ต่างกัน

๕. เพื่อศึกษาว่า คะแนนที่ได้จากวิธีการให้คะแนนทั้ง ๕ วิธี มีความสัมพันธ์ซึ่งกันและกันมากน้อยเพียงใด

๖. เพื่อศึกษาว่า โครงสร้างความรู้ที่เกิดจากข้อเรื่องและภาพประกอบเรื่องซึ่งมีคำศัพท์ที่มีความสำคัญต่อเรื่องที่อ่าน และความคุ้นเคยกับเนื้อหาสาระของเรื่องที่อ่าน มีผลต่อความสามารถทั่วไปในการอ่านเข้าใจความภาษาอังกฤษของนิสิต ของจุฬาลงกรณ์มหาวิทยาลัยที่เรียนภาษาอังกฤษเป็นภาษาต่างประเทศมากน้อยเพียงใด

๗. เพื่อศึกษาว่า ความคุ้นเคยต่อเรื่องที่อ่าน และความยากง่ายของเรื่อง มีความสัมพันธ์กันมากน้อยเพียงใด

ตัวแปรในการวิจัย

ก. ตัวแปรต้น มี ๒ จำพวก ได้แก่

๑. วิธีการให้คะแนน ๕ วิธี คือ

๑.๑ วิธีใช้ค่าเดลต้า (Δ) ซึ่งเป็นค่าความยากง่ายที่ได้จากการวิเคราะห์ข้อสอบรายข้อวิธีประเพณีนิยม เป็นค่าน้ำหนักคะแนนแต่ละข้อ

๑.๒ วิธีใช้ค่าเบต้าแปลง (β_L) ซึ่งเป็นค่าระดับความยากง่ายที่ได้จากการวิเคราะห์ข้อสอบตามรูปแบบบราสซ์และแปลงแล้ว เป็นค่าน้ำหนักคะแนนแต่ละข้อ

๑.๓ วิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) ซึ่งเป็นค่าน้ำหนักคะแนนที่เหมาะสม ที่คำนวณจากค่าพารามิเตอร์ต่าง ๆ ที่ได้จากการวิเคราะห์ข้อสอบตามรูปแบบโลจิสติกใช้พารามิเตอร์ ๓ ตัว เป็นค่าน้ำหนักคะแนนแต่ละข้อ

๑.๔ วิธีใช้ค่าผลบวกความแปรปรวนร่วม (h^2) ซึ่งเป็นค่าผลบวกของความแปรปรวนร่วมต่อสิ่งที่มีงทดสอบ ซึ่งได้จากการวิเคราะห์ตัวประกอบ เป็นค่าน้ำหนักคะแนนรายข้อ

๑.๕ วิธีประเพณีนิยม ซึ่งให้คะแนนข้อที่ตอบถูก = ๑ คะแนน และข้อที่ตอบผิด = ๐ คะแนน

๒. โครงสร้างความรู้ในการอ่านที่แตกต่างกัน ๔ อย่าง คือ

๒.๑ ชื่อเรื่องและภาพประกอบเรื่อง ซึ่งมีคำศัพท์ที่มีความหมายสำคัญต่อเรื่อง และการที่ไม่มีชื่อเรื่องและภาพประกอบเรื่อง ซึ่งมีคำศัพท์ที่มีความหมายสำคัญต่อเรื่อง (Context and Transparency VS No-Context and Transparency)

๒.๒ ความคุ้นเคยกับเนื้อหาสาระของเรื่องที่ย่านว่า คุ้นเคยกับเรื่องใดมากกว่ากัน (Familiarity VS Novelty)

ข. ตัวแปรตาม ได้แก่

๑. ความสามารถทั่วไปในการอ่าน เข้าใจความภาษาอังกฤษ
๒. ความตรงเชิงพยากรณ์ของแบบสอบอิงกลุ่ม
๓. ความเที่ยงของแบบสอบอิงกลุ่ม
๔. ความเที่ยงของแบบสอบอิง เกณฑ์
๕. อันดับที่ยกคะแนนสอบ

สมมุติฐานในการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยมีสมมุติฐานเพื่อกำเนินการวิจัยดังนี้

๑. ในกรณีการสอบแบบอิงกลุ่ม แบบสอบที่ให้คะแนนรายข้อโดยใช้วิธีค่าน้ำหนักคะแนนที่เหมาะสม (w) มีค่าความตรงเชิงพยากรณ์สูงกว่าค่าความตรงเชิงพยากรณ์ของแบบสอบที่ให้คะแนนวิธีประเพณีนิยม และสูงกว่าค่าความตรงเชิงพยากรณ์ของแบบสอบที่ให้คะแนนโดยวิธีอื่น ๆ อย่างมีนัยสำคัญ ไม่ว่าจะแบบสอบนั้นจะมีโครงสร้างความรู้แบบใดก็ตาม

๒. ในกรณีการสอบแบบอิงกลุ่ม แบบสอบที่ให้คะแนนรายข้อโดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) มีค่าความเที่ยงสูงกว่าค่าความเที่ยงของแบบสอบที่ให้คะแนนวิธีประเพณีนิยม และสูงกว่าค่าความเที่ยงของแบบสอบที่ให้คะแนนโดยวิธีอื่น ๆ อย่างมีนัยสำคัญ ไม่ว่าจะแบบสอบนั้นจะมีโครงสร้างความรู้แบบใดก็ตาม

๓. ในกรณีการสอบแบบอิง เกณฑ์ แบบสอบที่ให้คะแนนรายข้อโดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (w) มีค่าความเที่ยง (ค่าดัชนีความเชื่อถือ) สูงกว่าค่าความเที่ยงที่คำนวณจากคะแนนโดยวิธีอื่น ๆ ไม่ว่าแบบสอบนั้นจะมีโครงสร้างความรู้แบบใดก็ตาม

๔. การให้คะแนนต่างกันแต่ละวิธีทำให้อันดับที่ของคะแนนการสอบแบบสอบการอ่าน เข้าใจความภาษาอังกฤษเปลี่ยนไปจากอันดับที่ของคะแนนที่ให้โดยวิธีประเพณีนิยม

๕. คะแนนที่ได้จากรีการให้คะแนนรายข้อทั้ง ๕ วิธี ต่างมีความสัมพันธ์ซึ่งกันและกันอย่างมีนัยสำคัญ

๖. โครงสร้างความรู้ที่เกิดจากชื่อ เรื่องและภาพประกอบเรื่อง ซึ่งมีคำศัพท์ที่มีความหมายสำคัญต่อเรื่องที่อ่าน และความคุ้นเคยกับเนื้อหาสาระของเรื่องที่อ่านต่างมีผลต่อความสามารถทั่วไปในการอ่าน เข้าใจความภาษาอังกฤษของนิสิตจุฬาลงกรณ์มหาวิทยาลัย ที่เรียนภาษาอังกฤษเป็นภาษาต่างประเทศอย่างมีนัยสำคัญ

๗. ความคุ้นเคยต่อเรื่องที่อ่านและความยากง่ายของเรื่อง มีความสัมพันธ์กันอย่างมีนัยสำคัญ

ขอบเขตของการวิจัย

การวิจัยเรื่องนี้ ผู้วิจัยกำหนดขอบเขตของการวิจัยไว้ดังนี้

๑. ข้อความเพื่อการอ่านเข้าใจความภาษาอังกฤษ มีความยาวประมาณ ๓๐๐ คำ และเป็นเรื่องทั่ว ๆ ไปทางสังคมศาสตร์และวิทยาศาสตร์ ที่มีความยากง่าย (Readability) ระดับใช้เพื่อการเรียนการสอนได้และเหมาะสมกับนิสิตชั้นปีที่ ๑ ในมหาวิทยาลัยเท่านั้น

๒. การอ่านเข้าใจความ มุ่งเน้นที่การอ่านไม่ออกเสียงเท่านั้น

๓. พลวิชัย (Subjects) ที่ใช้เพื่อการวิจัยครั้งนี้ ได้แก่ นิลิตชั้นปีที่ ๑ ของจุฬาลงกรณ์มหาวิทยาลัย ที่ลงทะเบียนเรียนวิชาภาษาอังกฤษพื้นฐานเฉพาะการอ่าน II (Foundation English : Reading II) ภาคปลายปีการศึกษา ๒๕๒๗ เท่านั้น

ข้อตกลงเบื้องต้น

การวิจัยครั้งนี้ ผู้วิจัยถือว่า

๑. ความสามารถในการอ่านเข้าใจความภาษาอังกฤษมีลำดับขั้น ผู้ที่อ่านข้อความได้เข้าใจมาก ย่อมมีและใช้ความสามารถเพื่อทำความเข้าใจงานในการอ่านมาก ส่วนผู้ที่อ่านข้อความเดียวกันแล้ว เข้าใจได้น้อยกว่า แสดงว่ามีและ/หรือใช้ความสามารถเพื่อการอ่านน้อยกว่า

๒. ระดับความสัมพันธ์ของข้อสอบกับสิ่งที่มุ่งสอบ เป็นปัจจัยหนึ่งที่แสดงถึงระดับของความเข้าใจในเนื้อหาสาระของสิ่งที่อ่านได้ด้วย ดังนั้น ข้อที่มีค่า h^2 สูง จึงมีความสำคัญ และมีค่าน้ำหนักคะแนนมากกว่าข้อที่มีค่าดังกล่าวต่ำกว่า

๓. การอ่านเข้าใจความเป็นกระบวนการที่เกี่ยวข้องผสมผสานกลมกลืนกันทางจิตสังคมภาษาศาสตร์

ความจำกัดของการวิจัย

แม้ว่าแบบสอบของการวิจัยครั้งนี้ จะเป็นส่วนหนึ่งของแบบสอบเพื่อการวัดผลจริงประจำภาคของกลุ่มตัวอย่าง แต่งานวิจัยนี้อาจคลาดเคลื่อนบ้าง เพราะอาจมีนิสิตบางคนทำแบบสอบไม่เต็มความสามารถ เนื่องจากเหตุผลทางจิตวิทยา เช่น ความเมื่อยล้า ความวิตกกังวล ความเจ็บป่วย หรือเพราะค่าน้ำหนักคะแนนในการสอบที่ไม่มากนัก (๑๐%) เป็นต้น

คำจำกัดความที่ใช้ในการศึกษาค้นคว้า

นิสิต หมายถึง นิลิตชั้นปีที่ ๑ ของจุฬาลงกรณ์มหาวิทยาลัย ที่ลงทะเบียนเรียนรายวิชาภาษาอังกฤษพื้นฐานเฉพาะการอ่าน II ภาคปลายปีการศึกษา ๒๕๒๗

ความสามารถทั่วไปในการอ่าน หมายถึง คะแนนจากแบบสอบแบบ เลือกตอบ ซึ่งผู้วิจัยสร้างขึ้นเป็นแบบสอบที่มีความยาว ๕๐ ข้อ เพื่อใช้วัดความรู้ความสามารถทั่วไปในการอ่านเข้าใจความภาษาอังกฤษ

แบบสอบอิงกลุ่ม หมายถึง แบบสอบวัดความสามารถทั่วไปในการอ่านที่ผู้วิจัยสร้างขึ้น เมื่อนำผลของการสอบของผู้สอบแต่ละคนมาตัดสิน โดยการเปรียบเทียบกับผลการสอบของผู้อื่นในกลุ่ม เป็น เกณฑ์

แบบสอบอิงเกณฑ์ หมายถึง แบบสอบวัดความสามารถทั่วไปในการอ่านที่ผู้วิจัยสร้างขึ้น เมื่อนำผลของการสอบของผู้สอบแต่ละคนมาตัดสิน โดยการใช้จุดตัดที่กำหนดขึ้น เป็นเกณฑ์สำคัญในการแปลความหมาย (interpretation) ของคะแนนการสอบ (Brennan, ๑๙๘๔: ๓๒๔) ซึ่งเป็นคุณลักษณะ เฉพาะของแบบสอบชนิดนี้มากกว่าคุณลักษณะอื่น เช่น การสร้างข้อสอบและการเลือกข้อสอบ เพราะลักษณะดังกล่าวของแบบสอบอิง เกณฑ์และแบบสอบอิงกลุ่มไม่แตกต่างกันอย่างเด่นชัด (Block and Dockrell, ๑๙๘๔: ๓๖)

ความตรงเชิงพยากรณ์ หมายถึง ความสามารถของผลการสอบที่อาจพยากรณ์พฤติกรรม หรือสิ่งที่ต้องการวัดที่จะเกิดขึ้นในอนาคตได้ ค่านี้ได้จากค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน โพรดักต์โมเมนต์ (Pearson Product-moment Correlation Coefficient) ระหว่างคะแนนการสอบจากแบบสอบความสามารถทั่วไปในการอ่าน กับคะแนนรวมการสอบภาคกลางและภาคปลายของวิชาภาษาอังกฤษพื้นฐานเฉพาะการอ่าน

ความเที่ยงของแบบสอบอิงกลุ่ม หมายถึง ความคงที่ของคะแนนการสอบของผู้สอบ ต่อรายข้อว่า มีมากน้อยเพียงใด ซึ่งในที่นี้ได้แก่ ค่า Cronbach α

ความเที่ยงของแบบสอบอิง เกณฑ์ หมายถึง ความคงที่ของคะแนนการสอบของผู้สอบแต่ละบุคคลว่า แตกต่างจากจุดตัดที่ใช้ เป็นเกณฑ์มากน้อยเพียงใด ซึ่งในที่นี้ได้แก่ค่าดัชนีความเชื่อถือได้

โครงสร้างความรู้ หมายถึง มโนทัศน์และโครงสร้างของความรู้รวมทั้งการคาดหวังที่เกิดขึ้นจากชื่อ เรื่องและภาพประกอบเรื่อง ศัพท์ที่มีความหมายสำคัญต่อ เรื่อง และความคุ้นเคยกับ เนื้อหาสาระของ เรื่องที่อ่าน

ลำดับชั้นความเข้าใจในการอ่าน หมายถึง ระดับความสามารถของผู้อ่านที่จะเข้าใจเนื้อหาสาระของเรื่องทีอ่านได้ลึกซึ้งเพียงใด ตั้งแต่ขั้นการถ่ายโอน (Transfer) การสื่อความหมาย (Communication) และการวิพากษ์วิจารณ์ (Criticism)

แบบสอบที่ไม่มีชื่อเรื่องและภาพประกอบเรื่อง หมายถึง แบบสอบวัดความสามารถทั่วไปในการอ่านเข้าใจความภาษาอังกฤษที่ไม่มีชื่อเรื่องและภาพประกอบเรื่อง แต่มีคำศัพท์ที่มีความหมายสำคัญต่อเรื่อง (No-context and Transparency) เป็นแบบสอบแบบเลือกตอบ มี ๔ ตัวเลือก ยาว ๕๐ ข้อ และมีเรื่องสำหรับอ่าน ๒ เรื่อง

แบบสอบที่มีชื่อเรื่องและภาพประกอบเรื่อง หมายถึง แบบสอบวัดความสามารถทั่วไปในการอ่านเข้าใจความภาษาอังกฤษที่มีชื่อเรื่องและภาพประกอบเรื่อง และมีคำศัพท์ที่มีความหมายสำคัญต่อเรื่อง (Context and Transparency) เป็นแบบสอบแบบเลือกตอบ มี ๔ ตัวเลือก ยาว ๕๐ ข้อ และมีเรื่องสำหรับการอ่าน ๒ เรื่อง

วิธีใช้ค่าเดลต้า (Δ) หมายถึง การให้คะแนนรายข้อที่ตอบถูก โดยใช้ค่าความยากง่าย (p) ที่ได้จากการวิเคราะห์แบบสอบรายข้อวิธีประเพณีนิยม แล้วแปลงค่า p ให้เป็นค่า Δ เพื่อให้เป็นค่าคะแนนมาตรฐาน และสื่อความหมายได้เหมาะสมยิ่งขึ้น คือ ข้อที่ง่ายมีค่า Δ น้อยกว่าข้อที่ยากกว่า ข้อที่ตอบผิดมีค่า = ๐

วิธีใช้ค่าเบต้าแปลง (β_c) หมายถึง การให้คะแนนรายข้อที่ตอบถูกโดยใช้ค่าความยากง่าย (β) ที่ได้จากการวิเคราะห์แบบสอบรายข้อตามรูปแบบบราสซ์ แล้วแปลงค่า β ให้เป็นค่า β_c เพื่อให้เป็นค่าคะแนนมาตรฐานที่มีค่าบวก ข้อที่ง่ายมี β_c น้อยกว่าข้อที่ยากกว่า ข้อที่ตอบผิดมีค่า = ๐

วิธีใช้ค่านำหนักคะแนนที่เหมาะสม (w) หมายถึง การให้คะแนนรายข้อที่ตอบถูกโดยใช้ค่านำหนักคะแนนที่เหมาะสม (w) ที่คำนวณจากค่าอำนาจจำแนก (a) ค่าความยากง่าย (p) และค่าการเดา (c) ที่ได้จากการวิเคราะห์แบบสอบรายข้อตามรูปแบบโลจิสติกส์ใช้พารามิเตอร์ ๓ ตัว ข้อที่ตอบผิดมีค่า = ๐

วิธีใช้ค่าผลบวกของความแปรปรวนร่วม (h^2) หมายถึง การให้คะแนนรายข้อที่ตอบถูก โดยใช้ค่าผลบวกของความแปรปรวนร่วมต่อสิ่งที่มีงทดสอบ ซึ่งได้จากการวิเคราะห์ตัวประกอบ ข้อสอบที่มีความสัมพันธ์กับสิ่งที่มีงทดสอบน้อย มีค่า h^2 น้อยกว่าข้อที่มีความสัมพันธ์มากกว่า ข้อที่ตอบผิดมีค่า = ๐

วิธีประเพณีนิยม หมายถึง การให้คะแนนรายข้อที่ตอบถูกมีค่า = ๑ ข้อที่ตอบผิดมีค่า = ๐

ประโยชน์ที่คาดว่าจะได้รับ

๑. สามารถทราบได้ว่า การให้คะแนนข้อสอบแต่ละข้อ วิธีใดทำให้แบบสอบมีความตรงเชิงพยากรณ์และความเที่ยงมากที่สุด เพื่อให้ผู้ใช้สามารถวัดผลแบบอิงกลุ่มได้ถูกต้องและคงเส้นคงวามากยิ่งขึ้นเมื่อเลือกใช้วิธีดังกล่าว และสามารถลดจำนวนข้อสอบลงได้ ซึ่งจะเป็นการลดค่าใช้จ่าย และแรงงานในการวัดและประเมินผลได้เป็นอย่างดีวิธีหนึ่ง

๒. สามารถทราบได้ว่า การให้คะแนนข้อสอบแต่ละข้อ วิธีใดทำให้แบบสอบมีความเที่ยง (ค่าดัชนีความเชื่อถือ) สูงที่สุด และเมื่อเลือกใช้วิธีการให้คะแนนวิธีดังกล่าวนี้ จะทำให้การวัดผลแบบอิงเกณฑ์มีความคงเส้นคงวามากยิ่งขึ้น

๓. สามารถบอกได้ว่า การให้คะแนนข้อสอบแต่ละข้อ วิธีใดดีที่สุดในด้านทฤษฎีและปฏิบัติ

๔. สามารถทราบได้ว่า โครงสร้างความรู้บางอย่างของเรื่องที่อ่านมีผลต่อการอ่านมากน้อยเพียงใด ซึ่งจะเป็นประโยชน์ในการเรียนการสอน การผลิตเอกสารประกอบการสอนและการสอบและการประเมินผลต่อไปว่า ควรจะเลือกใช้เรื่องหรือข้อความชนิดใดจึงจะเหมาะสมกับผู้อ่าน