

# การเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบภาษาอังกฤษ ที่ตรวจให้คะแนนโดยวิธีธรรมดา และวิธีกำหนดน้ำหนักของข้อสอบ

นุชวรา เหลืองอังกร \*

## ที่มาและความสำคัญของปัญหา

การวัดและประเมินผลการศึกษา แต่เดิมนั้นเป็นการทดสอบแบบปากเปล่า และการทดสอบแบบรายบุคคล จนกระทั่งกลางศตวรรษที่ 19 จึงได้เริ่มมีแบบทดสอบข้อเขียนสำหรับทดสอบแบบกลุ่ม ได้มีผู้ทำการเปรียบเทียบคะแนนจากการสอบของกลุ่มตัวอย่างขนาดใหญ่ เช่น Rice (อ้างถึงใน Stanley and Wang, 1970) ซึ่งได้ใช้แบบทดสอบการเขียนศัพท์ภาษาอังกฤษจำนวน 50 คำ วิธีตรวจให้คะแนนคือให้ 2 คะแนนสำหรับผู้ที่จะกคคำศัพท์ถูก และให้ 0 คะแนนสำหรับผู้ที่จะกคคำศัพท์ผิดคั้งนั้นช่วงคะแนนสำหรับผู้สอบจะอยู่ระหว่าง 0-100 คะแนน ซึ่งคล้ายกับวิธีตรวจให้คะแนนโดยทั่ว ๆ ไปในปัจจุบัน Rice ได้นำคะแนนเฉลี่ยด้านการสะกคคำศัพท์คั้งกล่าวมาเปรียบเทียบระหว่างไปโรงเรียนเพื่อศึกษาค่าความแตกต่างของคะแนนเฉลี่ยระหว่างโรงเรียนต่าง ๆ Stanley and Wang (1970) มีความเห็นว่าถ้า Rice ให้คะแนนการสะกคคำศัพท์แต่ละคำไม่เท่ากัน กล่าวคือให้คำที่ยากมีคะแนนมากกว่าคำที่ง่ายแล้ว จะทำให้คะแนนเฉลี่ยของแต่ละโรงเรียนแตกต่างกันมากกว่าการที่ให้คะแนนคำศัพท์ละ 2 คะแนนเท่ากันหมดทุกคำ อันนับได้ว่าเป็นจุดเริ่มสนใจปัญหาการกำหนดน้ำหนักข้อสอบ สำหรับ

การตรวจให้คะแนนการสอบนั้น แต่เดิมใช้วิธีการตรวจให้คะแนนโดยวิธีธรรมดา (The conventional method) ก็น้อย่างกว้างขวางโดยถ้าผู้สอบเลือกคำตอบที่ถูกให้ 1 คะแนน ถ้าเลือกคำตอบที่ผิดให้ 0 คะแนน ข้อดีของวิธีนี้คือ เป็นวิธีการตรวจให้คะแนนที่ง่ายโดยนับคะแนนตามคำตอบที่ถูก อย่างไรก็ตามได้มีการค้นคว้าหาวิธีอื่น ๆ อีกหลายวิธี ตัวอย่างเช่น การกำหนดน้ำหนักของข้อสอบ (item weighting) คือการให้คะแนนแก่ข้อสอบแต่ละข้อแตกต่างกัน หรือการตรวจให้คะแนนตามส่วนความรู้ (partial knowledge) ที่ผู้สอบแสดงให้เห็นในการตอบข้อสอบ กล่าวคือนอกจากจะให้คะแนนสำหรับคำตอบที่ถูกต้องแล้ว ก็ยังมีคะแนนให้กับผู้สอบที่สามารถชี้ว่าตัวใดผิดให้ตามระดับความถูกต้องของตัวเลือกที่เลือก วิธีใดวิธีหนึ่งเป็นต้น การค้นคว้าหาวิธีตรวจให้คะแนนแบบใหม่ก็เพื่อพยายามแปลงคะแนนสอบสำหรับบุคคลให้มีค่าความเชื่อมั่น (reliability) สูงขึ้น ซึ่งจะส่งผลกระทบต่อความคลาดเคลื่อนในการวัดผลให้ลดลง และจะทำให้คะแนนสอบนั้นเป็นตัวแทนที่บ่งระดับความสามารถของผู้สอบได้ใกล้เคียงยิ่งขึ้น

วิธีการกำหนดน้ำหนักของข้อสอบ (item weighting) มีจุดมุ่งหมายเพื่อปรับปรุงค่าความเชื่อมั่นของคะแนนสอบและความตรง (validity)

\* หัวหน้าภาควิชาพื้นฐานของการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ  
มหาสารคาม

ของแบบทดสอบ ในทางทฤษฎีการกำหนดน้ำหนักของข้อสอบจะทำให้ความตรงเชิงพยากรณ์ (predictive validity) หรือความตรงตามคุณลักษณะ (construct validity) ของแบบทดสอบสูงขึ้น แต่ในทางปฏิบัติจะได้ผลเพียงเล็กน้อยเท่านั้น โดยทั่วไปการกำหนดน้ำหนักของข้อสอบ คือ กระบวนการตัดสินใจสัดส่วนของคะแนนแต่ละข้อ ต่อคะแนนรวมของแบบทดสอบ หรือในกรณีที่เป็นข้อสอบแบบเลือกตอบ (multiple choice) อาจเป็นการกำหนดสัดส่วนของน้ำหนักแต่ละตัวเลือกต่อคะแนนของข้อสอบ การกำหนดน้ำหนักของข้อสอบอาจทำกับข้อสอบวัดผลสัมฤทธิ์ทางการเรียน วัดความถนัด วัดความสนใจ หรือวัดบุคลิกภาพก็ได้ (Wang and Stanley, 1970) ในการตรวจให้คะแนนโดยวิธีการกำหนดน้ำหนักของข้อสอบนั้น เราอาจกระทำได้ 3 วิธีดังนี้คือ

1. การกำหนดน้ำหนักแบบธรรมชาติ (natural weight) หรือเรียกอีกชื่อหนึ่งว่าการกำหนดน้ำหนักแบบสุ่ม (random weight) คือ การที่ไม่มีการควบคุมการกำหนดน้ำหนัก คำว่า "สุ่ม" ในที่นี้มิได้แสดงว่าความแตกต่างระหว่างน้ำหนักเกิดขึ้นแล้วแต่โอกาส (by chance) แต่จะสุ่มจากช่วงคะแนนที่เป็นความแตกต่างที่แท้จริงของความแปรปรวนของข้อสอบและความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบ

เป็นที่น่าสังเกตว่าการกำหนดน้ำหนักแบบธรรมชาติมีส่วนคล้ายกับไม่มีการกำหนดน้ำหนักแต่ความจริงแล้วการไม่มีการกำหนดน้ำหนักหมายถึงกรณีที่กำหนดให้ข้อสอบทุกข้อมีน้ำหนักเท่ากัน

2. การกำหนดน้ำหนักก่อนการสอบ (priori weight) คือการกำหนดน้ำหนักโดยใช้การพิจารณาตัดสินใจของผู้เชี่ยวชาญ โดยให้กลุ่มผู้เชี่ยวชาญกำหนดน้ำหนักแก่ข้อสอบตามความคิดเห็นของแต่ละคน แล้วนำมาเฉลี่ยเป็นน้ำหนักของข้อสอบ หรือใช้การประมาณค่า (ratings) ของผู้เชี่ยวชาญ โดยให้ผู้

เชี่ยวชาญเป็นผู้ประมาณค่าข้อสอบแต่ละข้อแล้วให้ตัวเลขแสดงระดับความสำคัญของข้อสอบนั้น ๆ เพื่อเป็นตัวแทนน้ำหนักของข้อสอบหรือการใช้วิธีการอื่นที่คล้ายคลึงกันเพื่อกำหนดน้ำหนักของข้อสอบ (Wang and Stanley, 1970)

3. การกำหนดน้ำหนักจากผลของการสอบ (empirical weight) คือการกำหนดน้ำหนักจะกระทำหลังจากการนำแบบทดสอบไปใช้สอบแล้วนำผลการสอบมาวิเคราะห์แล้วจึงกำหนดน้ำหนักโดยอาจพิจารณาน้ำหนักด้วยวิธีต่าง ๆ ดังนี้ (Wang and Stanley, 1970)

3.1 การกำหนดน้ำหนักตามค่าสัมประสิทธิ์การถดถอย (regression coefficient) ของสมการถดถอยพหุคูณ (multiple regression equation) คือการนำคะแนนสอบจากแบบทดสอบ 2 ฉบับคือ คะแนนของแบบทดสอบที่เป็นเกณฑ์และคะแนนรวมของแบบทดสอบ ซึ่งเป็นตัวพยากรณ์ มาสร้างสมการถดถอย เพื่อจะหาตัวพยากรณ์ที่จะทำให้อาศัยพื้นที่ระหว่างคะแนน 2 ชุดมีค่าสูงสุดแล้วนำค่าสัมประสิทธิ์การถดถอยมากำหนดเป็นน้ำหนักแก่ข้อสอบ

3.2 การกำหนดน้ำหนักตามค่าสหสัมพันธ์แบบพ้อยท์ไบซีเรียล (point biserial correlation coefficient) คือการใช้ค่าสหสัมพันธ์แบบพ้อยท์ไบซีเรียล ระหว่างคะแนนสอบแต่ละข้อกับคะแนนรวมที่ตรวจให้คะแนนด้วยวิธีธรรมดา มากำหนดเป็นน้ำหนัก (Bejar and Weiss, 1977)

3.3 การกำหนดน้ำหนักตามค่าส่วนกลับของค่าความเบี่ยงเบนมาตรฐาน (reciprocal of item standard deviation) ทำได้โดยกำหนดน้ำหนักข้อสอบเท่ากับ  $\frac{1}{\sigma_i}$  ของข้อสอบที่ได้จากการสอบแต่ละครั้ง

3.4 การกำหนดน้ำหนักตามความยาวของแบบทดสอบ (Test length) คือการกำหนด

น้ำหนักตามจำนวนข้อสอบในแบบทดสอบ แบบทดสอบที่มีข้อสอบมากข้อ จะมีน้ำหนักมากกว่าแบบทดสอบที่มีข้อสอบน้อยข้อ เมื่อวัดในเนื้อหาวิชาแบบเดียวกัน โดยในแต่ละแบบทดสอบจะมีน้ำหนักเท่ากันทุกข้อ ตัวอย่างเช่น แบบทดสอบชุดที่หนึ่งมี 50 ข้อ ชุดที่สองมี 100 ข้อ แบบทดสอบชุดที่สองจะมีน้ำหนักเป็น 2 เท่าของชุดที่หนึ่ง สำหรับแบบทดสอบที่วัดต่างวิชากันไม่สามารถนำมาเปรียบเทียบกันได้นอกจากนี้แบบทดสอบที่ต่างรูปแบบกัน เช่น แบบทดสอบแบบเลือกตอบกับแบบทดสอบแบบถูก-ผิด ก็ไม่สามารถเปรียบเทียบกันได้

วิธีกำหนดน้ำหนักแบบนี้ บางครั้งก็จัดในพวกการกำหนดน้ำหนักก่อนการสอบ

3.5 การกำหนดน้ำหนักตามค่าความยากของข้อสอบ (item difficulty) เป็นการกำหนดน้ำหนักแก่ข้อสอบตามสัดส่วนของจำนวนคนที่ตอบข้อนั้นผิด บางครั้งใช้แบบทดสอบมาตรฐานนำมาสอบกับกลุ่มตัวอย่าง แล้วใช้ค่าความยากจากการสอบมาเป็นน้ำหนักของข้อสอบ หลักการของวิธีนี้ขึ้นกับความเชื่อที่ว่าคนที่ทำแบบทดสอบที่ยากได้ถูก จะมีความสามารถหรือมีผลสัมฤทธิ์ทางการเรียนสูงกว่าคนที่ทำข้อง่ายได้ถูก

3.6 การกำหนดน้ำหนักตามค่าความตรงเชิงพยากรณ์ (predictive validity) ของแบบทดสอบ เป็นการให้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบทดสอบที่เป็นตัวจริงกับแบบทดสอบที่เป็นเกณฑ์ นำมาเป็นน้ำหนักของข้อสอบ หรือเป็นการใช้สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบทดสอบ 2 ชุด คือแบบทดสอบที่เป็นเกณฑ์กับแบบทดสอบที่เป็นตัวพยากรณ์ นำมาเป็นน้ำหนักของข้อสอบเมื่อเราไม่สามารถหาตัวพยากรณ์ที่เหมาะสมในการสร้างสมการถดถอยได้

ผลงานวิจัยเกี่ยวกับการตรวจให้คะแนนในระยะที่ผ่านมาส่วนใหญ่ใช้ค่าสหสัมพันธ์ระหว่างคะแนน

แบบกำหนดน้ำหนักและแบบไม่กำหนดน้ำหนักเป็นตัวบ่งชี้ความสัมพันธ์ระหว่าง 2 วิธี ผลงานวิจัยส่วนใหญ่พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนแบบกำหนด น้ำหนักและแบบไม่กำหนดน้ำหนักมีค่าสูงถึง .90 หรือ มากกว่า (Douglas and Spencer, cited in Wang and Stanley, 1970; Odell, cited in Masang, 1973; Pothoff and Barnett, cited in Wang and Stanley, 1970; West, cited in Wang and Stanley, 1970) นอกจากนี้งานวิจัยอีกส่วนหนึ่งก็เปรียบเทียบค่าความเชื่อมั่นของคะแนน สอบที่ตรวจด้วยวิธีกำหนดน้ำหนักกับวิธีธรรมดา ซึ่งมักจะพบว่า ค่าความเชื่อมั่นของคะแนนสอบที่ตรวจด้วยวิธีธรรมดามีค่าสูงกว่าหรือแตกต่างกันเพียงเล็กน้อยจากที่ตรวจด้วยวิธีกำหนดน้ำหนัก (Clark, cited in Whitley, 1968; Douglas and Spencer, cited in Stanley and Wang, 1970; Guilford, Lovell and Williams, cited in Masang, 1973) จะเห็นได้ว่าผลงานวิจัยทั้งสอบกลุ่มสรุปได้คล้าย ๆ กันคือ การกำหนดน้ำหนักไม่ทำให้เกิดความแตกต่างจากวิธีตรวจให้คะแนนแบบธรรมดา แต่อย่างไรก็ตามยังมีนักวิจัยที่สนใจในด้านนี้ได้ดำเนินการวิจัยเพื่อหาเงื่อนไขที่จะทำให้วิธีการกำหนดน้ำหนัก มีประสิทธิภาพยิ่งขึ้น

ในการค้นคว้าเรื่องเงื่อนไขต่าง ๆ ที่จะทำให้การตรวจให้คะแนนแบบกำหนดน้ำหนักของข้อสอบที่มีประสิทธิภาพดีกว่าแบบธรรมดานั้น ได้มีข้อสรุปทั่ว ๆ ไปดังนี้คือ Wilks (อ้างถึง ใน Stanley and Wang, 1970) พบว่า ในแบบทดสอบที่มีจำนวนข้อมากและข้อสอบเหล่านั้นมีความสัมพันธ์กันการกำหนดน้ำหนักสำหรับแต่ละข้อจะทำให้ผลคะแนนแตกต่างกันเพียงเล็กน้อย ซึ่งแสดงว่าลำดับที่ตามคะแนนของแต่ละคนมีแนวโน้มจะคงที่ไม่ผันแปร แม้จะมีวิธีการตรวจให้คะแนนที่แตกต่างกัน Aiken (อ้างถึงใน Masang, 1973) ได้สรุปว่าการกำหนดน้ำหนักแตกต่าง

ต่างกันของคำตอบที่ถูกสำหรับข้อสอบแบบเลือกตอบ จะมีผลเพียงเล็กน้อยต่อลำดับที่ของคะแนนรวม เมื่อค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบแต่ละข้อ เป็นบวกและ Nunnally (1967) ได้สรุปเงื่อนไขที่จะทำให้การกำหนดน้ำหนักของข้อสอบที่มีความแตกต่างจากวิธีการเดิมไว้สองประการ คือ ประการแรก เมื่อข้อสอบมีจำนวนน้อย (น้อยกว่า 20 ข้อ) ประการที่สองเมื่อค่าสหสัมพันธ์ระหว่างคะแนนข้อสอบแต่ละข้อ กับคะแนนรวมของข้อสอบ (item-total correlation) แตกต่างกันมาก

นี่คือเงื่อนไขบางประการที่ทำให้การกำหนดน้ำหนักของข้อสอบให้ผลแตกต่างจากวิธีธรรมดา ทั้งในด้านค่าความเชื่อมั่นกับความตรง และความเชื่อมั่นหรือความตรงเพียงอย่างเดียวอย่างใดอย่างหนึ่ง ดังนั้นต่อคำถามในเรื่องการกำหนดน้ำหนักที่ว่า "การกำหนดน้ำหนักสามารถปรับปรุงค่าความเชื่อมั่นของคะแนนสอบให้สูงกว่าคะแนนสอบที่ได้จากการตรวจให้คะแนน แบบธรรมดาหรือไม่" ตอบได้ว่า "สามารถเป็นไปได้ภายใต้เงื่อนไขบางประการ"

เนื่องจากผู้วิจัยบางท่านได้เสนอแนะเงื่อนไขต่าง ๆ ที่จะให้คะแนนที่ได้จากการตรวจให้คะแนนแบบกำหนดน้ำหนักมีประสิทธิภาพสูงกว่าจากการตรวจให้คะแนนแบบธรรมดา ดังนั้น Masang (1973) ได้ทำการวิจัยเกี่ยวกับการกำหนดน้ำหนักตามค่าความยากของข้อสอบเพื่อหาเงื่อนไขที่ทำให้ค่าความเชื่อมั่นและค่าความตรงของแบบทดสอบสูงขึ้น โดยเปรียบเทียบวิธีตรวจให้คะแนน 2 วิธีคือวิธีนับจำนวนข้อถูก (the number right method) ซึ่งเป็นวิธีการตรวจให้คะแนนแบบตอบถูก ให้ 1 คะแนน ตอบผิดให้ 0 คะแนน และวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ (the  $q$  value method) ซึ่งกำหนดน้ำหนักตามค่า  $q$  เมื่อ  $q = 1-p$  และ  $p$  คือค่าความยากของข้อสอบ เมื่อจำนวนข้อสอบไม่เท่ากัน สัดส่วนของจำนวนข้อสอบที่ง่ายในแต่ละชุดไม่เท่ากัน

และเมื่อกลุ่มข้อสอบมีความเป็นเอกพันธ์กับไม่มีความเป็นเอกพันธ์ โดยได้กระทำกับข้อมูลสมมติ (hypothetical data) โดยการเลือกค่าความยาก (item difficulty) ของข้อสอบจากตารางเลขสุ่มมาจัดเป็นคะแนนของข้อสอบเป็นชุด ๆ ตามเงื่อนไขที่ต้องการ แล้วทำการวิจัย ซึ่งผลจากการวิจัยปรากฏว่าวิธีการตรวจให้คะแนนแบบกำหนดน้ำหนักตามค่า  $q$  จะให้ค่าความเชื่อมั่นของแบบทดสอบสูงกว่าวิธีตรวจให้คะแนนแบบนับจำนวนข้อถูกเมื่อแบบทดสอบมีลักษณะดังนี้คือ ประการที่หนึ่งข้อสอบมีจำนวนน้อย (10 ข้อ หรือน้อยกว่า) ประการที่สองแบบทดสอบมีจำนวนข้อยากมากกว่าข้อง่าย ประการที่สามข้อสอบข้อยากมีความเป็นเอกพันธ์ และประการที่สี่แบบทดสอบมีความยากมาก

จากผลงานวิจัยที่ผ่านมา แม้จะเห็นได้ว่าการตรวจให้คะแนนแบบกำหนดน้ำหนักของข้อสอบส่วนใหญ่ไม่ทำให้ค่าความเชื่อมั่นและความตรงของแบบทดสอบสูงขึ้นกว่าการตรวจให้คะแนนโดยวิธีธรรมดาก็ตาม แต่สิ่งที่น่าค้นคว้าคือ เงื่อนไขต่าง ๆ เกี่ยวกับการกำหนดน้ำหนักที่จะทำให้แบบทดสอบมีค่าความเชื่อมั่นสูงขึ้นนั้น มีองค์ประกอบใดบ้างในเรื่องนี้จะเห็นว่าการวิจัยเกี่ยวกับการกำหนดน้ำหนักตามค่าความยากมีค่อนข้างน้อยทั้ง ๆ ที่วิธีการกำหนดน้ำหนักทำได้ง่ายไม่ซับซ้อน Masang (1973) ได้ทำการวิจัยเพื่อหาเงื่อนไขที่จะทำให้ความเชื่อมั่นของแบบทดสอบสูงขึ้น เมื่อตรวจให้คะแนนด้วยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ โดยใช้ข้อสอบของนักศึกษาชั้นปีที่ 1 มหาวิทยาลัยขอนแก่น จากทุกคณะ ประจำปีการศึกษา 2524 จำนวน 1,042 คน ที่ได้สอบแบบทดสอบวัดความรู้เพื่อจัดกลุ่มเรียน (Placement test) วิชาภาษาอังกฤษ เพราะงานวิจัยเกี่ยวกับการกำหนดน้ำหนักที่ผ่านมามักจะมีแต่แบบวัดผล สมฤทธิ์ทางการเรียน (Corey, cited in Masang, 1973; Guilford, Lovell and Williams,

cited in Masang, 1973) แบบวัดความถนัดและแบบวัดความสนใจ (Lord, 1968; Echer-nacht, 1976) ส่วนงานวิจัยเรื่องการกำหนดน้ำหนักแบบทดสอบภาษา อังกฤษ เพื่อจัดกลุ่มเรียนและแบบทดสอบภาษาอังกฤษ สำหรับนักเรียนต่างชาติยังมีค่อนข้างน้อยมาก ผู้วิจัยจึงเลือกใช้แบบทดสอบภาษาอังกฤษในงานวิจัยครั้งนี้ เพื่อให้เป็นประโยชน์แก่วงการทดสอบภาษาอังกฤษ ต่อไป

### วัตถุประสงค์ของการวิจัย

ในการศึกษาค้นคว้าครั้งนี้ผู้วิจัยมีจุดมุ่งหมายดังนี้คือ

1. เพื่อศึกษาถึงเงื่อนไขลักษณะของแบบทดสอบที่จะทำให้ค่าความเชื่อมั่นของแบบทดสอบสูงสุดเมื่อตรวจให้คะแนนโดยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ โดยวิธีกำหนดน้ำหนัก  $C+q$  ตามค่า  $q$  ของข้อสอบและโดยวิธีธรรมดา
2. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบในแต่ละชุดที่ตรวจให้คะแนน โดยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ โดยวิธีกำหนดน้ำหนัก  $C + q$  ตามค่า  $q$  ของข้อสอบและโดยวิธีธรรมดา
3. เพื่อเพิ่มพูนความรู้ให้แก่วงการทดสอบในด้านเกี่ยวกับเงื่อนไขการให้คะแนนที่มีผลทำให้ค่าความเชื่อมั่นของแบบทดสอบเปลี่ยนไป

### สมมติฐานการวิจัย

จากผลงานวิจัยของ Masang (1973) ซึ่งได้ให้ข้อเสนอแนะสำหรับผู้ที่ยืนยันว่าเรื่องการกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบนั้น ทำให้ผู้วิจัยได้แนวทางในการตั้งสมมติฐานการวิจัยดังนี้คือ

1. สำหรับแบบทดสอบที่มีจำนวนข้อยากเป็นจำนวนเกินกว่าร้อยละ 50 จะทำให้ค่าความเชื่อมั่นของคะแนนสอบที่ตรวจให้คะแนนด้วยวิธีกำหนด

น้ำหนักตามค่า  $q$  ของข้อสอบและวิธีกำหนดน้ำหนัก  $C + q$  ตามค่า  $q$  ของข้อสอบ สูงกว่าที่ตรวจให้คะแนนด้วยวิธีธรรมดา

2. สำหรับแบบทดสอบที่ข้อสอบแต่ละข้อมีค่าอำนาจจำแนก ( $r$ ) เกินกว่า .20 และมีจำนวนข้อยาก เป็นจำนวนเกินกว่าร้อยละ 50 จะทำให้ค่าความเชื่อมั่นของคะแนนสอบที่ตรวจให้คะแนนด้วยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบและวิธีกำหนดน้ำหนัก  $C + q$  ตามค่า  $q$  ของข้อสอบ สูงกว่าที่ตรวจให้คะแนนด้วยวิธีธรรมดา

### ขอบเขตของการวิจัย

การวิจัยครั้งนี้ ได้กำหนดขอบเขตในการวิจัยดังนี้

1. กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ นักศึกษาชั้นปีที่ 1 จากทุกคณะของมหาวิทยาลัยขอนแก่น ประจำปีการศึกษา 2524 จำนวน 1,042 คน
2. เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบวัดความรู้ภาษาอังกฤษ เพื่อจัดกลุ่มเรียนของภาควิชาภาษาต่างประเทศ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น ซึ่งสร้างขึ้นเพื่อวัดความรู้ภาษาอังกฤษของนักศึกษาชั้นปีที่ 1 ก่อนจะจัดกลุ่มเรียน
3. ค่าความยากและค่าอำนาจจำแนกของข้อสอบแต่ละข้อ ได้มาจากผลการสอบของกลุ่มตัวอย่างในงานวิจัยครั้งนี้

### ข้อตกลงเบื้องต้นในการวิจัย

1. การตรวจให้คะแนนแบบกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบมีหลักการที่ว่าคนที่ตอบข้อสอบข้อยากได้ถูกต้องจะต้องมีความรู้มากกว่าคนที่ตอบข้อสอบข้อง่ายได้ถูกต้อง สำหรับแบบทดสอบที่วัดในเนื้อหาเดียวกัน

2. ในการตรวจให้คะแนนแบบกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ การกำหนดน้ำหนักจะขึ้นกับระดับความยากของข้อสอบแต่ละข้อ ข้อสอบที่ยากจะได้น้ำหนักมากเพราะข้อสอบที่ยากจะมีค่า  $p$  ต่ำ และทำให้ค่า  $q$  มาก ส่วนข้อสอบที่ง่ายจะได้น้ำหนักน้อย เพราะข้อสอบที่ง่ายมีค่า  $p$  สูง และทำให้ค่า  $q$  น้อย

### วิธีดำเนินการวิจัย

กลุ่มตัวอย่าง กลุ่มตัวอย่างของงานวิจัยครั้งนี้เป็นนักศึกษาชั้นปีที่ 1 ของมหาวิทยาลัยขอนแก่น ประจำปีการศึกษา 2524 ซึ่งได้สอบแบบทดสอบวัดความรู้ภาษาอังกฤษ เพื่อจัดกลุ่มเรียน เมื่อเริ่มเข้าเรียนในภาคการศึกษาแรกทุกคนจาก 10 คณะ รวมทั้งสิ้น 1,042 คน

เครื่องมือ เครื่องมือที่ใช้ในการวิจัยครั้งนี้ คือ แบบทดสอบวัดความรู้ภาษาอังกฤษเพื่อจัดกลุ่มเรียน ประจำปีการศึกษา 2524 ของภาควิชาภาษาต่างประเทศ คณะมนุษยศาสตร์และสังคมศาสตร์ มหาวิทยาลัยขอนแก่น ซึ่งใช้สำหรับวัดความรู้ของนักศึกษาชั้นปีที่ 1 เพื่อจัดกลุ่มเรียนภาษาอังกฤษตามระดับความสามารถ ลักษณะของแบบทดสอบเป็นแบบทดสอบปรนัยแบบเลือกตอบ (multiple choice test) จำนวน 140 ข้อ ประกอบด้วยแบบทดสอบย่อย 4 ชุด ดังนี้คือ

1. แบบทดสอบการฟัง (listening comprehension) มีข้อสอบจำนวน 30 ข้อ แบ่งเป็น 3 ส่วนคือ

ส่วนที่ 1 มี 10 ข้อ เป็นข้อสอบแบบ 3 ตัวเลือก

ส่วนที่ 2 มี 10 ข้อ เป็นข้อสอบแบบ 4 ตัวเลือก

ส่วนที่ 3 มี 10 ข้อ เป็นข้อสอบแบบ 4 ตัวเลือก

2. แบบทดสอบไวยากรณ์ (structure) มีข้อสอบจำนวน 40 ข้อ เป็นข้อสอบแบบ 4 ตัวเลือก

3. แบบทดสอบคำศัพท์ (vocabulary) มีข้อสอบจำนวน 30 ข้อ เป็นข้อสอบแบบ 4 ตัวเลือก

4. แบบทดสอบการอ่าน (reading comprehension) มีข้อสอบจำนวน 40 ข้อ เป็นข้อสอบแบบ 4 ตัวเลือก

### การวิเคราะห์ข้อมูล

1. นำข้อมูลจากกระดาษคำตอบของนักศึกษาชั้นปีที่ 1 ทั้งหมดจำนวน 1,042 คน ซึ่งสอบแบบทดสอบวัดความรู้ภาษาอังกฤษเพื่อจัดกลุ่มเรียนจำนวน 140 ข้อมาลงรหัสข้อมูล เพื่อใช้เครื่องคอมพิวเตอร์วิเคราะห์ข้อมูล

2. นำผลการสอบจากข้อ 1 มาดำเนินการตรวจให้คะแนนด้วยวิธีธรรมดาคือถูกให้ 1 ผิดให้ 0 คะแนน โดยใช้โปรแกรมสำเร็จรูปชื่อ Item Analysis ของสถาบันบริการคอมพิวเตอร์จุฬาลงกรณ์มหาวิทยาลัย ซึ่งดำเนินการวิเคราะห์ข้อสอบโดยใช้เทคนิคร้อยละ 27 เพื่อวิเคราะห์รายข้อทั้งตัวเลือกและตัวลวงหาค่าความยาก ( $p$ ) รายข้อและค่าอำนาจจำแนกรายข้อด้วยวิธีหาค่าสหสัมพันธ์พอยท์ไบเซรียล (point biserial correlation)

3. นำแบบทดสอบแต่ละฉบับย่อยมาตรวจให้คะแนน ด้วยวิธีธรรมดากำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ และวิธีกำหนดน้ำหนัก  $C + q$  ตามค่า  $q$  ของข้อสอบ ซึ่งมีรายละเอียดดังต่อไปนี้คือ

3.1 วิธีการตรวจให้คะแนนแบบธรรมดา (the conventional method) หมายถึง วิธีการตรวจให้คะแนนแบบที่ว่าถ้าผู้สอบเลือกคำตอบที่ถูกให้ 1 คะแนน ถ้าผู้สอบเลือกคำตอบที่ผิดให้ 0 คะแนน

3.2 วิธีการตรวจให้คะแนนแบบกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ (the  $q$  value

method) หมายถึง วิธีการตรวจให้คะแนนแบบที่ว่า ถ้าผู้สอบเลือกคำตอบที่ถูกให้  $q$  คะแนนตามค่า  $q$  ของข้อสอบแต่ละข้อ ถ้าผู้สอบเลือกคำตอบที่ผิดให้ 0 คะแนน

เมื่อค่า  $q = 1-p$  และ  $p$  คือค่าความยากของข้อสอบ (item difficulty index) โดยค่า  $p$  เป็นผลที่ได้มาจากการวิเคราะห์ข้อสอบโดยการสอบของกลุ่มตัวอย่างในครั้งนั้น

3.3 วิธีการตรวจให้คะแนนแบบธรรมดา รวมกับแบบกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ ใช้สัญลักษณ์ว่า  $C+q$  หมายถึง วิธีการตรวจให้คะแนน 2 วิธีคือ

3.3.1 วิธีการตรวจให้คะแนน  $C + q$  เฉพาะข้อถูก คือถ้าผู้สอบเลือกคำตอบที่ถูกให้  $1 + q$  คะแนนตามค่า  $q$  ของข้อสอบแต่ละข้อ ถ้าผู้สอบเลือกคำตอบที่ผิดให้ 0 คะแนน

3.3.2 วิธีการตรวจให้คะแนน  $C + q$  เฉพาะข้อถูกที่มีค่า  $P$  ต่ำกว่า .50 ข้อถูก

นอกนั้นให้ 1 คือถ้าผู้สอบเลือกคำตอบที่ถูกในข้อที่มีค่า  $p$  ต่ำกว่า .50 จะให้  $1+q$  คะแนน ถ้าผู้สอบเลือกคำตอบที่ถูกในข้ออื่น ๆ จะให้ 1 คะแนน และถ้าผู้สอบเลือกคำตอบที่ผิดให้ 0 คะแนน นอกจากนี้ยังได้คัดเลือกเฉพาะข้อที่มีค่าอำนาจจำแนกสูงกว่า + .20 มาจัดทำเป็นชุดย่อยเพื่อตรวจให้คะแนนตามวิธีดังกล่าวอีกครั้ง

4. คำนวณค่าสัมประสิทธิ์ความเชื่อมั่น (reliability coefficient) ของแบบทดสอบแต่ละชุด ซึ่งได้ทำการตรวจให้คะแนนโดยวิธีธรรมดา วิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบและวิธีกำหนดน้ำหนัก  $C + q$  ตามค่า  $q$  ของข้อสอบ โดยใช้สูตร KR-20 (Kuder - Richardson Formula 20) สำหรับวิธีธรรมดา ใช้สูตร Coefficient Alpha ซึ่งแปลงไปเป็นสูตรเฉพาะเพื่อใช้กับวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ และใช้สูตร Coefficient Alpha สำหรับวิธีกำหนดน้ำหนัก  $C+q$  ตามค่า  $q$  ของข้อสอบ ซึ่งมีขั้นตอนการคำนวณโดยละเอียดดังนี้

4.1 คำนวณค่าความเชื่อมั่นของแบบ ทดสอบที่ตรวจให้คะแนน โดยวิธีธรรมดาจากสูตร

$$r_{11} = \frac{k}{k-1} \left[ 1 - \frac{\sum_{i=1}^k p_i q_i}{\sigma^2(x)} \right] \quad (\text{Nunnally, 1967})$$

เมื่อ  $p$  คือสัดส่วนของจำนวนคนตอบถูกในแต่ละข้อ

$q$  คือ  $1-p$

$\sigma^2(x)$  คือความแปรปรวนของคะแนนสอบที่ตรวจ ให้คะแนนโดยวิธีธรรมดา

$k$  คือจำนวนข้อสอบ

4.2 คำนวณค่าความเชื่อมั่นของแบบ ทดสอบที่ตรวจให้คะแนนโดยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบจากสูตร

$$r_{11} = \frac{k}{k-1} \left[ 1 - \frac{\sum_{i=1}^k \sigma_i^2 q_i^2}{\sigma^2 \sum_{i=1}^k q_i x_i} \right] \quad (\text{Masang, 1973})$$

เมื่อ  $\sum_{i=1}^k \sigma_i^2 q_i^2$  เป็นผลรวมความแปรปรวนของข้อสอบ ที่ตรวจให้คะแนน โดยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ  
 $\sigma^2 \sum_{i=1}^k q_i x_i$  เป็นความแปรปรวนของคะแนน สอบที่ตรวจให้คะแนนโดยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ  
 $k$  เป็นจำนวนข้อสอบ

4.3 ค่าความเชื่อมั่นของแบบ ทดสอบ ที่ตรวจให้คะแนน โดยวิธีรวมค่ารวมกับวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ ( $C + q$  ตามค่า  $q$  ของข้อสอบ) จากสูตร

$$r_{11} = \frac{k}{k-1} \left[ 1 - \frac{\sum_{i=1}^k \sigma_i^2}{\sigma^2(x)} \right] \quad (\text{Nunnally, 1967})$$

$\sum_{i=1}^k \sigma_i^2$  เป็นผลรวมความแปรปรวนของข้อสอบแต่ละข้อที่ตรวจให้คะแนน  $C + p$  ตามค่า  $q$  ของข้อสอบ

$\sigma^2(x)$  เป็นความแปรปรวนของคะแนน สอบที่ได้จากการตรวจให้คะแนน  $C + q$  ตามค่า  $q$  ของข้อสอบ

$k$  เป็นจำนวนข้อสอบ

5. เปรียบเทียบค่าความเชื่อมั่นของคะแนน สอบที่ได้จากการตรวจให้คะแนนวิธีรวมค่า วิธีกำหนด

น้ำหนักตามค่า  $q$  ของข้อสอบและวิธีกำหนดน้ำหนัก  $C + q$  ตามค่า  $q$  ของข้อสอบ

ความเชื่อมั่นสูงที่สุดไม่ว่าจะตรวจให้คะแนนทั้งฉบับ หรือจัดชุดโดยคัดเลือกเฉพาะข้อที่มีค่า  $r$  สูงกว่า .20

2. ในเรื่องเงื่อนไขของแบบทดสอบที่จะทำหาค่าความเชื่อมั่นของคะแนนสอบที่ตรวจให้คะแนน "วิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ" "และวิธีกำหนดน้ำหนัก  $C+q$  ตามค่า  $q$  ของข้อสอบ" มีค่าสูงกว่า "วิธีรวมค่า" ถ้าตรวจสอบจากสมมติฐานการวิจัยจะพบว่า

### ผลการวิจัย

ผลการวิจัยสรุปได้ดังนี้คือ

1. เมื่อเปรียบเทียบค่าความเชื่อมั่นของคะแนนสอบที่ตรวจให้คะแนนด้วยวิธีรวมค่า วิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ และวิธีกำหนดน้ำหนัก  $C+q$  ตามค่า  $q$  ของข้อสอบ พบว่า สำหรับแบบทดสอบการฟัง แบบทดสอบไวยากรณ์ และแบบทดสอบการอ่านวิธีการตรวจให้คะแนน "วิธีรวมค่า" ให้ค่าความเชื่อมั่นสูงที่สุด ไม่ว่าจะตรวจให้คะแนนทั้งฉบับหรือจัดชุดโดยคัดเลือกเฉพาะข้อที่มีค่า  $r$  สูงกว่า .20 แต่สำหรับแบบทดสอบคำศัพท์ วิธีการตรวจให้คะแนน "วิธีให้  $q$  เฉพาะข้อถูก" ให้ค่า

2.1 สำหรับแบบทดสอบที่มีจำนวนข้อยาก เป็นจำนวนเกินกว่าร้อยละ 50 จะทำหาค่าความเชื่อมั่นของคะแนนสอบที่ตรวจให้คะแนนด้วยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ มีค่าสูงกว่าที่ตรวจด้วยวิธีรวมค่าเพียงกรณีเดียว คือแบบ ทดสอบคำศัพท์

2.2 สำหรับแบบทดสอบที่ข้อสอบแต่ละ

ข้อมีค่าอำนาจจำแนก ( $r$ ) เกินกว่า .20 พบว่า เมื่อแบบทดสอบมีจำนวนข้อยากเป็นจำนวนเกินกว่าร้อยละ 50 จะทำให้ค่าความเชื่อมั่นของคะแนน สอบที่ตรวจให้คะแนนด้วยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบ มีค่าสูงกว่าที่ตรวจด้วยวิธีธรรมดาเพียงกรณีเดียวคือแบบทดสอบค่าศัพท์

### อภิปรายผลและข้อเสนอแนะ

จากผลงานวิจัยครั้งนี้ สามารถสรุปเงื่อนไขบางประการที่จะทำให้ค่าความเชื่อมั่นของคะแนนสอบที่ตรวจ ด้วยวิธีกำหนดน้ำหนักตามค่า  $q$  ของข้อสอบและวิธีกำหนดน้ำหนัก  $C+q$  ตามค่า  $q$  ของข้อสอบมีค่าสูงกว่าที่ตรวจด้วยวิธีธรรมดาว่าเกี่ยวกับเงื่อนไขต่อไปนี้คือ

1. องค์ประกอบเกี่ยวกับจำนวนข้อสอบ สรุปได้ว่าแบบทดสอบต้องมีจำนวนน้อยข้อและต้องมีจำนวน

ข้อยากเกินกว่าร้อยละ 50 จากทั้งหมด

2. องค์ประกอบเกี่ยวกับดัชนีความยากของข้อสอบ พบว่าข้อสอบที่ยากแต่ละข้อต้องมีอำนาจจำแนกตั้งแต่ .20 ขึ้นไป จึงจะช่วยให้ความเชื่อมั่นสูง

3. องค์ประกอบเกี่ยวกับการเดาและความคลาดเคลื่อนในการวัด จะส่งผลทำให้ความเชื่อมั่นของแบบทดสอบต่ำกว่าที่ควรจะเป็น

4. องค์ประกอบเกี่ยวกับสูตรที่ใช้ในการคำนวณ ทั้งนี้เนื่องจากสูตร Coefficient Alpha เกี่ยวข้องกับความแปรปรวนของคะแนนสอบเมื่อค่าความแปรปรวนเพิ่มขึ้นจะทำให้ความเชื่อมั่นลดลง และในงานวิจัยครั้งนี้ เมื่อตรวจให้คะแนนด้วยวิธีต่าง ๆ แล้ว ทำให้ค่าความแปรปรวนของคะแนนสอบ เพิ่มขึ้น จึงทำให้ความเชื่อมั่นลดลง ดังนั้นสำหรับผู้ที่จะทำการวิจัยต่อไป ควรใช้วิธีแบ่งครึ่งข้อสอบ (Split-Half Method)

## เอกสารอ้างอิง

ชวาล แพ้วัดกุล. เทคนิคการวัดผล กรุงเทพฯ : วัฒนาพานิช 2518

Bejar, I.I. and Weiss, D.J. A Comparison of empirical differential option weighting scoring procedures as a function of inter-item correlation. **Educational and Psychological Measurement.** 37 : 335-340. 1977

Echternacht, G. Reliability and validity of item option weighting schemes. **Educational and Psychological Measurement.** 36 : 301-309. 1976

Lord, F.M. and Novick, M.R. **Statistical Theories of Mental Test Scores.** Mass : Addison Wesley. 1968

Masang, B **Some Properties of Test Scores Derived by Weighting Items According to their Difficulty.** Doctoral Dissertation, The Florida State University. 1973

Nunnally, J.C. **Psychometric Theory.** New York : McGraw-Hill. 1967

Stanley, J.C. and Wang, M.D. Weighting test items and test-item options. an overview of the analytical and empirical literature. **Educational and Psychological measurement.** 30 : 21-35. 1970

Wand, M.D. and Stanley, J.C. Differential weighting : a review of methods and empirical studies. **Review of Educational Research.** 40 (5) : 663-706. 1970

Whitla, D.K. **Handbook of Measurement and Assessment in Behavioral Sciences.** Reading, Mass : Addison-Wesley. 1968