

# มโนทัศน์เบื้องต้นของการเชื่อมโยงคะแนนระหว่างแบบสอบ (Basic Concept of Test Linking Scores)

ดร.พัชรี จันทร์เพ็ง\*

## บทนำ

สถานการณ์การทดสอบในปัจจุบันมีความจำเป็นที่จะต้องใช้แบบสอบหลายฉบับในการวัดครั้งเดียวกัน หรือใช้แบบสอบหลายฉบับในการประเมินการเรียนรู้ในเวลาที่ต่างกัน ทำให้ต้องมีการเปรียบเทียบผลที่ได้จากแบบสอบต่างฉบับ ให้อยู่ในมาตราวัดหรือสเกลเดียวกัน โดยเฉพาะอย่างยิ่งแบบสอบมาตรฐาน ซึ่งมีเป้าหมายสำคัญของการทดสอบเพื่อให้เกิดความยุติธรรมเกิดความเที่ยงธรรม และให้ข้อมูลสารสนเทศที่เป็นประโยชน์เกี่ยวกับผู้สอบ ดังนั้นจึงต้องมีการพัฒนาแบบสอบเพื่อใช้วัดความสามารถหรือทักษะของผู้สอบอย่างต่อเนื่อง ซึ่งคะแนนที่ได้จะเป็นข้อมูลสำคัญในการตัดสินใจทั้งระดับรายบุคคล ระดับสถาบัน รวมถึงระดับชาติ เพื่อนำไปเชื่อมโยงกับข้อมูลอื่น ซึ่งผลที่ได้ต้องมีความเสมอภาค และสามารถเปรียบเทียบกันได้ของคะแนน โดยถือว่าเป็นประเด็นสำคัญของสถานการณ์การทดสอบที่ได้รับความสนใจทั้งในอดีตและปัจจุบัน โดยเฉพาะ ความหลากหลายของรูปแบบการวัดที่ต่างกันแต่ต้องการวัดในสิ่งเดียวกัน (Kolen, 2001) ขณะเดียวกันการรักษาความลับของแบบสอบ (Security of Test) ทั้งที่มีการดำเนินการสอบเวลาเดียวกันหรือแตกต่างกัน ถือเป็นปัจจัยที่ควรให้ความสำคัญ เช่น การทดสอบทางการศึกษาระดับชาตินี้พื้นฐาน (O-NET) การทดสอบทางการศึกษาระดับชาตินี้ขั้นสูง (A-NET) การทดสอบวัดความสามารถทางด้านต่างๆ การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAT) หรือแม้กระทั่งการทดสอบวัดผลสัมฤทธิ์ทางการเรียนทั้งในระดับสถานศึกษาและเขตพื้นที่การศึกษา เป็นต้น อย่างไรก็ตาม

การเปรียบเทียบสเกลของแบบสอบเพื่อให้สามารถแปลความอย่างมีความหมาย เช่น ในกรณีที่มีระดับชั้นแตกต่างกัน ชนิดของแบบสอบแตกต่างกัน ดังที่กล่าวมาแล้วข้างต้น จำเป็นอย่างยิ่งที่ต้องอาศัยเทคนิคของการเชื่อมโยงคะแนนระหว่างแบบสอบ (Test Linking Scores)

ด้วยตระหนักในความสำคัญของการเชื่อมโยงคะแนนระหว่างแบบสอบ ผู้เขียนในฐานะผู้สอนวิชาทางด้านการวัดและประเมินผลการศึกษา ได้ติดตามความก้าวหน้าและมีโอกาสทำวิจัยเกี่ยวกับเทคนิคการเชื่อมโยงคะแนน บทความนี้เป็นผลจากการประมวลความรู้เกี่ยวกับแนวคิดพื้นฐานของการเชื่อมโยงคะแนนที่ได้จากการแลกเปลี่ยนเรียนรู้กับอาจารย์ที่มีความรู้และมีความเชี่ยวชาญ จากประสบการณ์ในการทำวิจัย รวมถึงการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง สาระในบทความนี้อาจไม่ครอบคลุมเทคนิคการเชื่อมโยงคะแนนระหว่างแบบสอบที่ได้รับการพัฒนาใหม่ๆ ทั้งหมด แต่ผู้เขียนหวังว่าจะมีสาระครอบคลุมการเชื่อมโยงคะแนนระหว่างแบบสอบมากพอที่จะนำไปเป็นแนวคิดสำคัญอันจะนำไปสู่การปฏิบัติ เนื่องจากศัพท์คำว่า “การเชื่อมโยงคะแนนระหว่างแบบสอบ” (Test Linking Scores) เป็นคำที่มีความหมายหมายกว้าง ครอบคลุมรูปแบบของการเชื่อมโยงคะแนนทั้งหมด ซึ่งถ้าหากไม่สามารถเข้าใจถึงความแตกต่างของรูปแบบการเชื่อมโยงคะแนนในแต่ละวิธี อาจทำให้เกิดมโนทัศน์ที่คลาดเคลื่อน อันนำไปสู่การนำรูปแบบการเชื่อมโยงคะแนนไปใช้ในสถานการณ์ที่ไม่เหมาะสม และผิดไปจากหลักความเป็นจริง ดังนั้น

\*อาจารย์ สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

ผู้เขียนจึงนำเสนอบทความแบ่งเป็นสามตอน ตอนแรกกล่าวถึงแนวคิดพื้นฐานของการเชื่อมโยงคะแนน ตอนที่สองให้รายละเอียดเกี่ยวกับเกณฑ์ที่ใช้ในการจำแนกเทคนิคการเชื่อมโยงคะแนน และในตอนสุดท้ายเป็นการสรุปรวมลักษณะภาพอนาคตของเทคนิคการเชื่อมโยงคะแนนที่สามารถนำไปประยุกต์ใช้

### แนวคิดพื้นฐานของการเชื่อมโยงคะแนน

คำว่า “การเชื่อมโยงคะแนนระหว่างแบบสอบ” ตรงกับศัพท์ภาษาอังกฤษว่า “Test Linking Scores” สำหรับประเทศไทย หลายคนคงไม่คุ้นกับศัพท์คำนี้ แต่จะคุ้นกับศัพท์คำว่า “การปรับเทียบคะแนนระหว่างแบบสอบ” หรือตรงกับภาษาอังกฤษว่า “Test Score Equating” เนื่องจากเป็นเทคนิคการแปลงคะแนนของแบบสอบต่างฉบับที่มุ่งวัดคุณลักษณะเดียวกันให้เป็นคะแนนที่สมมูลกัน (ศิริชัย กาญจนवासี, 2550) นอกจากนี้ Kolen และ Brennan (2004) ยังกล่าวไว้ว่าการปรับเทียบคะแนนเป็นวิธีการเชื่อมโยงคะแนนระหว่างแบบสอบสองฉบับวิธีหนึ่งซึ่งเป็นเทคนิคหนึ่งของการเชื่อมโยงคะแนน โดยมีการพิจารณาปรับเทียบแบบสอบสองฉบับที่มีค่าความยากแตกต่างกัน แต่การปรับเทียบคะแนนมิได้ปรับเทียบเนื้อหาของแบบสอบสองฉบับที่มีความแตกต่างกัน

“Equating adjusts for differences in difficulty, not differences in content.” (Kolen & Brennan, 2004)

เทคนิคการปรับเทียบคะแนน (Test Equating) นิยมนำมาใช้ในการศึกษาพัฒนาการภาคตัดขวางของผู้เรียนที่อยู่คนละระดับชั้น แต่อยู่ในภาคการศึกษาเดียวกันหรือปีเดียวกัน ให้เป็นคะแนนที่สมมูลกันเพื่อทำให้สามารถเปรียบเทียบคะแนนกันได้ และมีแนวโน้มที่จะนำไปใช้ในการปรับเทียบคะแนนเฉลี่ยสะสมของนักเรียนในแต่ละโรงเรียนที่มีมาตรฐานแตกต่างกัน และทำการปรับคะแนนให้อยู่ในสเกลที่สามารถเปรียบเทียบกันได้อย่างเที่ยงธรรม ซึ่งกำลังอยู่ในช่วงของการศึกษาวิจัย อย่างไรก็ตามมีคนจำนวนไม่น้อยที่ไม่ทราบแนวคิดที่ถูกต้องของศัพท์ดังกล่าว และเมื่อทำการศึกษาเอกสาร

และงานวิจัยที่เกี่ยวข้อง อาทิ Journal of Education Measurement, Applied Psychological Measurement, Applied Measurement in Education, Educational Measurement: Issues and Practice, Journal of Applied Measurement และ Psychometrika ซึ่งเป็นวารสารทางการวัดผลการศึกษาและจิตมิติ ได้เผยแพร่ประเด็นสำคัญของหัวข้อการศึกษาวิธีการเชื่อมโยงคะแนนในรูปแบบต่างๆ ซึ่งในช่วงเกือบสิบปีที่ผ่านมาไม่นิยมใช้ศัพท์คำว่า “การปรับเทียบคะแนน” (Test Score Equating) แต่ใช้คำกว้างๆ ว่า “การเชื่อมโยงคะแนนระหว่างแบบสอบ” (Test Linking Scores) ทั้งที่ใช้รูปแบบของการปรับเทียบคะแนนไว้อย่างชัดเจน

นอกจากนี้ในปัจจุบันนักวัดผลทางการศึกษาต้องเผชิญกับภารกิจที่ค่อนข้างท้าทายในเรื่องเกี่ยวกับการเชื่อมโยงคะแนนระหว่างแบบสอบสองฉบับที่วัดคุณลักษณะที่ต่างกัน (Different Tests) โดยมีผู้เชี่ยวชาญหลายท่านได้มีการถกเถียงกันมากถ้านำผลของคะแนนที่ได้มาจากแบบสอบสองฉบับที่วัดคุณลักษณะที่ต่างกันมาปรับเทียบคะแนน (Equating) เมื่อเป็นเช่นนี้จึงมีคำถามเกิดขึ้นว่า “ถ้าต้องการเชื่อมโยงระหว่างคะแนนในกรณีที่เป็นแบบสอบทั้งสองฉบับนั้น **ขาดความเป็นคู่ขนาน** (เช่น การสร้างข้อสอบที่มีความแตกต่างกันทางด้านโครงสร้าง/รูปแบบของแบบสอบ) ยังคงใช้การปรับเทียบคะแนน (Equating) ได้อยู่หรือไม่ ?” (สุวิมล กฤษศยาสา, 2551) จากคำถามที่เกิดขึ้นข้างต้นทำให้เกิดคำว่า **การเชื่อมโยงคะแนน** (Linking Scores) และมีการพัฒนาเทคนิคการเชื่อมโยงคะแนนเพิ่มเติมเพื่อให้สอดคล้องกับสภาพความเป็นจริงมากขึ้น เช่น เทคนิคการปรับคะแนนให้เป็นมาตรฐาน (Calibration) ซึ่ง Angoff (1971, cited in Kolen, 2004) ได้นิยามศัพท์คำว่า “การปรับคะแนนให้เป็นมาตรฐาน (Calibration)” ว่าเป็นการเชื่อมโยงคะแนนจากแบบสอบที่วัดโครงสร้างเดียวกันแต่มีระดับความยากของแบบสอบทั้งสองฉบับแตกต่างกัน เช่น การเชื่อมโยงคะแนนของแบบสอบวัดผลสัมฤทธิ์ระดับประถมศึกษาที่ถูกออกแบบมาเพื่อใช้ในระดับชั้นต่างๆ แต่ใช้เทคนิคดังกล่าวเข้ามาช่วยแปลงให้เป็นคะแนนมาตรฐาน ซึ่งโดยทั่วไปนิยมเรียกว่าการสร้างสเกลแนวตั้ง (Vertical scaling) นอกจากนี้ผู้เชี่ยวชาญบางท่านกล่าว

ว่า “แบบสอบที่วัดความรู้ที่มีลักษณะที่สัมพันธ์กันหรือมีส่วนร่วมกันอย่างใดอย่างหนึ่งสามารถนำคะแนนมาเปรียบเทียบกันได้โดยตรง ก่อให้เกิดเทคนิคของการเชื่อมโยงคะแนนเพิ่มเติมในกรณีแบบสอบที่วัดมีคุณลักษณะต่างกัน (Different Tests) ทั้งสองฉบับซึ่งจำเป็นต้องอาศัยกระบวนการทางสถิติเข้ามาช่วย เช่น เทคนิคการฉายภาพ (Projection) รวมถึงเทคนิคการสร้างคะแนนความสอดคล้อง (Concordances) โดยเป็นวิธีการในการเชื่อมโยงคะแนนหรือการแปลงคะแนนระหว่างแบบสอบสองฉบับที่มีความแตกต่างกันของแบบสอบทั้งในระดับของเนื้อหาที่นำมาใช้กำหนดคุณลักษณะเฉพาะและค่าความยากของแบบสอบ ซึ่ง Doran (2004) ถือเป็นผู้ที่บุกเบิกวิธีการดังกล่าว อย่างไรก็ตาม Angoff กล่าวว่าในการเชื่อมโยงคะแนนที่ได้จากแบบสอบที่มีโครงสร้างต่างกันผลของคะแนนที่ได้จะมีลักษณะของการผันแปรตามกลุ่มผู้สอบ (Variance Across Group)

จากที่กล่าวมาข้างต้นไม่มีเอกสารที่ระบุไว้อย่างชัดเจนเกี่ยวกับความหมายของการเชื่อมโยงคะแนน แต่เมื่อพิจารณาถึงเทคนิคของการเชื่อมโยงคะแนนในแต่ละวิธีข้างต้นสามารถสรุปได้ว่า

“การเชื่อมโยงคะแนนจากแบบสอบสองฉบับเป็นการดำเนินการโดยนำคะแนนสอบจากแบบสอบสองฉบับทั้งที่มีลักษณะของการวัดโครงสร้าง และเนื้อหาเดียวกันและต่างกัน มาเชื่อมโยงกันด้วยวิธีการสถิติ และสามารถนำคะแนนมาเปรียบเทียบกันได้อย่างสมเหตุสมผลและมีความยุติธรรม”

### เกณฑ์ที่ใช้ในการจำแนกเทคนิคการเชื่อมโยงคะแนน

แนวคิดเกี่ยวกับ “การเชื่อมโยงคะแนนระหว่างแบบสอบ” (Test Linking Scores) มีอยู่หลายลักษณะ Mislevy (1992 cited in Kolen & Brennan, 2004) & Linn (1993 cited in Kolen & Brennan, 2004) ได้จำแนกประเภทของการเชื่อมโยงคะแนนระหว่างแบบสอบ 2 ฉบับ โดยใช้ลักษณะร่วมกันของโครงสร้างเนื้อหาและความแกร่ง (Robustness) จากผลของคะแนนที่ได้จากการเชื่อมโยงเป็นเครื่องมือในการจำแนก

แบ่งออกเป็น 4 ชนิด คือ 1) การปรับเทียบคะแนน (Equating), 2) การปรับคะแนนให้เป็นมาตรฐาน (Calibration), 3) การฉายภาพ (Projection) และ 4) การปรับค่าทางสถิติ (Statistical Moderation) ซึ่งกระบวนการเชื่อมโยงคะแนนระหว่างแบบสอบ 2 ฉบับที่สำคัญ คือ การประเมินทั้งสองฉบับต้องมีเป้าหมายเหมือนกัน และมีเนื้อหาครอบคลุมลักษณะที่วัดเดียวกัน นอกจากนี้ Feuer และคณะ (1999) ได้เสนอเกณฑ์การจำแนกรูปแบบการเชื่อมโยงคะแนนเพิ่มเติมโดยจำแนกการเชื่อมโยงคะแนนระหว่างแบบสอบตามแบบแผนในการสร้างและพัฒนาแบบสอบ โดยยังคงไว้ซึ่งรูปแบบในการเชื่อมโยงทั้ง 4 ชนิดดังกล่าวข้างต้นที่เป็นไปตามแนวคิดของ Mislevy (1992) & Linn (1993) ซึ่งเกณฑ์ที่ Feuer และคณะ (1999) ใช้ในการจำแนกแบ่งเป็น 3 ขั้นตอน คือ

**ขั้นตอนที่ 1 การกำหนดโครงสร้างของแบบสอบ (Framework Definition)** สามารถประเมินจากขั้นตอนการกำหนดโครงสร้างของแบบสอบทั้งสองฉบับ โดยพิจารณาจากการระบุถึงขอบเขต (Scope) ปริมาณ (Extent) ของกรอบมวลพฤติกรรม (Domain) เพื่อให้ทราบว่าผู้เรียนมีความสามารถอะไรบ้างและมีความรู้ระดับใด

**ขั้นตอนที่ 2 การกำหนดคุณลักษณะเฉพาะของแบบสอบ หรือการสร้างผังข้อสอบ (Test Specification or Blueprint)** เป็นการเสนอรายละเอียดของการทดสอบแต่ละครั้งว่าวัดเนื้อหาอะไร หรือจุดมุ่งหมายเชิงพฤติกรรมที่ต้องการวัดจากผู้เรียน ขอบเขตของเนื้อหาวิชา ตลอดจนมีการกำหนดน้ำหนักความสำคัญ/สัดส่วนของจำนวนข้อกระทง/ข้อสอบที่จะสร้างขึ้นเป็นแบบสอบที่ใช้สำหรับวัดพฤติกรรมตามขอบเขตของเนื้อหาวิชาที่ต้องการทดสอบแต่ละครั้ง รวมถึงการตรวจให้คะแนนด้วย

**ขั้นตอนที่ 3 การคัดเลือกข้อสอบ (Item Selection)** เป็นการพิจารณาจากข้อสอบที่ถูกคัดเลือกมาบรรจุไว้ในแบบสอบ ข้อสอบที่คัดเลือกมาควรเป็นตัวแทนที่ดีของผังข้อสอบให้มากที่สุดเท่าที่จะเป็นไปได้

โดยปกติในการกำหนดโครงสร้างของแบบสอบ (Framework Definition) ส่วนมากจะมีการกำหนดใน

ลักษณะที่เป็นส่วนหนึ่งของกรอบมวลพฤติกรรม (Domain) เช่น ถ้าต้องการวัดกรอบมวลพฤติกรรมด้านการอ่านของผู้เรียน ฉะนั้นในการสร้างแบบสอบควรมีการกำหนดโครงสร้างให้ได้ตามกรอบมวลพฤติกรรมที่กำหนดไว้ให้ได้มากที่สุด

จากที่กล่าวมาข้างต้นเกี่ยวกับกรอบแนวคิดของกรอบมวลพฤติกรรมของแบบสอบสามารถสรุปได้ตามแนวคิดของ Feuer และคณะ (1999) ที่ได้เสนอเกณฑ์จำแนกการเชื่อมโยง 3 ลักษณะ คือ

**เกณฑ์ที่ 1** แบบสอบทั้งสองฉบับมีการกำหนดโครงสร้างของแบบสอบและมีการกำหนดคุณลักษณะเฉพาะของแบบสอบเดียวกัน

**เกณฑ์ที่ 2** แบบสอบทั้งสองฉบับมีการกำหนดโครงสร้างของแบบสอบที่มีลักษณะเดียวกัน แต่มีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่ต่างกัน

**เกณฑ์ที่ 3** แบบสอบทั้งสองฉบับมีการกำหนดโครงสร้างของแบบสอบและมีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่ต่างกัน

กรอบแนวคิดเกี่ยวกับเกณฑ์ในการจำแนกการเชื่อมโยงคะแนนของ Feuer และคณะ (1999) สามารถสรุปได้ดังตารางที่ 1

**ตารางที่ 1** การจำแนกประเภทของการเชื่อมโยงคะแนนตามเกณฑ์ของ Feuer และคณะ (1999)

ประเภทของการเชื่อมโยงคะแนน	การกำหนดโครงสร้างของแบบสอบ (Framework)	การกำหนดคุณลักษณะเฉพาะของแบบสอบของแบบสอบ (Test Specifications)
การปรับเทียบคะแนน (Equating)	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน
คะแนนความสอดคล้อง (Concordance/Moderation)	มีลักษณะเดียวกัน	มีลักษณะแตกต่างกัน
การใช้แบบสอบเป็นตัวปรับอยู่ภายใน (Moderation)	มีลักษณะแตกต่างกัน	มีลักษณะแตกต่างกัน

ต่อมาในปี ค.ศ. 2004 Kolen และ Brennan มีการขยายแนวคิดเกี่ยวกับความเหมือนและคล้ายคลึงกันของแบบสอบ โดยเสนอหลักเกณฑ์การเลือกชนิดการเชื่อมโยงคะแนน โดยพิจารณาจากความคล้ายคลึงกันของแบบสอบ ใน 4 ลักษณะ คือ

1. การนำผลของคะแนนมาสรุปอ้างอิง (Inferences) เป็นการตอบคำถามว่า “ปริมาณ (extent) ของความคล้ายคลึงกันระหว่างแบบสอบ 2 ฉบับ อยู่ในระดับใด?” นั่นคือ พิจารณาถึงความมีเป้าหมายในการวัดร่วมกันระหว่างแบบสอบสองฉบับว่าอยู่ในระดับใด

2. โครงสร้าง/คุณลักษณะ (Constructs) ของแบบสอบทั้งสองฉบับ เป็นการตอบคำถามว่า “แบบสอบสองฉบับ นั้นวัดคุณลักษณะเดียวกันหรือไม่?” นั่นคือ พิจารณาถึงคะแนนจริงของแบบสอบ 2 ฉบับที่เกี่ยวข้องกัน สำหรับในบริบทของการเชื่อมโยงคะแนน อาจมีคุณลักษณะหรือโครงสร้างร่วมกันในระดับหนึ่ง แต่มิได้เป็นคุณลักษณะหรือโครงสร้างเดียวกัน

3. กลุ่มประชากรผู้สอบ (Populations) เป็นการตอบคำถามว่า “แบบสอบสองฉบับถูกออกแบบเพื่อวัดผู้สอบกลุ่มเดิมหรือไม่?”

4. เงื่อนไข/สถานการณ์ในการวัด (Measurement Conditions) เป็นการตอบคำถามว่า “แบบสอบสองฉบับมีเงื่อนไขในการวัดร่วมกันหรือไม่?” เงื่อนไขในการวัด เช่น ความยาวของแบบสอบ รูปแบบของแบบสอบ และการบริหารการสอบ ซึ่งในทฤษฎีการสรุปอ้างอิง (Generalizability Theory) เรียกเงื่อนไขในการวัดนี้ว่า Facets นั้นเอง อย่างไรก็ตามการกำหนดแบบแผนของแบบสอบก็เป็นส่วนหนึ่งของเงื่อนไขในการวัดโดยพิจารณาจากลักษณะความสำคัญ/ความร่วมมือกันของแบบสอบทั้ง 4 ประการที่กล่าวมาข้างต้น

สำหรับในกรณีของการเชื่อมโยงคะแนน Kolen & Brennan (2004) กล่าวว่าไม่ควรพิจารณาเฉพาะระดับความคล้ายคลึงกันของแบบสอบเท่านั้น แต่ควรพิจารณาถึงหลักเกณฑ์อื่นๆ ด้วย จากแนวคิดของ Feuer และคณะ (1999) ที่ได้เสนอเกี่ยวกับการวัดในกรณีที่มีแบบสอบขาดคุณลักษณะของความร่วมมือกัน และจาก

การจำแนกประเภทตามแนวคิดของ Mislevy (1992) & Linn (1993) ทั้ง 2 แนวคิดใช้เกณฑ์ในการพิจารณาความร่วมมือกันของแบบสอบ คือ โครงร่างของแบบสอบที่มีลักษณะเดียวกันและมีการกำหนดคุณลักษณะเฉพาะของแบบสอบที่มีลักษณะเดียวกัน โดยที่โครงร่างของแบบสอบ คือแนวคิดในการตอบคำถามว่า แบบสอบนั้นมีโครงสร้างหรือคุณลักษณะที่คล้ายคลึงกันหรือไม่? ส่วนการกำหนดคุณลักษณะเฉพาะของแบบสอบ คือแนวคิดในการตอบคำถามว่า คุณลักษณะของแบบสอบทั้ง 2 ฉบับที่ต้องการวัดมีความคล้ายคลึงกันหรือไม่?

เมื่อนำแนวคิดของ Mislevy (1992) & Linn (1993), Feuer และคณะ (1999), Kolen & Brennan (2004) มาสังเคราะห์ร่วมกัน สามารถจำแนกประเภทของการเชื่อมโยงคะแนนในรูปแบบต่างๆ พร้อมทั้งหลักเกณฑ์ในการพิจารณา ดังตารางต่อไปนี้

ตารางที่ 2 ประเภทของการเชื่อมโยงคะแนนตามแนวคิดของ Mislevy (1992) & Linn (1993) และระดับของความคล้ายคลึงกันระหว่างแบบสอบสองฉบับ

ประเภท	การสรุปอ้างอิง (Inferences)	โครงสร้าง/คุณลักษณะ (Constructs)	กลุ่มประชากรผู้สอบ (Populations)	คุณลักษณะของแบบสอบที่ใช้ในการวัด (Measurement characteristics)
การปรับเทียบคะแนน (Equating)	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน
การสร้างสเกลในแนวตั้ง (Vertical Scaling)	มีลักษณะเดียวกัน	มีลักษณะเดียวกัน/ คล้ายคลึงกัน	ไม่มีความคล้ายคลึงกัน	มีลักษณะเดียวกัน/ คล้ายคลึงกัน
การฉายภาพ (Projection)	มี/ไม่มีความคล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน	มีความคล้ายคลึงกัน	ไม่มีความคล้ายคลึงกัน
การปรับค่าทางสถิติ (Statistical Moderation)	มี/ไม่มีความคล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน	มี/ไม่มีความคล้ายคลึงกัน	ไม่มีความคล้ายคลึงกัน

จากที่กล่าวมา สามารถจำแนกรูปแบบการเชื่อมโยงคะแนนได้เป็น 4 รูปแบบ คือ

### 1. การปรับเทียบคะแนน (Test Equating)

เป็นศัพท์เฉพาะที่นักจิตมิติ (Psychometrician) นำมาใช้ในกระบวนการวัดและประเมินผลทางพฤติกรรมศาสตร์และทางการศึกษา ช่วงแรกนักวัดผลโดยส่วนมากไม่ได้ให้ความสนใจเท่าที่ควร จะศึกษาเฉพาะกลุ่มนักจิตมิติเท่านั้น (Kolen & Brennan, 2004) ต่อมาในปี ค.ศ. 1980 ได้มีเอกสารเกี่ยวกับเรื่องนี้เพิ่มขึ้นทำให้นักวัดผลเริ่มเห็นความสำคัญและนำมาใช้ให้เป็นประโยชน์กับการทดสอบมากยิ่งขึ้น โดยเฉพาะอย่างยิ่งในช่วง 15 ปีต่อมามีการเพิ่มจำนวนและความหลากหลายของโปรแกรมการสอบที่มีการใช้แบบสอบหลายฉบับ ผู้รับผิดชอบในการทำแบบสอบมีความต้องการที่จะอ้างอิงบทบาทของการปรับเทียบคะแนนเพื่อวิจารณ์การทดสอบ ความเคลื่อนไหวในการตรวจสอบและติดตามผลการจัดการศึกษา และประเด็นความยุติธรรมเกี่ยวกับการทดสอบที่ปรากฏชัดเจนขึ้น ทำให้ศาสตร์ด้านการปรับเทียบคะแนนได้รับการศึกษาและพัฒนาขึ้นตามลำดับ โดยการปรับเทียบคะแนนต้องอาศัยการออกแบบการใช้แบบสอบและวิธีการที่เหมาะสม โดยแบบสอบที่ใช้ต้องวัดคุณลักษณะเด่นเดียวเหมือนกัน (Unidimensionality) มีความเป็นคู่ขนานด้านเนื้อเรื่อง โครงสร้าง รูปแบบชนิดข้อสอบ เวลาที่ใช้ และมีความเที่ยงสูงทัดเทียมกันสำหรับวิธีการปรับเทียบที่ใช้จะต้องมีความเสมอภาค (Equity) ไม่ผันแปรตามกลุ่ม (Invariance Across Groups) และมีความสมมาตร (Symmetry) นั่นคือแบบสอบฉบับต่างๆ ที่นำมาปรับเทียบคะแนนกัน ต้องไม่ก่อให้เกิดผลที่แตกต่างกับผู้สอบ กล่าวโดยสรุปคะแนนที่ผ่านการปรับเทียบแล้วจะต้องมีการแจกแจงเหมือนกับคะแนนที่กลุ่มนั้นทำได้จริง ซึ่งคะแนนที่ได้ไม่ผันแปรไม่ว่าจะปรับเทียบจากกลุ่มใด จึงสามารถนำผลไปใช้กับกลุ่มอื่นๆ ได้ และผลการปรับเทียบคะแนนจะต้องเหมือนกัน ไม่ว่าจะเป็นการปรับเทียบคะแนนจากแบบสอบฉบับที่ 1 ไปยังแบบสอบฉบับที่ 2 หรือปรับเทียบคะแนนจากแบบสอบฉบับที่ 2 ไปยังฉบับที่ 1 (ศิริชัย กาญจนวาสี, 2550) ซึ่งคะแนนสามารถแลกเปลี่ยนกันได้ (Interchangeable)

2) การปรับคะแนนให้เป็นมาตรฐาน (Calibration) หรือในทางปฏิบัตินิยมใช้คำว่า การสร้างสเกลแนวตั้ง (Vertical Scaling) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีลักษณะร่วมกันของโครงสร้างเนื้อหา แต่มีระดับความยากที่แตกต่างกัน ถือได้ว่าเป็นการเชื่อมต่อที่มีเงื่อนไขน้อยกว่าการปรับเทียบคะแนน ซึ่งผลที่ได้จากการแปลงคะแนนจะไม่มีคุณสมบัติของการสมมาตร นั่นคือ การเชื่อมโยงคะแนนจากแบบสอบฉบับที่ 1 ไปยังแบบสอบฉบับที่ 2 ค่าของคะแนนที่แปลงได้จะไม่เท่ากับการเชื่อมโยงคะแนนจากแบบสอบฉบับที่ 2 ไปยังฉบับที่ 1 ซึ่งคะแนนจะไม่สามารถแลกเปลี่ยนกันได้ (Interchangeable)

3) การฉายภาพ (Projection) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีโครงสร้างเนื้อหาต่างกัน โดยมีทิศทางของการเชื่อมโยงคะแนนเพียงทิศทางเดียว (Unidirection) จากคะแนนที่ได้จากแบบสอบฉบับหนึ่งไปยังคะแนนของแบบสอบอีกฉบับหนึ่งด้วยวิธีการทำนาย โดยเทคนิคการฉายภาพมีลักษณะที่แตกต่างจากการปรับเทียบคะแนน และการสร้างสเกลแนวตั้ง คือ 1) มีทิศทางของการเชื่อมโยงคะแนนเพียงทิศทางเดียว (Unidirection) 2) จำเป็นต้องมีการออกแบบวิธีการเก็บรวบรวมข้อมูลแบบกลุ่มเดียว (Single Group Design) นั่นคือ ใช้ผู้สอบกลุ่มเดียวทำแบบสอบทั้งสองฉบับ โดยให้ผู้สอบทำแบบสอบฉบับหนึ่งแล้วตามด้วยแบบสอบอีกฉบับหนึ่ง และ 3) ไม่จำเป็นต้องวัดโครงสร้างเนื้อหาเดียวกัน โดยทั่วไปจะนิยมใช้สมการถดถอย (Regressions) เป็นสถิติที่ใช้ในการทำนาย ดังนั้นจึงไม่มีคุณสมบัติของการสมมาตร

4) การปรับด้วยค่าสถิติ (Statistical Moderation) เป็นการเชื่อมโยงคะแนนจากแบบสอบที่มีโครงสร้างเนื้อหาต่างกัน มีตัวแปรปรับ (Moderator Variable) ผ่งอยู่ในแบบสอบแต่ละฉบับ ซึ่งมีความจำเป็นต้องมีการออกแบบการเก็บรวบรวมข้อมูลโดยใช้ผู้สอบกลุ่มเดียว (Single Group Design) เหมือนเทคนิคการฉายภาพ (Projection) อย่างไรก็ตามอาจมีการยืดหยุ่นให้สามารถออกแบบการเก็บรวบรวมข้อมูลโดยใช้รูปแบบผู้สอบกลุ่มสุ่ม (Random Group Design) โดยใช้กลุ่มผู้สอบซึ่งได้จากการสุ่ม แล้วให้ผู้สอบแต่ละ

กลุ่มทำแบบสอบฉบับเดียว และรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกัน (Nonequivalent Group Design) โดยใช้ผู้สอบกลุ่มไม่เท่าเทียมกัน แล้วให้ผู้สอบแต่ละกลุ่มทำแบบสอบฉบับเดียว กลุ่มละฉบับ ซึ่งเทคนิคดังกล่าวไม่เป็นที่นิยมเนื่องจากความน่าเชื่อถือของผลที่ได้จากการเชื่อมโยงคะแนนมีความคลาดเคลื่อนมากกว่าวิธีอื่นอย่างเห็นได้ชัด อย่างไรก็ตามสหรัฐอเมริกาถือเป็นประเทศแรกที่ได้นำมาใช้ในการเปรียบเทียบสเกลของคะแนนที่ได้จากแบบสอบมาตรฐานต่างๆ ซึ่งนำมาใช้จริงในทางปฏิบัติ เช่น การปรับด้วยค่าสถิติของคะแนน ACT (American Colleague Testing) และ SAT I (Scholastic Aptitude Test I) เพื่อให้ได้คะแนนที่เทียบเท่าและให้ผู้เรียนสามารถเลือกสอบแบบสอบมาตรฐานต่างๆ ได้ตามความสมัครใจ (Voluntary National Tests) ในรายวิชาคณิตศาสตร์และด้านภาษา ซึ่งคะแนนที่ได้นำไปใช้ในการเข้าศึกษาต่อในระดับอุดมศึกษา (Kolen and Brennan, 2004) โดยเริ่มเกิดขึ้นในยุคของประธานาธิบดีคลินตัน เป็นต้น

ในการพิจารณาค่าทั้ง 4 ค่าข้างต้น จะมีความแตกต่างกันในข้อตกลงเบื้องต้นที่กำหนดโดยมีความแปรตามข้อตกลงเบื้องต้นลดน้อยลงตามลำดับ การปรับด้วยค่าสถิติมีข้อตกลงเบื้องต้นที่ผ่อนปรนมากที่สุด ส่วนข้อกำหนดของการเทียบคะแนนจะมีข้อตกลงที่เข้มงวดมากที่สุด โดยการเทียบคะแนนเป็นการเชื่อมโยงคะแนนเมื่อข้อสอบวัดในโครงสร้างเนื้อหาเหมือนกัน ในขณะที่การสร้างสเกลแนวตั้งใช้เมื่อโครงสร้างอาจจะไม่จำเป็นต้องเหมือนกันแต่ต้องมีลักษณะที่คล้ายคลึงกัน แต่การฉายภาพและการปรับด้วยค่าสถิติเป็นการเชื่อมโยงคะแนนเมื่อข้อสอบวัดในโครงสร้างเนื้อหาต่างกัน

### อนาคตของการประยุกต์ใช้เทคนิคการเชื่อมโยงคะแนน

เมื่อพิจารณาเทคนิคการเชื่อมโยงคะแนนที่นำเสนอมาทั้งหมด ถือว่าเป็นวิธีการที่มีความสำคัญมากสำหรับนำมาใช้ในบริบทของประเทศไทย ซึ่งสามารถสรุปรวมลักษณะภาพอนาคตของการประยุกต์ใช้เทคนิคการเชื่อมโยงคะแนน ดังนี้

### การปรับเทียบคะแนนเฉลี่ยสะสมของผู้เรียน

การนำผลการเรียนเฉลี่ยสะสมรวมของนักเรียนระดับมัธยมศึกษาตอนปลาย 6 ภาคการศึกษา (GPAX) มาเป็นองค์ประกอบสำคัญในคัดเลือกเข้าศึกษาในสถานศึกษาระดับอุดมศึกษาด้วยระบบกลาง (Admission) โดยในปัจจุบันให้น้ำหนักความสำคัญร้อยละ 20 ในขณะที่แต่ละโรงเรียนจะมีลักษณะของการให้เกรดที่แตกต่างกัน ทั้งปล่อยเกรด ปกติ และกดเกรด ซึ่งสะท้อนให้เห็นว่าแต่ละโรงเรียนมีมาตรฐานการประเมินผลการเรียนรู้ของผู้เรียนที่แตกต่างกัน และระดับคุณภาพของโรงเรียนที่แตกต่างกัน ดังนั้นจึงจำเป็นต้องมีนวัตกรรมการเชื่อมโยงคะแนนโดยอาศัยเทคนิคที่เหมาะสม เช่น การพัฒนาวิธีการปรับเทียบผลการเรียนเฉลี่ยสะสมตามกลุ่มสาระการเรียนรู้ของนักเรียนมัธยมศึกษาตอนปลาย เพื่อให้เป็นมาตรฐานเดียวกัน และสามารถเปรียบเทียบคะแนนกันได้อย่างยุติธรรม (ศิริชัย กาญจนवासี และคณะ, 2551) อย่างไรก็ตาม การนำเทคนิคการปรับเทียบคะแนนมาใช้ต้องมีความระมัดระวังในข้อตกลงเบื้องต้น โดยเฉพาะอย่างยิ่งความเป็นเอกมิติของเนื้อหาที่นำมาใช้ในการปรับเทียบคะแนน และเกณฑ์ที่นำมาในการปรับเทียบต้องมีความเป็นมาตรฐาน ซึ่งถ้ามองในบริบทของประเทศไทยปัจจุบันคะแนนที่สามารถนำมาใช้เป็นเกณฑ์ได้คือ คะแนน O-NET เนื่องจากเป็นผลการสอบได้มาจากแบบสอบมาตรฐานในระดับชาติ

### การเทียบโอนผลการเรียนระหว่างสถานศึกษา

สำหรับกรณีที่ผู้เรียนต้องการย้ายสถานศึกษา จำเป็นอย่างยิ่งที่ต้องมีการเทียบโอนผลการเรียนระหว่างสถานศึกษา เนื่องจากการเทียบโอนผลการเรียนเป็นแนวทางสำคัญที่ทำให้เกิดการศึกษาดลอดชีวิต โดยเป็นกลไกสำคัญในการนำผลการเรียนเดิมมาเป็นส่วนหนึ่งของผลการเรียนรู้สำหรับผู้เรียน ดังนั้นการเทียบโอนผลการเรียนที่มีความแตกต่างของหลักสูตรสถานศึกษาแต่ละแห่งส่งผลให้ผู้เรียนจากสถานศึกษาต่างกันได้รับการเรียนการสอนที่มีความแตกต่างทั้งในด้านของเนื้อหาสาระการเรียนรู้ ตลอดจนงานด้านการวัดและประเมินผลการศึกษา (กระทรวงศึกษาธิการ, 2549) จึงเกิดคำถามประเด็นหนึ่งว่า “ผลการเทียบโอนมีความยุติธรรมและ

เกิดประโยชน์สูงสุดแก่ผู้เรียนหรือไม่?” ซึ่งแนวทางหนึ่ง ที่ช่วยในแก้ปัญหาดังกล่าว คือ การใช้เทคนิคการสร้าง คะแนนความสอดคล้องเข้ามาช่วยแก้ปัญหาดังกล่าว เพื่อนำคะแนนสอบหรือผลการเรียนที่ได้จากการตรวจ ให้คะแนนและการแปลผลคะแนนจากต่างสถานศึกษาที่มีการสร้างแบบสอบโดยยึดโครงสร้างเนื้อหา จุดเน้นของ เนื้อหาและค่าความยากที่แตกต่างกันมาเปรียบเทียบกันได้ พร้อมทั้งสามารถทำให้คะแนนที่แตกต่างกันอยู่บน มาตรฐานเดียวกันได้ (สูวิมล กฤษศยาสา, 2551) ดังนั้น จึงถือเป็นเรื่องท้าทายอีกเรื่องหนึ่งในการพัฒนารูปแบบ การเทียบโอนผลการเรียนให้มีความเหมาะสม ง่ายต่อ การทำความเข้าใจ มีความยุติธรรม และเชื่อถือได้

#### การศึกษาพัฒนาภาคตัดขวางของผู้เรียน

ประเด็นสำคัญอีกประการหนึ่งซึ่งถือเป็นหัวใจ สำคัญของการวัดและประเมินผลการเรียนรู้อของผู้เรียน คือ มุ่งเน้นการพัฒนาผู้เรียนมากกว่านำผลที่ได้ไปตัดสิน ผลการเรียนเพียงอย่างเดียวเหมือนเช่นในอดีตที่ผ่านมา ดังนั้นการศึกษาเชิงพัฒนาเพื่อพิจารณาความก้าวหน้า หรือพัฒนาการเรียนรู้อของผู้เรียนจึงเป็นสิ่งจำเป็น ของสถานศึกษาทุกแห่ง โดยการศึกษาภาคตัดขวาง (Cross Sectional Study) ซึ่งเป็นการศึกษาจากกลุ่ม ตัวอย่างต่างๆ กันหลายกลุ่มที่เป็นตัวแทนของแต่ละช่วง พัฒนาการขณะใดขณะหนึ่งพร้อมกันถือเป็นการศึกษาที่ มีความเหมาะสมอย่างยิ่ง เนื่องจากมีความรวดเร็วใน การรู้ผลการวิจัย มีปัญหาน้อยในเรื่องการสูญหายของ กลุ่มตัวอย่างและความคลาดเคลื่อนในการวัดจากกลุ่ม ตัวอย่าง โดยให้สารสนเทศเพื่อหาข้อบกพร่องในกระบวนการเรียนการสอน อันจะนำไปสู่การเร่งแก้ปัญหาได้ทัน ท่วงทีก่อนที่นักเรียนจะจบการศึกษาออกไปจากโรงเรียน และเป็นประโยชน์ต่อการกำหนดนโยบาย หรือการพัฒนา การเรียนการสอนในระดับหนึ่ง แต่ก็ยังนับว่าไม่เพียงพอ ในการสรุปโยงความสัมพันธ์ระหว่างผลการประเมินต่าง ระดับชั้นเหล่านั้กันอย่างน่าเชื่อถือได้ เพราะการประเมิน ในช่วงเวลาใดเวลาหนึ่ง โดยใช้กลุ่มตัวอย่างที่ต่างกัน และใช้เครื่องมือวัดที่ต่างกั้มาเปรียบเทียบกันโดยตรง ทำให้ผลการเปรียบเทียบมีความคลาดเคลื่อนสูง ขาดความแม่นยำของสารสนเทศอันส่งผลสืบเนื่อง

ให้การตัดสินใจสรุปผลผิดพลาดได้ (ศิริชัย กาญจนาวาสี, 2550; พัชรี จันทร์เพ็ง, 2547) ดังนั้นการแก้ปัญห การเปรียบเทียบผลการเรียนของบุคคลต่างกลุ่มที่ได้ จากเครื่องมือวัดที่ต่างกัน ในปีการศึกษาเดียวกัน เพื่อนำมาเปรียบเทียบและศึกษาพัฒนาการในการเรียนได้ จึงควรใช้เทคนิคการเชื่อมโยงคะแนนที่เรียกว่า “การปรับเทียบแนวตั้ง (Vertical Equating) ในปีการศึกษา เดียวกันระหว่างแบบสอบต่างฉบับ ได้จากนักเรียน ต่างกลุ่มกัน แล้วจึงนำผลที่ได้มาเปรียบเทียบกัน การปรับเทียบคะแนนดังกล่าว เป็นวิธีการเปลี่ยนระบบ หน่วยคะแนนของแบบสอบฉบับหนึ่ง ไปสู่ระบบหน่วย คะแนนของแบบสอบอีกฉบับหนึ่ง เพื่อให้ได้คะแนนที่ แปลงจากแบบสอบทั้งสองฉบับมีความเท่าเทียมกัน (Angoff, 1971) ผลการเปรียบเทียบจึงให้สารสนเทศที่ ถูกต้อง แม่นยำ และเพียงพอต่อการตัดสินใจ อย่างไรก็ตามด้วยข้อตกลงเบื้องต้นที่สำคัญของการใช้ เทคนิคการปรับเทียบคะแนน นั่นคือแบบสอบที่นำมา มาใช้ในการปรับเทียบต้องมีโครงสร้างเนื้อหาเหมือนกัน ซึ่ง อาจไม่สอดคล้องกับสภาพความเป็นจริงที่โครงสร้าง เนื้อหาของแบบสอบที่นำมาใช้ในการปรับเทียบจะ เหมือนกัน แต่อาจจะมีลักษณะที่คล้ายคลึงกัน (Similar) ดังนั้นจึงจำเป็นต้องใช้เทคนิคของ “การสร้างสเกล แนวตั้ง” (Vertical Scaling) ซึ่งในปัจจุบันนักวัดผล กำลังให้ความสำคัญกับการใช้เทคนิคดังกล่าวเพิ่มมากขึ้น เพื่อนำมาใช้ในการศึกษาพัฒนาการภาคตัดขวางของ ผู้เรียน

#### การเชื่อมโยงคะแนนสำหรับแบบสอบที่มีการวัดหลายมิติ (Multidimensional)

อย่างไรก็ตามถือเป็นเรื่องท้าทายสำหรับ การศึกษาเทคนิคของการเชื่อมโยงคะแนนโดยเฉพาะใน บริบทของการปรับเทียบคะแนนและการสร้างสเกลแนวตั้ง ที่อยู่ภายใต้ของข้อตกลงเบื้องต้นของความเป็นเอกมิติ ของแบบสอบที่นำมาใช้ในการเชื่อมต่อ ความจริงแล้ว ความเป็นเอกมิติของแบบสอบเกี่ยวกับเนื้อหาวิชา ที่ต้องการวัดนั้นออกจะกว้างมากจนแทบจะบอกความเป็น เอกมิติไม่ได้ และยังในปัจจุบันนี้การเรียนการสอนใน ห้องเรียนที่เน้นถึงลักษณะของการบูรณาการความรู้

มากขึ้น ทำให้การวัดมีวัตถุประสงค์เปลี่ยนแปลงไปจากอดีตโดยมุ่งเน้นการวัดสมรรถนะที่มีความซับซ้อนมากขึ้น ซึ่งในอนาคตมีความพยายามในการที่จะนำเทคนิคของการเชื่อมโยงคะแนนมาใช้วัดในหลายมิติ เพื่อให้สอดคล้องกับสภาพจริงมากขึ้น ที่อาจมีหลากหลายมิติเมื่อเป็นเช่นนี้ แนวปฏิบัติเกี่ยวกับการเรียนการสอนของครูจึงสวนทางกับแนวปฏิบัติของทฤษฎีทางด้านวัดผล (Torre & Patz, 2006) นิยามของเอกมิตติก็น่าจะต้องหลากหลายจนไม่อาจเป็นเอกมิตติที่แท้จริงได้ โดย Warm (1978) ได้กล่าวว่า ข้อตกลงเบื้องต้นของเอกมิตติมีความซับซ้อนมากที่สุด และยังเป็นข้อตกลงที่เข้มงวดที่สุดของทฤษฎี การตอบสนองข้อสอบ (Item Response Theory; IRT) เมื่อเป็นเช่นนี้แล้ว ในฐานะที่เป็นนักวัดผลควรจะทำอย่างไร จึงได้เกิดมีข้อตกลงในเรื่องการวัดแบบสอบที่มีลักษณะ พหุมิติ (Multidimensional) ซึ่งคะแนนที่ได้จะเป็นข้อมูลสำคัญเพื่อนำไปสู่การตัดสินใจทั้งระดับรายบุคคล ระดับสถาบัน รวมถึงระดับชาติ ในการนำไปเชื่อมโยงกับข้อมูลอื่น ซึ่งผลที่ได้ต้องมีความเสมอภาค และสามารถเปรียบเทียบกันได้ของคะแนน โดยถือว่าเป็นประเด็นสำคัญของสถานการณ์การทดสอบที่ได้รับความสนใจทั้งในอดีตและปัจจุบัน (Cook & Eignor, 1991) โดยเฉพาะความหลากหลายของรูปแบบการวัดที่แตกต่างกันแต่ต้องการวัดในสิ่งเดียวกัน (Kolen, 2001) อย่างไรก็ตามการเปรียบเทียบสเกลของแบบสอบเพื่อให้สามารถแปลความอย่างมีความหมาย เช่น ในกรณีที่มีระดับชั้นแตกต่างกัน ชนิดของแบบสอบแตกต่างกัน ในยุคที่ให้ความสำคัญกับการบูรณาการความรู้ โดยมีความพยายามที่จะนำเทคนิคของการเชื่อมโยงคะแนน ในกรณีที่แบบสอบมีการวัดในหลายมิติตั้งแต่ปี ค.ศ. 1989 อย่างไรก็ตามในอดีตยังอยู่ในลักษณะการศึกษาเชิงทดลอง ต่อมาในปีตั้งแต่ปี ค.ศ. 2000 เป็นต้นมา เริ่มมีการนำมาใช้ในสถานการณ์จริงเพิ่มมากขึ้น โดยเฉพาะการศึกษาพัฒนาการภาคตัดขวางของผู้เรียน (พัชรินทร์เพ็ง, 2550) อย่างไรก็ตามวิธีการดังกล่าวค่อนข้างมีความสลับซับซ้อน ดังนั้นจึงจำเป็นต้องมีการพัฒนาเทคนิคดังกล่าวให้ชัดเจนและเกิดผลในทางปฏิบัติเพิ่มมากขึ้น

การเชื่อมโยงคะแนนของแบบสอบที่มีลักษณะผสมระหว่างแบบสอบปรนัยและอัตนัย (Mixed Format Test)

ประการสุดท้าย อาจเป็นประเด็นที่ยังไม่เกิดผลชัดเจนในทางปฏิบัติแต่มีแนวโน้มในการที่สามารถนำมาใช้ปฏิบัติได้ในอนาคต นั่นคือ การมุ่งประเด็นไปยังเครื่องมือที่จะใช้วัดผลการเรียนรู้ของผู้สอบที่มีการวัดผลการเรียนรู้ของผู้เรียนที่นิยมใช้แบบสอบอัตนัยร่วมกับแบบสอบปรนัยเพิ่มมากขึ้นไม่มุ่งเน้นเฉพาะแบบสอบปรนัยเหมือนในอดีตที่ผ่านมาเนื่องจากสามารถวัดความคิดขั้นสูงของผู้เรียนได้ โดยเฉพาะอย่างยิ่งแบบสอบมาตรฐานในระดับประเทศ นิยมใช้แบบสอบที่มีลักษณะผสม (Mixed Format Test) ดังกล่าวมาแล้วข้างต้น ดังนั้นจึงมีความจำเป็นอย่างยิ่งในการพัฒนาเทคนิคการเชื่อมโยงคะแนน เพื่อให้สอดคล้องกับกระบวนทัศน์มุ่งเน้นการวัดผลการเรียนรู้ของผู้เรียนโดยใช้แบบสอบที่มีลักษณะผสมระหว่างปรนัยและอัตนัย

## บทส่งท้าย

เมื่อพิจารณาถึง แนวคิดพื้นฐานของการเชื่อมโยงคะแนนและเกณฑ์ที่ใช้ในการแบ่ง และภาพอนาคตของการประยุกต์ใช้เทคนิคการเชื่อมโยงคะแนน จะเห็นได้ว่าข้อตกลงเบื้องต้นของการเชื่อมโยงคะแนน มีลักษณะยืดหยุ่นและผู้ใช้สามารถนำไปประยุกต์ใช้ให้เหมาะสมกับสถานการณ์ในการทดสอบที่มีความหลากหลายและสอดคล้องกับสภาพเป็นจริงในการทดสอบ ซึ่งในแต่ละแนวคิดยังคงไม่ใช่วิธีที่ดีที่สุดที่สามารถนำมาประยุกต์ใช้ในการเชื่อมโยงคะแนนได้ ดังนั้นในการเชื่อมโยงคะแนนระหว่างแบบสอบ 2 ฉบับควรมีการกำหนดกรอบการดำเนินงานให้ชัดเจนและควรพิจารณาถึงการแปลผลของคะแนนที่ได้ด้วยว่าสามารถแสดงถึงความสามารถที่แท้จริงและคะแนนที่ได้เป็นตัวแทนที่ดีของผู้เรียนได้มากน้อยเพียงไร

## เอกสารอ้างอิง

- พัชรี จันทรเพ็ญ. (2547). การศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนโดยวิธีการปรับเทียบแนวตั้งตามวิธีทฤษฎีตอบสนองข้อสอบร่วมกับวิธีการปรับเทียบเชิงเส้นตรง. วิทยานิพนธ์ปริญญาโทบริหารศึกษาด้านจิตวิทยาวิจัยและจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- พัชรี จันทรเพ็ญ. (2550). การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติภายใต้การหมุนแกน โครงสร้างเชิงมิติและระดับความสัมพันธ์ที่แตกต่างกัน. วิทยานิพนธ์ปริญญาโทศึกษาด้านจิตวิทยาวิจัยและจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- สุวิมล กฤษศยาสา. (2551). การพัฒนารูปแบบการเทียบโอนผลการเรียนภายในกลุ่มสาระการเรียนรู้: การประยุกต์ใช้เทคนิคการสร้างคะแนนสอดคล้อง. วิทยานิพนธ์ปริญญาโทศึกษาด้านจิตวิทยาวิจัยและจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2550). ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theories). พิมพ์ครั้งที่ 3. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี และคณะ. (2551). การพัฒนาวิธีการปรับเทียบผลการเรียนเฉลี่ยสะสมตามกลุ่มสาระการเรียนรู้โดยใช้คะแนน ONET ของนักเรียนมัธยมศึกษาตอนปลาย. เอกสารอัดสำเนา  
ศึกษาธิการ, กระทรวง, สำนักทดสอบทางการศึกษา. (2549). แนวปฏิบัติเกี่ยวกับการเทียบโอนผลการเรียนเข้าสู่อการศึกษาในระบบการศึกษาขั้นพื้นฐาน. (ม.ป.ท.)
- Angoff, W. H. (1971). Scales, Norm and Equivalent Score. In R.L. Thordike(Ed.), *Education Measurement*. (2nd ed.). Washington, D.C.: American Council on Education, 508-600.
- Dorans, N. J. (2004). Equating, Concordance, and Expectation. *Applied Psychological Measurement*, 28(4): 227-246.
- Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Hemphill, F. C. (1999). *Uncommon Measures: Equivalence and Linkage among Educational Tests*. Washington, DC: National Academy Press.
- Kolen M. L., & Brennan R. L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices*. New York: Springer Science Business Media.
- Kolen, M. J. (2001). Linking Assessments Effectively: Purpose and Design. *Education Measurement: Issues and Practice*, 20(1): 5-9.
- Mislevy, R.J. (1992). *Linking Educational Assessments: Concept, Issues, Methods and Prospects*. [Online]. Available from: <http://eric.ed.gov>. [2005, October 27]