มโนทัศน์ที่คลาดเคลื่อนในการวัดและประเมินทางการศึกษา
Misconceptions in Educational Measurement and Evaluation

ศิริชัย กาญจนวาสี*
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อ

การวัดและประเมินเป็นศาสตร์ที่มีความเข้มแข็งในวิชาชีพทางการศึกษามาอย่างยาวนาน อย่างไรก็ตาม การใช้หลักการ แนวคิด ทฤษฎีทางการวัดและประเมินทางการศึกษา ก่อให้เกิดมโนทัศน์ที่คลาดเคลื่อนอยู่หลายประการ ความเข้าใจผิดในบางอย่างย่อมเป็นอุปสรรคต่อปฏิบัติการทางการวัดและประเมิน ทำให้ได้ผลหรือสารสนเทศ ที่ไม่ถูกต้อง อันเป็นปัญหาต่อการตัดสินใจและการพัฒนาทางการศึกษา มโนทัศน์คลาดเคลื่อนที่สำคัญเกิดขึ้นทั้ง 3 ระดับ ได้แก่ ระดับความหมาย โมเดล และทฤษฎี

คำสำคัญ : มโนทัศน์ที่คลาดเคลื่อนการวัดและประเมิน โมเดลการวัดและประเมิน ทฤษฎีการวัดและประเมิน

Abstract

Measurement and evaluation have been developed in the field of education for a long time. However, practicing measurement and evaluation induce misconception in many way. Misconceptions cause misconduct, misinterpretation, and misuse of the results. Misconceptions in educational measurement and evaluation occurred at three levels : meaning level, model level, and theory level.

Keyword : Misconceptions in Educational, Measurement and Evaluation, Measurement and Evaluation Models, Measurement and Evaluation Theory

* ผู้ประสานงานหลัก (Corresponding Author)
e-mail: skanjanawasee@hotmail.com

บทนำ

การวัดและประเมินเป็นศาสตร์ที่มีความเข้มแข็ง และแข็งแกร่งในสายวิชาชีพทางการศึกษามาอย่างยาวนาน เพราะการศึกษาเป็นกระบวนการพัฒนาทรัพยากรมนุษย์ให้เป็นมนุษย์ที่สมบูรณ์ ซึ่งทุกประเทศทั่วโลกต่างใช้การศึกษาเป็นยุทธศาสตร์สำคัญในการพัฒนาประเทศ ทั้งทางด้านเศรษฐกิจ สังคม การเมือง และสิ่งแวดล้อม ธรรมชาติของการจัดการศึกษาเป็นการจัดประสบการณ์ให้ผู้เรียนเกิดการเรียนรู้ โดยใช้หลักสูตร กระบวนการจัดกิจกรรม และการวัดประเมินผู้เรียน เพื่อติดตามสถานภาพการเรียนรู้และการพัฒนาผู้เรียน ทั้งด้านความรู้ ทักษะ และคุณลักษณะที่พึงประสงค์ การวัดและประเมินการเรียนรู้ซึ่งเป็นคุณลักษณะภายในของมนุษย์ เป็นลักษณะทางจิตมิติเป็นเรื่องที่สลับซับซ้อนและมีความละเอียดอ่อน นักวิชาการทางการศึกษาได้ทำการศึกษา ค้นคว้า และวิจัยกันมาอย่างต่อเนื่อง ทำให้ศาสตร์การวัดและประเมินมนุษย์ จึงมีความเข้ม แกร่งในวิชาชีพทางการศึกษามากกว่าวิชาชีพอื่นๆ

อย่างไรก็ตาม การใช้หลักการ แนวคิด ทฤษฎีทางการวัดและประเมินในแวดวงประชาคมการศึกษาก่อให้เกิดมโนทัศน์ที่คลาดเคลื่อนอยู่หลายอย่าง **มโนทัศน์ที่คลาดเคลื่อน (Misconception)** เป็นแนวคิดในเชิงนามธรรมที่เข้าใจผิด หรือเข้าใจคลาดเคลื่อนจากความถูกต้อง นักการศึกษา นักวิชาการ นักวิจัย จำเป็นจะต้องศึกษา เพื่อให้เข้าใจ จะได้เข้าถึง และนำไปใช้ประโยชน์ในการพัฒนาผู้เรียน จำเป็นจะต้องรู้ถึงรากฐานของแนวคิด ข้อตกลงเบื้องต้นของโมเดล ทฤษฎีการวัดและประเมินอย่างถูกต้อง จึงจะสามารถนำผลไปใช้ได้อย่างเหมาะสม อันจะนำไปสู่การพัฒนาผู้เรียนได้อย่างแท้จริง

มโนทัศน์ที่คลาดเคลื่อนในการวัดและประเมินทางการศึกษา มีอยู่หลายประการ ความเข้าใจผิดในบางอย่างเป็นอุปสรรคต่อปฏิบัติการทางการวัดและประเมิน ทำให้ได้รับผลที่ไม่ถูกต้อง อันเป็นปัญหาต่อการตัดสินใจและพัฒนา มโนทัศน์คลาดเคลื่อนที่สำคัญเกิดขึ้นทั้ง 3 ระดับ ดังต่อไปนี้

1. การวัดและประเมิน : ความหมาย

ทางการศึกษาเราคุ้นเคยกับคำว่า “การวัดผล” “การประเมินผล” และ “การวัดและประเมินผล” มาเป็นเวลานาน และได้ยินอยู่ตลอดเวลา ถ้าเทียบเคียงทางภาษา “การวัดผล” น่าจะแปลมาจากคำว่า “Measurement or Measuring” ส่วน “การประเมินผล” น่าจะแปลมาจาก “Evaluation or Evaluating” การแปลคำดังกล่าวเป็นประโยชน์ในการสื่อสารกิจกรรมทางการวัดและประเมิน แต่ในขณะเดียวกันก็ทำให้เกิดความเข้าใจที่คลาดเคลื่อนจากสิ่งที่ควรจะเป็น นิยามถูกตีวงแคบลง อันนำไปสู่ปฏิบัติการในวงจำกัด ทำให้ความหมายของ Measurement & Evaluation ถูกจำกัดขอบเขตไปด้วย

การแปล “Measurement” ว่าเป็น “การวัดผล” ทำให้นักการศึกษาหมกมุ่นอยู่กับคำว่า “ผล” การวัดผลจึงเป็นกิจกรรมที่เข้าใจกันโดยทั่วไปว่า เป็นการวัดที่ต้องรอให้การดำเนินงานเสร็จสิ้นลงเสียก่อน แล้วจึงทำการวัดผลที่เกิดขึ้น แนวคิดของการวัดผลจึงถูกพัฒนามาอย่างเข้มข้นในการ “วัดผลสรุป (Summative Measurement)” แล้วละเลยการวัดสิ่งต่างๆ ที่นอกเหนือจากการวัดผล (Formative measurement) ทำให้เสียโอกาสของการพัฒนา การวัดบริบท การวัดปัจจัย การวัดกระบวนการ การวัดพัฒนาการซึ่งมีความสำคัญไม่ยิ่งหย่อนไปกว่าการวัดผลที่เกิดขึ้นหลังการดำเนินงาน

ในทำนองเดียวกัน การแปล “Evaluation” ว่าเป็น “การประเมินผล” จึงเป็นการประเมินที่ติดหล่มของการประเมินโดยเน้นผลที่เกิดหลังการดำเนินงานสิ้นสุดลง การประเมินจึงเป็นกระบวนการที่เข้าใจกันโดยทั่วไปว่าเป็นการตัดสินคุณค่าของผลการดำเนินงาน (Summative Evaluation) เลยก้าวข้ามการประเมินระหว่างทางหรือการประเมินความก้าวหน้า (Formative Evaluation) ซึ่งมีบทบาทสำคัญมากในการเสนอสารสนเทศย้อนกลับ (Feedback) เพื่อการปรับเปลี่ยน ปรับปรุง แก้ไขให้ทันท่วงที ก่อนที่จะสายเกินกาล ถ้าการประเมินความก้าวหน้าเป็นไปอย่างมีประสิทธิภาพ น่าจะเป็นหลักประกันของผลสำเร็จในการประเมินผลสรุป ดังนั้นเพื่อการหลีกเลี่ยงความเข้าใจที่คลาดเคลื่อนในความหมายของ Measurement & Evaluation น่าจะแปลความหมายของคำว่า “Measurement” ว่า “การวัด” และแปลความหมายของคำ “Evaluation” ว่า “การประเมิน” ความหมายหมายของคำไทยจึงจะตรงและมีความครอบคลุมถูกต้อง

2. การวัดและประเมิน : โมเดล

2.1 โมเดลการวัด (Measurement Models)

มีความเชื่อในหมู่นักการศึกษาบางกลุ่มว่า การใช้โมเดลการวัดขั้นสูง ซึ่งมีความสลับซับซ้อนและต้องใช้การวิเคราะห์ด้วยสถิติขั้นสูงนั้น นอกจากแสดงถึงความทันสมัยแล้ว ยังจะช่วยเยียวยาปัญหาการวัดขั้นพื้นฐานได้อย่างมีคุณภาพ ความเชื่อดังกล่าวถูกต้อง จริงหรือ ?

ก่อนอื่นคงต้องมาทำความเข้าใจเกี่ยวกับโมเดลการวัดกันเสียก่อนว่า **โมเดลการวัด** เป็นแบบจำลองปรากฏการณ์หรือสถานการณ์ของการวัดตามแนวความคิดทางทฤษฎีที่สมเหตุสมผล นักทฤษฎีมักแสดงแนวคิดในรูปของ**โมเดลทางสถิติ** ซึ่งเป็นฟังก์ชั่นความสัมพันธ์ระหว่างคุณลักษณะภายในที่อยู่ในตัวมนุษย์ (คุณลักษณะแฝงซึ่งเป็นนามธรรม) กับสิ่งที่คุณลักษณะนั้นแสดงออกมาให้ปรากฏ (พฤติกรรมซึ่งเป็นรูปธรรมที่สามารถสังเกตได้) ความสัมพันธ์ระหว่างคุณลักษณะแฝงกับพฤติกรรมดังกล่าวอยู่บนพื้นฐานของข้อตกลงเบื้องต้น (Assumptions) ของโมเดลที่เชื่อว่าเป็นความจริง ฉะนั้นการนำโมเดลมาใช้วัดคุณลักษณะใดของมนุษย์คนใดก็ตาม จะต้องยอมรับเงื่อนไขที่ตกลงไว้นั้นแต่ถ้าข้อตกลงเบื้องต้นนั้นไม่เป็นจริง หรือไม่น่าเชื่อถือเสียแล้ว ผลการวัดที่ได้ก็จะไม่ถูกต้องหรือไม่น่าเชื่อถือ เพราะอยู่บนพื้นฐานของสิ่งที่ไม่น่าเชื่อถือตั้งแต่ต้น ดังนั้นการเลือกใช้โมเดลการวัด ไม่ว่าจะใช้โมเดลการวัดขั้นพื้นฐานแบบดั้งเดิม โมเดลการวัดขั้นสูง หรือโมเดลการวัดขั้นสูงมากๆ แบบแนวใหม่ล่าสุด ถ้าสถานการณ์หรือองค์ประกอบเงื่อนไขของการวัดไม่เป็นไปตามข้อตกลงเบื้องต้นของโมเดลการวัดนั้น ผลการวัดก็จะไม่ถูกต้อง ไม่เป็นจริงหรือไม่น่าเชื่อถือ โมเดลการวัดขั้นสูงจึงไม่สามารถชดเชย เยียวยาปัญหาเชิงคุณภาพของการวัดขั้นพื้นฐานได้ถ้าการใช้โมเดลการวัดขั้นสูงนั้นไม่เป็นไปตามเงื่อนไขหรือข้อตกลงของโมเดล

ตัวอย่างเช่น โมเดลการวัดแบบดั้งเดิม (Classical Measurement Model) $X = T + E$ คะแนนที่สังเกตได้ (X) เกิดจากผลรวมเชิงเส้นตรงของคะแนนจริง (T) กับคะแนนความคลาดเคลื่อน (E) โดย E มีความเป็นอิสระจาก T และ E มีค่าเฉลี่ยเป็น 0 ในการสอบแต่ละครั้ง จึงมีค่า E เกิดขึ้นทั้งในทาง + และ - เมื่อรวมค่า E ทางบวกและทางลบแล้วมีค่าเท่ากัน เด็กเก่งหรือเด็กอ่อนสามารถมีค่า E เป็น + หรือ - ได้ไม่แตกต่างกัน และค่าความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of

Measurement, SEM) เป็นค่าคงที่สำหรับทุกคน เช่นถ้าการสอบคณิตศาสตร์มีความคลาดเคลื่อนเท่ากับ ± 5 คะแนน คะแนนจริงของเด็กแต่ละคนจะอยู่ในช่วง X ± 5 คะแนนโดยไม่สามารถบอกได้ว่าความคลาดเคลื่อนจะเกิดในทางบวกหรือลบ

โมเดลการวัดตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory, G-Theory) โมเดลการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดอยู่บนพื้นฐานข้อตกลงเบื้องต้นว่าคะแนนจริงเป็นคะแนนเฉลี่ยจากการสอบหรือการวัดภายใต้สถานการณ์หรือเงื่อนไขของการวัดที่ยอมรับได้ทั้งหมด ความคลาดเคลื่อนมีแหล่งที่มาอย่างเป็นระบบอย่างน้อย 1 แหล่ง (FACET) ที่สามารถระบุและทำการศึกษาผลของมันที่มีต่อความน่าเชื่อถือของผลการวัดได้ (G-Coefficient)

โมเดลการตอบสนองข้อสอบ (Item Response Models) เป็นระบบความสัมพันธ์ระหว่างโอกาสตอบข้อสอบถูก (P_i) กับความสามารถที่มีอยู่ภายในผู้ตอบ (θ) ในรูปของโค้งลักษณะข้อสอบ (ICC) ซึ่งเป็นฟังค์ชั่นโลจิส (Logistic Function) ที่แสดงความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องขึ้นกับความสามารถของผู้สอบและคุณลักษณะของข้อสอบ โมเดลการตอบสนองข้อสอบตั้งอยู่บนความเชื่อหรือข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกมิติ (Unidimensionality) ของแบบสอบ ความเป็นอิสระระหว่างข้อสอบและผู้สอบ (Local Independence) และการสอบที่ไม่แข่งขันด้านเวลา (Nonspeeded Test Administration)

2.2 โมเดลการวัดเชิงสุ่ม (Ramdom - Effect Model)

การวิเคราะห์คุณภาพของเครื่องมือวัดตามโมเดลการวัดแบบดั้งเดิม ไม่ว่าจะเป็นการวิเคราะห์คุณภาพของข้อคำถาม ด้วยค่าความยาก (p) และค่าอำนาจจำแนก (r) หรือการวิเคราะห์คุณภาพของแบบวัด ด้วยค่าความเที่ยง (Reliability) และค่าความตรง (Validity) นักวิชาการมีการถกเถียงกันว่าผลการวิเคราะห์ที่ได้แสดง**คุณภาพของเครื่องมือ** (Measurement Tool) หรือ**คุณภาพของคะแนน**การทดสอบครั้งนั้น (Test scores)

การตอบคำถามนี้จะต้องทำความเข้าใจเกี่ยวกับลักษณะของโมเดลที่ใช้ในการวัดโดยการเปรียบเทียบความแตกต่างระหว่าง “โมเดลอิทธิพลเจาะจง (Fixed - EffectModel)” กับ “โมเดลอิทธิพลเชิงสุ่ม (Random - Effect Model)” โมเดลอิทธิพลเชิงสุ่ม เป็นโมเดลเชิงสถิติที่กำหนดให้ตัวแปรอิสระที่สนใจ **เป็นตัวแปรสุ่ม** โดยถือว่ากลุ่มตัวอย่างข้อมูลของตัวแปรที่นำมาวิเคราะห์ได้มาอย่างสุ่มจากประชากร ข้อมูลจึงมีหลายชุดที่สุ่มมาจากประชากรเดียวกัน ในโมเดลอิทธิพลเชิงสุ่ม ค่าสัมประสิทธิ์ที่คำนวณได้จึงเป็นค่าที่มีองค์ประกอบของค่าคงที่ (ค่าเฉลี่ยของกลุ่ม) และค่าความผันแปรระหว่างกลุ่ม จึงมีค่าความคลาดเคลื่อนจากค่าเฉลี่ยเป็นส่วนประกอบอยู่ในโมเดลทำให้การประมาณค่าสัมประสิทธิ์ต่างๆ มีความถูกต้องตามความเป็นจริงมากขึ้น ดังนี้

$$\beta_{ij} = \overline{\beta}_j + e_{ij}$$

ในขณะที่โมเดลอิทธิพลเจาะจงเป็นโมเดลเชิงสถิติที่กำหนดให้ตัวแปรอิสระที่สนใจเป็นตัวแปรคงที่ โดยถือว่ากลุ่มตัวอย่างข้อมูลของตัวแปรที่นำมาวิเคราะห์ประกอบด้วยข้อมูลชุดเดียว ถึงแม้จะเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างใหม่ก็จะได้ข้อมูลที่ไม่แตกต่างจากชุดเดิม (ซึ่งไม่น่าเป็นจริง) ในโมเดลอิทธิพลเจาะจง ค่าสัมประสิทธิ์ที่คำนวณได้จึงเป็นค่าคงที่ของทุกกลุ่ม จึงไม่มีค่าความผันแปรระหว่างกลุ่ม ไม่มีเทอมของความคลาดเคลื่อนจากค่าเฉลี่ยของกลุ่ม [$\beta_{ij} = \beta_j$]

โมเดลการวัดแบบดั้งเดิม เป็นโมเดลอิทธิพลเจาะจงทั้งผู้สอบ และข้อสอบ (examinee fixed และ item fixed) ค่าสัมประสิทธิ์ของข้อสอบ และแบบสอบที่คำนวณได้ จึงเป็นค่าคงที่เฉพาะของการสอบครั้งนั้น ไม่สามารถสรุปอ้างอิงไปยังผู้สอบกลุ่มอื่น หรือข้อสอบชุดอื่นได้ ถ้ามีการเปลี่ยนชุดของคำถามหรือใช้ชุดของคำถามชุดเดิมแต่นำไปใช้กับผู้สอบกลุ่มใหม่ ค่าสถิติของข้อสอบและแบบสอบก็จะเปลี่ยนแปลงไปตามกลุ่มตัวอย่างที่ใช้ ดังนั้นผลการวิเคราะห์คุณภาพของข้อคำถามหรือแบบสอบ จึงเป็นคุณภาพของคะแนนสอบในการทดสอบครั้งนั้น มิได้เป็นคุณภาพประจำเครื่องมือนั้น

2.3 โมเดลการประเมิน

1) โมเดลการประเมินหนึ่งเดียว : Goal - Based Evaluation?

นักวิชาการทั่วไปเมื่อถึงคราวจำเป็นต้องทำหน้าที่ประเมิน ส่วนใหญ่จะมีแนวความคิดที่สอดคล้องกันว่า การประเมินเป็นกระบวนการ**ตัดสินคุณค่าของผลสรุป**ว่าบรรลุผลตามจุดมุ่งหมายหรือวัตถุประสงค์ที่กำหนดไว้ล่วงหน้าหรือไม่ เพียงไร แนวคิดดังกล่าวอาจเรียกว่าเป็นแนวคิดการประเมินแบบ Goal-Based Evaluation (GBE) ซึ่งประชาคมวิชาการได้ยึดถือแนวคิดแบบนี้มาโดยตลอด การประเมินที่ยึดติดกับแนวคิดเดียวแบบ GBE นี้ น่าจะเป็นแนวคิดที่มีความคลาดเคลื่อน ขาดความยืดหยุ่นที่เหมาะสมกับสถานการณ์ เพราะว่ามีอยู่บ่อยครั้งที่การเขียนจุดมุ่งหมายหรือวัตถุประสงค์ของนโยบาย/แผนงาน/โครงการ ขาดความรอบคอบ ไม่ครอบคลุมความต้องการจำเป็นตามมาตรฐานของการกำหนดจุดมุ่งหมาย/วัตถุประสงค์ที่ดี ในสถานการณ์เช่นนั้น GBE น่าจะขาดความชอบธรรมในการเป็นรูปแบบการประเมินที่เหมาะสม เพราะจุดมุ่งหมาย/วัตถุประสงค์ที่เขียนไว้ มิได้เป็นจุดหมายที่สำคัญอีกต่อไป จึงไม่เหมาะที่จะใช้เป็นเกณฑ์ผลลัพธ์ที่คาดหวังให้เกิดขึ้นเพื่อแสดงผลสำเร็จของนโยบาย/แผนงาน/โครงการ

แนวคิดของการประเมินนโยบาย/แผนงาน/โครงการ ในสถานการณ์ดังกล่าวข้างต้นที่มีความเหมาะสมกว่า น่าจะเป็นแนวคิดการประเมินแบบไม่จำเป็นต้องล้อผลตามจุดมุ่งหมาย/วัตถุประสงค์ (ที่ไม่เหมาะสม) แต่ทำการประเมินหรือตัดสินคุณค่าของผลจริงที่เกิดขึ้นทั้งหมด (Total Effects) ทั้งผลผลิต (output) ผลลัพธ์ (outcome) และผลกระทบ (impact) ซึ่งเรียกการประเมินตามแนวคิดนี้ว่า Goal - Free Evaluation (GFE) แต่ถ้ามีการกำหนดจุดมุ่งหมาย/วัตถุประสงค์ได้อย่างเหมาะสม ครอบคลุมผลลัพธ์ที่จำเป็นได้ทั้งหมด GBE กับ GFE น่าจะเป็นโมเดลเดียวกันศึกษารายละเอียดเพิ่มเติมได้จาก Scriven (1973)

2) โมเดลการประเมินในฝัน : CIPP ?

หากถามผู้สนใจการประเมินว่า โมเดลการประเมินที่คนทั่วโลกรู้จักกันมากที่สุดคืออะไร คนส่วนใหญ่คงจะตอบเป็นเสียงเดียวกันว่า “CIPP Model” พร้อมทั้งอธิบายต่อได้ว่า จะต้องทำการประเมิน

Context, Input, Process, และ Product อันเป็นสูตรสำเร็จของการประเมิน ผู้เขียนยอมรับได้ว่า CIPPModel เป็นโมเดลการประเมินที่คนทั่วโลกรู้จักกันมากที่สุด แต่ขอบอกว่า CIPP Model เป็นโมเดลการประเมินที่คนทั่วโลกเข้าใจผิดมากที่สุด เช่นกัน

CIPP Model เป็นโมเดลการประเมินที่ Stufflebeam et.al. (1971) พัฒนาขึ้นมาเพื่อช่วยให้ผู้บริหารตัดสินใจได้อย่างเหมาะสมในแต่ละขั้นตอนของการดำเนินงาน การประเมินบริบท (Context Evaluation) เพื่อช่วยผู้บริหารตัดสินใจเกี่ยวกับการเลือกโครงการ/เป้าหมาย/จุดมุ่งหมาย การประเมินปัจจัยเบื้องต้น (Input Evaluation) เพื่อช่วยผู้บริหารกำหนดยุทธวิธี/แผนงาน/การดำเนินงาน การประเมินกระบวนการ (Process Evaluation) เพื่อช่วยผู้บริหารปรับเปลี่ยนยุทธวิธี/แผนงาน/การดำเนินงานให้เหมาะสม และการประเมินผลผลิต (Product Evaluation) เพื่อช่วยให้ผู้บริหารตัดสินใจเกี่ยวกับการยุติ/คง/ขยายโครงการ (ศิริชัย กาญจนวาสี, 2554) CIPP จึงเป็นโมเดลการประเมินที่ดีมากในการเสนอสารสนเทศให้ผู้บริหารใช้ตัดสินใจก่อนเริ่มดำเนินการ เพื่อเลือกนโยบาย/แผนงาน/โครงการที่ดี เพื่อเลือกยุทธวิธี/วิธีดำเนินการที่ดี ระหว่างการดำเนินงาน เพื่อปรับเปลี่ยน ปรับปรุงยุทธวิธี/วิธีดำเนินงานให้เหมาะสม และหลังการดำเนินงาน เพื่อตัดสินใจเกี่ยวกับอนาคตของนโยบาย/แผนงาน/โครงการ

3. การวัดและประเมินผล : ทฤษฎี

ทฤษฎีการวัดเป็นองค์ความรู้ที่มีนัยทั่วไปเกี่ยวกับการวัด วิธีการแก้ปัญหาการวัด และการพัฒนาเครื่องมือสำหรับวัดการศึกษาทฤษฎีก็เพื่อใช้เป็นแหล่งความรู้สำหรับทำความเข้าใจโมเดลการวัด ข้อตกลงเบื้องต้น การพัฒนาเครื่องมือให้มีคุณภาพ การวิเคราะห์ผล การแปลผลการวัดได้อย่างถูกต้อง ตลอดจนการใช้เป็นสารสนเทศสำหรับการตัดสินใจทางการศึกษาได้อย่างเหมาะสม ในการประยุกต์ความรู้ทางทฤษฎีไปใช้ทางปฏิบัติ หากขาดความรู้ความเข้าใจอย่างถ่องแท้แล้ว ความผิดพลาดคลาดเคลื่อนอาจเกิดขึ้นได้

3.1 ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory)

- ความเที่ยง (Reliability)

ความเที่ยง เป็นคุณลักษณะความคงเส้นคงวาของคะแนนสอบ เมื่อทำการสอบซ้ำในทางทฤษฎี ความเที่ยงเป็นสัดส่วนของ $\frac{\sigma_T^2}{\sigma_X^2}$ หรือ $1-\frac{\sigma_E^2}{\sigma_X^2}$ ดังนั้นถ้าคะแนนจริงของผู้สอบมีความแตกต่างกันมาก (σ_T^2 มีค่าสูง) หรือคะแนนที่สังเกตได้มีความแปรปรวนสูง (σ_X^2 มีค่าสูง) เนื่องจากกลุ่มมีลักษณะวิวิธพันธ์ (Heterogeneous) ย่อมส่งผลให้ความเที่ยงมีค่าสูง “แต่มีความเข้าใจที่คลาดเคลื่อนว่าถ้าใช้กลุ่มตัวอย่างขนาดเล็ก ค่าความเที่ยงจะต่ำ ถ้าเพิ่มขนาดกลุ่มตัวอย่างให้ใหญ่ขึ้น จะทำให้ค่าความเที่ยงสูงตามไปด้วย” คำกล่าวนี้ไม่น่าจะถูกต้อง เพราะขึ้นอยู่กับว่าลักษณะของกลุ่มเมื่อมีจำนวนขนาดตัวอย่างเพิ่มขึ้น แล้วจะทำให้ความผันแปรของคะแนนในกลุ่มเปลี่ยนแปลงไปในทิศทางใด สมมุติว่ากลุ่มเดิมมีขนาดเล็ก เมื่อเพิ่มจำนวนกลุ่มตัวอย่างเข้ามา มีผลทำให้ความแปรปรวนของคะแนนในกลุ่มใหม่เพิ่มขึ้นหรือลดลง ถ้าความแปรปรวนของคะแนนในกลุ่มเพิ่มขึ้น จะทำให้ค่าความเที่ยงเพิ่มขึ้น แต่ถ้าความแปรปรวนของคะแนนในกลุ่มลดลง จะทำให้ค่าความเที่ยงลดลงเช่นเดียวกัน

- สัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha)

ค่าสัมประสิทธิ์แอลฟาของครอนบาค (α) เป็นดัชนีวัดความเป็นเอกพันธ์ของเนื้อหาในแบบสอบ (Homogeneity) ว่าวัดเนื้อหาเดียวกันเพียงใด มีนักวิชาการจำนวนมากเข้าใจผิดว่าเป็นการวัดความเป็นคุณลักษณะเดียว หรือมิติเดียว (Unidimensionality) ของแบบสอบ

การตรวจสอบความสอดคล้องภายในของแบบสอบเป็นวิธีการประมาณค่าความเที่ยงของแบบสอบโดยใช้การทดสอบเพียงครั้งเดียวด้วยการใช้แบบสอบฉบับเดียวทำการทดสอบผู้สอบกลุ่มเดียวในการตรวจสอบความสอดคล้องภายในนั้นเป็นการวัดระดับความเป็นเอกพันธ์ (Homogeneity) ของข้อสอบในแบบสอบนั้นว่าวัดเนื้อเรื่องเดียวกันเพียงใด ถ้าแบบสอบวัดในเรื่องเดียวกัน เมื่อทำการวัดซ้ำๆ ก็น่าจะมีความคงที่หรือสอดคล้องในผลการวัดสูง

การประมาณค่าความเที่ยงแบบวัดความสอดคล้องภายในสามารถคำนวณได้จากสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของกลุ่มข้อสอบที่มีการแยกส่วนแบบต่างๆ ผลการวัดจัดเป็นความเป็นเอกพันธ์ของแบบสอบ (Test homogeneity) ซึ่งแสดงถึงการวัดมวลเนื้อเรื่องเดียวกัน (Content domain) แต่ถ้าข้อสอบข้อต่างๆ วัดเนื้อเรื่องที่ไม่เกี่ยวข้องสัมพันธ์กัน คะแนนข้อสอบเหล่านั้นจะไม่สอดคล้องกัน ทำให้แบบสอบมีความสอดคล้องภายในต่ำ (ศิริชัย กาญจนวาสี, 2552)

การใช้สูตรคำนวณค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha) มีข้อตกลงเบื้องต้นว่า แบบสอบถูกแบ่งออกเป็น k ส่วน แต่ละส่วนอาจเป็นกลุ่มข้อสอบหรือข้อสอบแต่ละข้อก็ได้ ถ้าแต่ละส่วนมีความทัดเทียมกัน (คู่ขนานกัน) หรือมีคะแนนจริงของแต่ละส่วนเป็นฟังก์ชั่นเชิงบวกต่อกัน ค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha) จึงจะเป็นค่าที่ถูกต้องของความเที่ยงของคะแนนสอบ แต่ถ้าส่วนต่างๆ ไม่ทัดเทียมกัน (คู่ขนานกัน) ค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha) จะเป็นค่าที่ต่ำกว่าค่าความเที่ยงที่แท้จริงของแบบสอบ คะแนนสอบสามารถมีค่า α สูง (วัดเนื้อเรื่องเดียวกัน) แต่มีคุณลักษณะที่วัดหลายด้านหรือหลายมิติ (Multidimension) เช่น แบบวัดคณิตศาสตร์มีการวัดความสามารถด้านการคำนวณพื้นฐาน ด้านการวิเคราะห์ขั้นสูงด้านการคิดแก้ปัญหาโจทย์ เป็นต้น

- ความตรง (Validity)

ความตรง เป็นคุณลักษณะความถูกต้องแม่นยำของคะแนนสอบ ถ้าคะแนนสอบมีความถูกต้องแม่นยำ หรือสอดคล้องกับคะแนนผลการวัดคุณลักษณะเดียวกันของเครื่องมือมาตรฐานหรือเครื่องมืออื่นที่เชื่อถือได้ เราเรียกว่า ความตรงตามเกณฑ์สัมพันธ์ (Criterion - Related Validity) ถ้าคะแนนเกณฑ์นั้นมาจากการวัดในสภาพปัจจุบัน เรียกว่า "ความตรงตามสภาพ (Concurrent Validity)" แต่ถ้าคะแนนเกณฑ์นั้นเป็นการวัดในอนาคต เรียกว่า "ความตรงเชิงทำนาย (Predictive Validity)"

ความตรงตามเกณฑ์สัมพันธ์ (Υ_{xy}) สามารถประมาณค่าได้จากการคำนวณค่าสหสัมพันธ์ระหว่างคะแนน X จากเครื่องมือที่สร้างขึ้นกับคะแนนเกณฑ์ Y จากเครื่องมือมาตรฐานหรือเครื่องมืออื่นที่เชื่อถือได้ การประมาณค่าความตรง มิใช่ค่าถาวรที่แสดงคุณภาพประจำเครื่องมือวัด แต่เป็นค่าที่แสดงคุณภาพของคะแนนที่ได้จากการวัดในครั้งนั้น ค่าความตรงของการวัดสามารถผันแปรได้ตามสภาพของการวัด ขนาดและลักษณะของกลุ่มตัวอย่าง

1) การอ่อนตัวของค่า r_{xy}

เมื่อสภาพการวัดคะแนน X และ Y มีความคลาดเคลื่อนเกิดขึ้นในการวัด X และ Y ทำให้ค่าความตรง (r_{xy}) อ่อนตัวลง ในทางทฤษฎี เราสามารถปรับแก้หาค่าความตรง ($r_{T_xT_y}$) ได้ดังนี้

$$r_{T_xT_y} = \frac{r_{xy}}{\sqrt{r'_{xx}}\sqrt{r'_{yy}}}$$

เมื่อ และ เป็นค่าความเที่ยงของการวัด x และ y ตามลำดับ

2) ผลของขนาดและลักษณะของกลุ่มตัวอย่างต่อ r_{xy}

ในการประมาณค่า r_{xy} ถ้ากลุ่มตัวอย่างที่ใช้ทดสอบเป็นกลุ่มคัดสรร เช่น กลุ่มเก่ง หรือกลุ่มอ่อน หรือการทดสอบเกิดกรณี Floor Effect หรือ Ceiling Effect จะทำให้เกิดลักษณะคะแนนช่วงจำกัด (Range restricted) ส่งผลต่อการอ่อนตัวลงของ r_{xy}

นอกจากนี้การใช้กลุ่มตัวอย่างที่ขนาดเล็ก ซึ่งจะมีความคลาดเคลื่อนจากการสุ่ม (Sampling error) สูงจะมีผลให้ค่า r_{xy} ต่ำลง เพื่อแก้ปัญหานี้ ขนาดกลุ่มตัวอย่างควรมีขนาดใหญ่ เพื่อให้การคำนวณค่าสหสัมพันธ์มีความเสถียรมากขึ้น เช่น ไม่ต่ำกว่า 100 คน เป็นต้น

3.2 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory)

G-Theory เป็นทฤษฎีทางสถิติของการวิเคราะห์ความน่าเชื่อถือของผลการวัดในสถานการณ์ของการวัดผลลักษณะต่างๆ ที่เป็นเป้าหมายของการนำเครื่องมือไปใช้ ความน่าเชื่อถือของผลการวัด หมายถึง ความถูกต้องของการสรุปอ้างอิง (Generalization) จากคะแนนที่สังเกตได้ไปยังคะแนนจริงของบุคคล โดยคะแนนจริงเป็นคะแนนเฉลี่ยที่พึงได้ของผู้สอบแต่ละคน จากการสอบภายใต้สถานการณ์หรือเงื่อนไขของการวัดที่ยอมรับได้ทั้งหมด G-Theoryใช้โมเดลการวัดเชิงสุ่ม (Random-Effect Model) ดังนั้น ระดับหรือเงื่อนไขของแหล่งความคลาดเคลื่อนอย่างเป็นระบบที่นำมาศึกษาจะต้องเป็นเชิงสุ่ม (Random Facet) เพื่อให้การสรุปอ้างอิงสู่เอกภพของการสรุปอ้างอิงและการคำนวณได้ค่าความเที่ยง หรือ G-coefficient ที่มีความน่าเชื่อถือและถูกต้องในกรณีที่มีความจำเป็นจะต้องใช้เงื่อนไขของแหล่งความคลาดเคลื่อนแบบเจาะจง (Fixed Facet) เช่น การศึกษาคุณภาพของเครื่องมือวัด เจาะจงเฉพาะกลุ่มสาระวิชาที่สนใจ หรือเจาะจงเฉพาะกลุ่มกรรมการตรวจให้คะแนนเท่าที่มีอยู่ เป็นต้น นักวิจัยจะต้องคำนวณองค์ประกอบเชิงสุ่ม เพื่อปรับแก้ให้เป็นองค์ประกอบแบบเจาะจงที่สอดคล้องกับเงื่อนไขที่เป็นจริง

3.3 ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

ทฤษฎีการตอบสนองข้อสอบ ได้เสนอแนวทางอธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ภายในตัวบุคคล กับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกหรือได้คะแนนมากน้อยเพียงไร ทฤษฎีนี้มีพื้นฐานความเชื่อว่าพฤติกรรมการตอบสนอง

ต่อข้อสอบของผู้สอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิด หรือได้คะแนนมากน้อยเพียงใด จะถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่อยู่ภายในตัวบุคคลซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง ทฤษฎีนี้ได้อธิบายความสัมพันธ์ดังกล่าวในรูปของฟังก์ชันการตอบสนองข้อสอบ ซึ่งแสดงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบแต่ละข้อได้ถูก $[P_i(\theta)]$ กับระดับความสามารถของผู้สอบที่วัดได้โดยแบบสอบฉบับนั้น (θ) เมื่อนำมาเขียนเป็นกราฟ จะได้โค้งลักษณะข้อสอบ (Item Characteristic Curve; ICC) โค้งลักษณะข้อสอบมีหลายลักษณะ ขึ้นอยู่กับโมเดล (Model) หรือแบบจำลองที่ใช้อธิบายความสัมพันธ์ดังกล่าว โมเดลที่นิยมใช้กันคือ โมเดลแบบหนึ่งพารามิเตอร์ (One-Parameter Model) โมเดลแบบสองพารามิเตอร์ (Two-Parameter Model) และโมเดลแบบสามพารามิเตอร์ (Three-Parameter Model) ซึ่งจะต้องเลือกใช้ให้เหมาะสม และจะต้องทำการทดสอบให้แน่ใจว่าข้อมูลผลการทดสอบที่มีอยู่หรือข้อมูลเชิงประจักษ์มีความสอดคล้อง (fit) กับโมเดลการตอบสนองข้อสอบที่เลือกใช้นั้นจึงจะทำให้การแปลผลที่ได้จากการวิเคราะห์มีความน่าเชื่อถือและถูกต้อง

เอกสารอ้างอิง

ศิริชัย กาญจนวาสี. (2554). (พิมพ์ครั้งที่ 8). *ทฤษฎีประเมิน (Evaluation Theories).* กทม. : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2552). (พิมพ์ครั้งที่ 6). *ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory).* กทม. : คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2555). (พิมพ์ครั้งที่ 4). *ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theories).* กทม. : คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

Huck, S.W. (2009). Statistical Misconceptions. New York : Taylor & Francis Group. LLC.

Scriven, M. (1973). *"Goal - Free Evaluation." In E.R. House (Ed.), School Evaluation : The Politics and Process.* Berkeley, CA : McCutchan.

Stufflebeam, D.L., et.al (1971). *Educational Evaluation and Decision - Making.* Itasca, Illinois : Peacock Publishing.

ผู้เขียน

ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

e-mail: skanjanawasee@hotmail.com