

วิธีวิทยาขั้นสูงด้านการวัดและประเมินผลการศึกษา

ศิริชัย กาญจนวาสี

การพัฒนาทำให้เกิดความเจริญงอกงาม ความมั่งคั่งที่เกิดขึ้นเราทุกคนคาดหวังให้เป็น การเปลี่ยนแปลงไปในทิศทางที่พึงปรารถนา แต่ในขณะที่เดียวกันการเปลี่ยนแปลงย่อมมีผลกระทบทั้งทางบวกและทางลบต่อสังคม การเปลี่ยนแปลงที่เกิดจากการวางแผนและปฏิบัติชอบ ย่อมก่อให้เกิดความเจริญก้าวหน้า แต่ผลที่เกิดจากการเปลี่ยนแปลงอาจทำให้ปัญหาเปลี่ยนรูปแบบใหม่ จึงจำเป็นต้องการแนวคิดและวิธีการแก้ปัญหาแบบใหม่ที่แตกต่างไปจากเดิม

ขณะนี้เราเข้าสู่สังคมโลกยุคใหม่ ที่แต่ละสังคมถูกเชื่อมโยงเป็นสังคมโลกเดียวกันใน ด้านสารสนเทศ กระแสของข่าวสารและเทคโนโลยีเข้ามามีส่วนในการผลักดันการเคลื่อนไหว ของโลกอย่างรวดเร็ว และรุนแรง ซึ่งจะช่วยย่นย่อระยะเวลาของการสร้างสมองค์ความรู้ของ ศาสตร์ทุกแขนง ให้มีอัตราเร่งของความก้าวหน้าอย่างไม่เคยเป็นมาก่อน การพัฒนาองค์ความรู้ ของศาสตร์สาขาต่าง ๆ ได้เข้าสู่กระแสของการสื่อสารและใช้เทคโนโลยีเข้ามาเป็นเครื่องมือใน การศึกษาค้นคว้าวิจัย ซึ่งเป็นสิ่งที่เราจะต้องรับรู้เร่งความรู้เท่าทัน ปรับตัวและใช้ประโยชน์ใน ขณะที่ของเก่าที่ดีทรงคุณค่า จะต้องรักษาไว้ให้ได้และพัฒนาให้เหมาะสมกับสภาพการ

วิธีวิทยาทางด้านการวัดและประเมินผลการศึกษา ได้มีการศึกษามาเป็นเวลานาน และ เริ่มศึกษากันมาตั้งแต่ศตวรรษที่ 19 ซึ่งในระยะนั้นได้มีการเสนอทฤษฎีการทดสอบแบบคลาสสิก (classical test theory) ($X=T+E$) ก็ยังเป็นที่ยอมรับและยังใช้อย่างกว้างขวาง นับถึงปัจจุบัน นี้ ทฤษฎี classical มีอายุเกือบจะครบหนึ่งร้อยปีในอีกไม่กี่ปีข้างหน้า

บทความที่ถอดคำบรรยายจากการนำเสนอในการสัมมนาคุษฎีบัณฑิตสาขาการวัดและการประเมินผลการศึกษา ภาค วิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย “เรื่องวิธีวิทยาขั้นสูงด้านการวิจัย สถิติ การวัดและประเมินผลการศึกษา” วันจันทร์ที่ 21 สิงหาคม 2538 ณ อาคารศปฐศาลา จุฬาลงกรณ์มหาวิทยาลัย

แต่เนื่องจากข้อจำกัดบางประการของทฤษฎี classical จึงทำให้มีการพัฒนาปรับปรุงวิธีวิทยาด้านการวัดและประเมินผลการศึกษาเรื่อยมา รูปธรรมที่ปรากฏชัดคือในปี ค.ศ. 1980 ได้มีการประยุกต์ทฤษฎีการทดสอบแนวใหม่ ซึ่งเป็นทฤษฎีการวัดที่สามารถอธิบายความสัมพันธ์ระหว่างคุณลักษณะภายใน (trait) หรือความสามารถที่มีอยู่ภายในตัวบุคคล (true ability) กับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงใด

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) นี้มีพื้นฐานความเชื่อว่า พฤติกรรมการตอบสนองต่อข้อสอบของผู้สอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิด จะถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง

ทฤษฎี IRT นี้ จะอธิบายความสัมพันธ์ของคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคล กับพฤติกรรมการตอบสนองข้อสอบในรูปของฟังก์ชันทางคณิตศาสตร์ (ฟังก์ชันโลจิสติก ฟังก์ชันปกติสะสม) ซึ่งเรียกว่า โมเดลการตอบสนองข้อสอบซึ่งมีหลายรูปแบบ เช่น โมเดล 1, 2, 3 พารามิเตอร์ เป็นต้น

ทฤษฎี IRT นับได้ว่าเป็นวิธีวิทยาขั้นสูงของการวัดผล ทั้งนี้เพราะการวัดตามทฤษฎี IRT นี้สามารถประมาณค่าคุณลักษณะภายในของมนุษย์ หรือความสามารถของมนุษย์ได้อย่างน่าเชื่อถือ และสามารถอธิบายเชื่อมโยงกับพฤติกรรมภายนอกของมนุษย์ได้ ซึ่งเป็นสิ่งที่นักวัดผลทั้งหลายควรจะทำการศึกษาทฤษฎีนี้ให้เข้าใจอย่างลึกซึ้ง ทั้งนี้เพราะทฤษฎีทางการทดสอบที่จะเกิดขึ้นใหม่ ๆ ก็จะต้องเข้ามาเชื่อมโยงกับทฤษฎีดังกล่าวนี้ รวมทั้งการประยุกต์ IRT ในการศึกษา การทำหน้าที่ต่างกันของผู้สอบ (Differential Item Functioning; DIF) การเปรียบเทียบประสิทธิภาพระหว่างข้อสอบและแบบสอบ เป็นต้น ประกอบกับความก้าวหน้าของเทคโนโลยีด้านคอมพิวเตอร์ ก็จะมีส่วนสำคัญอย่างยิ่งต่อการพัฒนาและประยุกต์ทฤษฎีการวัดและประเมินผลทางด้านสังคมศาสตร์ คอมพิวเตอร์จะเข้าไปมีบทบาทในการประมาณสิ่งที่เป็นคุณลักษณะภายในของมนุษย์ได้ใกล้เคียงกับความเป็นจริงมากที่สุด ถึงแม้ว่าปัจจุบันจะยังไม่สามารถบอกได้ว่าถูกต้องที่สุดก็ตาม

ปัญหาที่เป็นพื้นฐานที่สุดที่ทำให้ศาสตร์ด้านการวัดและประเมินผลทางสังคมศาสตร์ ก้าวหน้าไปได้ช้ากว่าการวัดทางวิทยาศาสตร์ ซึ่งส่วนใหญ่เป็นการวัดทางด้านกายภาพ ก็คือปัญหาในเรื่อง validity ของการวัด trait ที่เราสนใจ ปัญหา validity ของการวัด เป็นปัญหา

ที่พื้นฐานที่สุด สำคัญที่สุด และยากที่สุด การวัดที่ดีจะต้องมีทฤษฎีที่ดีที่เป็นตัวนำว่าสิ่งที่เราต้องการวัดคืออะไร ประกอบด้วย component อะไรบ้าง และแต่ละ component มีอะไรเป็นตัวบ่งชี้ เพื่อนำไปสู่การสร้างข้อคำถามของแต่ละตัวบ่งชี้ เมื่อนำมาสร้างเป็นเครื่องมือแล้ว จะทำการทดสอบ validity ของเครื่องมือนั้นได้อย่างไร

ในการทดสอบ validity นั้น ในแนวของ classical validity จะทำการตรวจสอบ content validity, concurrent validity, predictive validity และ construct validity การตรวจสอบ validity ที่สำคัญที่สุดก็คือ construct validity หรือการตรวจสอบว่ามีความตรงตามทฤษฎีหรือไม่ เทคนิคที่นำมาใช้ในการตรวจสอบ construct validity ทาง classical ใช้วิธี Multitrait Multimethod (MTMM) หรือใช้ Exploratory Factor Analysis (EFA) สำหรับการตรวจสอบขั้นสูงควรใช้ confirmatory factor analysis (CFA) software ที่มีใช้อยู่ในปัจจุบัน ได้แก่ LISREL, LISCOM, EQS เป็นต้น

เทคนิค confirmatory factor analysis นี้ เป็นเทคโนโลยีของการวิเคราะห์ทางสถิติที่ดีที่สุดในปัจจุบัน ที่สามารถยืนยันได้ว่าเป็น construct validity ตามทฤษฎีหรือไม่ ทั้งนี้เพราะ construct validity ตามวิธีการนี้สามารถเชื่อมโยงไปถึงคุณลักษณะภายในของมนุษย์ได้ ถึงแม้ว่าปัจจุบันเรายังไม่แน่ใจว่าเป็นคุณลักษณะที่มุ่งวัดแท้จริงหรือไม่ แต่ถ้ามีการศึกษาและ confirm ด้วยเทคนิคนี้บ่อย ๆ ก็น่าจะยืนยันความเชื่อในความตรงของการวัดนั้นน่าจะมีความถูกต้องเหมาะสมมากขึ้น

นอกจากวิธีวิทยาการตรวจสอบความตรงแล้ว วิธีการศึกษาความเที่ยงทั่วไปเพื่อระบุถึงค่าความเที่ยงภายใต้สภาพการวัดต่าง ๆ กัน ก็เป็นสิ่งที่น่าสนใจมิใช่น้อย จากทฤษฎีการสรุปอ้างอิงทางการทดสอบ (Generalizability Theory; GT) สามารถที่จะช่วยให้นักวัดผลใช้ข้อมูลจากกลุ่มตัวอย่าง สรุปอ้างอิงถึงนัยทั่วไปของค่าความเที่ยงของการวัดภายใต้เงื่อนไขของการวัดต่าง ๆ กัน อันเป็นจุดอ่อนของทฤษฎี Classical Reliability เพื่อตอบคำถามว่าความเที่ยงควรจะเป็นเท่าไรเมื่อสภาพการวัดมีความหลากหลายในระดับต่าง ๆ กันของเงื่อนไขการวัด เช่น ความยาวของข้อสอบ ลักษณะเนื้อหาที่มุ่งวัด จำนวนครั้งของการสอน วิธีการวัด จำนวนผู้ตรวจให้คะแนน ระดับการวัด เป็นต้น ซึ่งจะช่วยนักวัดผลในการตัดสินใจทางการทดสอบว่า ถ้าต้องการใช้เครื่องมืออย่างมีคุณภาพ ควรจะใช้ภายใต้สถานการณ์การวัดลักษณะใด แนวทางของทฤษฎีนี้จะนำไปสู่การสรุปอ้างอิงนัยทั่วไปของความตรงต่อไปได้

แนวโน้มของการวัดผลในอนาคต สังคมเราจะต้องมีการสอบมากขึ้น และบ่อยขึ้น โดยจะยังคงใช้ concept ของ formative evaluation และ summative evaluation ปัญหาสำคัญอีกอย่างหนึ่งของการวัดผลจึงอยู่ที่การสร้างแบบสอบ ซึ่งปัญหาที่เป็นอยู่ในปัจจุบันก็คือ “การสร้างที่ใช้ที่” และไม่ค่อยจะได้วิเคราะห์ข้อสอบกัน แนวทางการแก้ไขปัญหานี้ก็คือการนำเทคโนโลยีมาช่วยในการสร้างข้อสอบที่ดีขึ้นมาครั้งเดียว แล้วสามารถทำการ generate ข้อสอบที่ดีมาใช้ได้บ่อย ๆ ขึ้น

ในการสร้าง item generator นั้น จะต้องทำการวิเคราะห์ construct ของสิ่งที่ต้องการวัดออกมาก่อนว่าประกอบด้วย component อะไรบ้าง และแต่ละ component จะประกอบด้วยตัวบ่งชี้อะไรบ้าง เมื่อกำหนด Test Specification และ Item Specification และสร้าง Item Forms แล้วใช้คอมพิวเตอร์ช่วยในการ generate item ออกมา ซึ่งก็จะได้ข้อสอบที่มีลักษณะคล้าย ๆ กัน จัดเป็นชุดของแบบสอบเพื่อวัดตามลักษณะของตัวแปรที่ต้องการได้

นอกจากนี้ยังสามารถกำหนดเป็น table of specification ขึ้น เมื่อต้องการสร้างแบบสอบเพื่อวัดคุณลักษณะนี้มี test spec. อย่างนี้ ก็สามารถ generate ออกมาเป็นฉบับได้เลย สร้างให้มีแบบสอบคู่ขนาน A, B, C ก็สามารทำได้ ซึ่งขณะนี้ก็มีเทคโนโลยีด้านคอมพิวเตอร์รองรับอยู่แล้ว เพียงแต่คนที่สร้างระบบนี้จะต้องมีพื้นฐานด้าน programming พอสมควร

เมื่อมีการ generate ข้อสอบและนำไปใช้ ทุกครั้งของการใช้ควรมีการวิเคราะห์ผลเพื่อเก็บเข้าคลังข้อสอบ (item bank) จากพื้นฐานความรู้เกี่ยวกับ Item Information ตามทฤษฎี IRT สามารถนำไปสู่ระบบการคัดเลือกข้อสอบเพื่อสร้างเป็นแบบสอบคู่ขนานที่น่าเชื่อถือ รวมทั้งระบบการคัดเลือก

ปัญหาประการต่อมาของการวัดผลก็คือ ในการสอบวัดผลสิ่งเดียวกัน แต่ทำการสอบต่างเวลา ต่างสถานที่ และต่างกลุ่มกัน คะแนนที่ได้จะมีการปรับเทียบกันได้อย่างไรจึงจะทำให้เกิดความยุติธรรมสำหรับผู้สอบ ซึ่งปัญหานี้สามารถแก้ไขได้โดยใช้เทคนิค test equating ซึ่งจะต้องศึกษาและกำหนด design ของการ equate ให้เป็นไปอย่างยุติธรรม มีสิ่งที่ต้องศึกษาอยู่ 3 ประเด็น คือ การสร้างแบบสอบให้มีความทัดเทียมกัน วิธีการเก็บข้อมูลและเทคนิคการปรับแก้คะแนนระหว่างแบบสอบ ซึ่งมีสูตร model ต่าง ๆ ที่ต้องศึกษาในรายละเอียดทั้ง Horizontal Equating และ Vertical Equating เพื่อให้ Standard Error of Equating มีค่าจำกัดอยู่ใน

ระดับที่ยอมรับได้

ปัญหาอีกประการหนึ่งของการวัดและประเมินผล ก็คือ การวัดผลในสิ่งที่อยู่ต่าง level กัน จะวัดอย่างไร และจะทำการวิเคราะห์ความสัมพันธ์ของผลการวัดต่างระดับกันอย่างไร เช่น ภาวะผู้นำของผู้บริหารโรงเรียนและพฤติกรรมการสอนของครูในชั้น ไปมีผลกระทบต่อผลสัมฤทธิ์ทางการเรียนของผู้เรียนได้อย่างไร เป็นต้น การวัดในลักษณะนี้เป็นลักษณะของการวัดที่เป็น multilevel จำเป็นต้องใช้ multilevel analysis ซึ่งมีเทคนิคและ software เช่น HLM ที่จะต้องทำการศึกษาในรายละเอียดต่อไป

สรุปประเด็นในเรื่องของวิธีวิทยาขั้นสูงของการวัดและประเมินผลทางการศึกษาที่น่าสนใจ และควรที่จะศึกษาวิธีการและนำมาใช้อย่างจริงจังได้แก่

1. การศึกษาการพัฒนาและการประยุกต์ทฤษฎีการตอบสนองข้อสอบ (IRT)
2. การศึกษาการพัฒนาและการประยุกต์ทฤษฎีการสรุปอ้างอิงทางการทดสอบ (GT)
3. concept และการตรวจสอบในเรื่องของ validity
4. การสร้างและการพัฒนาระบบ item generator/item bank
5. การสร้างและพัฒนาระบบ test equating
6. การวัดและการวิเคราะห์ multilevel analysis