ความรู้ชายแดนด้านการวัดผลการศึกษา

อุทุมพร จามรมาน

1. แหล่งและขอบเขต

ความรู้ชายแดนที่ปรากฏในบทความนี้ ได้มาจากการอ่านวารสาร 4 ฉบับ คือ (1) Journal of Educational Measurement (2) Educational Measurement : Issues and Practice (3) Applied Psychological Measurement และ (4) Psychometicka ในช่วง ค.ศ. 1993-1995

2. หัวข้อที่เกี่ยวข้อง

จากการอ่านวารสารดังกล่าวในช่วงเวลาที่ระบุ พบว่า มีการพูดถึงหัวข้อที่น่าจะยังคงต้องการการศึกษาค้นคว้าและวิจัยเพิ่มเติมอยู่ 7 หัวข้อ คือ

1. พื้นฐานและปัญหาในการวัด
2. การวัดแบบเดิม (Classical Test)
3. การวัดการตอบสนองรายข้อ (Item Response)
4. การทำคลัง
5. ทฤษฎีเปิดของคูมส์ (Coomb's unfolding)
6. การวัดความลำเอียง
7. การอ้างอิงสรุป

ดังมีรายละเอียดต่อไป

บทความนำเสนอในการสัมมนาดุษฎีบัณฑิตสาขาการวัดและประเมินผลการศึกษา เรื่องวิธีวิทยาขั้นสูงด้านการวิจัย สถิติ การวัดและประเมินผลการศึกษา วันจันทร์ที่ 21 สิงหาคม 2538 ณ ศศปฐศาลา จุฬาลงกรณ์มหาวิทยาลัย

3. พื้นฐานและปัญหาในการวัด

ประเด็นสำคัญของการวัดมี 2 ประเด็น คือ วัดอะไร วัดอย่างไร จึงจะตรงและเชื่อถือได้

3.1 การวัดอะไรนั้น อุทุมพร จามรมาน (2538) ได้ระบุว่า มีการวัดอยู่ 6 ประเภท คือ

1. วัดลักษณะทางร่างกาย เช่น อวัยวะ สุขภาพอนามัย
2. วัดลักษณะทางจิต และบุคลิกภาพ เช่น ทัศนคติ ค่านิยม บุคลิกภาพ ความรู้สึก
3. วัดความสามารถทางสมอง คือ สติปัญญา ความถนัด และผลสัมฤทธิ์ทางการเรียน
4. วัดลักษณะความสัมพันธ์ระหว่างคนและกลุ่มคน
5. วัดทักษะทางกล้ามเนื้อ
6. วัดทักษะทางปัญญา เช่น ทักษะทางภาษาและการแก้ปัญหาคณิตศาสตร์

ความรู้ชายแดนในประเด็นนี้ได้แก่

1. การวัดสติปัญญา

ทฤษฎีกระบวนการประมวลผลทางปัญญา (Cognitive Information Processing Theories) ได้ระบุว่าสติปัญญามี 3 ประการ คือ สติปัญญาที่เกี่ยวกับพันธุกรรม ประสบการณ์ และสิ่งแวดล้อม

2. การวัดภาคปฏิบัติ (Performance) ในโรงเรียนให้เน้นการวัดตามสภาพที่แท้จริงและการวัดประเด็นที่เกี่ยวข้องกับการดำรงชีวิตจริง โครงการวิจัยใหญ่ ๆ ที่กำลังดำเนินอยู่ได้แก่ โครงการวัดแบบ Authentic Assessment Portfolio Assessment และ Performance Based Assessment โครงการดังกล่าวเน้นที่สาระที่วัด วิธีการวัด ตัวชี้ของผลภาคปฏิบัติ และความถี่ของการวัดหลาย ๆ วิธี

3.2 การวัดอย่างไรจึงจะได้ข้อมูลที่เชื่อถือได้ เป็นประเด็นสำคัญในสาขาการวัดผลมานาน และยังคงเป็นประเด็นที่อภิปรายถกเถียงกันอยู่ว่า "ข้อมูล" ที่ได้นั้น สะท้อนธรรมชาติที่แท้จริงของผลการวัดหรือไม่ และมากน้อยเพียงใด

อุทุมพร จามรมาน (2538) ได้ระบุวิธีการวัดโดยอาศัยประสาทสัมผัสของคนไว้ 6 วิธี คือ (1) การสังเกตด้วยตา (2) การฟังด้วยหู (3) การถามด้วยปาก (4) การดมด้วยจมูก

(5) การจับต้องด้วยมือ (6) การผสมหลาย ๆ วิธี ซึ่งวิธีการดังกล่าวนำมาสู่การวัดในชื่อเรียกต่าง ๆ เช่น การสัมภาษณ์ การสังเกต การทดสอบ การสอบถาม

การวัดที่ให้ได้ข้อมูลที่เชื่อถือได้ ต้องอาศัยความชัดเจนของสาระที่วัด (วัดอะไร) กับความสอดคล้องของวิธีการกับสาระที่วัด ดังนั้น การวัด (Measurement) จึงเน้นที่ข้อมูลที่ได้ว่ามีความเที่ยง (Reliability) และความตรง (Validity) มากกว่าเครื่องมือ (Measure) หรืออีกนัยหนึ่งก็คือการเน้นที่ความเที่ยงและความตรงของข้อมูล**มากกว่า**ความเที่ยงและความตรงของเครื่องมือ

ความรู้ชายแดนในประเด็นนี้คือ

1. ทำอย่างไรจึงจะจัดวิธีการวัดให้ตรงกับสาระที่วัด แล้วให้ได้ข้อมูลที่เชื่อถือได้ เช่น การวัดความสัมพันธ์ของคน 2 คน จะวัดอย่างไรจึงจะได้ข้อมูลที่สะท้อนความสัมพันธ์ของคน 2 คนจริง
2. สูตร KR-20,21 เป็นสูตรที่วัดความสอดคล้องภายในได้จริงหรือไม่
3. การหาค่าความเที่ยง (Reliability) ของข้อมูลที่เป็นภาคบรรยาย (Qualitative หรือ Polytomous) จะทำอย่างไร
4. การหาค่าความเที่ยงของข้อมูลที่ได้จากวิธีการวัดแบบ Complex Multiple Choice (CMC) หรือ Multiple True False (MTF) จะทำอย่างไร (หมายเหตุ ข้อความแบบ MTE หมายถึง ตัวเลือกทั้งหลายไม่เป็นอิสระต่อกัน)

4. การวัดแบบเดิม (Classical Test Theory)

การวัดแบบเดิมตั้งอยู่บนฐานของการพิจารณาค่าคะแนนที่ได้รับ(X) คะแนนจริง (T) และคะแนนความคลาดเคลื่อน (E) ภายใต้สมการเชิงเส้นตรง คือ

$$X_i = T_i + E_i$$

โดยที่ i หมายถึงคนที่ i

ประเด็นที่เป็นความรู้ชายแดนได้แก่

1. สมการที่ไม่ใช่เส้นตรง (Non-Linear) น่าจะสะท้อนค่าที่ได้รับได้ดีกว่า
2. การวัดค่าที่แท้จริง (True) ทำอย่างไร ซึ่งเป็นปัญหาของการวัดในข้อ 3 ที่กล่าวมาแล้ว

3. สูตรความเที่ยงแบบ KR-20,21 ยังคงเป็นสูตรที่ระบุค่าความสอดคล้องภายในใช่หรือไม่

4. จะระบุความเชื่อถือได้ของผลการวัดได้อย่างไร

5. ข้อตกลงเบื้องต้นที่นำไปสู่การได้สูตรความเที่ยง เช่น ความคลาดเคลื่อนเป็นอิสระต่อกัน ค่าจริงมีความสัมพันธ์เป็น 0 ยังคงจำเป็น และสะท้อนความเป็นไปได้มากน้อยเพียงไร การฝ่าฝืนข้อตกลงเบื้องต้นดังกล่าว จะทำให้มีวิธีหาค่าความเที่ยงแบบอื่นได้หรือไม่

5. การวัดการตอบสนองรายข้อ

ทฤษฎี Item Response Theory ได้เน้นที่การวัดรายข้อ โดยมุ่งที่การประมาณคุณลักษณะของข้อ (item) 3 ประการ คือ ความยาก อำนาจจำแนก และการเดาถูก ซึ่งถ้าประมาณค่า parameter ของข้อได้ ก็สามารถคำนวณค่าความสามารถ (θ) ของผู้ตอบได้

ทฤษฎีนี้ตั้งอยู่บนข้อตกลงเบื้องต้น เช่น ลักษณะที่วัดเป็นมิติเดียว (Unidimension) และความเป็นอิสระภายในของข้อ (Local Independence) ซึ่งในปัจจุบันพบว่า ลักษณะที่วัดมิได้มีมิติเดียว หากแต่มีหลายมิติ เช่น ความสามารถทางสมอง ลักษณะทางจิตใจ ทักษะทางร่างกาย ดังนั้น การฝ่าฝืนข้อตกลงเบื้องต้น จึงเป็นสิ่งที่น่าสนใจว่าจะยังคงใช้ทฤษฎีได้หรือไม่

ความรู้ชายแดนที่เกี่ยวข้องได้แก่

1. การวัดมิติพหุ (Multidimension) โดยอิงทฤษฎี IRT จะทำอย่างไร

2. ความไม่เป็นอิสระภายใน (Local Dependence) จะกระทบทฤษฎีนี้มากน้อยเพียงใด

3. ทฤษฎีนี้ให้คำตอบที่เกี่ยวข้องกับลักษณะของข้อ 3 อย่างคือ ความยาก อำนาจ-จำแนก และการเดาถูก (3 parameter) ซึ่งในเชิงคณิตศาสตร์ parameter ที่เกี่ยวข้องกับเส้นโค้ง (ICC) มีมากกว่า 3 อย่าง การนำ parameter ตัวที่ 4,5 มาประยุกต์กับลักษณะของข้อจะทำได้หรือไม่ และให้ความหมายถึงการวัดรายข้อหรือไม่ (parameter ตัวที่ 4,5 คืออะไรบ้าง)

4. ข้อความที่นำมาใช้กับทฤษฎี IRT ใช้กับ Dichotomous Item ซึ่งไม่เพียงพอต่อการวัดลักษณะบางอย่าง ซึ่งไม่จำเป็นต้องเป็นข้อความประเภทนี้เท่านั้น การวัดลักษณะเชิงบรรยาย ซึ่งให้ค่าและคะแนนที่มิใช่ 0,1 ถ้าจะนำมาใช้กับทฤษฎีนี้ได้ และควรทำอย่างไร (ข้อความเหล่านี้บางทีเรียกว่า logic item หรือ polytomous item)

6. การทำคลัง (Banking)

การทำคลังข้อสอบและคลังแบบทดสอบ โดยอิงคอมพิวเตอร์กับทฤษฎี IRT เป็นสิ่งที่เป็นไปได้ ได้มีโปรแกรมสำเร็จรูปที่สามารถจัดเก็บข้อสอบเพื่อการเรียกใช้และปรับเปลี่ยนค่าสถิติ หลังการใช้แล้ว

ความรู้ชายแดนในเรื่องนี้ ได้แก่

1. การทำคลังแบบทดสอบ (Test Bank) เพื่อการใช้ให้ตรงกับจุดประสงค์การวัด และเนื้อหาที่วัดในหลักสูตร ยังไม่ได้มีผู้ศึกษาค้นคว้าไว้ ผู้ที่ให้ความคิดในเรื่องนี้คือ Eva Baker แต่การดำเนินงานยังไม่ได้ทำจริงจัง

2. การสร้างข้อสอบ (Theory Based Item Construction) ตามโครงสร้างเดิม พบว่าทำได้ในเนื้อหาที่เกี่ยวข้องกับความสามารถทางสมอง (Cognitive) เช่น การสร้างข้อสอบภายใต้เนื้อหาที่เกี่ยวกับทรง (shape) จำนวน (number) ทำได้แล้ว การสร้างข้อสอบในเนื้อหาอื่นจะทำอย่างไร

3. การทดลองใช้ข้อสอบ แล้วนำข้อมูลมาคำนวณค่าพารามิเตอร์ของข้อสอบรายข้อ มักมีปัญหาในเรื่องจำนวนคน ความเป็นตัวแทนของประชากร วิธีสุ่มคน ซึ่งถ้าใช้ทฤษฎี CTT ในการวิเคราะห์ต้องระมัดระวังในเรื่องลักษณะกลุ่มที่เป็นตัวแทน (Norm group) และถ้าใช้ทฤษฎี IRT ในการวิเคราะห์ต้องเน้นความระมัดระวังในเรื่องจำนวนคนที่จะนำมาประมาณค่าพารา-มิเตอร์ของข้อ

ความแตกต่างระหว่างทฤษฎีทั้ง 2 ในการวิเคราะห์ข้อสอบ มีดังนี้

CTT	IRT
1. จำเป็นต้องใช้กลุ่มปรกติวิสัย (norm group)	1. ไม่จำเป็น
2. ให้ค่า p,r	2. ให้ค่า a,b,c
3. ค่าที่ได้เปลี่ยนแปลงตามกลุ่ม	3. ค่าที่ได้ไม่เปลี่ยนแปลงตามกลุ่ม ถ้าได้จำนวนคนมากพอ
4. คำนวณง่ายและคำนวณจากเครื่องคิดเลขได้	4. คำนวณจากคอมพิวเตอร์

CTT	IRT
5. ค่าความสามารถของคนเป็นค่าที่ได้จากการสังเกต (observed score)	5. ค่าความสามารถของคน คำนวณจากคุณลักษณะของข้อ 6. สามารถสร้างกลุ่มปรกติวิสัยได้หลังจากการทดลองใช้หลาย ๆ ครั้ง (สามารถ build up กลุ่มปรกติวิสัยได้) 7. สามารถสร้างลักษณะการแจกแจงข้อมูลตามลักษณะที่กำหนดได้ โดยการ Simulate data 8. ใช้ได้กับข้อความแบบ polytomous 9. สามารถทำ item calibration ได้โดยวิธี (ก) link (ข) pool data 10. ใช้ได้กับลักษณะมิติพหุ 11. ให้สารสนเทศของข้อ (item information) 12. สามารถระบุความลำเอียงของข้อ (item bias) ได้จาก Differential Item Function (DIF) โดยวิธี SIBTEST ของ Shealy and Stout

6. ทฤษฎีเปิดของคูมส์ (Coombs' Theory of Unfolding)

คูมส์ (1950) ได้เสนอว่า คนทุกคนมีความพึงพอใจ (Preference) ของตนอยู่เมื่อให้ตัดสินสิ่งเร้าทีละคู่ เช่น เขาจะพึงพอใจ x มากกว่า y จุดของความพึงพอใจของเขาบนเส้นต่อเนื่องจะอยู่ใกล้ x มากกว่า y

ดังนั้นทฤษฎีนี้ระบุว่า

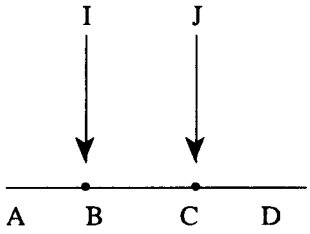
1. ตัวแปรสิ่งเร้าตัวหนึ่งที่เกี่ยวข้องกับการตัดสินเรื่องความพึงพอใจ
2. ตัวแปรนี้เป็นตัวแปรเชิงปริมาณ
3. คนแต่ละคนจะมีค่าของความพึงพอใจสูงสุงบนตัวแปรนั้น

4. ความพึงพอใจในตัวแปรสิ่งเร้าลดลง ระยะห่างระหว่างสิ่งเร้ากับจุดของความพึงพอใจจะเพิ่มขึ้น

5. คนทั้งหมดเห็นด้วยกับตำแหน่งของสิ่งเร้าบนมิติที่เกี่ยวข้อง

คูมส์เรียกจุดของความพึงพอใจสูงสุดของคนว่า จุดอุดมคติของเขาคนนั้น และเรียกความพึงพอใจในการจัดลำดับสิ่งเร้าว่า มาตร I (I คือ ลำดับความพึงพอใจของเขา)

ตัวอย่างเช่น มีสิ่งเร้า A,B,C,D และมีค่าความพึงพอใจ ซึ่งแสดงด้วยจุดบนเส้นตรงเส้นหนึ่ง



คน I มีจุดบนเส้นตรงต่อเนื่องที่แสดงถึงความพึงพอใจสูงสุดของเขา ดังนั้นลำดับที่ของความพึงพอใจสิ่งเร้าจึงเป็น A B C D อีกคนหนึ่ง มีความพึงพอใจสูงสุดอยู่ที่ J ลำดับความพึงพอใจของเขาจึงเป็น C D A B

ทฤษฎีของคูมส์ มีจุดอ่อนคือ

1. ดูเหมือนว่าจะเป็นทฤษฎีที่เกี่ยวกับมิติเดียว ซึ่งในกรณีที่สิ่งเร้ามีหลายมิติ ทฤษฎีนี้จะใช้ไม่ได้

2. การระบุว่า ทุกคนมีความพึงพอใจสูงสุดเพียง 1 ค่า อาจไม่จริง เพราะบางคนพอใจเลือก x แทนที่จะเลือก y เพราะเงื่อนไขที่ต่างกัน เช่น คนที่เลือกกาแฟไม่ใส่น้ำตาลและนม แทนกาแฟที่ใส่น้ำตาลและนมอาจเป็นเพราะเขาอยู่ในช่วงลดความอ้วน หรือเขาอาจชอบกาแฟแบบนี้จริงก็ได้

3. การระบุให้ความมากน้อยของความพอใจเกี่ยวข้องกับระยะจากจุดที่เขาพอใจสูงสุดเป็นการระบุที่ไม่มีทิศทาง (direction) ว่าข้างซ้ายหรือข้างขวาของจุด และอัตราการเพิ่มหรือลดจะเท่ากันหรือไม่

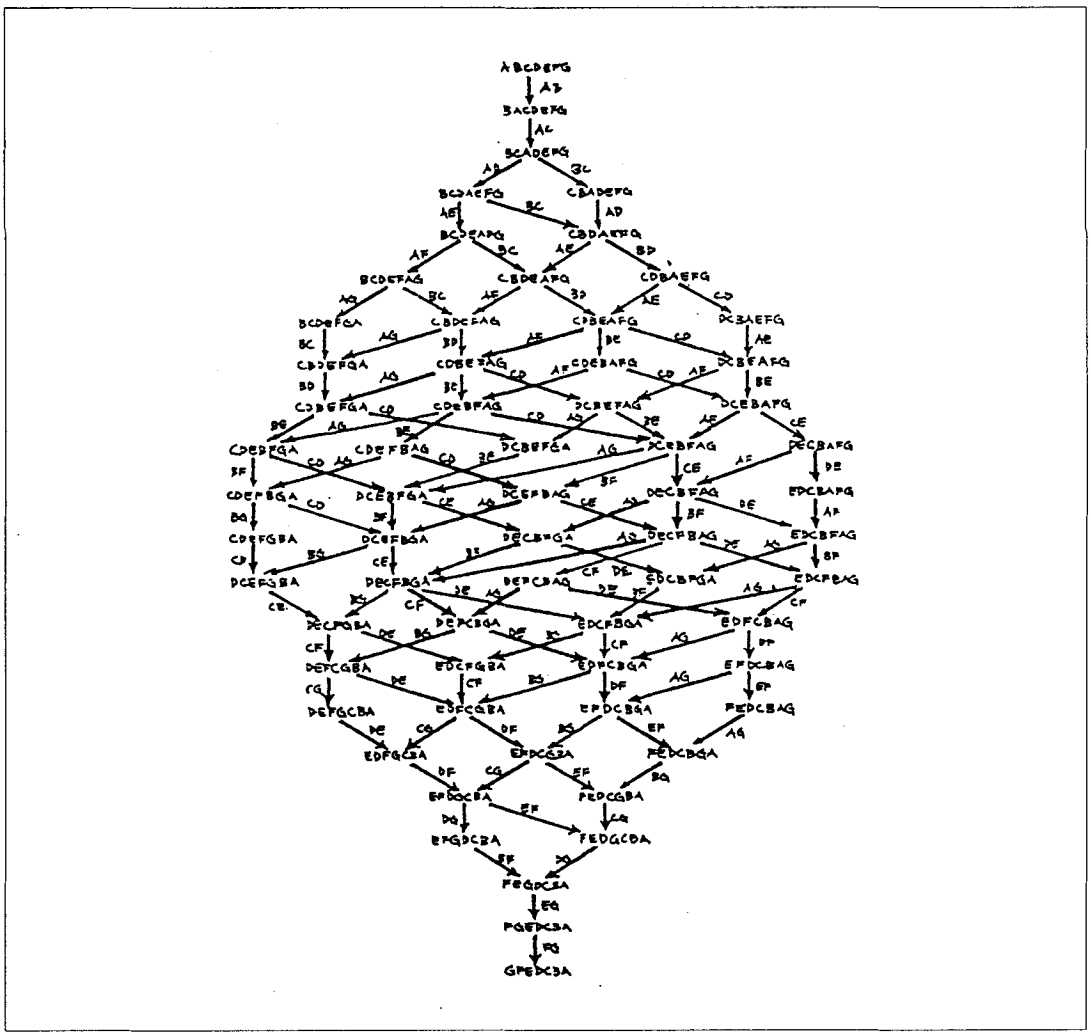
4. ตำแหน่งของสิ่งเร้าบนมิติที่เกี่ยวข้องได้มาจากการตัดสินของคนกลุ่มหนึ่ง ซึ่งคนในกลุ่มอาจมีความเห็นที่ไม่สอดคล้องกัน ทำให้ตำแหน่งของสิ่งเร้านั้นไม่เป็นตัวแทนของกลุ่ม

ตัวอย่าง คูมส์ได้ให้นักศึกษา 62 คนตัดสินเกี่ยวกับเกรด (A^+, A, A^-, B^+, B, B^-, C^+) โดยให้พิจารณาเกรดทีละคู่แล้วระบุว่าพอใจเกรดใดมากกว่า

เพื่อจะได้ไม่ต้องเกี่ยวข้องกับเครื่องหมายจะใช้สัญลักษณ์ A, B, C, D, E, F, G แทนเกรดดังกล่าวในการจัดลำดับความพึงพอใจต่อเกรด 7 ตัวนี้จะมี $2^6 = 64$ แบบ

การวิเคราะห์ให้พิจารณาคำตอบที่สอดคล้องในแนวเดียวกันกับลำดับของ A, B, C, D, E, F, G ของคนได้ 58 คน จาก 62 คน ที่จัดลำดับไปในทางเดียว (Transitive Order)

ผลการวิเคราะห์ของคูมส์เปิดเผยให้เห็นลำดับที่ของความพึงพอใจในเกรดทั้ง 7 ตัว ดังนี้



7. การวัดความลำเอียงของข้อ (Item Bias)

การที่ข้อสอบเดียวกันวัดกลุ่มคน 2 กลุ่ม แล้วให้ลักษณะของข้อที่แตกต่างกัน อาจมาจากการลำเอียงของผู้ตอบ ซึ่งสามารถตรวจสอบได้ โดยอิงความรู้เรื่อง Differential Item Information (DIF) กับข้อมูล Dichotomous เพื่อเทียบ (equate) ค่า/คะแนนที่ได้

ความรู้ชายแดนที่เกี่ยวข้องคือ

1. วิธีการของ Shealy and Stout ชื่อ SIBTEST เป็นวิธีที่ได้รับการแนะนำว่าเหมาะสมกว่าวิธีของคนอื่น การหาวิธี detect ความลำเอียงของข้อมูล Polytomous ยังเป็นสิ่งที่จะต้องกระทำต่อไป
2. ปัจจัยที่มีผลต่อความลำเอียงของข้อ เช่น เพศ ภาคภูมิศาสตร์ เขต ภูมิหลัง เป็นปัจจัยที่ก่อให้เกิดความลำเอียงจริงหรือไม่ และยังมีปัจจัยอื่นอีกหรือไม่

8. การอ้างอิงสรุป (Generalizability Theory)

ได้มีการระบุว่า แหล่งที่เกี่ยวข้องกับผลสัมฤทธิ์ทางการเรียนมี 5 แหล่ง คือ

1. all possible tasks
2. all possible occasions
3. all possible raters
4. all possible measurement methods
5. all possible level units (คน โรงเรียน)

ซึ่งเมื่อประยุกต์กับทฤษฎีอ้างอิงสรุป (Generalizability Theory) พบว่าเป็นปัจจัยที่มีผลต่อค่าที่สะท้อนผลสัมฤทธิ์ทางการเรียน

ความรู้ชายแดนในที่นี้ได้แก่

1. ยังคงมีปัจจัยอื่นที่เกี่ยวข้องอีกหรือไม่ เช่น item formats, variable types, sample size
2. ปัจจัยดังกล่าวส่งผลกระทบด้วยค่า effect size มากน้อยเพียงใด

สรุป

จากการระบุความรู้ชายแดน (Borderline of knowledge) ในสาขาการวัดผล พบว่ายังมีประเด็นที่ต้องการการศึกษาค้นคว้าอีกมาก บางประเด็นที่เคยคิดว่าเป็นความรู้จริงก็กลับมาสู่ข้อสงสัย เช่น สูตรคำนวณค่าความเที่ยงแบบ KR-20,21 หรือความเข้าใจพื้นฐานที่เกี่ยวกับความเที่ยงตรงของเครื่องมือหรือของข้อมูล

บทความนี้ต้องการกระตุ้นให้นักวิชาการวัดผลได้เกิดการตื่นตัว และเพิ่มการคิดวิเคราะห์ในสาระที่เกี่ยวข้อง เพื่อสร้างศาสตร์การวัดผลให้เป็นศาสตร์ที่แข็งแกร่งกว่านี้