

“การพัฒนาธนาคารข้อทดสอบ”

อัจฉรา วงศ์โสธร

แนวคิดเกี่ยวกับการจัดทำธนาคารข้อสอบมีมาตั้งแต่ ค.ศ. 1968 ดังที่ปรากฏในเอกสารเรื่อง Item Banking เขียนโดย R. wood และ L.S. Skurnik แห่ง National Foundation for Educational Research และบทความของ R. wood เรื่อง “The Place and Value of Item Banking” ในวารสาร Educational Research ค.ศ. 1968 ประโยชน์สำคัญเบื้องต้นของธนาคารข้อสอบ คือ เพื่อให้ผู้สอนใช้ทดสอบประจำภาค สอบระหว่างเรียน และสำหรับกรรมการสอบภายนอกได้ใช้ในการวัดระดับความรู้ความสามารถของผู้เรียนนอกเหนือจากการเรียนการสอน

ในประเทศไทย สำนักงานทดสอบของกรมวิชาการ กระทรวงศึกษาธิการก็ได้มีการเก็บรวบรวมข้อทดสอบต่าง ๆ เข้าไว้เป็น Item Pool เพื่อนำมาใช้ในภายหลัง และยังได้มีการตั้งคณะกรรมการเพื่อจัดสร้างข้อทดสอบมาตรฐานภาษาอังกฤษระดับมัธยมปลายขึ้นใน พ.ศ. 2518 เพื่อให้โรงเรียนต่าง ๆ นำไปใช้สอบเทียบระดับความสามารถภาษาอังกฤษของนักเรียนของตน และเพื่อทราบเกี่ยวกับประสิทธิภาพในการจัดการเรียนการสอนภาษาอังกฤษของโรงเรียนอีกด้วย การจัดทำ Item Pool นี้เป็นในลักษณะเก็บรวบรวมแบบทดสอบมาตรฐานไว้เฉพาะเรื่อง โดยระวังรักษาความปลอดภัย ป้องกันการรั่วไหลของข้อสอบ และมีระเบียบการในการใช้อย่างเคร่งครัด

แนวคิดข้างต้นทำให้เกิดการพัฒนาคลังข้อสอบและธนาคารข้อสอบขึ้น

คลังข้อสอบ หมายถึง ที่เก็บสะสมข้อสอบไว้โดยไม่กำหนดพารามิเตอร์ หรือเกณฑ์มาตรฐานของข้อสอบที่เก็บ ตลอดจนไม่คำนึงถึงการพัฒนาและการถ่ายเปลี่ยนข้อสอบ

ธนาคารข้อสอบ หมายถึง ระบบที่ประกอบขึ้นด้วยขั้นตอนต่าง ๆ ได้แก่

(1) การนำข้อสอบเข้าเก็บตามพารามิเตอร์ที่กำหนด เช่น วัตถุประสงค์ของหลักสูตร พฤติกรรมในการสอบของผู้เรียน ซึ่งบ่งถึง สมรรถนะของการสื่อสารทางภาษา (communicative competence)

(2) การนำเข้^{ขึ้น}เป็นไปตามเกณฑ์ที่กำหนด เช่น ข้อสอบต้องมีค่าความยากง่ายค่าอำนาจจำแนกในระดับที่ประสงค์และเป็นที่ยอมรับ เช่น ค่าความยากง่าย 25 %-75 % และค่าอำนาจจำแนก .30 ขึ้นไป

(3) การจัดหมวดหมู่ของข้อสอบเป็นไปตามแต่ละพารามิเตอร์ เช่น การอ่านเพื่อความเข้าใจ ในระดับท้ายโอน การกำหนดพารามิเตอร์ของข้อสอบอาจใช้ตามแบบของ Valette และ Disick ดังน^{ี้} :-

วัตถุประสงค์	วัตถุประสงค์ทั่วไป		วัตถุประสงค์จำเพาะภาษา	
	การฟัง	การอ่าน	การพูด	การเขียน
1. ทักษะเขียน	<ul style="list-style-type: none"> - การจำแนกเสียงที่เป็นอิสระต่างประเภทย่อยจากเสียงของภาษาแม่ 	<ul style="list-style-type: none"> - การจำแนกเสียงที่ปรากฏในประโยคหรือบทจากเสียงที่แยกออกจากตนเอง - ความสัมพันธ์ระหว่างเสียงและสัญลักษณ์ 	<ul style="list-style-type: none"> - การออกเสียง - การออกจำคำพูด 	<ul style="list-style-type: none"> - การทำงานของมือในการเขียน - การประสานงานระหว่างสายตาและกล้ามเนื้อในการเขียน - การคัดลอก การลากเส้น ความเร็วประ และ การเขียนขีดตัวอักษร
2. ความเข้าใจเขียน	<ul style="list-style-type: none"> - ความสัมพันธ์ระหว่างเสียงและความหมาย - ความเข้าใจคำพูดที่เขียน 	<ul style="list-style-type: none"> - ความสัมพันธ์ระหว่างสัญลักษณ์และความหมาย - เข้าใจประโยคที่เขียน 	<ul style="list-style-type: none"> - การตอบคำถามที่เขียน - การใช้รูปแบบทางภาษาที่ถูกต้อง 	<ul style="list-style-type: none"> - การเขียนประโยคที่เขียนและเปลี่ยนแปลงหรือต่อเติมประโยคที่เขียน
3. ทักษะโอน (บริบทจำกัด)	<ul style="list-style-type: none"> - เข้าใจคำพูดที่ไม่เขียน 	<ul style="list-style-type: none"> - เข้าใจข้อความที่ไม่เขียน - รู้ศัพท์ที่ไม่เขียนโดยอาศัยบริบทที่เขียนเป็นเครื่องช่วย - การถ่ายทอดความหมายเดิม โดยใช้รูปแบบประโยคที่แตกต่างออกไป - การเข้าใจเชิงอรรถในบางกรณี 	<ul style="list-style-type: none"> - การเติมประโยคเข้าด้วยกัน หรือการเปลี่ยนแปลงรูปแบบประโยคได้ตามคำสั่ง เช่น เปลี่ยนรูปแบบประโยคออกแล้วเติมประโยคคำถาม 	<ul style="list-style-type: none"> - การเขียนและเปลี่ยนแปลงประโยคให้เป็นข้อความที่สื่อความหมาย
4. สื่อความหมาย (บุคลิก)	<ul style="list-style-type: none"> - เข้าใจข้อความที่มีเนื้อความต่อเนื่องที่ไม่เขียน เช่น บทสนทนา สบรยายบทเพลง 	<ul style="list-style-type: none"> - เข้าใจข้อความที่มีเนื้อความต่อเนื่องที่ไม่เขียน - เข้าใจรายละเอียดปลีกย่อย สรุปความเหตุและผล ที่ปรากฏในเนื้อความ 	<ul style="list-style-type: none"> - การพูดเพื่อสื่อความหมายโดยไม่มีข้อจำกัด 	<ul style="list-style-type: none"> - การเขียนเพื่อสื่อความหมาย โดยไม่มีข้อจำกัด
5. ทักษะวิจารณ์ (บุคลิก)	<ul style="list-style-type: none"> - ความละเอียดอ่อนต่อความแตกต่างของรูปแบบ สัทภาพ และท่าเขียนของภาษา - โทณณ์เจตนาที่แฝงอยู่ของผู้พูด - ผลกระทบทางจิตตต่าง ๆ ที่เขียนทางประและไม่เป็นทางการ การแยกแยะความหมายที่ต่างกัน โดยใช้การฟัง สัทภาพที่ต่างกัน 	<ul style="list-style-type: none"> - อารมณ์ความรู้สึก ความคิดเห็นและทัศนคติของผู้เขียนที่แฝงอยู่ในวิธีการเขียน - ทัศนคติในการวิเคราะห์และประเมินค่าของภาษา 	<ul style="list-style-type: none"> - พูดโต้แย้งกับผู้อื่น และโต้แย้งกับตนเอง 	<ul style="list-style-type: none"> - เขียนโต้แย้งกับผู้อื่น ภาษาต่าง ๆ และถ้อยคำอารมณ์และทัศนคติต่าง ๆ

ข้อสอบที่นำเข้า เก็บรักษา พัฒนา และนำออกใช้ได้ในธนาคารข้อสอบนั้น สามารถเป็นได้ทั้งข้อสอบแบบปรนัย และอัตนัย

ข้อสอบปรนัย : True/False

Multiple Choice

Matching

นำเข้าไว้ได้พร้อมทั้งตัวเลือก และค่าสถิติของตัวลงและตัวเลือกที่ถูก

ข้อสอบอัตนัย : Essay Dictation

Translation

Cloze

Oral Cue

นำเข้าได้ในส่วนที่เป็นคำสั่ง หัวเรื่องที่ให้เขียนบท Dictation และข้อความที่ใช้ทดสอบแบบ cloze และ cue คำถาม หรือบทสนทนาสำหรับการสอบพูด

ค่าสถิติเป็นของข้อทดสอบทั้งชุดหรือทั้งตอน ไม่ใช่แต่ละข้อแบบข้อสอบปรนัย

แนวคิดในการพัฒนาธนาคารข้อทดสอบ อาศัยวิธีการสุ่ม (sampling) อย่างมีระเบียบวิธี โดยที่จำนวนของข้อทดสอบในธนาคารจำต้องมีมากกว่าจำนวนที่จำเป็นที่จะต้องใช้ในแต่ละครั้งอย่างน้อยมากกว่า 1 เท่าขึ้นไป และมีเป็นอัตราส่วนตามความสำคัญ และความจำเป็นของเนื้อหาระดับทักษะทางภาษา ตลอดจนทักษะทางภาษา

การพัฒนาธนาคารข้อทดสอบพัฒนาขึ้นได้ตามหลักสูตรภาษาเพื่อใช้วัดและประเมินผล การเรียนการสอนภาษา ซึ่งอาศัยหลักสูตร ปริมาณข้อทดสอบในคลังจะมีตามอัตราความสำคัญของเนื้อหา และทักษะที่บรรจุไว้ในหลักสูตร ซึ่งระบุไว้ในตารางวิเคราะห์หลักสูตร เช่น หลักสูตรรายวิชาภาษาอังกฤษพื้นฐาน I ซึ่งเป็นหลักสูตรแบบ ทักษะสัมพันธ์ (integrated skills)

ตัวอย่าง ตารางวิเคราะห์หลักสูตรแบบทักษะสัมพันธ์

ทักษะ (skill)	เนื้อหา (contents)	หน้าที่ภาษา (functions)	ความสำคัญ
ฟัง-พูด			ฟัง-พูด 40%
ฟัง-พูด	- บทสนทนา	- ถาม/ตอบเพื่อให้ได้ข้อมูล	10%
ฟัง-อ่าน	- คำบรรยาย		
ฟัง-เขียน	- โฆษณาทางวิทยุและโทรทัศน์	- พูดให้ข้อมูล อธิบายขยายความ	10%
	- ประกาศทางวิทยุและโทรทัศน์ เช่น พยากรณ์อากาศข่าวในและนอกประเทศ	- พูดเพื่อโน้มน้าวใจ	5%
		- ฟังเพื่อจับใจความ	7%
		- เขียนสรุปคำบรรยาย	4%
		- อ่านเพื่อจับใจความ ข้อความที่เกี่ยวข้องกับบทสนทนา/ โฆษณา/ประกาศและรายการที่ได้ฝึกฟัง-พูดแล้ว	4%
อ่าน-เขียน			อ่าน-เขียน 60%
อ่าน-เขียน	หนังสือพิมพ์	- อ่านเพื่อจับใจความ	30%
อ่าน-พูด	สารานุกรม	- อ่านเพื่อพูดสรุป	5%
อ่าน-ฟัง	ประกาศ	- อ่านเพื่อเขียนสรุป	10%
	บันทึกข้อความ	- อ่านเพื่อเขียนอธิบายขยายความ	7%
	จดหมายส่วนตัว		
	จดหมายธุรกิจ	- อ่านและฟังข้อความที่เกี่ยวข้อง	4%
	ข้อความแบบบรรยายโวหาร	- อ่านเพื่อพูดอธิบายขยายความ	4%
	ความยาว 200-300 คำ		

จากตารางจะเห็นว่า อัตราส่วนของข้อทดสอบในธนาคารจะเป็นฟัง-พูด 40% และอ่าน-เขียน 60% ซึ่งอัตรานี้ใช้เป็นเกณฑ์ในการพัฒนาธนาคารข้อทดสอบให้มีจำนวนข้อในธนาคารที่สอดคล้องกับหลักสูตร

นอกจากนี้ธนาคารข้อทดสอบยังอาจพัฒนาขึ้นตามการใช้ภาษาในชีวิตจริงโดยมีลักษณะเป็นข้อทดสอบที่พัฒนาขึ้นเพื่อใช้กับธนาคารข้อทดสอบที่อิงเกณฑ์ ซึ่งก็จำเป็นต้องอาศัยตารางวิเคราะห์เช่นกัน แต่ไม่ใช่ตารางวิเคราะห์หลักสูตรหากแต่เป็นตารางวิเคราะห์ภาษาที่ใช้ในชีวิตประจำวัน ซึ่งจะใช้เป็นเกณฑ์ในการพัฒนาข้อทดสอบวัดความสามารถการใช้ภาษา (proficiency test) ขึ้น ผู้ที่จะสามารถกำหนดตารางขึ้นได้ คือ นักวิชาการทางการใช้ภาษา ผู้ที่ใช้ภาษาและนักวิจัยทางภาษา

ตัวอย่าง ตารางวิเคราะห์การใช้ภาษาในชีวิตประจำวัน

ทักษะ	เนื้อหา	หน้าที่ภาษา	ความสำคัญ
อ่าน	เอกสาร วารสาร หนังสือพิมพ์ บทความ เรื่องสั้น นวนิยาย	จับใจความ สรุปความ	50%
ฟัง	บทสนทนา เพลง ข่าว	เพื่อการสังคม เพื่อความบันเทิง เพื่อความรู้อ	25%
เขียน	จดหมาย บันทึก รายงาน	ถาม/ตอบเพื่อให้ได้ข้อมูล เพื่อใช้อ้างอิง เพื่อเตือนความจำ	20%
พูด	เรื่องส่วนตัว เรื่องธุรกิจ บรรยาย/อธิบาย บอกทิศทาง	ถาม/ตอบเพื่อให้ได้ข้อมูล	5%

ปริมาณของข้อทดสอบที่พัฒนาขึ้น และบรรจุในธนาคารจะมีส่วนสัมพันธ์ตามความสำคัญของการใช้ในชีวิตจริง

ดังนั้น ผู้ที่จะนำข้อทดสอบจากธนาคารมาใช้ ก็จะใช้วิธีสุ่มอย่างมีระเบียบวิธีโดยอาศัยหลักเกณฑ์ข้างต้นว่า ควรเป็นข้อทดสอบทักษะใดในเนื้อหา และหน้าที่ของภาษาแบบใด เป็นจำนวนมากน้อยเท่าใด

ข้อทดสอบที่นำเข้าเก็บไว้ในธนาคารจะต้องมีความเป็นมาตรฐาน คือ มีค่าสถิติที่เป็นที่ยอมรับว่า ดีมาก ดี หรือใช้ได้ ข้อสอบใดที่ดีกว่ามาตรฐานที่กำหนดไว้ก็ต้องมีการปรับปรุงจนได้มาตรฐานเสียก่อนจึงจะนำมาเข้าธนาคาร

หลักการในการพัฒนาการข้อสอบขึ้นมีอยู่ 3 แบบ ซึ่งเป็นอิสระต่อกัน ดังนี้

1. การพัฒนาการข้อสอบตามหลักการเดิม (traditional หรือ classical model)

ตามหลักการเดิม การวิเคราะห์ข้อสอบอาศัยแนวคิดของ Chung Teh Fan ในการหาค่าความยากง่าย และค่าอำนาจจำแนกของข้อสอบ ซึ่งเป็นการวิเคราะห์แบบอิงกลุ่มซึ่งยึดความสามารถของกลุ่มที่ทำการทดสอบ และมีข้อตกลงเบื้องต้นว่า เป็นกลุ่มปกติ คือ คะแนนที่ได้มีการกระจายแบบปกติ (normal distribution) และมีค่าที่ได้จากการวิเคราะห์ตามเกณฑ์ที่ยอมรับคือค่าอำนาจจำแนก $.30 >$ และค่าความยากง่าย 25%–75%

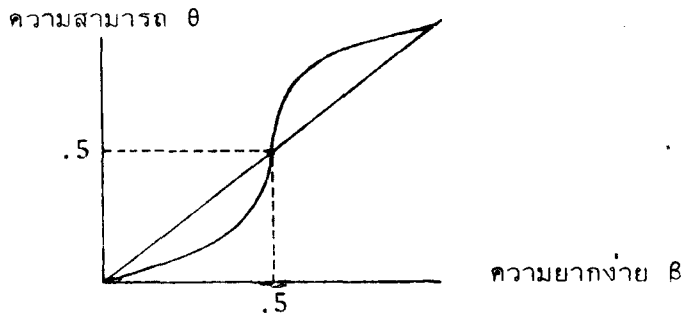
2. การพัฒนาการข้อสอบโดยใช้หลักการของการทดสอบอิงเกณฑ์ (criterion referenced model)

ข้อทดสอบที่มีมาตรฐานตามทฤษฎีอิงเกณฑ์ไม่ต้องมีคะแนนที่มีการกระจายปกติ แต่จะมีการกระจายแบบเบ้ไปทางขวา (negative skew) และวิเคราะห์ตามเกณฑ์ได้/ตก ข้อทดสอบที่พัฒนาขึ้นตามหลักการนี้ใช้เป็นข้อสอบระหว่างเรียน เช่น ข้อสอบประจำบท (unit test)

3. การพัฒนาการข้อสอบโดยหลักการของทฤษฎีความสามารถแฝง (latent trait model)

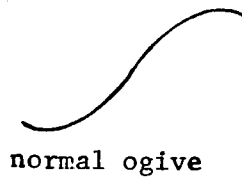
ทฤษฎีความสามารถแฝงพัฒนาขึ้นมาจากแนวคิดที่ว่า คุณสมบัติของข้อสอบควรจะขึ้นอยู่กับคุณสมบัติของตัวข้อสอบเอง ไม่ใช่อิงอยู่กับความสามารถของกลุ่มบุคคลที่ทำการสอบ ดังนั้นจึงสามารถใช้เป็นหลักการพื้นฐานของการทดสอบอิงเกณฑ์ได้เป็นอย่างดี

Lord (1952) ได้กำหนดว่า ในการทดสอบมีปัจจัย 2 อย่างที่เกี่ยวข้อง คือ ความสามารถของบุคคล (θ) และความยากง่ายของข้อทดสอบ (β) และกำหนดความสัมพันธ์เป็นเส้นตรง (linear) และมีสมมติฐานว่า ถ้า $\theta = .5$ β ก็จะมี $= .5$ คือ ความสัมพันธ์น่าจะเป็น 1

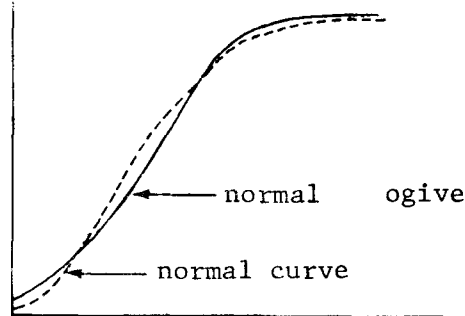


แต่ Lord ไม่สามารถพิสูจน์สมมติฐานนี้ได้

นักคณิตศาสตร์ ชื่อ Birnbaum (1968) ซึ่งเป็นผู้ให้กำเนิดแนวคิดใหม่ทางความสามารถแฝง ตั้งสมมติฐานว่า ความสัมพันธ์ควรจะเป็นเส้น normal ogive มากกว่าเส้นตรง



Lord และ Novick (1968) ได้เพิ่มพารามิเตอร์อีก 2 ตัว คือ c = การเตาและ a = slope คือ ค่า inflection point หรือค่าที่จุดตัด ซึ่งจะเป็นค่าอำนาจจำแนกในขณะที่ β คือ ค่าความยากง่ายของข้อสอบ



การปรับค่า normal ogive เป็น logistic normal curve โดยใช้ค่า $d = 1.7$ และใช้การคำนวณค่า normal curve แทนค่า normal ogive ซึ่งมีความแตกต่างกันเพียงประมาณ 0.01 แต่คำนวณได้ง่ายกว่า

Albert Rasch (1960) นักคณิตศาสตร์ชาวสวีเดนเป็นผู้สร้าง Rasch model ขึ้น ใน model นี้ ไม่มีค่า a และ c มีแต่ค่าพารามิเตอร์ β หรือความยากง่ายเท่านั้น การคำนวณหาค่า β ที่ได้โดยใช้ \log ดังนั้น model นี้เรียกว่า logistic model จึงถูกขนานนามว่าเป็น one-parameter logistic model ในขณะที่ model ของ Lord ได้ชื่อว่าเป็น three-parameter logistic model เพราะประกอบด้วยค่า θ , β และค่า c

Ben Wright แห่ง University of Chicago เป็นผู้ให้นำ Rasch Model มาเผยแพร่ และได้เขียนตำราชื่อ Best Test Design (1979) โดยอาศัย one-parameter logistic model ของ Rasch เป็นบรรทัดฐานของการวิเคราะห์ข้อสอบ Model นี้ มีแต่ β เพราะมีแนวคิดว่าการเดาเป็นความสามารถอย่างหนึ่งของผู้ทำแบบทดสอบมีแฝงอยู่ในค่า ϕ แล้ว

เช่น ถ้าข้อสอบที่มีตัวเลือก 4 ตัว ค่า c ของ 1 ตัวเลือก = .25 ของ 2 ตัวเลือก = .25² และของ 3 ตัวเลือก = .25³ ดังนั้น จึงไม่ต้องคิดค่า c (ค่าที่ผู้สอบที่ไม่มีความสามารถอาจเดาได้ถูก) เพราะมีน้อยมาก และค่า a คือ ค่าอำนาจจำแนก ก็ไม่จำเป็นเพราะเป็นการวัดความสามารถในการทำสอบ จำเป็นแต่ค่าความยากง่าย β

ค่าความสัมพันธ์ของคะแนนแต่ละข้อ และคะแนนรวมของแบบทดสอบ (r_{xy}) เป็นดัชนีของความเที่ยงตรง

สูตรในการหาค่าของความเป็นไปได้ที่คน ๆ หนึ่งจะตอบข้อสอบถูกหรือผิด คือ

$$P \{ x_{ve} \mid B_v, \theta_e \} = \text{Exp} [x_{ve} (B_v - \theta_e)] / [1 + \text{Exp} (\beta_v - \theta_e)]$$

p = probability

x = คะแนน

ve = ถูกหรือผิด มีค่า 1 และ 0

$/$ = ทหาร

Exp = Exponential

$|$ = given

หมายความว่า โอกาสที่จะทำถูกหรือผิดขึ้นอยู่กับความยากง่ายของข้อสอบ และความสามารถของคน ๆ หนึ่งในการทำสอบ

ในการพัฒนาการข้อทดสอบที่ใช้ Rasch Model เป็นสิ่งจำเป็นที่นักทดสอบทางภาษาพึงทราบเกี่ยวกับ ข้อตกลงเบื้องต้น (assumptions) ของ Rasch Model ซึ่งกำหนดคุณสมบัติของข้อทดสอบไว้ว่าต้องมี :-

(1) ความเป็นมิติเดียว (unidimensionality) คือ ข้อสอบต้องวัดสิ่ง ๆ เดียว หรือคุณลักษณะเดียว คือ ความเป็น homogenous แต่ในกลุ่มของข้อสอบที่วัดคุณลักษณะหลาย ๆ ด้าน อาจมีการจัด cluster เดียว โดยวิธีการ factor analysis ก่อน จึงนำมาพัฒนาเข้า วิชาการข้อสอบที่ใช้หลักการของ Rasch

(2) ต้องมีความเป็นอิสระในตำแหน่งที่อยู่ (local independent) และความเป็นอิสระทางสถิติ (statistical independent) ข้อทดสอบทางภาษาแบบเติมความ เติมคำ และโคลง ขนานแท้ (real cloze) และแบบเลือกตอบต่าง ๆ สามารถใช้กับ model นี้ได้ แต่ถ้าเป็นข้อสอบ multiple point เช่น การเขียนเรียงความ ซึ่งคะแนนที่ได้ไม่ใช่ 1/0 สำหรับถูก-ผิด แล้ว ข้อสอบ โคลงแบบจับคู่ (matching cloze) ไม่สามารถใช้กับ Rasch model ได้ เพราะแต่ละข้อไม่เป็น อิสระต่อกัน เพราะตัวเลือกที่ใช้แล้วไม่สามารถใช้ได้อีก แต่ ตามวิธีการของคณิตศาสตร์แล้ว พบว่าข้อตกลงเบื้องต้นข้อที่ 1 ครอบคลุมข้อที่ 2 ด้วย ความเป็นมิติเดียวจึงสำคัญที่สุด ดังนั้น การวิเคราะห์ข้อทดสอบชนิดดังกล่าวก็สามารถใช้ logistic model ได้

(3) θ หรือความสามารถต้องมีการกระจายแบบปกติ แต่ว่าถ้ามีการสุ่มมามากพอควร ก็สามารถใช้กับ Rasch model ได้

Latent trait model มีประโยชน์ดังต่อไปนี้ คือ

(1) ใช้วิเคราะห์ข้อสอบโดยไม่ต้องอาศัยกลุ่มของผู้ทดสอบเป็นเกณฑ์ (sample free)

(2) ขจัดความลำเอียงของข้อทดสอบ (item bias) ที่เกิดจากตัวแปรต่าง ๆ เช่น ทางเศรษฐกิจ-สังคม และวัฒนธรรม โดยหาค่าเฉลี่ยของแต่ละกลุ่ม เช่น ของคนผิวขาวและคนผิวดำ และพิสูจน์ว่า จะเข้ากับเส้น ogive ได้อย่างมีนัยสำคัญหรือเปล่า

(3) การจัดข้อทดสอบให้เป็นรายบุคคล (tailoring test) ทั้งนี้เพราะคุณลักษณะสำคัญของ Rasch model คือ การที่ข้อทดสอบเป็นอิสระต่อกัน คือ ไม่ว่าจะอยู่ที่ใด ก็จะมีค่าคงที่

และการที่ข้อทดสอบเป็นอิสระไม่ขึ้นกับบุคคล คุณลักษณะที่ใช้ประโยชน์ได้ดีในการพัฒนาการ ข้อทดสอบสำหรับวัดเป็นรายบุคคล

- (4) ใช้ในการทำข้อทดสอบเพื่อวินิจฉัยปัญหาได้เพราะ ใช้เกณฑ์ความยากง่าย/ความสามารถ
- (5) ใช้กับการทดสอบแบบอิงเกณฑ์ได้เพราะใช้เกณฑ์การได้-ตก ซึ่งขึ้นกับความยากง่าย/ความสามารถของแต่ละบุคคล
- (6) ใช้กับข้อทดสอบหลายวัตถุประสงค์ได้ โดยสามารถใช้วิธีการ factor analysis จัดกลุ่มข้อสอบ และหา loading factor ได้ก่อนแล้วจึงนำเข้าธนาคารข้อทดสอบ
- (7) ใช้กำหนดจุดตัดของคะแนน (cut-off score) ได้
- (8) ใช้ในการทำสมการเทียบคะแนนของข้อทดสอบต่างชุดกัน
- (9) ใช้ในการสร้างข้อทดสอบได้ โดยจำแนกประเภทข้อสอบว่า ข้อใดพวกที่เรียนอ่อนมีแนวโน้มทำได้ ข้อใดพวกที่เรียนดีมีแนวโน้มทำได้ และข้อใดที่พวกความสามารถระดับกลางสามารถทำได้ และทำข้อสอบที่ผู้ที่มีความสามารถสูงทำได้เป็นข้อสอบจัดระดับ หรือยกเว้นการเรียน และข้อสอบที่พวกความสามารถปานกลางและอ่อนมีแนวโน้มทำได้เป็นข้อสอบวินิจฉัย เพื่อจัดการเรียนการสอนซ่อมเสริม

ธนาคารข้อทดสอบมีประโยชน์ในการประยุกต์ใช้ทั้งทางด้านบริหารและด้านวิชาการ ดังนี้ :-

ประโยชน์ทางด้านบริหาร

- (1) ทำให้ไม่เกิดการซ้ำซ้อนในการลงแรงงาน และการสิ้นเปลืองเวลาและทุนทรัพย์ ในการสร้างและพัฒนาข้อสอบตามวัตถุประสงค์เดียวกัน เพราะธนาคารข้อสอบเป็นแหล่งรวบรวมข้อสอบภาษาไว้ตามพารามิเตอร์
- (2) ทำให้ผู้บริหารการศึกษา ครู ผู้ปกครอง และผู้เรียนได้ทราบถึงผลที่เที่ยงตรงของการเรียนการสอน ซึ่งวัดได้โดยข้อสอบในธนาคารในเวลาอันรวดเร็วกว่าการวัดโดยจำต้องกำหนด ทำให้ผลที่ได้อาจไม่เที่ยงตรงเท่าที่ควร

ประโยชน์ทางด้านวิชาการ

- (1) ทำให้รักษามาตรฐานของการทดสอบและประเมินผลการเรียนการสอนภาษา
- (2) ทำให้ได้มีข้อสอบที่ได้มาตรฐาน และสร้างตามพารามิเตอร์ที่พึงประสงค์เป็นจำนวนมาก
- (3) ทำให้การทดสอบเป็นไปได้อย่างมีประสิทธิภาพ คือ ผู้เรียนสามารถสอบได้เมื่อพร้อม (individualized testing)
- (4) ทำให้การทดสอบไม่จำกัดระดับ (flexi-level testing) โดยมีข้อสอบทั้งง่ายปานกลาง และยากต่อผู้เรียนที่อยู่ระดับต่าง ๆ กันบรรจุไว้ในธนาคาร
- (5) ทำให้มีการพัฒนาข้อสอบให้เหมาะสมกับความสามารถของผู้เรียนแต่ละคน (tailored testing)
- (6) ทำให้เพิ่มพูนศักยภาพในการเรียนรู้ภาษาและสมรรถวิสัยในการใช้ภาษาเพื่อการสื่อสารของผู้เรียนโดยการวัดและประเมินผลที่สม่ำเสมอและมีมาตรฐาน
- (7) ทำให้ได้ข้อสอบชนิดต่าง ๆ จำนวนมากที่สร้างขึ้นตามพารามิเตอร์ต่าง ๆ เพื่อใช้ในการเรียนการสอนและการสอบ ทั้งนี้เป็นเพราะว่าข้อสอบที่นำเข้าไปมีต่างชนิดกัน
- (8) ทำให้แน่ใจได้ว่า การสอบแต่ละครั้งมีมาตรฐานเท่าเทียมกัน
- (9) ทำให้ผู้สอนได้จัดเตรียมการเรียนการสอนซ่อมเสริมสำหรับผู้เรียนเป็นรายบุคคลได้
- (10) ทำให้ผู้เรียนได้ทราบความก้าวหน้าในการเรียนของตนอย่างสม่ำเสมอ ตลอดจนได้มีโอกาสแก้ไขปัญหาในการเรียนที่ละชั้น ๆ

แต่กระนั้นก็ตาม การใช้ธนาคารข้อสอบโดยปราศจากความระมัดระวังอาจทำให้เกิดผลเสียได้เนื่องจาก

การใช้ธนาคารข้อสอบอาจนำไปสู่การสอนเพื่อการสอบ เช่น การสอนตามโรงเรียนกวดวิชา ไม่ใช่การสอนเพื่อการเรียนรู้ที่แท้จริง ซึ่งจะทำให้เกิดผลเสียดังนี้

1. นักเรียนมีแรงจูงใจและทัศนคติที่ผิดต่อการเรียน ซึ่งถ้าสะสมมากขึ้นจะทำให้เป็นผู้ที่หวังผลเร็ว ขาดความพยายามที่แท้จริง
2. นักเรียนจะขาดพื้นฐานความรู้ซึ่งเป็นสิ่งค่อยสะสม และพัฒนาขึ้นตามการเรียนรู้ ความรู้ที่มิจากการท่องจำข้อสอบจะเป็นความรู้ที่ขาดพื้นฐานมั่นคง

3. ผู้สอนอาจจะใช้วิธีการสอนที่ผิด เพราะสอนตามข้อทดสอบการเรียนรู้ และไม่ได้เป็นไปตามขั้นตอนจิตวิทยา

4. ผู้บริหารการศึกษาที่ยึดแนวการเรียนการสอนตามข้อทดสอบในธนาคารจะแปลงสภาพสถานศึกษาเป็นสถานกวดวิชา

ถึงแม้ว่าธนาคารข้อทดสอบอาจนำไปสู่ผลเสียที่พึงระวัง การพัฒนาธนาคารข้อสอบเป็นสิ่งจำเป็นของการเรียนการสอนที่มีการประเมินความก้าวหน้าในการเรียนด้วยการใช้การวัดผลที่สม่ำเสมอ และการเพิ่มพูนประสิทธิภาพของการทดสอบและประเมินผลด้วยการใช้ข้อทดสอบที่ได้มาตรฐาน มีความเที่ยงตรง เชื่อถือได้ และเป็นไปตามพารามิเตอร์ของการใช้ภาษาเพื่อการสื่อความหมาย

เอกสารอ้างอิง

- Birnbaum, A. 1968. "Some Latent Trait Models and Their Use in Inferring an Examinee's Ability." In F.M. Lord and M.R. Novick (eds.) *Statistical Theories of Mental Test Scores*. Reading, Mass. : Addison-Wesley.
- Fan, C.T. 1952. *Item Analysis Table*. Princeton, N.J. : Educational Testing Service.
- Jaeger, R.M. (ed.) 1975. "Applications of Latent Trait Models." *Journal of Educational Measurement*. East Lansing, Mich : National Council on Measurement in Education, Inc.
- Lord, F.M. 1952. "A Theory of Test Scores." *Psychometric Monographs*, No. 7.
- Lord, F.M. and M.R. Novick, 1968. *Statistical Theories of Mental Test Scores*. Menlo Park, Calif. : Addison-Wesley.
- Lord, F.M. 1980. *Applications of Item Response Theory to Practical Testing Problems*. New Jersey : Lawrence Erlbaum-Associates, Inc.
- Macintosh, H.G. (ed.) 1974. *Techniques and Problems of Assessment: A Practical Handbook for Teachers*. London: Edward Arnold, Ltd.
- Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmark : Danmarks Paedagogiske Institute.
- Valette, R.M. and R.S. Disick, 1972. *Modern Language Performance Objectives and Individualization*. New York : Harcourt Brace Jovanovich, Inc.
- Wood, R. 1968. "The Place and Value of Item Banking." *Educational Research*.
- Wood, R. and L.S. Skurnik, 1969. *Item Banking*. Slough : NFER.
- Wright, B.D., and M.H. Stone, 1979. *Best Test Design*. Chicago, ; Mesa Press.