

การพัฒนาแบบทดสอบ

การประเมินลักษณะมนุษย์

สันทัต พรประเสริฐมานิต

โครงร่างการนำเสนอ

- แนวคิดของแบบทดสอบ
- การสร้างแบบทดสอบ
- การทดลองใช้แบบทดสอบ
- การวิเคราะห์ข้อคำถาม
- การทบทวนแบบทดสอบ

แนวคิดของแบบทดสอบ

- วัตถุประสงค์ของแบบทดสอบ (เช่น วิจัย, ประเมิน, ทำนาย, จำแนก)
- กลุ่มเป้าหมายของแบบทดสอบ (เช่น ผู้ใหญ่, เด็ก, ผู้ป่วย)
- มีกระบวนการดำเนินการแบบทดสอบอย่างไร (เช่น ใช้ผู้ประเมิน, วัดแบบกลุ่ม, ใช้คอมพิวเตอร์)

การสร้างแบบทดสอบ

- การสร้างมาตร (Scaling) คือ การสร้างกฎในการสร้างตัวเลขจากข้อความ
- มีหลายรูปแบบ แต่วิธีที่ใช้มากที่สุดคือมาตรแบบผลรวมจากการประเมิน (Summated Rating Scale)
 - มาตรวัดแบบลิเคิร์ต
 - ผลรวมของข้อความที่ถูก
 - วิธีนี้เป็นวิธีที่สถิติที่พัฒนาขึ้นมาส่วนใหญ่ ถูกออกแบบให้นำมาใช้ เช่น การวิเคราะห์องค์ประกอบ (Factor analysis)
- ดังนั้น ในวิชานี้จะพูดถึงเฉพาะการสร้างมาตรรูปแบบนี้

มาตรลิเคิร์ตมีลักษณะดังนี้

1. ตัวเลือกมีการเรียงอันดับ
2. ช่องว่างระหว่างตัวเลือกในอันดับใกล้เคียงกัน สามารถเหมาว่าระยะห่างเท่ากันได้

How satisfied are you with*

	Very Unsatisfied	Unsatisfied	Neutral	Satisfied	Very Satisfied
The quality of our brand	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The prices we offer	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The speed of service we provide	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The customer support offered	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Are you basically satisfied with your life? yes NO

Have you dropped many of your activities and interests? YES no

Do you feel that your life is empty? YES no

Do you often get bored? YES no

Are you in good spirits most of the time? yes NO

Are you afraid that something bad is going to happen to you? YES no

Do you feel happy most of the time? yes NO

Do you often feel helpless? YES no

Do you prefer to stay at home, rather than going out and doing new things? YES no

Do you feel like you have more problems with memory than most? YES no

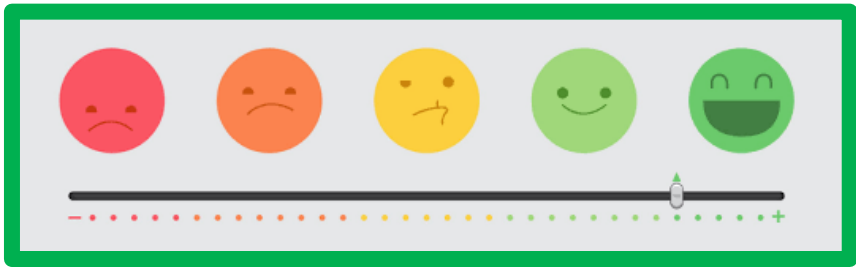
Do you think it is wonderful to be alive now? yes NO

Do you feel pretty worthless the way you are now? YES no

Do you feel full of energy? yes NO

Do you feel that your situation is hopeless? YES no

Do you think that most people are better off than you are? YES no



มาตรการตอบถูกผิด
แม้อาจไม่สามารถมองได้ว่าเป็น
มาตรลิเคิร์ตเต็มที่ แต่สถิติส่วนใหญ่
ที่มาตรลิเคิร์ตใช้ สามารถใช้กับมาตร
ตอบถูกผิดได้

Q3

How often do you conduct survey:

- Weekly
- Monthly
- Quarterly
- Annually

Likert-Type Scale Response Anchors

Citation:

Vagias, Wade M. (2006). *Likert-type scale response anchors*. Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management. Clemson University.

<p>Level of Acceptability</p> <ul style="list-style-type: none"> • 1 – Totally unacceptable • 2 – Unacceptable • 3 – Slightly unacceptable • 4 – Neutral • 5 – Slightly acceptable • 6 – Acceptable • 7 – Perfectly Acceptable <p>Level of Appropriateness</p> <ul style="list-style-type: none"> • 1 – Absolutely inappropriate • 2 – Inappropriate • 3 – Slightly inappropriate • 4 – Neutral • 5 – Slightly appropriate • 6 – Appropriate • 7 – Absolutely appropriate <p>Level of Importance</p> <ul style="list-style-type: none"> • 1 – Not at all important • 2 – Low importance • 3 – Slightly important • 4 – Neutral • 5 – Moderately important • 6 – Very important • 7 – Extremely important <p>Level of Agreement</p> <ul style="list-style-type: none"> • 1 – Strongly disagree • 2 – Disagree • 3 – Somewhat disagree • 4 – Neither agree or disagree • 5 – Somewhat agree • 6 – Agree • 7 – Strongly agree <p>Knowledge of Action</p> <ul style="list-style-type: none"> • 1 – Never true • 2 – Rarely true • 3 – Sometimes but infrequently true • 4 – Neutral • 5 – Sometimes true • 6 – Usually true • 7 – Always true <p>Reflect Me?</p> <ul style="list-style-type: none"> • 1 – Very untrue of me • 2 – Untrue of me • 3 – Somewhat untrue of me • 4 – Neutral • 5 – Somewhat true of me • 6 – True of me • 7 – Very true of me 	<p>My beliefs</p> <ul style="list-style-type: none"> • 1 – Very untrue of what I believe • 2 – Untrue of what I believe • 3 – Somewhat untrue of what I believe • 4 – Neutral • 5 – Somewhat true of what I believe • 6 – True of what I believe • 7 – Very true of what I believe <p>Priority:</p> <ul style="list-style-type: none"> • 1 – Not a priority • 2 – Low priority • 3 – Somewhat priority • 4 – Neutral • 5 – Moderate priority • 6 – High priority • 7 – Essential priority 	<p>Level of Support/Opposition</p> <ul style="list-style-type: none"> • 1 – Strongly oppose • 2 – Somewhat oppose • 3 – neutral • 4 – Somewhat favor • 5 – Strongly favor <p>Level of Probability</p> <ul style="list-style-type: none"> • 1 – Not probable • 2 – Somewhat improbable • 3 – Neutral • 4 – Somewhat probable • 5 – Very probable <p>Level of Agreement</p> <ul style="list-style-type: none"> • 1 – Strongly disagree • 2 – Disagree • 3 – Neither agree or disagree • 4 – Agree • 5 – Strongly agree <p>Level of Desirability</p> <ul style="list-style-type: none"> • 1 – Very undesirable • 2 – Undesirable • 3 – neutral • 4 – Desirable • 5 – Very desirable
	<p>Level of Concern</p> <ul style="list-style-type: none"> • 1 – not at all concerned • 2 – Slightly concerned • 3 – Somewhat concerned • 4 – Moderately concerned • 5 – Extremely concerned <p>Priority Level</p> <ul style="list-style-type: none"> • 1 – Not a priority • 2 – Low priority • 3 – Medium priority • 4 – High priority • 5 – Essential <p>Level of Problem</p> <ul style="list-style-type: none"> • 1 – Not at all a problem • 2 – Minor problem • 3 – Moderate problem • 4 – Serious problem 	<p>Level of Participation</p> <ul style="list-style-type: none"> • 1 – No, and not considered • 2 – No, but considered • 3 – Yes <p>Frequency – 5 point</p> <ul style="list-style-type: none"> • 1 – Never • 2 – Rarely • 3 – Sometimes • 4 – Often • 5 – Always
	<p>Affect on X</p> <ul style="list-style-type: none"> • 1 – No affect • 2 – Minor affect • 3 – Neutral • 4 – Moderate affect • 5 – Major affect <p>Level of Consideration</p> <ul style="list-style-type: none"> • 1 – Would not consider • 2 – Might or might not consider • 3 – Definitely consider 	<p>Frequency</p> <ul style="list-style-type: none"> • 1 – Never • 2 – Rarely • 3 – Occasionally • 4 – A moderate amount • 5 – A great deal <p>Frequency of Use</p> <ul style="list-style-type: none"> • 1 – Never • 2 – Almost never • 3 – Occasionally/Sometimes • 4 – Almost every time • 5 – Every time

มาตรวิเคิรต์ตอจมีทั้งแบบด้านบวกและลบ
หรือด้านบวกเพียงอย่างเดียว

มาตรด้านบวกเพียงอย่างเดียว จะเหมาะสม
รวมว่าระยะห่างเท่ากันยากกว่า และ
คะแนนรวมไม่สามารถบอกได้ว่าคนตอบ
มองเป็นด้านลบหรือด้านบวก

Check (✓) the statements with which you agree.

- 1. Abortion brings happiness to some, unhappiness to others. (5.1)
- 2. Abortion should be encouraged for all unwanted pregnancies. (10.4)
- 3. Abortion prevents an unwanted child growing up in an orphanage. (8.0)
- 4. Abortion is allowable if there is no financial cost to the public. (6.0)
- 5. Abortion weakens the moral fiber of our society. (1.6)
- 6. We must allow abortions for certain situations. (8.7)
- 7. Abortion is never justified. (1.1)
- 8. It doesn't make any difference to me whether we allow abortions or not. (5.6)
- 9. Most people are too lax about allowing abortions. (3.2)
- 10. People who get abortions deserve our understanding, but not special treatment. (7.6)
- 11. Abortions should be easier to obtain than they are now. (9.1)
- 12. Abortion is justified only in instances of rape. (3.6)
- 13. We can't call ourselves civilized as long as we allow abortions. (2.5)
- 14. Abortion may be unacceptable, but people deserve our sympathy. (4.6)
- 15. If a woman wants an abortion, it is her right. (10.1)
- 16. Abortion may be the best option in certain cases. (7.1)
- 17. People who get abortions should be treated like everybody else. (9.5)
- 18. Abortion cannot be regarded as a sane method of dealing with unwanted pregnancies. (2.2)
- 19. I think allowing abortion is necessary, but I wish it weren't. (6.6)
- 20. Giving up a child for adoption is a good alternative to abortion. (4.1)

ตัวอย่างมาตรวัดแบบอื่น เช่น การให้คะแนนแบบ Thurstone ที่การตอบถูกในแต่ละข้อแสดงถึงระดับของเจตคติในเรื่องดังกล่าว (ซึ่งเป็นแนวการให้คะแนนในสถิติแบบ Item Response Theory และ Computer Adaptive Test)

การสร้างแบบทดสอบ

- การเขียนข้อคำถาม ให้นึกถึงความตรงเชิงเนื้อหาเป็นหลัก
- รูปแบบของข้อคำถามมีจำนวนมาก ดูข้อเปรียบเทียบที่ตาราง 8.1
- เป็นส่วนที่ยากที่สุด เพราะต้องใช้จินตนาการเพื่อให้ข้อคำถามที่วัดภาวะสันนิษฐานให้ตรงจุด
- ให้นึกถึงกลุ่มตัวอย่างที่จะอ่านคำถามเสมอ ว่าอ่านรู้เรื่องหรือไม่ มีจุดที่ทำให้สับสนหรือมีนหรือไม่มี เนื้อหาที่ไม่เกี่ยวข้องหรือไม่ ก่อให้เกิดอคติหรือไม่
- ตรวจสอบความถูกต้องเสมอ
- หากเป็นข้อคำถามปลายเปิด ต้องมีแนวทางในการให้คะแนนเสมอ

Score Points for 1-point Items

1 Point

A one-point response is correct. The response indicates that the student has completed the task correctly.

0 Points

A zero-point response is completely incorrect, irrelevant, or incoherent.

Score Points for 2-point Items

2 Points (Full credit)

A two-point response is complete and correct. The response demonstrates a thorough understanding of the concepts and/or procedures embodied in the task.

- Indicates that the student has completed all aspects of the task, showing the correct application of concepts and/or procedures
- Contains clear, complete explanations, supporting work, or evidence when required

1 Point (Partial Credit)

A one-point response is only partially correct. The response demonstrates only a partial understanding of the concepts and/or procedures embodied in the task.

- Addresses some elements of the task correctly but may be incomplete
- May contain a correct answer but with an incomplete explanation when required
- May contain an incorrect answer but with an explanation or supporting work indicating a correct understanding of the concepts

0 Points

A zero-point response is inaccurate or inadequate, irrelevant, or incoherent.

From TIMSS & PIRLS (2011)

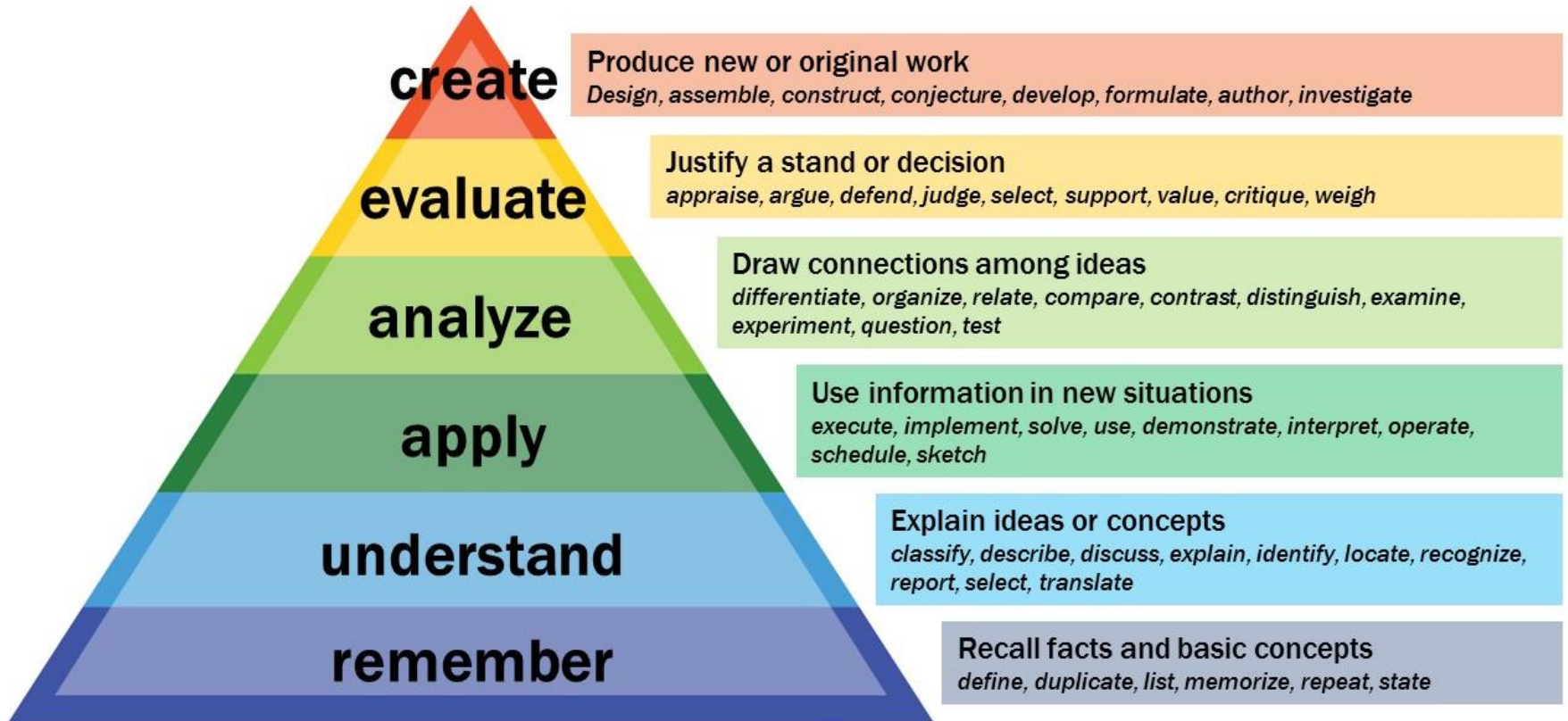
การสร้างแบบทดสอบ

- ข้อเสนอแนะในการเขียนข้อคำถาม
 - ไม่ใช่คำนิเสธ โดยเฉพาะอย่างยิ่งนิเสธซ้อนนิเสธ เช่น “ฉันไม่ชอบชีวิตที่ไม่สนุก”
 - ไม่ใช่ข้อคำถามที่มีสองเนื้อหา เช่น “ฉันชอบคนน่ารักและมีเสน่ห์”
 - ไม่ใช่ข้อคำถามที่มีเนื้อหาที่ไม่เกี่ยวข้อง เช่น “แม้สังคมจะวุ่นวาย ฉันชอบชีวิตสงบสุข”
 - สั้น กระชับ

การสร้างแบบทดสอบ

- ข้อเสนอแนะในการเขียนข้อคำถาม (จำเพาะสำหรับข้อสอบ)
 - สำหรับข้อคำถามที่มีคำตอบถูกต้อง ข้อคำถามควรเป็นประโยคคำถามหรือการเติมคำในช่องว่าง
 - ตัวเลือกควรมีอำนาจในการลวง มีเนื้อหาใกล้เคียงกัน
 - ตัวลวงไม่ควรใช้ “ถูกต้องทั้งหมด” หรือ “ไม่ถูกต้องทั้งหมด”
 - จำนวนตัวลวงอาจเพิ่มหรือลดได้ ถ้าทุกข้อมีความเป็นไปได้
 - เนื้อหาจากข้อคำถามหนึ่งไม่ควรไปอีกข้อคำถามหนึ่งได้

Bloom's Taxonomy



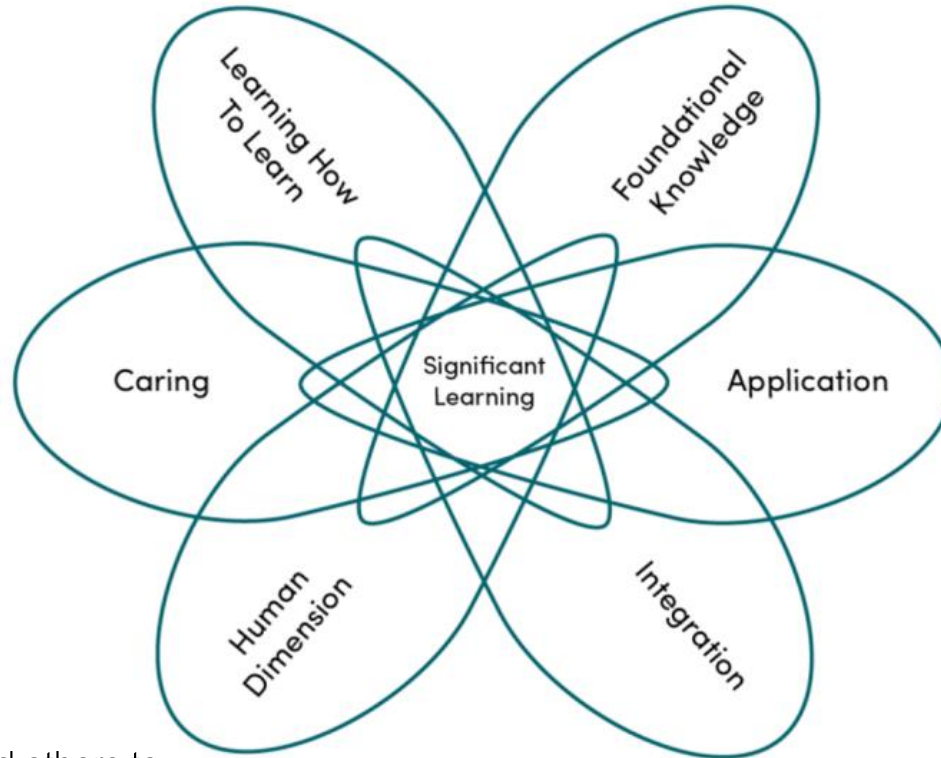
From Center for Teaching, Vanderbilt University

Fink's (2013) Taxonomy

Improving learning skills to continue learning beyond the course

Remembering and understanding information

Developing interest or value within students



Learning a new action, whether a new skill, way of thinking or how to manage projects

Learning about yourself and others to function and interact effectively (e.g., students learn their abilities, strengths; students understand others; students learn how to interact with others)

Making connection between ideas, learning experiences or subject areas

การสร้างแบบทดสอบ

- ข้อคำถามแบบจัดอันดับ หรือบังคับเลือก (Ipsative data) เป็นวิธีการลดการตอบสนองตามความคาดหวังของสังคมได้จริงหรือ?

I am the sort of person who...

			M	L
1	A	I try out new activities.	<input checked="" type="radio"/>	<input type="radio"/>
	B	I consider other people's feelings.	<input type="radio"/>	<input type="radio"/>
	C	I like to understand the underlying theory.	<input type="radio"/>	<input checked="" type="radio"/>

ข้อคำถามแบบกึ่งบังคับเลือก (Semi-Ipsative data)

Gordon Personal Profile-Inventory



Test Instructions

On the following screens are a number of descriptions of personal characteristics of people. These descriptions are grouped in sets of four. You are to examine each set and find the one description that is *most like you*. Then select the circle following the statement, in the column headed **Most**.

Next, examine the other three statements in the set and find the one description that is *least like you*; then select the circle following that statement, in the column headed **Least**. Leave the remaining two statements in the set unselected.

Here is an example set:

	Most	Least
has an excellent appetite	<input type="radio"/>	<input type="radio"/>
watches too much television	<input type="radio"/>	<input checked="" type="radio"/>
follows a well-balanced diet	<input type="radio"/>	<input type="radio"/>
doesn't get enough exercise	<input checked="" type="radio"/>	<input type="radio"/>

Suppose that you have read the four descriptive statements in the example and decided that, although several of the statements may apply to you to some degree, "doesn't get enough exercise" is *more like you* than any of the others. You would select the circle following the statement in the column headed **Most**, as shown in the example above.

You would then examine the other three statements to decide which one is *least like you*. Suppose that "watches too much television" is *less like you* than the other two. You would select the circle following that statement in the column headed **Least**, as shown in the example above.

For every set you should have *one and only one* selection in the **Most** column and *one and only one* selection in the **Least** column. There should be nothing selected for two of the statements.

In some cases it may be difficult to decide which statements you should select. Make the best decisions you can. There are no right or wrong answers. In each set you are to select two statements in the way in which they most *nearly apply to you*. Be sure to select one statement as being *most like you* and one as being *least like you*. Do this for every set. Do *not* skip any set.

Click **Next Page** to start the assessment.

Previous Page

Next Page

I am the sort of person who...

Most

Least

1

A

I try out new activities.



B

I consider other people's feelings.

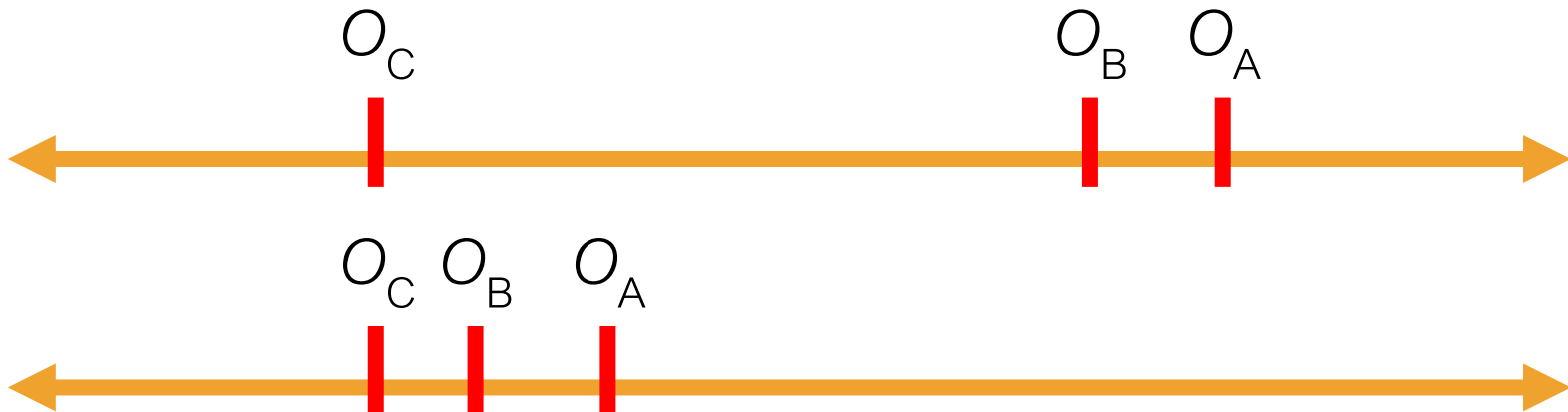


C

I like to understand the underlying theory.



จากข้อมูลข้อนี้ ค่าแฝงในแต่ละข้อ (O_A , O_B , O_C) จะอยู่ตำแหน่งใดก็ได้ ขอให้ลำดับถูกต้อง



การสร้างแบบทดสอบ

- เนื่องจากเราไม่ทราบตำแหน่งของ O_A , O_B , O_C ดังนั้นการนำคะแนนจากการจัดอันดับไปใช้โดยตรง จึงไม่ถูกต้อง
 - เช่น แทนค่า $A = 3$, $B = 2$, $C = 1$
- การนำคะแนนจากการจัดอันดับไปใช้โดยตรง เป็นการบังคับว่า คะแนนข้อหนึ่งสูง อีกข้อหนึ่งต้องต่ำ ทั้งที่ไม่จำเป็นต้องเป็นเช่นนั้น



- ดู Brown & Maydeu-Olivares (2013) สำหรับวิธีการวิเคราะห์เพิ่มเติม

$$\text{อย่าลืมว่า } O = T + E$$

การทดลองใช้แบบทดสอบ

- เมื่อได้ข้อคำถามแล้ว ควรทดลองเก็บข้อมูลทดลองใช้ เพื่อวิเคราะห์ในเชิงคุณภาพและปริมาณ
 - การวิเคราะห์คุณภาพ จะเน้นความเข้าใจเนื้อหาของผู้ตอบ
 - การวิเคราะห์เชิงปริมาณ จะดูคุณภาพในเชิงสถิติ ของแต่ละข้อคำถาม
- ควรเก็บข้อมูลกับกลุ่มตัวอย่างที่ใกล้เคียงกับกลุ่มตัวอย่างที่ต้องการจริง

การวิเคราะห์ข้อคำถามเชิงคุณภาพ

- การวิเคราะห์ข้อคำถามเชิงคุณภาพ (Qualitative Item Analysis) เพื่อเข้าใจลักษณะของข้อคำถามมากขึ้น ในมุมมองของผู้ทดสอบ (ตาราง 8.3)
- การทำโดยคิดดังๆ (Think aloud) เป็นวิธีที่ทำให้รู้กระบวนการคิดในการตอบคำถามต่างๆ ทำให้ป้องกันการคิดบางรูปแบบล่วงหน้าในการทำแบบทดสอบ (เช่น ผ่านสูตรลัด ไม่ได้แสดงถึงลักษณะที่ต้องการ)
- จำนวนกลุ่มตัวอย่างไม่จำเป็นต้องมาก แต่ให้แน่ใจว่าได้ข้อมูลที่อิ่มตัวแล้ว

การวิเคราะห์ข้อคำถามเชิงปริมาณ

- เป็นการตรวจสอบว่า ข้อคำถามแต่ละข้อสอดคล้องกับข้อคำถามข้ออื่นหรือไม่
- การวิเคราะห์ข้อคำถาม แบ่งเป็น
 - การวิเคราะห์แบบง่าย จะครอบคลุมในบทนี้
 - การวิเคราะห์ผ่านตัวแปรแฝง เช่น การวิเคราะห์องค์ประกอบ (Factor Analysis) และการวิเคราะห์ทฤษฎีการตอบสนองต่อข้อคำถาม (Item Response Theory)
- ควรเก็บข้อมูลให้เหมาะสมกับการวิเคราะห์องค์ประกอบตั้งแต่ต้น แต่ถ้าวิเคราะห์อย่างง่าย (ดังที่สอนในบทนี้) ให้ใช้เกณฑ์ 5 หรือ 10 คนต่อข้อ

การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย

ความยากรายข้อ (Item Difficulty)

- ความยากรายข้อ (Item difficulty) คือ สัดส่วนของกลุ่มตัวอย่างที่ตอบข้อดังกล่าวถูกต้อง ในข้อคำถามแบบลิเคิร์ต จะหมายถึงค่าเฉลี่ยของข้อคำถาม
- ค่าความยากรายข้อเฉลี่ยรวมทุกข้อที่ดี ควรอยู่กึ่งกลางระหว่างสัดส่วนที่ได้จากการสุ่มและคะแนนเต็ม เช่น $(.25 + 1) / 2 = .625$ สำหรับข้อทางเลือก 4 ข้อ (ในกลุ่มเป้าหมายที่ต้องการใช้จริง)
- ควรมีการออกแบบส่วนผสมของความยาก (หรือค่าเฉลี่ย) ระหว่างข้อคำถาม
 - ถ้าใช้ในการแบ่งแยกความแตกต่างระหว่างบุคคล ความยากควรผสมกัน
 - ถ้าใช้ในการตัดสินว่าผ่านเกณฑ์หรือไม่ ไม่ควรดูความยาก แต่ควรดูความแตกต่างระหว่างความยากรายข้อของกลุ่มคนที่ผ่านและไม่ผ่านเกณฑ์ (กล่าวถึงภายหลัง)

การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย

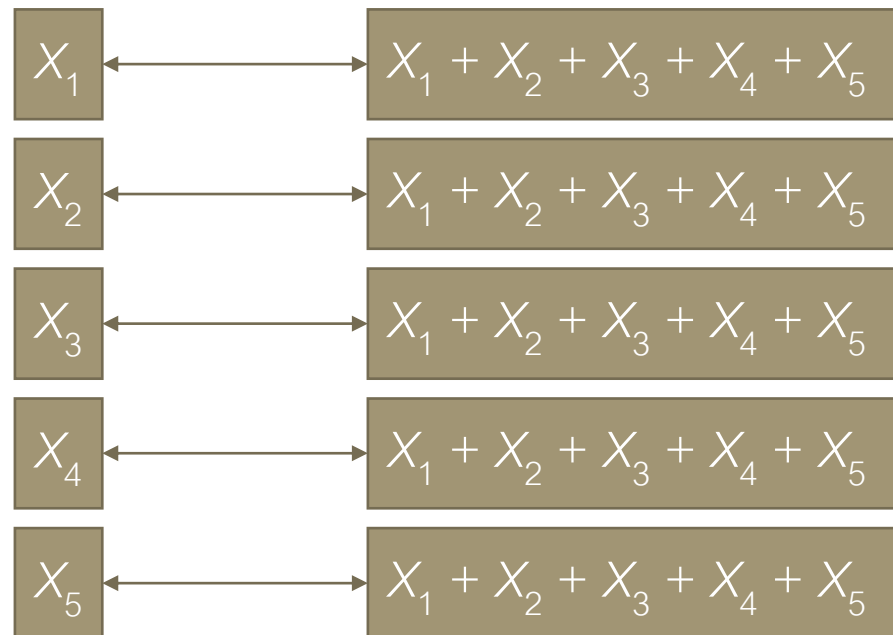
อำนาจจำแนก (Item Discrimination)

- อำนาจจำแนก (Item discrimination) คือ ความสามารถของข้อคำถามในการแบ่งระหว่างคนมีภาวะสันนิษฐานสูงและต่ำ
- ในที่แท้จริง ควรจะเป็นน้ำหนักองค์ประกอบ (Factor loading) ในการวิเคราะห์องค์ประกอบ (Factor Analysis) กล่าวคือ เมื่อภาวะสันนิษฐานเปลี่ยนแปลง 1 หน่วยแล้วคะแนนข้อคำถามเปลี่ยนแปลงไปเท่าไร
- ในอดีต การวิเคราะห์องค์ประกอบทำได้ยาก จึงใช้การประมาณค่าด้วย 3 ค่า คือ ความสัมพันธ์ระหว่างข้อคำถามและคะแนนรวม (Item total correlation), อำนาจจำแนกกลุ่มสูงต่ำ, ค่าอัลฟาเมื่อข้อคำถามถูกลบ

การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย

อำนาจจำแนก (Item Discrimination)

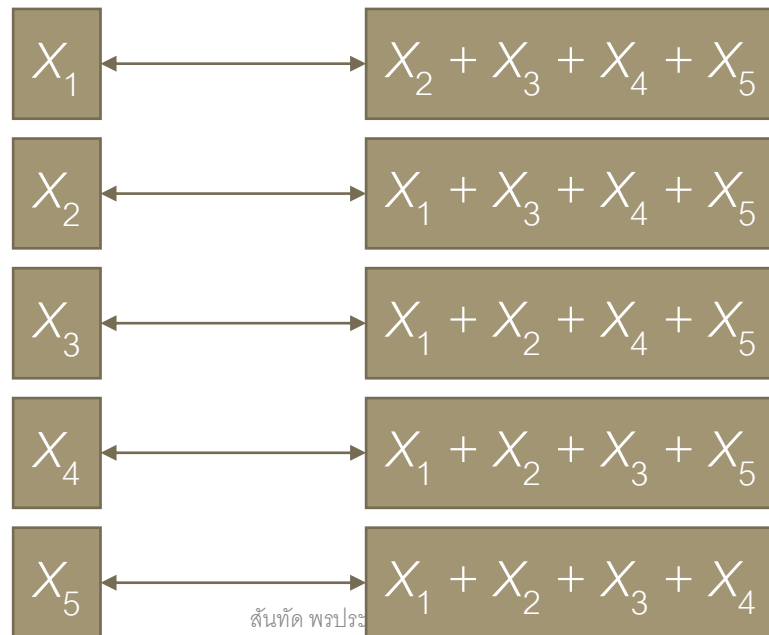
- ความสัมพันธ์ระหว่างข้อคำถามและคะแนนรวม (Item total correlation)
 - สมมติว่าข้อคำถามมี 5 ข้อ ค่านี้ของแต่ละข้อคือความสัมพันธ์ของแต่ละข้อกับคะแนนรวม



การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย

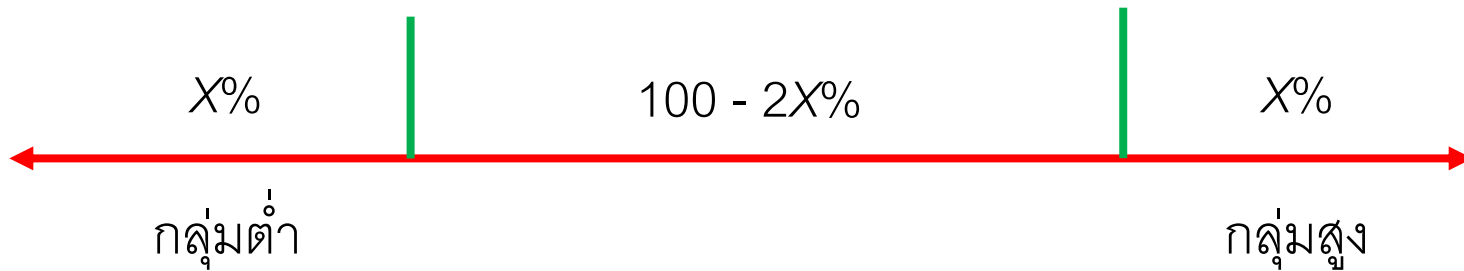
อำนาจจำแนก (Item Discrimination)

- ความสัมพันธ์ระหว่างข้อคำถามและคะแนนรวม (Item total correlation)
 - เนื่องจากคะแนนรวม มีคะแนนข้อคำถามที่ต้องการหาอำนาจจำแนกในการรวมคะแนนด้วย จึงตัดคะแนนข้อคำถามนั้นออกจากคะแนนรวมเพื่อลดการเพ้อของค่าสหสัมพันธ์ เรียกว่า Corrected item total correlation (CITC)



การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย อำนาจจำแนก (Item Discrimination)

- การเปรียบเทียบกลุ่มสูงกลุ่มต่ำ



$$d = \frac{M_{\text{กลุ่มสูง}} - M_{\text{กลุ่มต่ำ}}}{2}$$

- การเปรียบเทียบนี้ ให้ผลไม่แตกต่างจาก CITC (Preacher et al., 2005) ซึ่งยังทำให้กำลังในการวิเคราะห์ทางสถิติต่ำลงด้วย

การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย

อำนาจจำแนก (Item Discrimination)

- ค่าอัลฟา ถ้าข้อคำถามนั้นถูกลบออกไป (Alpha if item deleted)
 - เชื่อว่า ยิ่งค่านี้ยิ่งสูง ข้อคำถามดังกล่าวควรจะถูกลบออกจากมาตรวัด
 - แท้จริงแล้ว ค่านี้เป็นฟังก์ชันของ CITC

ให้ X_i เป็นคะแนนข้อที่ i , Y_i เป็นผลรวมของคะแนนยกเว้นข้อที่ i และ α_i เป็นค่าอัลฟาเมื่อลบข้อที่ i

$$r_{X_i Y_i} = \frac{Cov(X_i, Y_i)}{\sqrt{Var(X_i)Var(Y_i)}} = \frac{\sum_{j=1, j \neq i}^p Cov(X_i, X_j)}{\sqrt{Var(X_i)Var(Y_i)}}$$

$$\alpha_i = \left(\frac{p-1}{p-2} \right) \frac{\sum_{j=1, j \neq i}^p \sum_{k=1, k \neq i, j}^p Cov(X_j, X_k)}{Var(Y_i)}$$

จาก

$$\sum_{j=1, j \neq i}^p \sum_{k=1, k \neq i, j}^p Cov(X_j, X_k) + 2 \sum_{j=1, j \neq i}^p Cov(X_i, X_j) + \sum_{j=1}^p Var(X_j) = Var(Y)$$

ดังนั้น

$$\sum_{j=1, j \neq i}^p \sum_{k=1, k \neq i, j}^p Cov(X_j, X_k) + 2r_{X_i Y_i} \sqrt{Var(X_i) Var(Y_i)} + \sum_{j=1}^p Var(X_j) = Var(Y)$$

$$\alpha_i = \left(\frac{p-1}{p-2} \right) \frac{Var(Y) - \sum_{j=1}^p Var(X_j) - 2r_{X_i Y_i} \sqrt{Var(X_i) Var(Y_i)}}{Var(Y_i)}$$

$$\alpha_i = \left(\frac{p-1}{p-2} \right) \frac{Var(Y) - \sum_{j=1}^p Var(X_j)}{Var(Y_i)} - 2 \left(\frac{p-1}{p-2} \right) r_{X_i Y_i} \sqrt{\frac{Var(X_i)}{Var(Y_i)}}$$

แปรรูปผกผันกัน

การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย

อำนาจจำแนก (Item Discrimination)

- แสดงให้เห็นว่าค่าอัลฟาเมื่อข้อคำถามถูกลบไป แปรผกผันกับ CITC
- เมื่อทราบค่าหนึ่งแล้วจะรู้อีกค่าหนึ่ง ดังนั้นไม่จำเป็นต้องพิจารณาจากทั้งสองค่า (แต่อาจรายงานทั้งสองค่าได้)
- ค่า CITC ดีกว่า เนื่องจากค่าสหสัมพันธ์สามารถอ่านค่าได้โดยตรง มีการทดสอบระดับนัยสำคัญชัดเจน

การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย

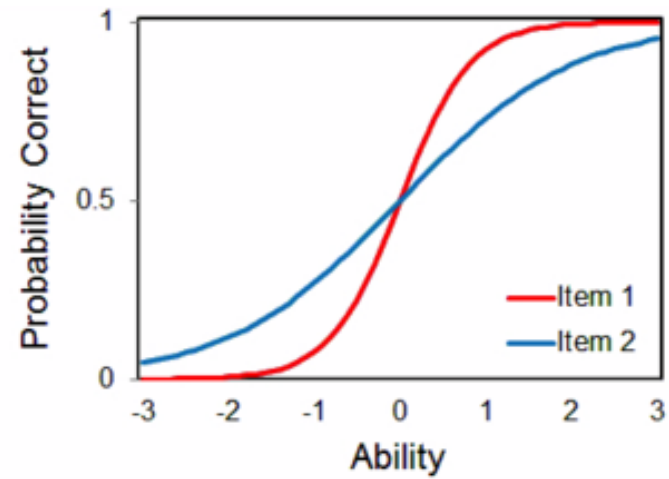
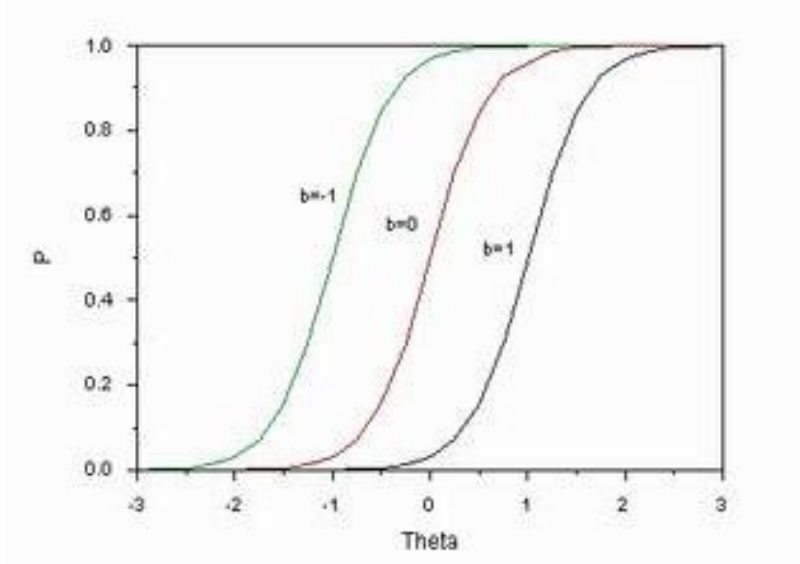
การวิเคราะห์ตัวลวง (Analysis of Item Alternatives)

- การวิเคราะห์ตัวลวง (Analysis of item alternatives)
 - สัดส่วนที่ผู้ตอบเลือกตัวลวง สัดส่วนนี้จะมีค่าเท่าไรก็ได้ แต่ไม่ใช่ 0 มิเช่นนั้นตัวลวงจะไม่ได้ทำหน้าที่เป็นตัวลวงเลย
 - ข้อคำถามที่ดี ผู้ที่มีภาวะสันนิษฐานต่ำต้องมีแนวโน้มที่จะเลือกตัวลวงมากกว่าผู้ที่มีภาวะสันนิษฐานสูง
 - กล่าวคือ มีความสัมพันธ์ทางลบกับภาวะสันนิษฐาน ซึ่งอาจวิเคราะห์ผ่าน CITC หรือการวิเคราะห์กลุ่มสูงต่ำ

การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย

ความยากรายข้อตามระดับภาวะสันนิษฐาน

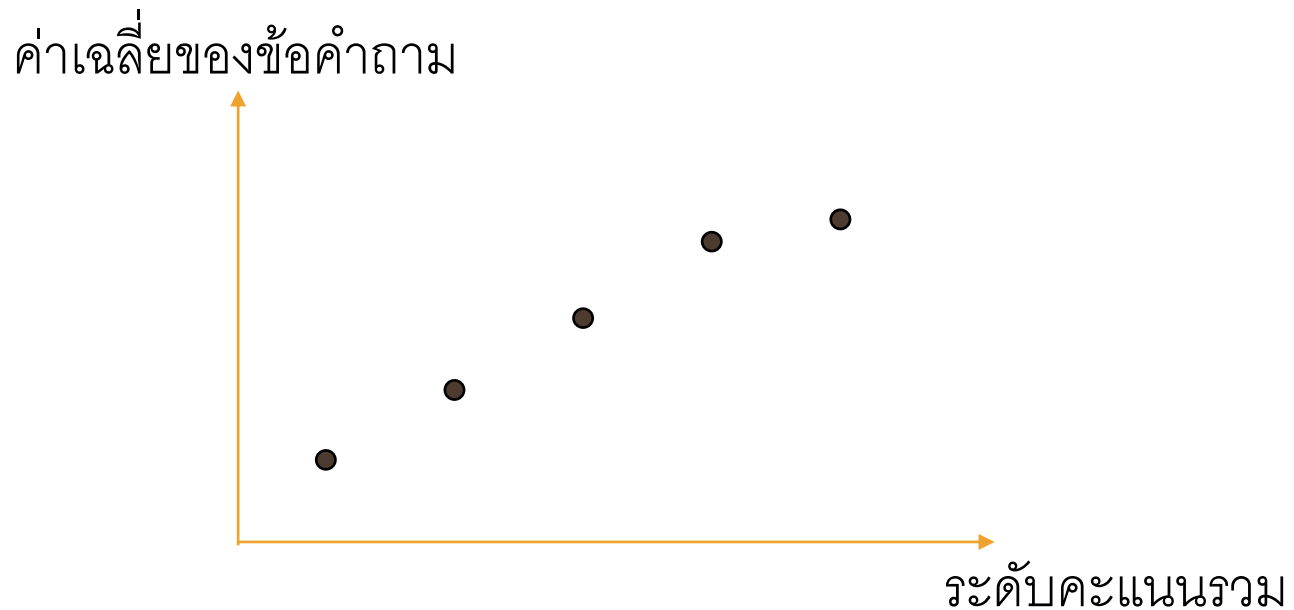
- วิธีทางหนึ่งในการวิเคราะห์ข้อคำถาม คือ นำความยากรายข้อ (Item difficulty) มาสร้างความสัมพันธ์กับระดับภาวะสันนิษฐาน
- เรียกว่า เส้นโค้งลักษณะของข้อคำถาม (Item Characteristic Curve)



การวิเคราะห์ข้อความเชิงปริมาณแบบง่าย

ความยากง่ายข้อตามระดับภาวะสันนิษฐาน

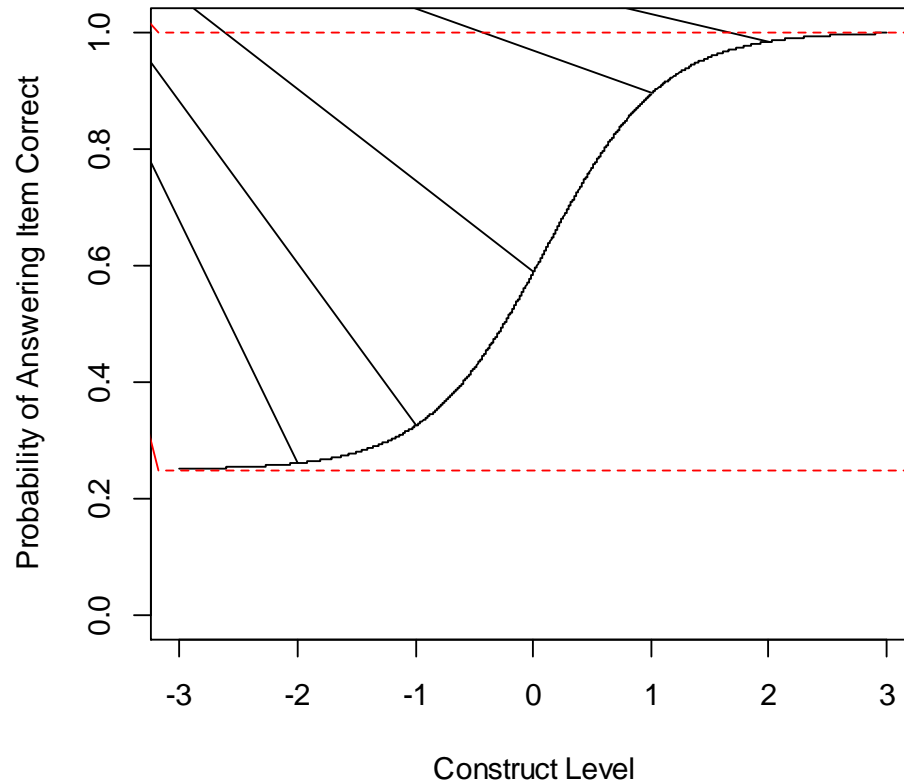
- วิธีอย่างง่ายที่สุด คือ ให้แบ่งคะแนนรวมออกเป็นกลุ่มๆ แล้วสังเกตค่าเฉลี่ยในแต่ละระดับของคะแนนรวม

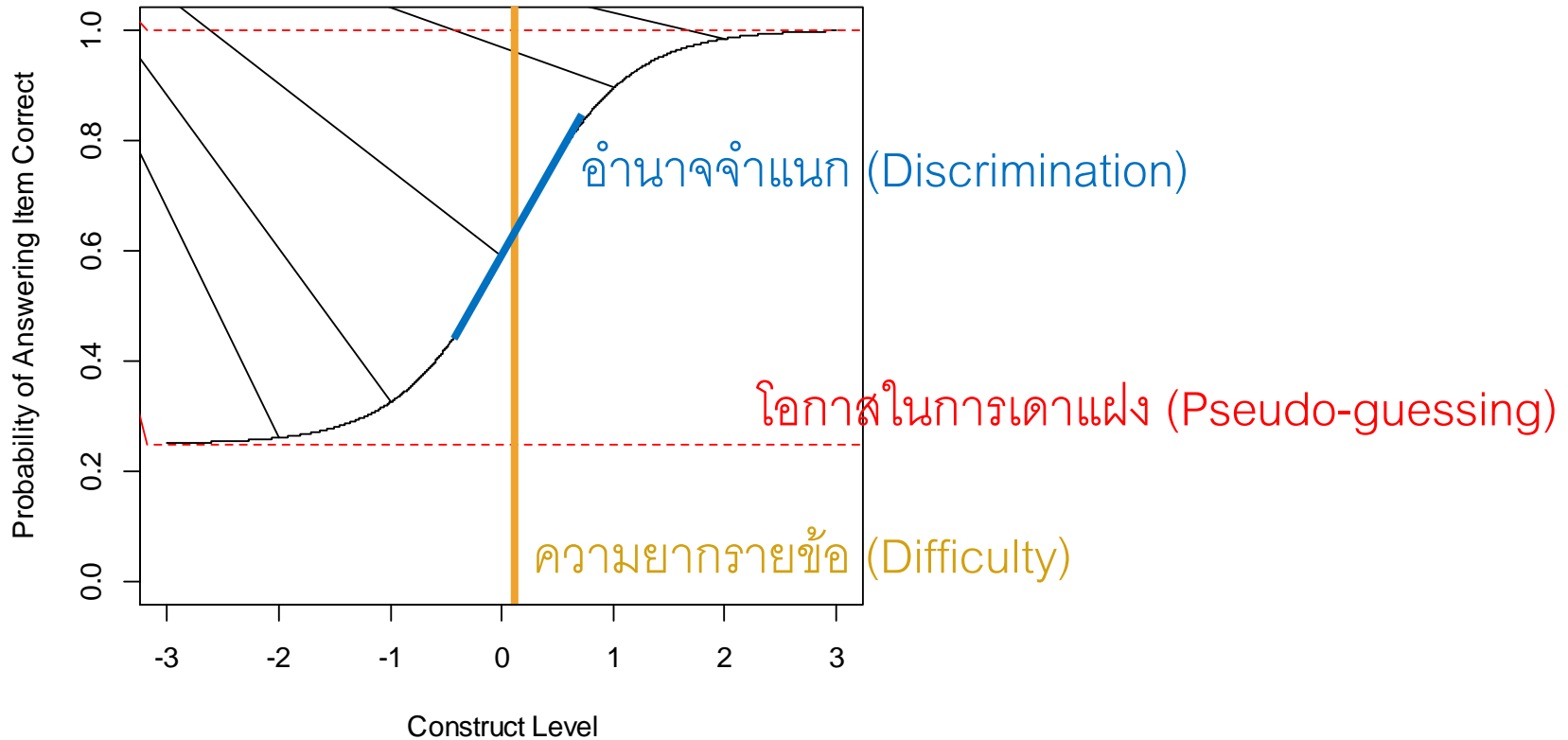


การวิเคราะห์ข้อคำถามเชิงปริมาณแบบง่าย

ความยากง่ายข้อตามระดับภาวะสันนิษฐาน

- ทฤษฎีการตอบสนองของข้อคำถาม (Item response theory) สามารถสร้างโมเดลแสดงความสัมพันธ์ระหว่างข้อคำถามและระดับภาวะสันนิษฐาน (คะแนนรวม)





$$P(X = 1|\theta) = 0.25 + (1 - 0.25) \left(\frac{e^{2\theta - 0.2}}{1 + e^{2\theta - 0.2}} \right)$$

การทบทวนแบบวัด

- ทบทวนแบบวัดที่เสร็จแล้วว่ามีคุณสมบัติความเที่ยง ความตรง ตามที่พึงประสงค์หรือไม่
 - ถ้ามีคุณสมบัติที่ดี ก็เขียนคู่มือการใช้แบบวัด และสร้างรายงานคุณภาพของแบบวัด
 - ถ้าไม่มีคุณสมบัติที่ดี ต้องทบทวนแบบทดสอบใหม่ กลับไปขั้นตอนสร้างข้อคำถาม

การทบทวนแบบวัด

- บางครั้ง แบบวัดที่พัฒนามานานแล้ว ต้องมีการทบทวนคุณภาพของแบบทดสอบใหม่
- ข้อคำถามบางข้ออาจหลุดออกไปยังคนทั่วไปแล้ว หรือข้อคำถามบางข้อไม่ทันสมัยแล้ว
- หากเป็นมาตรวัดที่ใช้ในการประเมินรายบุคคล ต้องมีการเทียบคะแนนระหว่างแบบวัดชุดเก่าและชุดใหม่ เรียกว่า การทำให้แบบทดสอบเท่าเทียมกัน (Test Equating)

การปรับปรุงแบบทดสอบ

- ก่อนสร้างแบบวัดใหม่ ควรตรวจสอบแบบวัดเก่าก่อน ว่ามีผู้พัฒนาแล้วหรือไม่
- หาแบบทดสอบเก่าได้อย่างไร
 - อินเทอร์เน็ต
 - งานวิจัย (เช่น งานวิจัยในวารสาร Assessment, วิทยานิพนธ์)
 - Tests in Print (ฉบับล่าสุดปี 2022)
 - Mental Measurements Yearbook (ฉบับล่าสุดปี 2021)
- แบบทดสอบเก่าเหมาะสมกับวัตถุประสงค์ของคุณหรือไม่
- ถ้าไม่เหมาะสม จะสามารถนำมาปรับปรุงได้หรือไม่ แล้วประเมินสิ่งที่ปรับปรุงแล้วได้อย่างไร

การปรับปรุงแบบทดสอบ

- แบบวัดอยู่ในภาษาอังกฤษ

- การแปลไปกลับ (Translation and Back Translation) ให้แก้ไขจนกว่าผู้วิจัยเชื่อว่า ข้อคำถามใกล้เคียงกันทั้งสองภาษา
- ทดสอบความเข้าใจภาษาของแบบวัดที่แปลแล้ว
- หากต้องเก็บข้อมูลจากทั้งสองกลุ่ม ต้องทดสอบความใกล้เคียงกันของแบบวัด (Measurement Invariance)
 - อำนาจจำแนกของแต่ละข้อคำถามต้องใกล้เคียงกัน
 - คนสองกลุ่มที่มีระดับภาวะสันนิษฐานเท่ากัน ต้องมีค่าเฉลี่ย (หรือสัดส่วนการตอบถูก [ความยากรายข้อ]) ของข้อคำถามทุกข้อใกล้เคียงกัน

การปรับปรุงแบบทดสอบ

- แบบวัดอยู่ในภาษาอังกฤษ

- หากเก็บข้อมูลจากแบบทดสอบที่แปลแล้วเพียงอย่างเดียว อาจวิเคราะห์ความเที่ยงและความตรงเพิ่มเติม โดยเฉพาะอย่างยิ่ง การวิเคราะห์องค์ประกอบ แล้วดูว่าน้ำหนักองค์ประกอบใกล้เคียงกับแบบวัดดั้งเดิมหรือไม่
- นำน้ำหนักองค์ประกอบที่หามาได้ มาหาความสัมพันธ์กับน้ำหนักองค์ประกอบจากมาตรวัดเดิม ค่าสหสัมพันธ์ที่ได้ควรมีค่า .8 ขึ้นไป
- ไม่ควรตัดข้อคำถาม แต่ควรเก็บข้อมูลทั้งหมดด้วยข้อคำถามทั้งหมด แล้ววิเคราะห์แยกระหว่างมีข้อคำถามที่แย่ (มาตรเดิม) และไม่มีข้อคำถามที่แย่ (ตัดข้อที่คุณภาพไม่ดีออก)

การปรับปรุงแบบทดสอบ

- กลุ่มตัวอย่างที่แตกต่างจากเดิม
 - ทดสอบความเข้าใจภาษาใหม่
 - วิเคราะห์น้ำหนักองค์ประกอบใหม่ แล้วหาสหสัมพันธ์ระหว่างน้ำหนักองค์ประกอบของกลุ่มตัวอย่างเก่าและกลุ่มตัวอย่างใหม่ ควรมีค่า .8 ขึ้นไป
 - เช่นเดิม ไม่ควรตัดข้อคำถาม แต่ควรวิเคราะห์แยกระหว่างข้อคำถามทั้งหมดและตัดข้อคำถามแล้ว

การปรับปรุงแบบทดสอบ

- แบบวัดใช้เวลานานเกินไป
 - มีผลอำนาจจำแนกหรือมีผลการวิเคราะห์องค์ประกอบจากแบบทดสอบเดิม
 - ให้ใช้ผลเดิม ตัดข้อคำถามที่มีอำนาจจำแนกหรือน้ำหนักองค์ประกอบต่ำออก
 - ถ้าใช้การวิเคราะห์องค์ประกอบ ควรให้มีข้อคำถามจากแต่ละองค์ประกอบย่อยอย่างน้อย องค์ประกอบละ 3 ข้อ
 - ไม่มีผลการวิเคราะห์องค์ประกอบเดิม
 - ตรวจสอบให้แบบทดสอบชุดใหม่ มีความครอบคลุมของเนื้อหาใกล้เคียงกับแบบทดสอบเดิม
 - วิเคราะห์องค์ประกอบ แล้วตัดข้อคำถามผ่านการวิเคราะห์องค์ประกอบ