

ความตรง

การประเมินลักษณะมนุษย์

สันทัต พรประเสริฐมานิต

โครงร่างการนำเสนอ

- นิยามความตรง
- แหล่งหลักฐานของความตรงตามแนวคิดของ Standards (2014)
- เปรียบเทียบกับการแบ่งความตรงสามองค์ (Trinitarian view of validity)

นิยามความตรง

- ระดับของหลักฐานและทฤษฎีที่สนับสนุนการแปลความหมายของคะแนนที่ได้จากแบบทดสอบ (ทางจิตวิทยา) ในการใช้รูปแบบต่างๆ เช่น สนับสนุนว่าคะแนนที่ได้
 - สะท้อนภาวะสันนิษฐานที่ต้องการ เช่น เซาว์นปัญญา, บุคลิกภาพ, ความเครียด
 - ทำนายตัวแปรในอนาคต เช่น ผลงาน อาชญากรรม
 - สะท้อนความบกพร่องในสมอง
 - แสดงความเสี่ยงในการเกิดภาวะความผิดปกติทางจิต

นิยามความตรง

- ไม่ใช่หลักฐานพิสูจน์ว่าแบบทดสอบวัดได้ตรงหรือไม่
 - แต่เป็นหลักฐานว่าการแปลความหมายสองคลังตามทฤษฎี
 - เป็นหลักฐานว่าคะแนนสูงต่ำของแต่ละคนที่ได้ สอดคล้องตามทฤษฎี
- ยิ่งหลักฐานมาก ยิ่งน่าเชื่อถือ
- เป็นหลักฐานที่จำเพาะในกลุ่มตัวอย่างหนึ่ง ในสถานการณ์หนึ่ง เวลาหนึ่ง ไม่ใช่คุณสมบัติของแบบวัดที่ใช้ได้ทุกสถานการณ์ ทุกกลุ่มตัวอย่าง ทุกเวลา
 - การอ้างอิงในสถานการณ์อื่น กลุ่มตัวอย่างอื่น เวลาอื่น อาจต้องใช้หลักเหตุผลในการอ้างอิง เช่น การนำหลักฐานความตรงจากต่างประเทศมาใช้ในประเทศไทย

การแปลความหมาย

ที่แบบทดสอบ

สะท้อนออกมา

ภาวะสันนิษฐาน

ที่ต้องการวัด

ส่วนที่ไม่เกี่ยวข้องกับ
ภาวะสันนิษฐาน
(Construct
Irrelevance)

ภาวะสันนิษฐานที่ขาดไป
(Construct
Underrepresentation)

เช่น ทักษะทางภาษาอังกฤษ
ในข้อสอบความถนัดในอาชีพนักบิน

เช่น การประเมินผลงานผ่านการประเมิน
คะแนน ที่ไม่สะท้อนผลงานที่นายไม่เห็น

แหล่งของหลักฐานความตรง

หลักฐานจากเนื้อหา
แบบทดสอบ

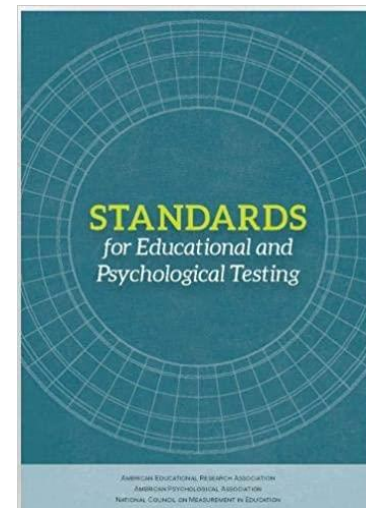
หลักฐานจากกระบวนการ
ตอบคำถาม

หลักฐานผลที่ตามมาจาก
การทดสอบ

หลักฐานจากโครงสร้าง
ภายในแบบทดสอบ

หลักฐานความสัมพันธ์กับ
ตัวแปรอื่น

อ้างอิงจาก Standards (2014)



หลักฐานจากเนื้อหาแบบทดสอบ

- เป็นการแสดงหลักฐานว่าแบบวัดที่ครอบคลุมเนื้อหาทั้งหมดของสิ่งที่ต้องการวัด
- เหมือนการสุ่ม (Random sample) ที่สุ่มตัวอย่างของข้อคำถามมาได้เป็นตัวแทนของสิ่งที่เป็นไปได้ทั้งหมดในสิ่งที่ต้องการวัด
- ความตรงตามเนื้อหา จะชัดเจนมากในการวัดผลสัมฤทธิ์หรือผลการปฏิบัติงาน

หลักฐานจากเนื้อหาแบบทดสอบ

- สัดส่วนความตรงตามเนื้อหา (Content Validity Ratio; CVR) ใช้การประเมินจากผู้เชี่ยวชาญว่าข้อคำถามแต่ละข้อ “สำคัญ” “เกี่ยวข้องแต่ไม่สำคัญ” และ “ไม่จำเป็น” ต่อภาวะสันนิษฐาน
- เช่น ข้อคำถามเหล่านี้เป็นสิ่งสำคัญในการวัดความสุขหรือไม่

	สำคัญ	เกี่ยวข้องแต่ไม่สำคัญ	ไม่จำเป็น
1. ฉันพึงพอใจในความสัมพันธ์ของฉัน			
2. ฉันมีความสุขในชีวิต			
3. ฉันมีเพื่อนที่ดี			

อาจตัดตัวเลือกกลางได้

หลักฐานจากเนื้อหาแบบทดสอบ

$$CVR = \frac{\#สำคัญ - N/2}{N/2}$$

- N = จำนวนผู้ประเมินในแต่ละข้อ
- มีค่าได้ตั้งแต่ 1 (ทุกคนเห็นตรงกันว่าสำคัญ) ไปจน 0 (ครึ่งหนึ่งเห็นว่าคุณสำคัญ) ไปจน -1 (ทุกคนเห็นตรงกันว่าไม่สำคัญ)
- คูตารางระดับนัยสำคัญได้จาก Ayre & Scally (2014)

Table 1. CVR_{critical} One-Tailed Test ($\alpha = .05$) Based on Exact Binomial Probabilities.

N (Panel Size)	Proportion Agreeing Essential	CVR _{critical} Exact Values	One-Sided p Value	N _{critical} (Minimum Number of Experts Required to Agree Item Essential)—Ayre and Scally, This Article	N _{critical} Calculated From CRITBINOM Function—Wilson et al. (2012)
5	1	1.00	.031	5	4
6	1	1.00	.016	6	5
7	1	1.00	.008	7	6
8	.875	.750	.035	7	6
9	.889	.778	.020	8	7
10	.900	.800	.011	9	8
11	.818	.636	.033	9	8
12	.833	.667	.019	10	9
13	.769	.538	.046	10	9
14	.786	.571	.029	11	10
15	.800	.600	.018	12	11
16	.750	.500	.038	12	11
17	.765	.529	.025	13	12
18	.722	.444	.048	13	12
19	.737	.474	.032	14	13
20	.750	.500	.021	15	14
21	.714	.429	.039	15	14
22	.727	.455	.026	16	15
23	.696	.391	.047	16	15
24	.708	.417	.032	17	16
25	.720	.440	.022	18	17
26	.692	.385	.038	18	17
27	.704	.407	.026	19	18
28	.679	.357	.044	19	18
29	.690	.379	.031	20	19
30	.667	.333	.049	20	19
31	.677	.355	.035	21	20
32	.688	.375	.025	22	21
33	.667	.333	.040	22	21
34	.676	.353	.029	23	22
35	.657	.314	.045	23	22
36	.667	.333	.033	24	23
37	.649	.297	.049	24	23
38	.658	.316	.036	25	24
39	.667	.333	.027	26	25
40	.650	.300	.040	26	25

ทดสอบ $H_0: CVR \leq 0$

แม้ระดับนัยสำคัญจะไม่สำคัญเท่ากับขนาดอิทธิพล แต่หากจำนวนผู้เชี่ยวชาญไม่เกิน 15 คน ค่าตัวเลขที่ถึงระดับนัยสำคัญ ก็ถือว่ามีความหมายขนาดอิทธิพลระดับสูงอยู่ (เกิน 80%) สามารถใช้ตารางนี้ได้

หลักฐานจากเนื้อหาแบบทดสอบ

- ปัญหาก็คือ ใครเป็นผู้เชี่ยวชาญ
- นอกจากนี้ CVR ไม่ได้วัดความตรงตามเนื้อหาทั้งหมด เพราะไม่ได้วัดความครอบคลุมของเนื้อหา
- การวัดความครอบคลุมจากเนื้อหาให้ใช้การวิเคราะห์ผ่านตารางลักษณะของภาวะสันนิษฐาน (Table of Specification)

หลักฐานจากเนื้อหาแบบทดสอบ

- ประเมินผลสัมฤทธิ์วิชาคณิตศาสตร์หลัก ม.4 ภาคต้น

เนื้อหา	จำนวนคาบที่สอน	จำนวนข้อ
1. เซต	3	15
2. การให้เหตุผล	3	15
3. จำนวนจริง	5	25
4. เลขยกกำลัง	3	15

คุณต้องคำนึงถึงตารางนี้ทั้งในขั้นตอนการสร้างข้อคำถาม และแบบวัดฉบับสมบูรณ์

หลักฐานจากเนื้อหาแบบทดสอบ

- ย้ำอีกครั้ง ว่าเป็นการแสดงความหลักฐานว่าแบบวัดที่ครอบคลุมเนื้อหาทั้งหมดของสิ่งที่ต้องการวัด
- การแสดงความหลักฐานไม่จำเป็นต้องให้ผู้เชี่ยวชาญเสมอไป สำคัญที่แสดงหลักการคิด ว่ากระบวนการสร้างคำถามมีหลักการแบบใด สุ่มเนื้อหาให้ครอบคลุมอย่างไร ข้อคำถามตรงกับภาวะสันนิษฐานอย่างไร
- เช่น ข้อสอบวัดการบวกเลขสองหลัก แบ่งโจทย์เป็น 6 กลุ่ม
 - ข้อคำถามที่ผลลัพธ์มี 2 หลัก vs. 3 หลัก
 - ข้อคำถามที่ไม่ต้องทดไปหลักที่สอง vs. ข้อที่ต้องทด
 - ข้อไม่ต้องทด แบ่งเป็น ตัวใดตัวหนึ่งลงท้ายด้วยศูนย์ vs. ไม่มีตัวใดลงท้ายด้วยศูนย์

ความตรงตามที่เห็น

- ความตรงตามที่เห็น (Face Validity) คือ ความคิดของผู้รับการประเมิน (หรือผู้ตรวจสอบแบบวัด) ว่าวัดในสิ่งที่ต้องการวัดหรือไม่
- ไม่ใช่หลักฐานความตรง
- มีความสำคัญส่วนหนึ่ง เพราะทำให้ผู้ประเมินให้ความร่วมมือในการทำแบบวัด

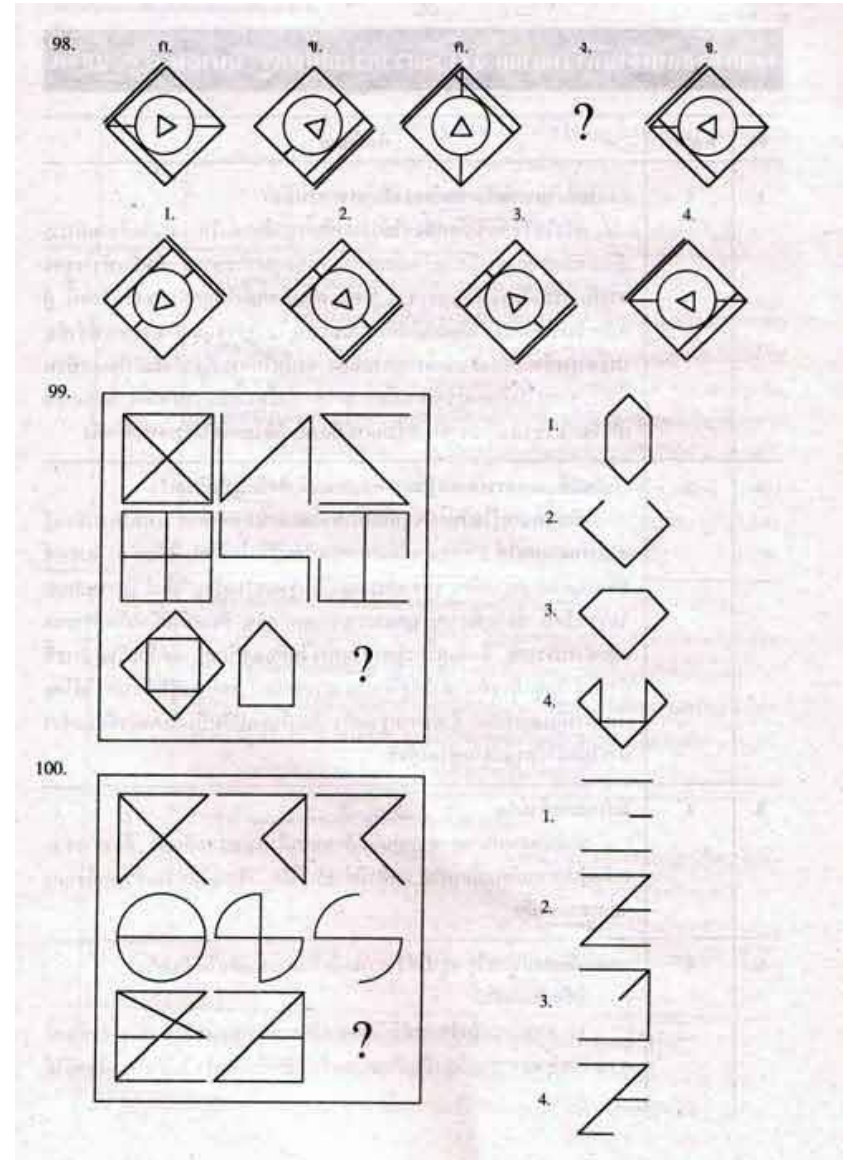
1. พฤติกรรมที่เป็นปัญหาของนักเรียนเกิดจากสาเหตุใด? *

- ก. ไม่ได้เรียนในโรงเรียนที่มีชื่อเสียง
- ข. มีพ่อแม่คอยเคี่ยวเข็ญ
- ค. สิ่งแวดล้อมชักจูง
- ง. มีจริยธรรมต่ำ

2. คุณธรรมสำหรับครูคือข้อใด? *

- ก. คุณงามความดีของคนที่เป็นครู
- ข. มีพฤติกรรมที่ดีเป็นที่ยอมรับของสังคม
- ค. มีความเสียสละมีความยุติธรรม
- ง. ถูกทุกข้อ

ข้อสอบวัดแววความเป็นครู



หลักฐานจากกระบวนการตอบคำถาม

- วิเคราะห์การตอบเชิงทฤษฎี หรือเชิงประจักษ์ต่อกระบวนการคิดในการสร้างพฤติกรรมตอบสนองของสิ่งเร้า ว่าถูกต้องตามภาวะสันนิษฐานที่ต้องการวัดหรือไม่ เช่น
 - การวัดความจำ ถ้าต้องการวัดหน่วยความจำ (Chunks) ของตัวเลข ตัวเลขควรมีไม่สามารถจับกลุ่มกันได้

Digits Forward

I am going to say some numbers. Listen carefully, and when I'm through, say them right after me. For example, if I say, 4-3, then you say, 4-3. Ready?

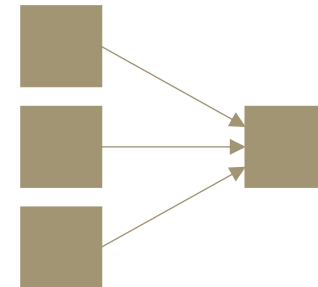
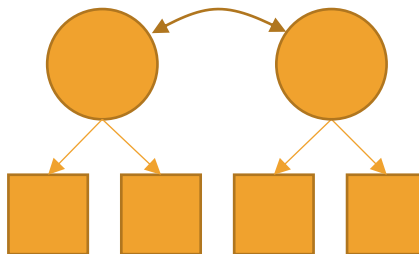
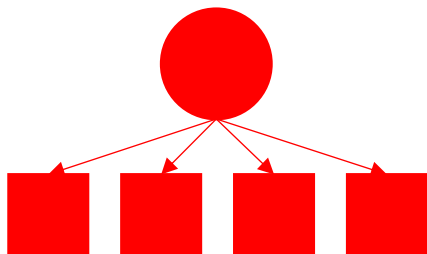
Series	Time 1	Time 2
3	7-4-9	1-7-4
4	8-5-2-1	5-2-9-7
5	2-9-6-8-3	6-3-8-5-1
6	5-7-1-9-4-6	2-9-4-7-3-8
7	8-1-5-9-3-6-2	4-1-9-2-7-5-1
8	3-9-8-2-5-1-4-7	8-5-3-9-1-6-2-7
9	7-2-8-5-4-6-7-3-9	2-1-9-7-3-5-8-4-6

หลักฐานจากกระบวนการตอบคำถาม

- วิเคราะห์การตอบเชิงทฤษฎี หรือเชิงประจักษ์ต่อกระบวนการคิดในการสร้างพฤติกรรมตอบสนองสิ่งเร้า ว่าถูกต้องตามภาวะสันนิษฐานที่ต้องการวัดหรือไม่ เช่น
 - ในการวัดความเครียด ถามว่า “คนอื่นชอบบอกว่าฉันเครียด” ซึ่งเครียดในที่นี้อาจหมายความว่าจริงจัง
 - ในการวัดความผูกพันต่อองค์กร ถามว่า “ฉันจะไม่ลาออก”
- การวิเคราะห์การตอบข้อคำถามต่างๆ สามารถใช้เหตุผลแล้ว ยังสามารถสัมภาษณ์ผู้ทำแบบทดสอบหลังตอบ หรือให้ผู้ทำแบบทดสอบพูดวิธีคิดระหว่างทำแบบทดสอบ
 - เช่น ถามผู้สัมภาษณ์ว่าทำไมถึงประเมินให้คะแนนต่างๆ แก่ผู้ถูกสัมภาษณ์

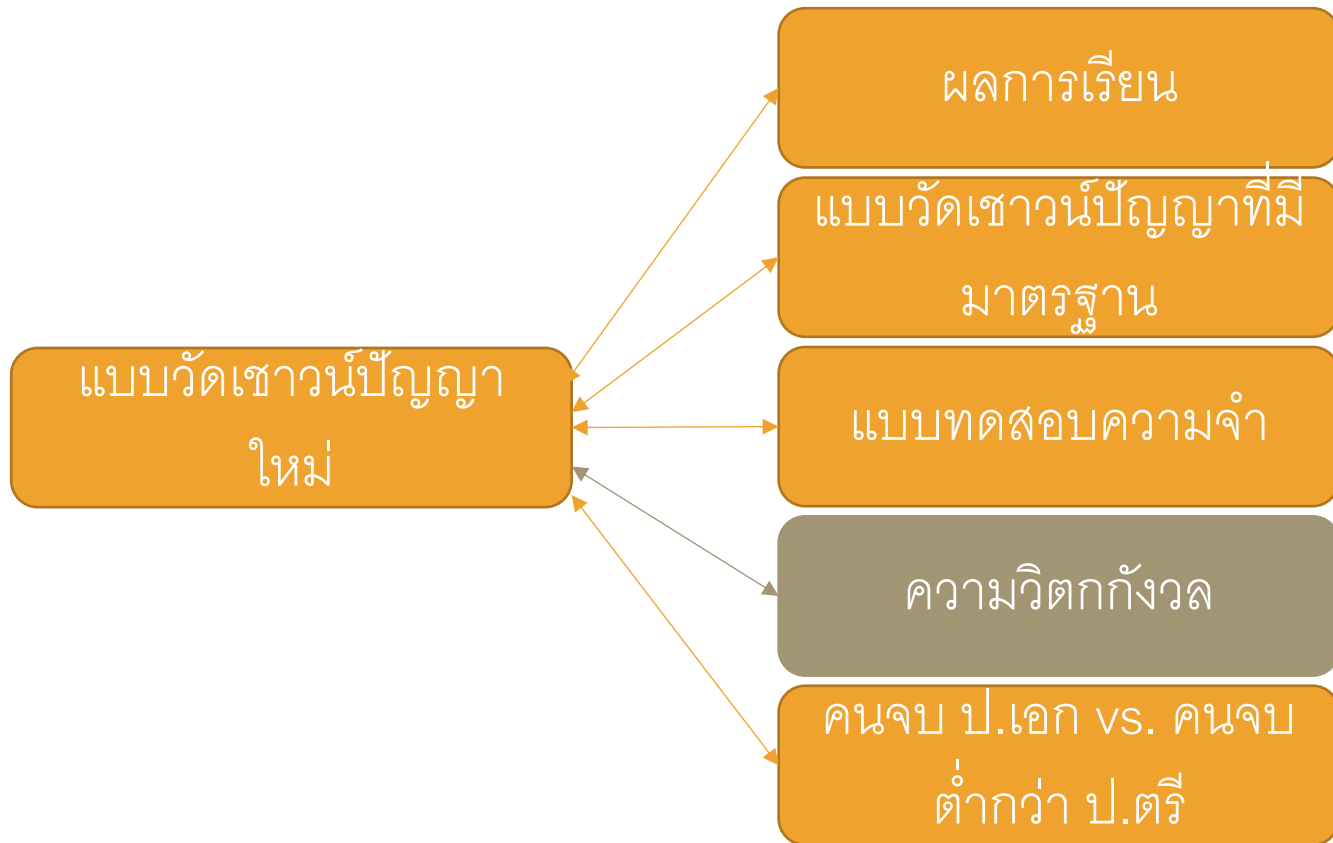
หลักฐานจากโครงสร้างภายใน

- ตรวจสอบความสัมพันธ์ระหว่างข้อความ คະแนบย่อยภายในแบบทดสอบ ต้องมีความสัมพันธ์กับสอดคล้องกับที่คาดหมายจากภาวะสันนิษฐาน
 - สอดคล้องเป็นหนึ่งเดียว เช่น วัดความพึงพอใจในชีวิตภาพรวม
 - สอดคล้องเป็นองค์ประกอบย่อย เช่น วัดความพึงพอใจในชีวิตด้านต่างๆ
 - ไม่จำเป็นต้องสอดคล้องกัน เช่น การสร้างมาตรเพื่อทำนายผลงาน
- หลักฐานโครงสร้างภายในอย่างเดียว ไม่เพียงพอในการอ้างความตรงของการแปลความหมาย



หลักฐานจากความสัมพันธ์กับตัวแปรอื่น

- เป็นหลักฐานที่สำคัญที่สุด และกว้างมาก พุดง่ายๆ คือ ควรมีความสัมพันธ์กับสิ่งที่ควรสัมพันธ์ เช่น
 - เกณฑ์ภายนอก เช่น ผลการปฏิบัติงาน ผลการเรียน
 - อาจวัดเกณฑ์พร้อมกับทำแบบวัด หรือวัดเกณฑ์ในภายหลังได้
 - แบบทดสอบที่วัดสิ่งเดียวกัน
 - แบบทดสอบอื่นที่วัดคนละภาวะสันนิษฐาน แต่มีความสัมพันธ์ทางทฤษฎี
 - อาจแสดงให้เห็นว่าไม่สัมพันธ์กับตัวแปรอื่น ตามทฤษฎี
 - การทดสอบว่าคะแนนแบบทดสอบแตกต่างกันระหว่างกลุ่ม



หลักฐานจากความสัมพันธ์กับเกณฑ์

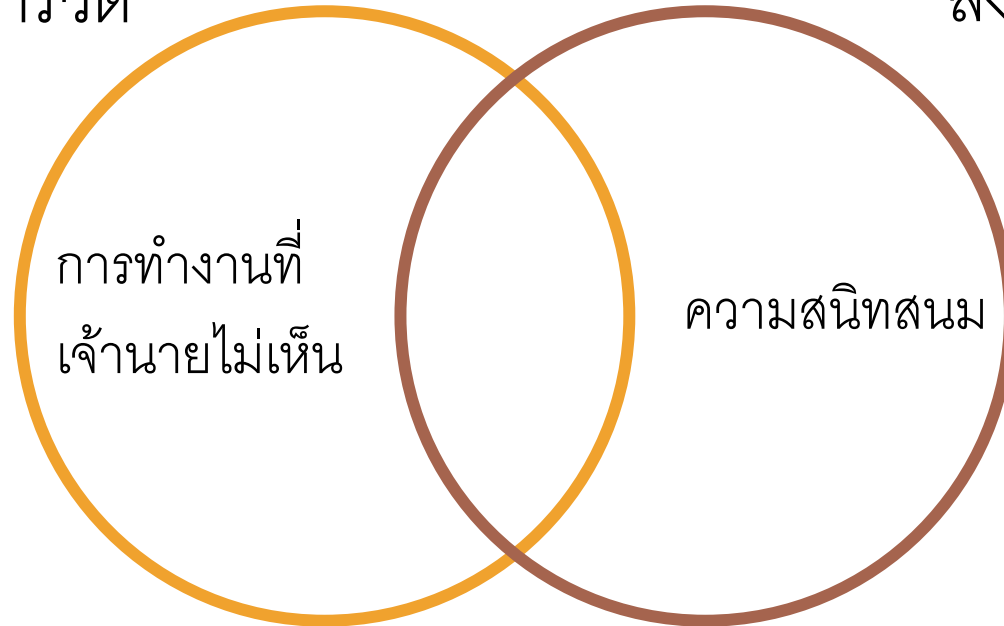
- เกณฑ์ (Criterion) คือ สิ่งที่แบบวัดต้องการใช้ในการเปรียบเทียบกับ เช่น ผลการปฏิบัติงาน ผลการเรียน ฯลฯ
- หลักฐานนี้เป็นการตรวจสอบว่าคะแนนที่ได้จากแบบวัดสัมพันธ์กับเกณฑ์ตามที่คาดหวังหรือไม่
- เรียกหลักฐานจากความสัมพันธ์มาตรที่สร้างขึ้นกับเกณฑ์ว่า ความตรงตามความสัมพันธ์กับเกณฑ์ (Criterion-related validity)
- สามารถแบ่งได้ 2 ประเภท คือ
 - ความตรงตามสภาพ (Concurrent Validity) ไม่เว้นระยะเวลาในการวัดเกณฑ์
 - ความตรงเชิงทำนาย (Predictive Validity) เว้นระยะเวลาในการวัดเกณฑ์

หลักฐานจากความสัมพันธ์กับเกณฑ์

- ความสัมพันธ์ที่ได้ควรถึงระดับนัยสำคัญ และมีขนาดตามที่คาดหวังด้วย
- **หมายเหตุ** เกณฑ์ต้องมีความตรง วัดในสิ่งที่ต้องการวัดด้วยเช่นกัน โดยมีคำศัพท์จำเพาะสำหรับเกณฑ์คือ
 - สิ่งบกพร่องในการวัดเกณฑ์ (Criterion deficiency)
 - สิ่งเจือปนในการวัดเกณฑ์ (Criterion contamination)

สิ่งที่ต้องการวัด

สิ่งที่วัดได้



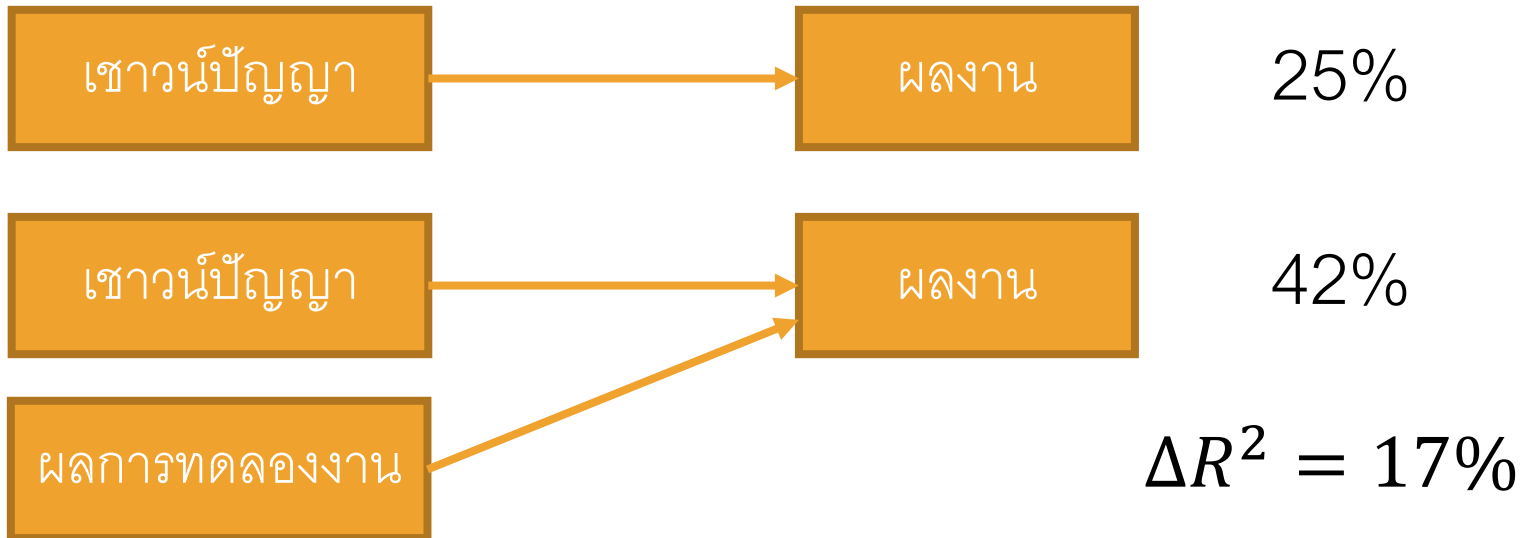
เช่น วัดผลการปฏิบัติงาน

ท่านคิดว่านาย A ทำงานได้ดีหรือไม่

ดีมาก	ดี	ปานกลาง	ไม่ดี	ไม่ดีมาก
-------	----	---------	-------	----------

หลักฐานจากความสัมพันธ์กับเกณฑ์

- ความตรงเพิ่มเติม (Incremental Validity) เป็นการตรวจสอบว่าแบบวัดที่มีนั้นสามารถทำนายเกณฑ์ได้เพิ่มเติมจากแบบวัดเดิมหรือไม่



Schmidt & Hunter (1998) การตรวจสอบความตรงเพิ่มเติมในการทำนายผลการปฏิบัติงาน เมื่อการวัดเชาวน์ปัญญาใช้เป็นตัวทำนาย

Table 1
*Predictive Validity for Overall Job Performance of General Mental Ability (GMA) Scores
Combined With a Second Predictor Using (Standardized) Multiple Regression*

Personnel measures	Validity (<i>r</i>)	Multiple <i>R</i>	Gain in validity from adding supplement	% increase in validity	Standardized regression weights	
					GMA	Supplement
GMA tests ^a	.51					
Work sample tests ^b	.54	.63	.12	24%	.36	.41
Integrity tests ^c	.41	.65	.14	27%	.51	.41
Conscientiousness tests ^d	.31	.60	.09	18%	.51	.31
Employment interviews (structured) ^e	.51	.63	.12	24%	.39	.39
Employment interviews (unstructured) ^f	.38	.55	.04	8%	.43	.22
Job knowledge tests ^g	.48	.58	.07	14%	.36	.31
Job tryout procedure ^h	.44	.58	.07	14%	.40	.20
Peer ratings ⁱ	.49	.58	.07	14%	.35	.31
T & E behavioral consistency method ^j	.45	.58	.07	14%	.39	.31
Reference checks ^k	.26	.57	.06	12%	.51	.26
Job experience (years) ^l	.18	.54	.03	6%	.51	.18
Biographical data measures ^m	.35	.52	.01	2%	.45	.13
Assessment centers ⁿ	.37	.53	.02	4%	.43	.15
T & E point method ^o	.11	.52	.01	2%	.39	.29
Years of education ^p	.10	.52	.01	2%	.51	.10
Interests ^q	.10	.52	.01	2%	.51	.10
Graphology ^r	.02	.51	.00	0%	.51	.02
Age ^s	-.01	.51	.00	0%	.51	-.01

หลักฐานจากความสัมพันธ์กับมาตรวัดสิ่งเดียวกัน

- แบบทดสอบที่สร้างขึ้นมา อาจไม่ใช่แบบทดสอบใหม่ที่วัดภาวะสันนิษฐานหนึ่ง อาจมีแบบทดสอบที่วัดภาวะสันนิษฐานนั้นไว้ก่อนแล้ว
- การหลักฐานเพียงแค่การไปทดสอบว่า แบบทดสอบที่สร้างขึ้นมานั้น มีความสัมพันธ์กับแบบทดสอบที่มีมาก่อนหน้านั้น
- เช่น แบบทดสอบบุคลิกภาพที่พัฒนาขึ้นมาใหม่ ที่ต้องการเปลี่ยนรูปแบบข้อคำถามให้เป็นแบบบังคับเลือก (Forced-choice) ได้ผลคะแนนสอดคล้องกับแบบทดสอบบุคลิกภาพที่โด่งดังในอดีต เช่น NEO-PI-R ที่วัดด้วยวิธีลิเคิร์ตปกติหรือไม่

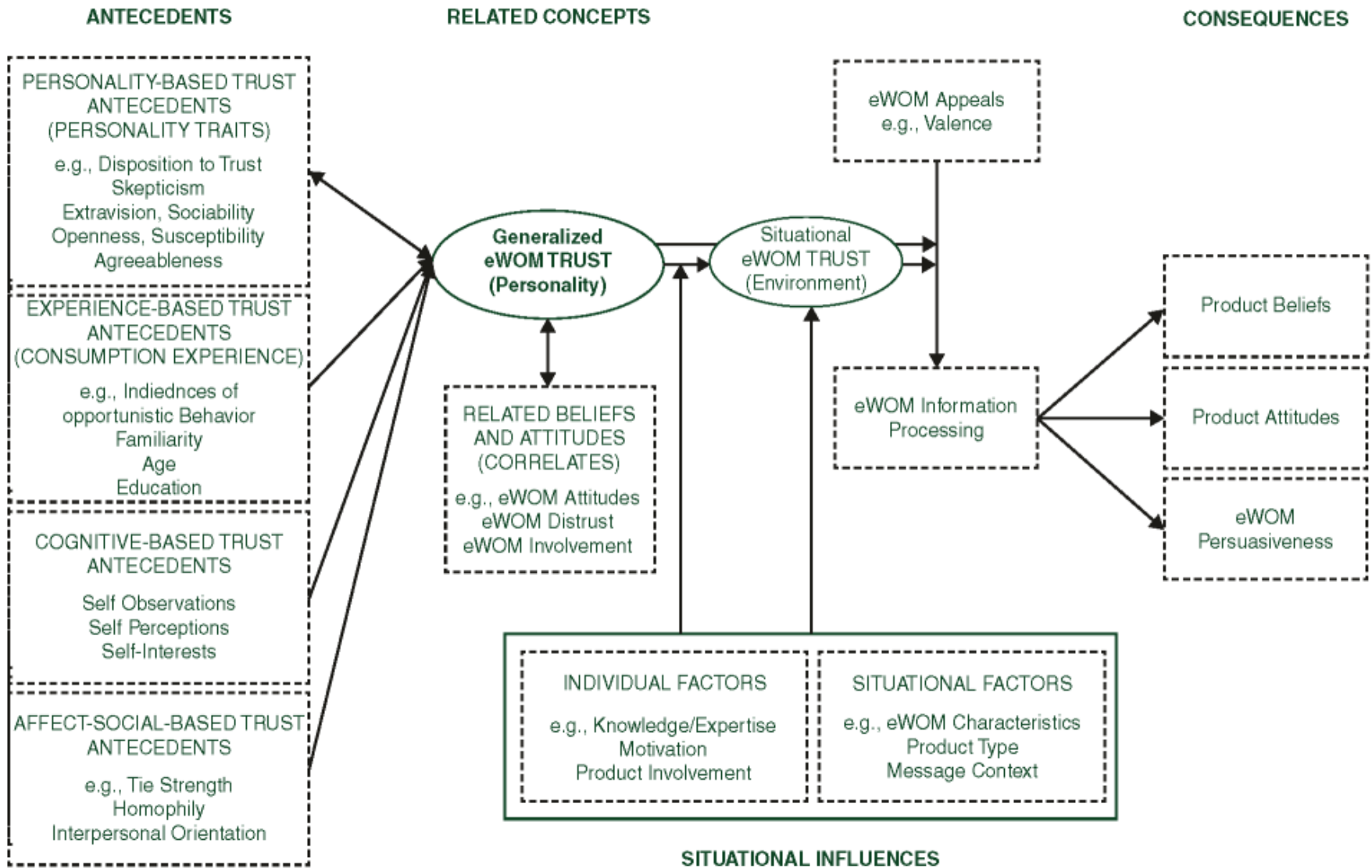
หลักฐานจากความสัมพันธ์กับแบบทดสอบอื่นที่วัดคนละ ภาวะสันนิษฐาน แต่มีความสัมพันธ์ทางทฤษฎี

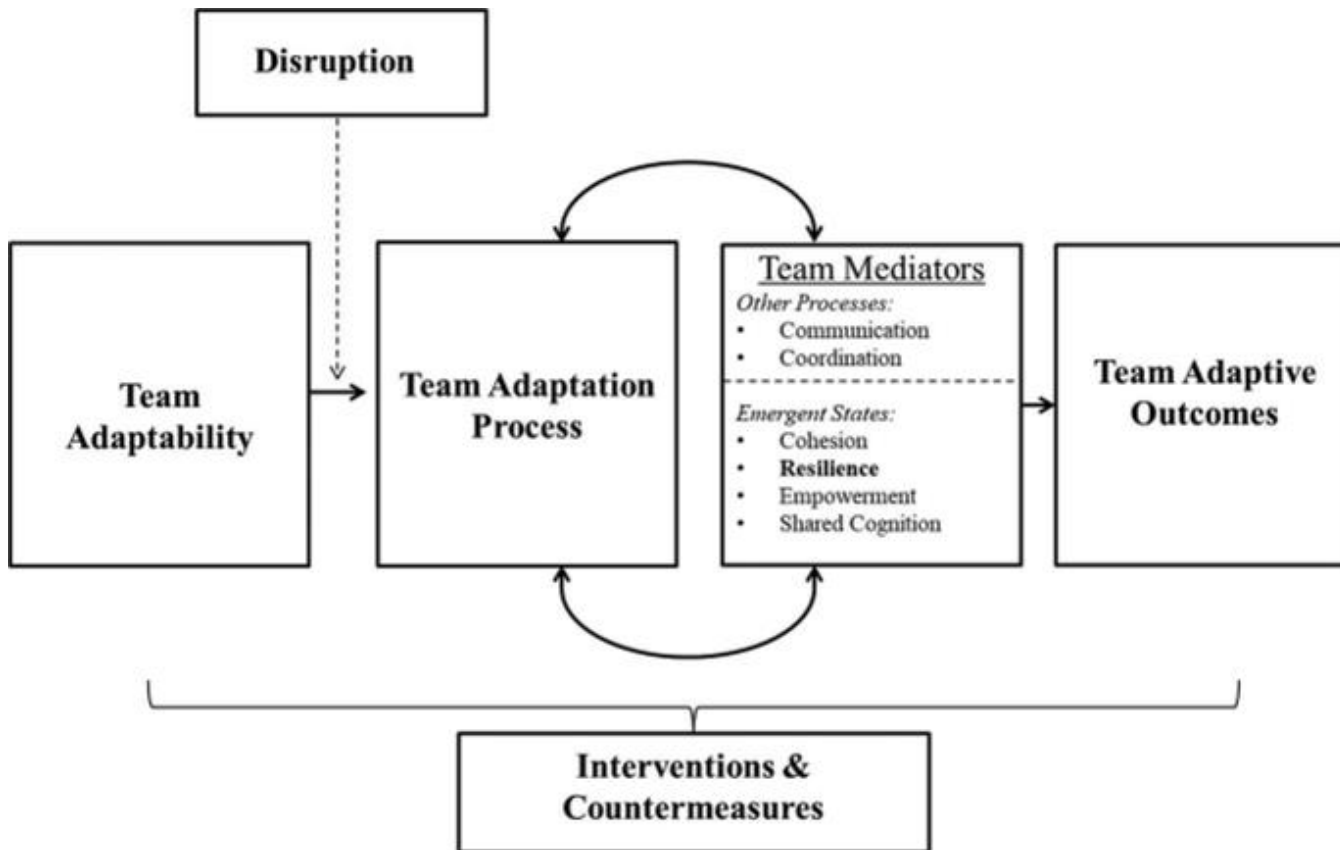
- ทฤษฎีว่าอย่างไร ถ้าแบบทดสอบที่สร้างขึ้นมา วัดสิ่งนั้นจริง ก็ต้องมี ความสัมพันธ์สอดคล้องกับทฤษฎี
 - ความสัมพันธ์กับอายุ พัฒนาการ (เช่น แบบทดสอบความสามารถทางภาษา)
 - เปลี่ยนแปลงตามการจัดกระทำ เช่น การเสียม (Priming)
 - <https://www.youtube.com/watch?v=NRTFPPBZGnl>
 - แตกต่างระหว่างกลุ่มตามทฤษฎี (Contrasted groups; Known groups) เช่น ตำรวจควรมีคะแนนแบบทดสอบความซื่อสัตย์มากกว่ากลุ่มอาชีพอื่น
 - สัมพันธ์กับตัวแปรอื่นตามทฤษฎี เช่น แบบทดสอบความหลงตัวเอง (Narcissism) ควรสัมพันธ์กับบุคลิกภาพแบบอ่อนไหวทางอารมณ์ (Neuroticism)

หลักฐานจากความสัมพันธ์กับแบบทดสอบอื่นที่วัดคุณลักษณะ ภาวะสันนิษฐาน แต่มีความสัมพันธ์ทางทฤษฎี

- รูปที่อธิบายความสัมพันธ์ระหว่างตัวแปรต่างๆ ที่เรียกว่า เครือข่ายกฎและเหตุผล (Nomological Network)
- วิธีการอธิบายความสัมพันธ์ระหว่างตัวแปรต่างๆ เป็นรูปภาพ ไม่มีมาตรฐานที่ชัดเจนนัก แต่จากที่ผมสังเกต สามารถแบ่งง่ายๆ เป็น 2 รูปแบบ
 - คล้ายกับกรอบแนวคิดในการวิจัย
 - คล้ายกับโมเดลความสัมพันธ์ในเชิงสถิติ

eWOM = electronic word of mouth

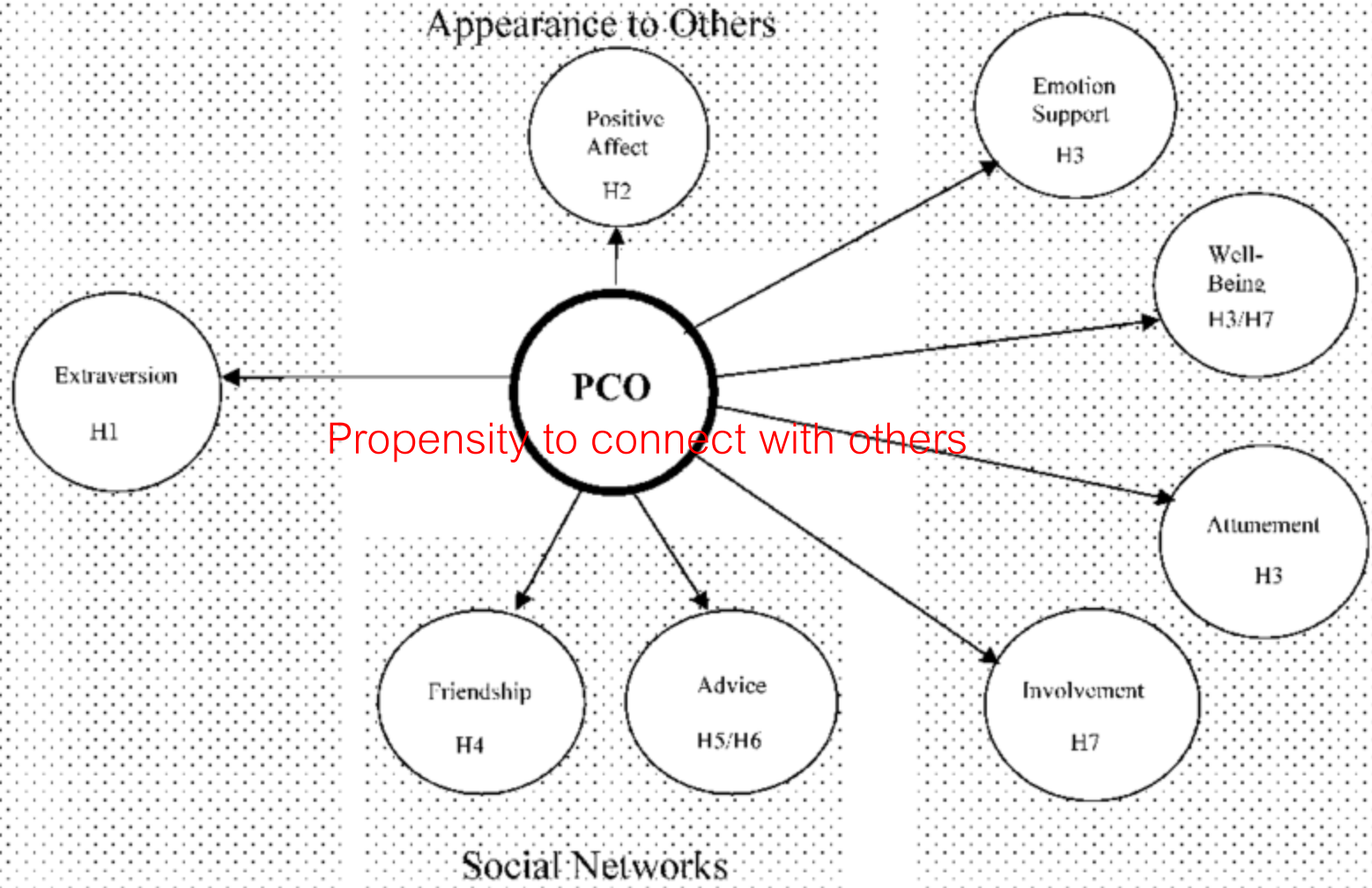




PERSONALITY
TRAITS

BEHAVIORAL
EXPRESSIONS

INDICATORS OF
ADJUSTMENT



Propensity to connect with others

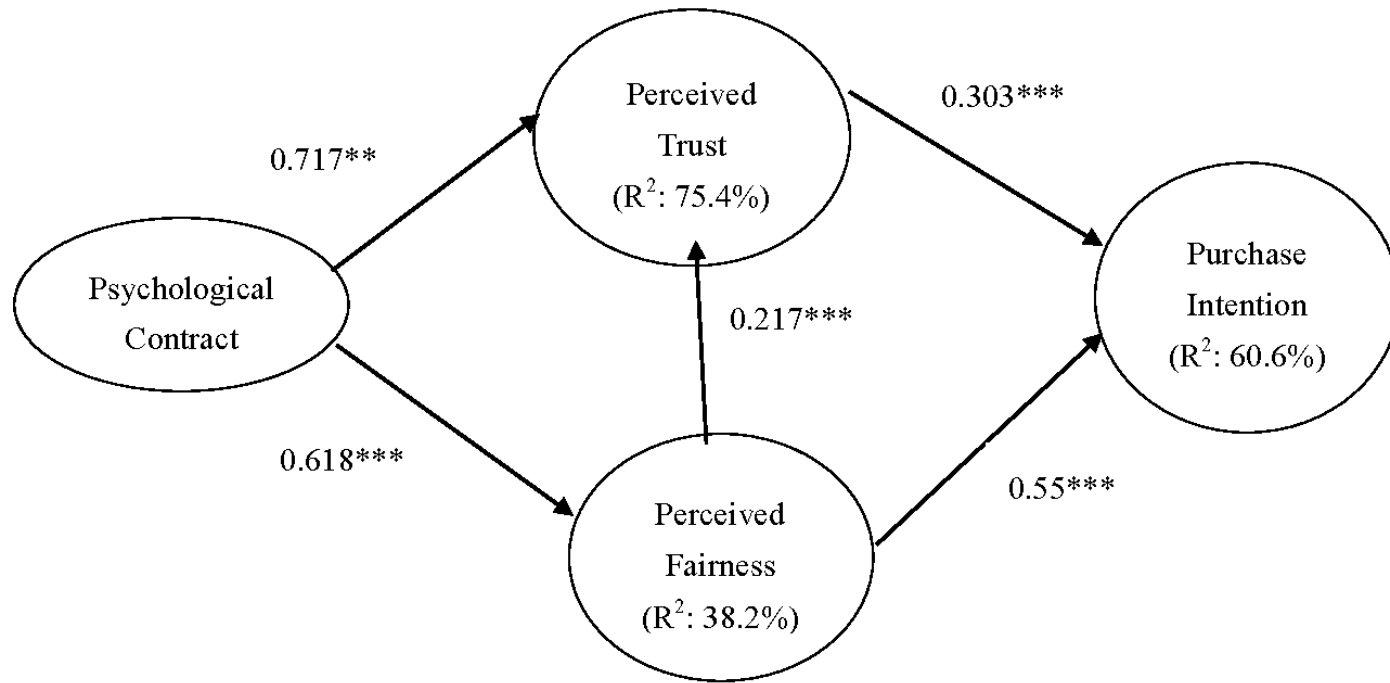


Figure 2. Results for the Proposed Research Model

หลักฐานจากความสัมพันธ์กับแบบทดสอบอื่นที่วัดคุณลักษณะ สันนิษฐาน แต่มีความสัมพันธ์ทางทฤษฎี

- คำว่าเครือข่ายกฎและเหตุผล (Nomological Network) มักถูกนำมาผูกกับการทดสอบความตรงของแบบทดสอบกับแบบทดสอบอื่นตามทฤษฎี
- ส่วนตัวมองว่า การคิดออกมาเป็นรูปภาพ ทำให้ผู้พัฒนาแบบทดสอบได้คิดถึงแบบทดสอบอื่นที่เกี่ยวข้องหลายมุม เช่น ตัวแปรนำ ตัวแปรตาม ฯลฯ
- รูปเป็นกรอบที่ช่วยให้คิด และช่วยอธิบายผู้อ่าน คล้ายกับการทำ Mind map แต่ไม่ใช่เทคนิคในการหาความตรง

หลักฐานจากความสัมพันธ์กับแบบทดสอบอื่นที่ไม่ควรสัมพันธ์กันทางทฤษฎี

- แบบทดสอบที่สร้างขึ้นมา ไม่ควรสัมพันธ์กับตัวแปรอื่นที่ไม่ควรสัมพันธ์กันทางทฤษฎี
- เช่น มาตรวัดความถนัด (Aptitude Test) ในการเป็นนักบิน ไม่ควรมีความสัมพันธ์กับความสามารถในการใช้ภาษาอังกฤษ
- คำศัพท์ที่เกี่ยวข้อง
 - สัมพันธ์กับแบบวัดอื่นที่วัดในสิ่งเดียวกัน หรือที่เรียกว่าความตรงลู่เข้า (Convergent Validity)
 - ไม่สัมพันธ์กับแบบวัดอื่นที่ไม่ได้วัดในสิ่งเดียวกัน หรือที่เรียกว่าความตรงลู่ออก (Discriminant Validity)

หลักฐานจากความสัมพันธ์กับตัวแปรอื่น

- การแผ่ขยายของความตรง (Validity Generalization) หมายถึงปรากฏการณ์ที่หลักฐานความสัมพันธ์กับตัวแปรอื่น ได้ถูกทดสอบมาบ่อยครั้งแล้ว ถูกนำมาพิจารณาว่าจะสามารถแผ่ขยายไปสถานการณ์อื่นได้ดีอย่างไร เช่น
 - ความสามารถทางปัญญา (Cognitive ability) สามารถทำนายผลการปฏิบัติงานได้ทุกอาชีพ
 - บุคลิกภาพแบบเปิดตัว ทำนายผลการปฏิบัติงานได้บางอาชีพ เช่น พนักงานขาย แต่ทำนายอาชีพอื่นไม่ได้
- การแผ่ขยายจะยิ่งน่าเชื่อถือ หากได้หลักฐานจากผลวิเคราะห์อภิมาน (Meta-analysis) ที่ใช้หลักการทางสถิติสังเคราะห์ผลการวิจัยจากหลายๆ งาน
- จากหลักฐานในวรรณกรรมที่แน่นอนหนาแล้ว ทำให้ไม่จำเป็นต้องทดสอบความตรงในการประยุกต์ใช้แบบทดสอบทุกครั้ง

หลักฐานจากความสัมพันธ์กับตัวแปรอื่น

- การจำกัดพิสัย (Range Restriction) ทำให้ค่าสหสัมพันธ์ที่ได้น้อยกว่าความเป็นจริง อาจส่งผลให้หลักฐานจากความสัมพันธ์ออกมาไม่ดี
- บางครั้ง นักวิจัยจะลองคำนวณดูว่า หากช่วงของคะแนนในงานของตัวเอง ถูกเปลี่ยนเป็นช่วงคะแนนตามประชากรแล้ว ความสัมพันธ์จะมีค่าเท่าไร สามารถใช้ฟังก์ชัน `rangeCorrection` ใน `psych` package ใน R ในการคำนวณได้

ค่าสหสัมพันธ์ที่ได้จากกลุ่มที่ถูกจำกัด ส่วนเบี่ยงเบนมาตรฐานของกลุ่มที่ถูกจำกัด

```
> rangeCorrection(r=.12, sdu=1, sdr=0.5)
```

[1] 0.2349781

ส่วนเบี่ยงเบนมาตรฐานของประชากร

หลักฐานผลที่ตามมาจากการทดสอบ

- เป็นการให้ผู้พัฒนาและผู้ใช้แบบทดสอบ พิจารณาว่าหลักจากใช้แบบทดสอบแล้ว มีผลที่ตามมาอย่างไร ซึ่งสามารถมองได้ 2 ด้าน
 1. การตีความหมาย อาจเป็นการตีความที่ไม่ได้ตรงกับเจตนารมณ์ของผู้พัฒนาแบบทดสอบ เช่น
 - แบบทดสอบ O-NET ไว้ประเมินคุณภาพโรงเรียน แต่ไปใช้ประเมินนักเรียน
 - แบบทดสอบบุคลิกภาพ ไปใช้คัดเลือกพนักงาน
 2. ผลที่ตามมา จากการใช้ทั้งแบบตั้งใจและไม่ตั้งใจ เช่น
 - การตีความ หากแบบทดสอบมีผลต่อชีวิตคนอย่างสูง
 - การทำนายผลการปฏิบัติงานด้วยข้อมูลอัตชีวประวัติ ส่งผลให้คนบางกลุ่ม ถูกคัดเลือกมากกว่า ทำให้แบบทดสอบดูไม่ยุติธรรม
 - การสร้างศูนย์ประเมิน ทำให้องค์กรมีภาพลักษณ์ที่ดี ดูเอาใจใส่พนักงาน และพนักงานระดับหัวหน้าจะมีมุมมองที่ดี ต่อพนักงานที่ตนเลือกมา
- กล่าวคือ เป็นการอธิบาย การตีความที่นอกเหนือจากเจตนารมณ์ของผู้พัฒนา และกล่าวถึงผลดี ผลเสีย ที่ตามมาจากการใช้

ความตรงสามองค์ (Trinitarian view of validity)

- ความตรงตามเนื้อหา (Content Validity)
- ความตรงตามความสัมพันธ์กับเกณฑ์ (Criterion-related Validity)
- ความตรงตามภาวะสันนิษฐาน (Construct Validity)
 - ความตรงส่วนนี้ครอบคลุมความตรงสองชนิดแรก
 - แบ่งเป็น ความตรงภายใน (Internal validity) และความตรงภายนอก (External validity)
- การแบ่งแบบนี้ ยังคงใช้ในตำราส่วนใหญ่ แต่ส่วนตัวมองว่าเป็นการแบ่งที่ไม่ค่อยมีประโยชน์



แหล่งหลักฐานความตรง

- มีหลักฐานความตรงจำนวนเท่าไรถึงเพียงพอ
 - เมื่อเรามั่นใจว่า การใช้คะแนนดังกล่าวพอมีหลักฐานพิสูจน์ ว่าตีความหมายประยุกต์ใช้แบบนั้นได้จริงๆ
 - แต่หากมีโอกาส ไม่ควรหยุดเก็บข้อมูลความตรง
- หากเก็บหลักฐานความตรงแล้ว พบว่าไม่ค่อยสนับสนุนการตีความของแบบทดสอบจะทำอย่างไร
 - หากกลุ่มตัวอย่างขนาดเล็กน้อย คุณอาจจะเจอความผิดพลาดจากการสุ่ม (sampling error) ควรเก็บข้อมูลมากกว่านี้
 - หากกลุ่มตัวอย่างคุณเยอะ ต้องพิจารณาการใช้แบบทดสอบนั้น ซึ่งหากวรรณกรรมมีหลักฐานแน่นหนา ให้สงสัยตัวเองไว้ก่อน ว่ามีข้อผิดพลาดอะไรหรือไม่