

## หลักการสร้างเครื่องมือในการทดสอบทางจิตวิทยา

สันทัต พรประเสริฐมานิต  
คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย

การสร้างเครื่องมือวัดลักษณะต่างๆ ในเชิงจิตวิทยาเป็นกระบวนการสร้างเครื่องมือในการวัดที่ทำให้แน่ใจได้ว่าเครื่องมือดังกล่าวสามารถวัดเป้าหมายที่ต้องการจริงๆ เช่น ความฉลาด (Intelligence) เครื่องมือการทดสอบความฉลาดในเชิงจิตวิทยาจะเป็นข้อคำถามหรือข้อสอบที่ทำให้แน่ใจได้ว่าวัดความฉลาดจริง และนำคะแนนของข้อคำถามแต่ละข้อไปคำนวณหาค่าคะแนนความฉลาดของผู้ที่ถูกทดสอบ

ก่อนที่จะเริ่มสร้างเครื่องมือการวัดเชิงจิตวิทยาได้นั้น จะต้องเข้าใจก่อนว่าเป้าหมายในการวัดคืออะไร ในทางจิตวิทยาเป้าหมายในการวัดมักจะเป็นลักษณะต่างๆ ของมนุษย์ เช่น ทักษะ ความฉลาด บุคลิกภาพ ความสามารถ ความสนใจ เป็นต้น ในกรณีของสมรรถนะหลักก็ถือว่าเป็นลักษณะทางจิตวิทยาอย่างหนึ่ง เช่นเดียวกัน จะต้องกำหนดว่าสมรรถนะหลักที่ต้องการคืออะไร

ลักษณะทางจิตวิทยาจะถือว่าเป็นลักษณะแฝงที่ไม่สามารถวัดได้โดยตรง อย่างเช่น ความฉลาด ไม่สามารถวัดได้โดยตรง ไม่เหมือนกับลักษณะทางกายภาพโดยทั่วไป เช่น ความเร็ว น้ำหนัก ส่วนสูง ที่สามารถวัดได้โดยตรง ลักษณะทางจิตวิทยาต้องวัดผ่านลักษณะพฤติกรรมที่สะท้อนให้เห็นถึงลักษณะนั้น เช่น ลักษณะของคนชอบเสี่ยง อาจจะมีตัวบ่งชี้เป็นข้อคำถามต่างๆ เช่น ฉันชอบเล่นการพนัน ฉันชอบความท้าทาย ฉันชอบลองสิ่งแปลกใหม่ เป็นต้น ถ้าบุคคลหนึ่งตอบว่ามีพฤติกรรมดังข้อคำถามทั้งสามสูง จะมีแนวโน้มที่จะเป็นคนชอบเสี่ยง เป็นต้น

แต่อย่างไรก็ตาม ทำอย่างไรจึงจะแน่ใจได้ว่าตัวบ่งชี้เหล่านี้ วัดลักษณะทางจิตวิทยาได้จริง ทำอย่างไรจึงจะแน่ใจได้ว่าจำนวนตัวบ่งชี้เพียงพอที่จะแสดงค่าลักษณะทางจิตวิทยาของแต่ละบุคคล ทำอย่างไรจึงจะแปลความหมายคะแนนของแบบทดสอบที่ได้ ทำอย่างไรจึงจะแน่ใจได้ว่าคะแนนที่ได้จากแบบทดสอบไม่ได้เกิดจากปัจจัยแทรกซ้อนอื่นๆ เช่น ความช่วยเหลือโดยไม่ตั้งใจของผู้จัดสอบ เป็นต้น ด้วยข้อคำถามเหล่านี้ทำให้เกิดกระบวนการวัดทางจิตวิทยาขึ้น เพื่อให้เครื่องมือวัดต่างๆ มีคุณภาพ กล่าวคือวัดได้ตรงกับสิ่งที่ต้องการ สามารถแบ่งแยกคนมีคะแนนสูงและคนมีคะแนนต่ำได้ วัดได้คงเส้นคงวา

กระบวนการสร้างแบบวัดเชิงจิตวิทยาอย่างง่าย สามารถแบ่งได้เป็นดังนี้

1. การเขียนข้อคำถาม (Item writing) เป็นการเขียนข้อคำถามเพื่อใช้ในแบบทดสอบ
2. การทดสอบนำร่อง (Try out) การทดสอบนำร่อง เป็นการนำเครื่องมือที่ได้เขียนข้อคำถามมา ไปเก็บข้อมูลเพื่อนำข้อมูลเหล่านี้ไปใช้หาคุณภาพ
3. การวิเคราะห์ข้อคำถาม (Item analysis) เมื่อได้ข้อมูลจากการทดสอบนำร่องแล้ว ก็จะนำข้อมูลเหล่านี้มาทดสอบดูว่าข้อคำถามแต่ละข้อสามารถแบ่งแยกคนได้คะแนนสูงและคะแนนต่ำหรือไม่เกี่ยวข้องกับสิ่งที่ต้องการวัดหรือไม่
4. การวัดความคงเส้นคงวา (Reliability) เป็นการทดสอบว่าข้อกระทงทั้งหมดนี้วัดในสิ่งๆ เดียวกันหรือไม่

---

บันทึกจากผู้เขียน

เอกสารนี้ได้ถูกเขียนขึ้นเมื่อ มิถุนายน 2006 เพื่อใช้ประกอบการสอนชั่วโมงปฏิบัติการ วิชาการวัดและการทดสอบทางจิตวิทยา คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย

ติดต่อผู้เขียน: Sunthud Pornprasertmanit. Email: [psunthud@gmail.com](mailto:psunthud@gmail.com)

5. การตรวจสอบความถูกต้อง (Validation) เป็นการตรวจสอบว่าสิ่งที่ได้วัดจากแบบสอบถามนั้นวัดในสิ่งที่ต้องการหรือไม่ กล่าวคือวัดเป้าหมายที่ได้ตั้งเอาไว้หรือไม่
6. การหาค่าเปรียบเทียบ (Norms) การสร้างเกณฑ์เพื่อใช้ในการเปรียบเทียบคะแนนของแต่ละบุคคล
7. การทำให้เป็นมาตรฐาน (Standardization) เป็นวิธีการใช้แบบทดสอบ เพื่อให้การทดสอบสามารถวัดความสามารถของบุคคลจริง ไม่มีปัจจัยแทรกซ้อนมาเกี่ยวข้อง เช่น ความคาดหวังของผู้จัดทดสอบ เป็นต้น

รายละเอียดของแต่ละขั้นตอนเป็นดังต่อไปนี้

### 1. การเขียนข้อคำถาม

ก่อนที่จะเขียนข้อคำถามได้ดั่งนั้น จะต้องให้คำนิยามเป้าหมายที่ต้องการวัดเสียก่อน เพื่อให้เป็นหลักในการเขียนข้อกระทงต่อไป คำนิยามเหล่านี้อาจได้มาจากทฤษฎี การที่นำคำนิยามที่ได้มาจากทฤษฎีจะมีข้อดีที่บุคคลอื่นได้ยืนยันว่าลักษณะทางจิตวิทยาดังกล่าวนั้นมีอยู่จริง สามารถหาความสัมพันธ์ของลักษณะเหล่านั้น กับสิ่งที่สนใจตามทฤษฎี เช่น ความฉลาดมีผลต่อผลงานหรือไม่ เป็นต้น และนอกจากนี้ยังสามารถหาวิธีการวัดลักษณะต่างๆ ทางจิตวิทยาได้อีกด้วย เช่น ความขยัน และความซื่อสัตย์ อาจมีเทคนิคการวัดที่มีลักษณะเฉพาะความขยันอาจจะถามอย่างตรงไปตรงมาได้ แต่ความซื่อสัตย์อาจไม่เป็นเช่นนั้น อย่างไรก็ตามก็อาจจะให้นิยามตามความต้องการของผู้สร้าง เช่น กำหนดคำนิยามของคนขยันขึ้นขึ้นมาเอง การสร้างคำนิยามแบบนี้ถ้าสร้างขึ้นมาแล้ว ควรจะมีการทดสอบว่าสามารถเป็นสมรรถนะหลักขององค์กรได้จริงหรือไม่ด้วย

การตั้งนิยามที่แตกต่างกัน ก็นำไปสู่วิธีการวัดที่แตกต่างกัน และการเขียนข้อคำถามที่แตกต่างกันเช่นกัน อาทิเช่น ความซื่อสัตย์ ถ้าให้คำนิยามว่าเป็นบุคคลที่มีความคิดไม่คดโกง และมีเจตนาที่จะไม่คดโกงไม่ว่าจะเป็นกรณีใดก็ตาม ก็จะมีลักษณะการเขียนข้อกระทงรูปแบบหนึ่ง แต่ถ้าให้คำนิยามว่าเป็นบุคคลที่มีความคิดไม่คดโกง แต่ว่าถ้าเกิดในกรณีที่ต้องคดโกงโดยไม่เจตนา จะเป็นที่ยอมรับได้ ก็จะทำให้การเขียนข้อกระทงเป็นในรูปแบบหนึ่ง โดยปกติแล้วในลักษณะทางจิตวิทยาจะมีคำนิยามหลายรูปแบบที่แตกต่างกัน ผู้ที่นำมาใช้จึงต้องพิจารณาว่าคำนิยามแบบใดที่สามารถเป็นสมรรถนะหลักขององค์กรได้ หรือคำนิยามใดที่มีหลักฐานยืนยันว่าสอดคล้องกับผลงานของบุคคลในองค์กร หรือพฤติกรรมทางบวกของบุคลากรในองค์กรจริง

ถึงแม้ว่าจะได้คำนิยามมาแล้ว จะต้องพิจารณาอีกครั้งว่าคำนิยามสามารถทำไปสร้างข้อกระทงได้จริงหรือไม่ คำนิยามที่ดีนั้นจะต้องเขียนให้ชัดเจน ครอบคลุมตามทฤษฎี มีลักษณะที่กว้าง สอดคล้องกับชื่อที่ได้ตั้งลงไป และบุคคลที่แตกต่างกันเมื่อมาอ่านคำนิยามดังกล่าวจะต้องเข้าใจตรงกัน

เมื่อได้นิยามเหล่านี้แล้ว จะนำนิยามมาเขียนเป็นข้อคำถาม ข้อคำถามที่ดีควรสอดคล้องกับคำนิยาม อาจจะเป็นข้อคำถามที่ถามตรงๆ เช่น ถ้าถามเรื่องความขยัน อาจถามตรงๆ ว่าคุณเป็นคนขยันขันแข็ง หรือลักษณะของคนที่มีคำนิยามเหล่านี้สูง เช่น ถามว่าฉันชอบทำงานหนัก เป็นต้น ข้อคำถามเหล่านี้จะถือว่าเป็นสิ่งกระตุ้นให้บุคคลตอบว่าบุคคลมีลักษณะดังกล่าวมาน้อยเพียงใด จะถือว่าข้อคำถามเหล่านี้ถูกสุ่มออกมาให้ครอบคลุมทุกสถานการณ์ที่คำนิยามดังกล่าวจะไปถึงได้ เช่น องค์กรต้องการบุคคลที่มั่นใจในตนเอง จะต้องสุ่มข้อคำถามในสถานการณ์ต่างๆ เช่น เวลานำเสนองาน เวลาทำงานที่ยากลำบาก เวลาพบปะผู้คนใหม่ๆ เวลาแก้ไขปัญหาในชีวิต เป็นต้น ถ้าคำนิยามครอบคลุมสถานการณ์เฉพาะเวลาทำงาน ก็ให้ครอบคลุมในสถานการณ์ทำงานทั้งหมด ถ้าคำนิยามครอบคลุมสถานการณ์ในชีวิตทั้งหมด ก็ต้องพยายามให้ข้อคำถามครอบคลุมสถานการณ์ในชีวิตทั้งหมดเช่นกัน

สิ่งที่ต้องคำนึงถึงในการเขียน คือ จะต้องอธิบายให้ชัดเจน ให้บุคคลที่แตกต่างกันอ่านแล้วเข้าใจตรงกัน ระดับของรูปแบบประโยค คำศัพท์ จะต้องเหมาะสมกับผู้ที่มาตอบข้อคำถาม เช่น บุคคลที่มีการศึกษาน้อย มาอ่านข้อคำถามที่ใช้คำศัพท์ยากๆ ก็จะทำให้ไม่เข้าใจ และตอบคำถามเหล่านั้นได้ไม่ถูกต้อง นอกจากนี้การเขียนจะต้องไม่ยาว เพราะว่ามนุษย์มีระดับความจำที่จำกัด อาจจะไม่สามารถจับประเด็นจากข้อคำถามที่ยาวๆ ได้ และสิ่งที่สำคัญอีกอย่างหนึ่งก็คือ จะต้องให้ข้อคำถามแต่ละข้อมีใจความเดียว ตัวอย่างข้อคำถามที่ไม่ได้มีใจความเดียว เช่น ฉันทันเป็นผู้ซึ่งถึงพระคุณและพร้อมจะตอบแทนพระคุณคน จะเห็นว่าคำถามนี้สามารถแบ่งได้เป็น 2 คำถามด้วยกัน คือ ฉันทันรู้ถึงพระคุณที่คนอื่นให้ และ ฉันทันพร้อมจะตอบแทนพระคุณที่คนอื่นให้ อาจจะเป็นไปได้ที่บางคนรู้ตระหนักถึงพระคุณ แต่ไม่ได้ตอบแทน ทำให้บุคคลนั้นอาจตอบข้อคำถามนี้ว่ามาก ถ้าสนใจเฉพาะความรู้คุณ หรืออาจตอบว่าน้อย ถ้าบุคคลนั้นคิดว่าใจความสำคัญของข้อความนี้คือการตอบแทนพระคุณ และในกรณีที่ข้อคำถามที่วัดความมากน้อยของลักษณะใดลักษณะหนึ่ง เช่น ความตรงต่อเวลา จะต้องมีย่อคำถามทางบวก คือถามว่าท่านไม่ตรงต่อเวลามากน้อยเพียงใด แบบสอบถามที่ดีควรมีย่อคำถามทางบวกและทางลบจำนวนใกล้เคียงกันด้วย เพื่อไม่ให้เกิดความโน้มเอียงไปด้านใดด้านหนึ่ง เช่น ถ้ามีแต่ข้อกระทงทางบวก จะทำให้คนส่วนใหญ่มีคะแนนสูง ข้อเสียที่ตามมาก็คือทำให้จำแนกคนตรงต่อเวลามาก และคนตรงต่อเวลาน้อยออกจากกัน ไม่ได้

ไม่ว่าจะเป็นการหาคำนิยาม หรือการเขียนข้อคำถามที่ดี ล้วนแต่มีความสำคัญอย่างยิ่งยวดต่อการสร้างแบบสอบถามที่ดี ถ้าจุดเริ่มต้นไม่ดีแล้ว ทำให้แทบจะไม่มีทางที่จะทำให้บรรลุเป้าหมายได้ด้วยดีได้ เพราะจะทำให้แบบสอบถามไม่มีคุณสมบัติที่ดี (คือไม่มีความคงเส้นคงวา และความตรง ซึ่งจะกล่าวถึงต่อไป)

จำนวนข้อคำถามที่ดีควรมีประมาณสองเท่าของจำนวนข้อที่ต้องการจะได้ และข้อคำถามในการวัดลักษณะแต่ละลักษณะควรมีอย่างน้อย 3 ข้อขึ้นไป หมายความว่าควรสร้างจำนวนข้อคำถามอย่างน้อย 6 ข้อ แต่พบว่าจำนวนข้อคำถามที่ส่งผลต่อคุณสมบัติของแบบทดสอบเหล่านี้ ซึ่งจะกล่าวถึงในส่วนของความคงเส้นคงวา และความตรง ถ้าต้องการแบบสอบถามที่มีคุณภาพมาก ตามทฤษฎีการวัดทางจิตวิทยาแล้ว จะต้องให้มีข้อคำถามจำนวนมาก แต่อย่างไรก็ดี ถ้ามีข้อคำถามมากเกินไปจะส่งผลต่อความตั้งใจของผู้ตอบ ยิ่งมากก็จะทำให้คำตอบที่ได้มา โดยเฉพาะอย่างยิ่งคำตอบที่ได้จากข้อคำถามท้ายๆ ไม่ตรงกับความเป็นจริง

รูปแบบที่ให้ผู้ทำแบบทดสอบตอบ จะมีลักษณะแตกต่างกันไปตามลักษณะทางจิตวิทยาที่เลือกตั้งที่กล่าวไว้ข้างต้น โดยทั่วไปจะแบ่งเป็น 2 แบบด้วยกันดังนี้

1) การทดสอบแบบมีข้อถูกและข้อผิด ข้อคำถามประเภทนี้จะเหมือนกับข้อสอบที่พบในบริบทของการศึกษา เช่น แบบทดสอบคัดเลือกรูปบุคคลเข้าเรียนระดับอุดมศึกษา จะเป็นข้อคำถามที่มีตัวเลือกให้ และจะมีตัวเลือกหนึ่งที่เป็นคำตอบที่ถูก บางครั้งก็เป็นข้อสอบที่ไม่มีตัวเลือกให้ ให้เขียนคำตอบลงไป แต่จะมีคำตอบที่ถูกต้องอยู่ อย่างไรก็ตาม ถ้าไม่มีตัวเลือกให้แล้ว จะต้องกำหนดคำตอบที่ถูกต้องให้ชัดเจน จนกระทั่งให้บุคคลที่แตกต่างกันมาตรวจข้อสอบ จะต้องได้คะแนนออกมาใกล้เคียงกันมาก หรือเหมือนกันให้ได้ รูปแบบการตอบประเภทนี้จะเหมาะสมกับการทดสอบเชาวน์ปัญญา ความสามารถทางสมอง ความถนัด เช่น ความถนัดทางเครื่องกล ความถนัดทางภาษา อย่างไรก็ตาม อาจมาใช้กับแบบทดสอบบุคลิกภาพได้ เช่น แบบทดสอบความฉลาดทางอารมณ์ อาจมีข้อคำถามว่า “ฉันไม่อยากจะเข้าใจว่าบุคคลอื่นต้องการอะไร” แล้วให้เลือกตอบว่าตรงหรือไม่ตรงกับผู้ตอบ ถ้าตอบว่าตรงแสดงว่าไม่ได้แสดงถึงความฉลาดทางอารมณ์ แต่ถ้าตอบว่าไม่ตรงแสดงว่าแสดงถึงความฉลาดทางอารมณ์

2) การทดสอบแบบให้ตอบว่ามากน้อย ข้อคำถามประเภทนี้จะมีความให้ แล้วให้ตอบว่าผู้ตอบมีลักษณะดังกล่าวมากหรือน้อยเพียงใด อาจจะช่องให้เลือกอยู่ 5 ระดับ เช่น “ฉันชอบเล่นกีฬา” แล้วมีคำตอบว่า เห็นด้วยอย่างยิ่ง เห็นด้วย เห็นด้วยและไม่เห็นด้วยพอๆ กัน ไม่เห็นด้วย และไม่เห็นด้วยอย่างยิ่ง คำถามประเภทนี้มักจะพบบ่อยมากในการวัดบุคลิกภาพ เจตคติ ความสนใจ ค่านิยม หรือตัวอย่างง่ายที่สุดที่มักจะเจอบ่อยครั้งก็คือ ความพึงพอใจต่อการใช้บริการ ก็มักจะพบรูปแบบประเภทนี้ แต่บุคคลที่เลือกใช้ข้อคำถามประเภทนี้จะต้องตัดสินใจว่าจะมีช่องให้เลือกกี่ระดับ บางครั้งอาจให้เลือก 4 ระดับ 10 ระดับ หรือให้คะแนนตั้งแต่ 0 ถึง 100 คะแนนเลยก็ได้ การเลือกยิ่งละเอียดมากยิ่งขึ้นทำให้แบ่งแยกบุคคลได้ดีขึ้น แต่ถ้ามีมากเกินไป บุคคลก็จะใช้ทางลัดในการตอบ กล่าวคือ ไม่ใส่ใจรายละเอียด เลือกที่ประมาณเอา เช่น ประมาณ 60 คะแนน เป็นต้น ซึ่งจริงๆ แล้วเขาอาจมีคะแนนอยู่ที่ระดับ 64 คะแนน

คำถามบางประเภทจะเป็นลักษณะผสมทั้งสองอย่างเข้าด้วยกัน เช่น รูปแบบการบังคับเลือก อาจมีตัวเลือกให้ 3 ตัวเลือก ตัวเลือกหนึ่งแสดงถึงความซื่อสัตย์ ตัวเลือกหนึ่งแสดงถึงการมีใจให้บริการ และอีกตัวเลือกหนึ่งแสดงถึงการมุ่งผลสำเร็จ ผู้ตอบต้องเรียงลำดับตัวเลือกทั้งสาม ว่าตัวเลือกใดตรงกับตนเองมากที่สุด ตัวเลือกใดรองลงมา และตัวเลือกใบน้อยที่สุด

ในการเลือกข้อคำถามแบบที่สอง ก็คือ ข้อคำถามแบบให้ตอบว่ามากน้อย จะต้องตัดสินใจว่า จะให้ผู้ใดเป็นคนตอบคำถาม อาจจะเป็นการตอบแบบรายงานตนเอง หรือให้ประเมินด้วยตนเอง หรือเป็นการประเมินโดยผู้อื่น อาจจะเป็นการสังเกตพฤติกรรม หรือไม่ผ่านก็ได้ ยกตัวอย่างเช่น เจ้าหน้าที่ประเมินผลงานลูกน้อง ครูประเมินความตั้งใจเรียนของนักเรียน หรือพ่อแม่ประเมินความซื่อสัตย์ของลูก เป็นต้น การให้ผู้อื่นตอบอาจจะเป็นการช่วยลดอคติในการเข้าข้างตนเองของผู้ตอบ แต่ในทางตรงกันข้าม ก็มีคำพูดที่ว่า “ไม่มีใครรู้ดีไปกว่าตัวของตัวเอง” และการประเมินด้วยผู้อื่น อาจมีข้อผิดพลาดได้ สังเกตได้โดยง่ายว่าแต่ละคนตีความนิยบบุคคลแตกต่างกัน คนสองคนอาจรับรู้ว่าเด็กคนหนึ่งมีนิสัยต่างกัน คนหนึ่งมองว่าเด็กชอบเล่น แต่อีกคนหนึ่งมองว่าเด็กน่ารำคาญ จะพบว่าทั้งการประเมินด้วยตนเองหรือผู้อื่น ก็มีข้อดีข้อเสียแตกต่างกัน ดังนั้นการเลือกใช้จะต้องคำนึงถึงความเป็นไปได้ของการให้ผู้อื่นประเมิน ความร้ายแรงของอคติที่เกิดขึ้นในแต่ละส่วน ค่าใช้จ่ายในการเลือกใช้วิธีการต่างๆ คุณภาพของคำตอบที่ได้จากการประเมินต่างแหล่ง มาใช้ในการพิจารณาประกอบการตัดสินใจ

## 2. การทดสอบนำร่อง (Try out)

การทดสอบนำร่อง เป็นการนำข้อคำถามที่ได้ตั้งขึ้นมาไปลองใช้ ขั้นตอนนี้นักวิจัยบางคนให้ความสำคัญมาก แต่บางคนไม่ค่อยจะใส่ใจกับการทดสอบนำร่องมากนัก คิดว่าเป็นเพียงการเก็บข้อมูลเพื่อใช้ในการวิเคราะห์ขั้นต่อไป ถึงอย่างไรก็ดี ถ้าใช้การทดสอบนำร่องให้เป็นประโยชน์แล้ว จะทำให้คุณภาพของแบบสอบถามมากขึ้นได้

การทดสอบนำร่องจะเป็นขั้นตอนที่ทำให้ผู้สร้างแบบสอบถามพบกับผู้ตอบแบบสอบถามได้โดยตรง ไม่ว่าจะเป็นการตอบด้วยตนเอง หรือการตอบโดยผู้อื่นก็ตาม ผู้สร้างสามารถถามถึงรูปแบบของข้อคำถาม ความชัดเจนของเนื้อความในแต่ละข้อคำถาม ความกำกวมของข้อคำถาม ระดับความยากของภาษาในแต่ละข้อคำถาม ความยาวของแบบสอบถามว่าเหมาะสมหรือไม่ จำนวนตัวเลือกมีมากเกินไปหรือไม่ มีคำคิดในข้อคำถามแต่ละข้อหรือไม่ คำแนะนำที่แสดงไว้ก่อนการตอบคำถามชัดเจนหรือเข้าใจตรงกันระหว่างผู้ตอบที่แตกต่างกันหรือไม่ ในกรณีของการสังเกตพฤติกรรม จะทำให้รับรู้ปัญหา อย่างเช่นมีพฤติกรรมที่กำกวมที่ไม่สามารถรู้ได้ว่าผู้ถูกสังเกตมีลักษณะอย่างไร เช่น การสังเกตความตั้งใจในการทำงาน ถ้าผู้ถูกสังเกตทำงานให้เสร็จภายใน 1 ชั่วโมง ภายใน

เวลา 2 ชั่วโมงที่กำหนด แล้วพัก จะถือว่าตั้งใจหรือไม่ตั้งใจ ตรงจุดนี้จะทำให้ผู้สร้างแบบสอบถามได้กลับ ทบทวนแก้ไขก่อนนำมาทดสอบอีกครั้งหนึ่ง

นอกจากนี้ผู้ที่นำมาทดสอบจะต้องสอดคล้องกับกลุ่มเป้าหมายของแบบทดสอบชุดนั้น เช่น ต้องการไป คัดเลือกบุคคลเข้าทำงาน จะต้องถามกับบุคคลเข้าทำงาน ถ้าต้องการใช้คัดเลือกกับพนักงานทุกตำแหน่ง ก็ควรจะ ทดสอบนำร่องกับพนักงานทุกตำแหน่งถ้าเป็นไปได้ แต่อย่างไรก็ตามเนื่องจากความสะดวกในการเก็บข้อมูลและ งบประมาณ ทำให้ไม่สามารถใช้ผู้สมัครมาตอบแบบสอบถามได้ อาจจะต้องใช้พนักงานปัจจุบันมาเป็นกลุ่ม ตัวอย่างในการทดสอบนำร่อง ทำให้ต้องระวังว่าพนักงานปัจจุบันอาจจะไม่ได้เป็นตัวแทนของผู้สมัคร เช่น พนักงานปัจจุบันอาจมีการศึกษาสูง แต่บุคคลที่มาสมัครมีตั้งแต่การศึกษาดีจนกระทั่งการศึกษาไม่ดี ทำให้เป็นไปได้ ที่ว่าไม่เจอปัญหาเรื่องภาษาในการเก็บข้อมูลนำร่อง แต่เจอปัญหาในการเก็บข้อมูลจริง ส่วนจำนวนของกลุ่ม ตัวอย่างนำร่อง ควรจะมีอย่างน้อย 5 คน ต่อ 1 ข้อคำถาม แต่จำนวนจะต้องปรับตามเทคนิคในการวิเคราะห์ คุณภาพของเครื่องมือวัดที่ใช้ และปรับตามงบประมาณที่มี

### 3. การวิเคราะห์ข้อคำถาม

การวิเคราะห์ข้อคำถามเป็นขั้นตอนที่จะยืนยันว่าข้อคำถามแต่ละข้อสามารถวัดนิยามได้จริงหรือไม่ ถ้า เป็นข้อคำถามที่มีข้อถูกหรือข้อผิด ข้อที่ยากมากๆ จนกระทั่งโอกาสที่คนจะถูกในข้อคำถามนี้พอๆ กับการเดา หรือข้อคำถามที่ง่ายมากๆ จนกระทั่งคนทุกคนตอบได้หมด ถือว่าเป็นข้อที่ไม่สามารถวัดลักษณะที่เป็นเป้าหมาย ของผู้สร้างได้ เพราะว่าข้อคำถามเหล่านี้ไม่สามารถจำแนกคนเก่งออกจากคนไม่เก่งได้ นอกจากนี้ข้อคำถามที่ดี คนเก่งจะต้องตอบข้อคำถามนั้นได้มากกว่าคนที่ไม่เก่งด้วย คุณสมบัติแรกก็คือความยากของข้อคำถาม และ คุณสมบัติที่สองคืออำนาจในการจำแนกคนเก่งหรือคนไม่เก่งออกจากกันของข้อคำถามดังกล่าว

ก่อนที่จะวิเคราะห์ข้อคำถามได้นั้น ต้องแปลผลคะแนนจากคำตอบของผู้ตอบแบบสอบถามก่อน ใน กรณีของข้อคำถามที่มีข้อถูกข้อผิด ให้ตรวจว่าแต่ละข้อถูกหรือผิด เช่น ข้อสอบข้อที่ 1 ตอบตัวเลือก 3 ถูกต้อง ดังนั้นถ้าตอบตัวเลือก 3 ให้คะแนน 1 คะแนน ถ้าตอบตัวเลือกอื่นให้ 0 คะแนน เป็นต้น แล้วนำคะแนนของแต่ละ ข้อมารวมกัน ส่วนข้อคำถามที่ให้ตอบว่ามากน้อย ให้ลงคะแนนในการตอบแต่ละช่อง เช่น ถ้าเป็นข้อคำถาม 5 ระดับโดยเป็นข้อกระทงทางบวก จะให้คะแนน 5 คือเห็นด้วยอย่างยิ่ง ไล่ไปจนกระทั่ง 1 คือไม่เห็นด้วยอย่างยิ่ง และข้อกระทงทางลบจะให้คะแนน 1 คือเห็นด้วยอย่างยิ่ง ไล่ไปจนกระทั่ง 5 คือไม่เห็นด้วยอย่างยิ่ง แล้วนำ คะแนนแต่ละข้อมารวมกัน

เมื่อได้คะแนนดังกล่าวออกมา คะแนนนี้จะเรียกว่าคะแนนดิบ การแปลความหมายคะแนนดังกล่าวต้อง ระวัง เช่น แบบสอบถามวัดเชาวน์ปัญญา ถ้าได้คะแนนรวม 13 คะแนน จากข้อคำถาม 20 ข้อ จะไม่สามารถบอก ได้ว่าผู้ตอบมีเชาวน์ปัญญาสูงหรือต่ำ บอกได้เพียงแต่ว่าผู้ตอบตอบถูก 13 จาก 20 ข้อเท่านั้น ถ้าบังเอิญ 13 ข้อที่ ผู้ตอบตอบถูกยาก อาจจะมีเชาวน์ปัญญาดี แต่ถ้า 13 ข้อนั้นง่าย ก็ตอบไม่ได้ว่าบุคคลนี้มีเชาวน์ ปัญญาสูงหรือต่ำ เช่นเดียวกับข้อคำถามแบบให้ตอบว่ามากน้อย เช่น คะแนนความซื่อสัตย์เท่ากับ 20 คะแนน โดยที่คะแนนที่เป็นไปได้เท่ากับ 100 คะแนน คะแนนเพียง 20 คะแนนบอกไม่ได้ว่าคนนี้ซื่อสัตย์หรือไม่ อาจจะ เป็นไปได้ว่าคนอื่นฯ ทำได้คะแนนมากกว่านี้ หรือทำได้คะแนนน้อยกว่านี้ ด้วยเหตุผลทั้งหมดที่กล่าวมาข้างต้น คะแนนดิบจะต้องแปลงเป็นคะแนนมาตรฐาน ซึ่งจะกล่าวถึงในส่วนของการสร้างค่าเปรียบเทียบ

ความยากของข้อคำถามจะมีวิธีการวัดง่ายๆ โดยการหาร้อยละของคนที่ตอบถูกต้องคนทั้งหมด ถ้ามีค่าเข้า ใกล้เคียงหนึ่งแสดงว่าข้อคำถามนั้นง่าย คนเกือบทุกคนตอบถูก แต่ถ้ามีค่าเข้าใกล้ศูนย์แสดงว่าข้อคำถามนั้นยาก แทบ จะไม่มีคนตอบถูก แต่ก่อนที่พิจารณาว่าค่าความยากเท่าไรที่ข้อคำถามแต่ละข้อควรมี จะต้องหาค่าความเป็นไป

ได้ของการเดาก่อน คำนี้หมายความว่าบุคคลไม่รู้คำตอบแล้วเขา จะมีความเป็นไปได้ที่จะตอบถูกเท่าไร เช่น ถ้ามี 4 ตัวเลือก แล้วมีตัวเลือกหนึ่งถูก จะมีโอกาสเดาถูก 0.25 หรือร้อยละ 25 ถ้าค่าความยากของข้อคำถามใกล้เคียงกับค่าความเป็นไปได้ที่เกิดจากการเดาแล้ว แสดงว่าข้อคำถามยากเกินไป ค่าที่เหมาะสมของข้อคำถามควรมีค่าเท่ากับจุดตรงกลางระหว่างง่ายที่สุด ก็คือค่าความยากเท่ากับ 1 และค่าความเป็นไปได้ที่เกิดจากการเดา เช่น ข้อคำถามที่มีตัวเลือก 4 ตัว ค่าที่เหมาะสมคือจุดกึ่งกลางของระยะห่างระหว่าง 1 และ 0.25 ก็คือ 0.625 นั่นเอง ถ้าค่าความยากของข้อคำถามได้น้อยกว่า 0.625 แสดงว่าข้อนั้นยากกว่าปกติ แต่ถ้าข้อคำถามใดมีค่าความยากมากกว่า 0.625 แสดงว่าข้อคำถามนั้นง่ายกว่าปกติ อาจจะมีค่าระหว่างข้อยากและข้อง่าย แต่ให้ตัดข้อยากมากๆ และข้อง่ายมากๆ ออก

ส่วนอำนาจจำแนกจะมีวิธีการวัดอย่างง่าย โดยคัดคนที่มีลักษณะดังกล่าวสูง และคนที่มีลักษณะดังกล่าวต่ำออกมา การระบุว่าใครมีลักษณะสูงหรือต่ำอาจจะมาจากคะแนนรวมจากข้อคำถามทุกข้อ แล้วตัดคนที่อยู่ในกลุ่มสูง เช่น ร้อยละ 25 ของคนที่ตอบได้สูงที่สุด และตัดคนที่อยู่ในกลุ่มต่ำ เช่น ร้อยละ 25 ของคนที่ตอบได้ต่ำที่สุดออกมา อาจจะใช้เกณฑ์ภายนอกแทนคะแนนรวมได้ เช่น ผลงาน คะแนนสอบที่เป็นมาตรฐาน เป็นต้น เมื่อได้คนกลุ่มสูง และคนกลุ่มต่ำแล้ว จะนำคะแนนของข้อคำถามของคนในกลุ่มสูงมาหาค่าเฉลี่ย ได้เป็นค่าเฉลี่ยของคะแนนข้อคำถามนั้นในกลุ่มสูง ทำเช่นเดียวกับกลุ่มต่ำ (ในกรณีที่มีข้อถูกหรือข้อผิด ค่าเฉลี่ยที่ได้ก็จะเป็นร้อยละของคนที่ตอบได้ในคนกลุ่มสูงและกลุ่มต่ำ) แล้วนำค่าเฉลี่ยทั้งสองกลุ่มมาเปรียบเทียบกัน ถ้ากลุ่มสูงมากกว่ากลุ่มต่ำ แสดงว่าจำแนกได้ ยิ่งค่ามาก ยิ่งแสดงว่าจำแนกได้ดี แต่ถ้าค่าทั้งสองกลุ่มใกล้เคียงกันแสดงว่าข้อคำถามนี้จำแนกคนได้คะแนนสูงออกจากคนได้คะแนนต่ำไม่ได้ ด้วยเหตุนี้จะต้องคัดข้อคำถามดังกล่าวออก แต่ถ้ากรณีค่าของกลุ่มต่ำมากกว่าของกลุ่มสูง แสดงว่าจำแนกได้ไม่ดี หรืออาจจะเป็นไปได้ว่าข้อคำถามเป็นข้อคำถามทางลบแล้วไม่ได้กลับคะแนน ส่วนเกณฑ์ที่ใช้ในการตัดสินใจได้ 2 วิธี คืออาจเลือกข้อที่มีอำนาจจำแนกสูงสุดที่ข้อตามที่ต้องการ หรือใช้สถิติในการทดสอบ เช่นการทดสอบค่าที หรือการทดสอบสัดส่วนด้วยค่าซี ถ้ากลุ่มสูงมากกว่ากลุ่มต่ำอย่างมีนัยสำคัญแสดงว่าควรนำข้อนั้นไว้ แต่การใช้สถิติจะต้องระมัดระวังเรื่องจำนวนกลุ่มตัวอย่าง ถ้ามากเกินไปทำให้ถึงระดับนัยสำคัญง่ายขึ้น ส่วนถ้าน้อยเกินไปจะทำให้ถึงระดับนัยสำคัญยาก

วิธีในการหาอำนาจจำแนกอีกวิธีหนึ่งคือการหาค่าสหสัมพันธ์ระหว่างคะแนนข้อคำถามและคะแนนรวมของข้อคำถามที่เหลือ (ถ้าใช้คะแนนรวมทุกข้อคำถาม จะทำให้ค่าสหสัมพันธ์สูงเกินจริง เพราะข้อมูลของข้อคำถามไปอยู่ในคะแนนรวมด้วย เช่น มีข้อคำถาม 5 ข้อ ถ้าหาอำนาจจำแนกระหว่างข้อ 1 และคะแนนรวมข้อ 1 ถึงข้อ 5 จะทำให้คะแนนรวมนั้นมีข้อมูลของข้อ 1 ด้วย ดังนั้นคะแนนรวมจึงควรรหาเฉพาะข้อ 2 ถึงข้อ 5 แทน) ถ้าข้อคำถามเป็นการตอบแบบมากน้อย จะใช้สถิติเป็นการหาค่าสหสัมพันธ์แบบเพียร์สัน (Pearson product moment correlation) แต่ถ้าข้อคำถามเป็นแบบมีข้อถูกและข้อผิด จะใช้สถิติเป็นการหาค่าสหสัมพันธ์แบบพอยท์ไบซีเรียล (Point biserial correlation) ซึ่งสามารถใช้การหาค่าสหสัมพันธ์แบบเพียร์สันแทนได้ เพราะว่าวิธีการค่าสหสัมพันธ์นี้ดัดแปลงมาจากค่าสหสัมพันธ์แบบเพียร์สันนั่นเอง ค่าสหสัมพันธ์จะมีค่าตั้งแต่ 1 ถึง -1 ถ้าค่ามากกว่า 0 แสดงว่าสัมพันธ์กันทางบวก แสดงว่าคนตอบคะแนนข้อคำถามสูง คะแนนรวมจะสูง และถ้าคนตอบคะแนนข้อคำถามต่ำ คะแนนรวมจะต่ำ ถ้าเท่ากับ 0 แสดงว่าไม่สัมพันธ์กัน หมายความว่าคะแนนข้อคำถามจะสูงหรือจะต่ำอย่างไร คะแนนรวมจะไม่เปลี่ยนแปลง และค่าน้อยกว่า 0 แสดงว่าค่าสัมพันธ์กันทางลบ กล่าวคือถ้าคนตอบคะแนนข้อคำถามสูง คะแนนรวมจะต่ำ และคนตอบคะแนนข้อคำถามต่ำคะแนนจะสูง ถ้าค่าเข้าใกล้ 1 หรือ -1 มาก แสดงว่ายังสัมพันธ์กันมาก ถ้าค่าเข้าใกล้ 0 แสดงว่าสัมพันธ์กันน้อย ค่าอำนาจจำแนกที่

ต้องการคือ ค่าสหสัมพันธ์ทางบวก เกณฑ์ในการตัดสินอาจจะตัดสินความสัมพันธ์เท่ากับ .20 ขึ้นไป หรือใช้วิธีการทดสอบนัยสำคัญ (ต้องระวังอย่างที่ว่ามาแล้วข้างต้น) หรือใช้วิธีเลือกค่าสูงสุดตามจำนวนข้อที่ต้องการ

ในกรณีของข้อคำถามที่มีข้อถูกและข้อผิด ตัวเลือกที่ไม่ได้เป็นคำตอบที่ถูกต้องหรือตัวลวงนั้นส่งผลต่อค่าความยากของข้อดังกล่าวด้วย ถ้าตัวลวงนั้นมีคนมาตอบเยอะ แสดงว่าตัวลวงนั้นเป็นตัวลวงที่ดี ถ้าตัวลวงใดไม่มีใครเลือกเลย แสดงว่าตัวลวงนั้นควรจะเปลี่ยนแปลง การที่ข้อคำถามง่ายอาจจะเป็นเพราะว่าตัวลวงไม่ดี เช่นคำถาม “ประเทศไทยมีประชากรเท่าไร” 1) 10,000 คน 2) 20,000 คน 3) 65,000,000 คน 4) 20,000,000,000 คน จะเห็นว่าคนที่ไม่รู้จริงอาจจะประมาณได้ เพราะว่าตัวลวงไม่มีอำนาจในการลวงนั่นเอง

#### 4. การวัดความคงเส้นคงวา

เมื่อเราวัดลักษณะทางจิตวิทยาแก่คนใดคนหนึ่งแล้ว ในเชิงทฤษฎีการวัดทางจิตวิทยาจะถือว่าแต่ละบุคคลจะมีค่าของลักษณะทางจิตวิทยาที่แท้จริงอยู่ค่าหนึ่ง แต่ทว่าเมื่อวัดในสภาพความเป็นจริงแล้ว ค่าที่ได้จะไม่ตรงกับค่าความเป็นจริงเสมอไป เพราะว่าจะมีสถานการณ์ในการวัดที่แตกต่างกัน เช่น บางครั้งอารมณ์ดี บางครั้งอารมณ์ไม่ดี บางครั้งอากาศร้อน บางครั้งอากาศเย็น บางครั้งใช้แบบสอบถามรูปแบบที่ 1 บางครั้งใช้แบบสอบถามอีกรูปแบบหนึ่ง สิ่งเหล่านี้จะทำให้ค่าที่วัดได้เบี่ยงเบนจากความเป็นจริง ค่าที่เบี่ยงเบนไปจากค่าความเป็นจริงเรียกว่าค่าผิดพลาด

ในทฤษฎีการวัดทางจิตวิทยาเชื่อว่า ถ้าวัดบุคคลใดบุคคลหนึ่งด้วยหลากหลายสถานการณ์ ใช้แบบทดสอบหลายชุด วัดหลายครั้ง ค่าเฉลี่ยของค่าที่ได้เหล่านี้จะถือเป็นค่าคะแนนที่แท้จริงของบุคคลนั้น แต่แนวคิดนี้จะเป็นจริงได้ก็ต่อเมื่อลักษณะของการวัดเป็นไปตามหลัก 2 ข้อ ดังนี้ 1) ค่าผิดพลาดต้องเกิดขึ้นโดยสุ่ม กล่าวคือ ในบางสถานการณ์ค่าผิดพลาดเป็นบวก (คะแนนที่วัดได้สูงกว่าค่าที่แท้จริง) และในบางสถานการณ์ค่าผิดพลาดเป็นลบ (คะแนนที่วัดได้ต่ำกว่าที่แท้จริง) ถ้าค่าความผิดพลาดมีรูปแบบที่ชัดเจน เช่น สอบคะแนนภาษาอังกฤษมาตรฐานชนิดหนึ่ง แล้วคะแนนของทุกคนสูงเกินจริง และมีคะแนนจุดตัดในการรับเข้าเรียนเท่ากับ 500 คะแนน จะทำให้มีคนผ่านเกณฑ์จำนวนมาก แม้ว่าความสามารถที่แท้จริงของบางคนในนั้นไม่ถึง 500 คะแนนก็ตาม และ 2) ถึงแม้จะมีการวัดหลายครั้งก็ตาม ค่าคะแนนที่แท้จริงต้องไม่เปลี่ยนแปลงไป

ถึงแม้ว่าจะสามารถหาค่าที่แท้จริงของบุคคลได้โดยการหาค่าเฉลี่ยจากการวัดหลายครั้งได้ แต่ในตามความเป็นจริงจะไม่สามารถวัดได้หลายครั้ง ส่วนใหญ่แล้วจะมีโอกาสวัดเพียงครั้งเดียว ดังนั้นต้องมีทดสอบว่า ถ้าเกิดวัดเพียงครั้งเดียวแล้ว ความสามารถที่จะวัดได้ตรงกับค่าที่แท้จริงมีเท่าไร ค่านี้เรียกว่าค่าความคงเส้นคงวา (จริงๆ แล้วในทางจิตวิทยาเรียกว่าค่าความเที่ยง แต่เพื่อให้ง่ายต่อความเข้าใจจึงใช้คำศัพท์นี้)

หลักการหาค่าความคงเส้นคงวาจะทดลองเปลี่ยนรูปแบบของการวัดไป เช่น วัดซ้ำหลายครั้งในคนๆ เดียวกัน วัดด้วยแบบทดสอบหลายๆ ชุดในคนๆ เดียวกัน ประเมินคนๆ เดียวกันด้วยผู้ประเมินหลายคน เป็นต้น จะเห็นว่ามี เวลา แบบทดสอบ และผู้ประเมิน ที่เปลี่ยนแปลงไป แล้วนำคะแนนมาดูว่าเปลี่ยนแปลงไปมากน้อยเพียงใด ถ้าเปลี่ยนแปลงไปไม่มากแสดงว่าความคงเส้นคงวาสูง แต่ถ้าคะแนนเปลี่ยนแปลงไปมากแสดงว่าไม่มีค่าความคงเส้นคงวา

วิธีการหาค่าความคงเส้นคงวา มีอยู่ 3 วิธีหลักด้วยกัน คือ

1) ความคงเส้นคงวาจากการวัดซ้ำ ถ้าวัดด้วยแบบทดสอบชุดเดียวกัน กับคนๆ เดียวกัน แล้ววัดครั้งแรกเสร็จ ทั้งระยะเวลา แล้ววัดครั้งที่ 2 ถ้านำคะแนนของทุกคนในครั้งที่ 1 มาหาค่าสหสัมพันธ์กับครั้งที่ 2 แล้ว จะได้เป็นค่าความคงเส้นคงวาจากการวัดซ้ำ ถ้าลำดับของบุคคลไม่เปลี่ยนแปลง คือ คนที่ได้คะแนนสูงครั้งแรก ก็ยัง

ได้คะแนนสูงในครั้งที่สอง และคนที่ได้คะแนนต่ำในครั้งแรกยังได้คะแนนต่ำในครั้งที่สอง แสดงว่าค่าความคงเส้นคงวาสูง

อย่างไรก็ตาม สิ่งที่เป็นปัญหาจากการหาความคงเส้นคงวาแบบวัดซ้ำก็คือ ผลของการฝึก เช่น จำข้อสอบเดิมได้ ถ้าเป็นแบบทดสอบความสามารถ (เช่น ความสามารถในการแยกประสาท) อาจดีขึ้นในครั้งที่สอง เพราะได้ฝึกจากครั้งแรกแล้ว ถ้าเป็นกรณีแรก คนจำได้ แล้วตอบเหมือนเดิม มีแนวโน้มทำให้ค่าความคงเส้นคงวาสูงขึ้นในกรณีที่สอง คะแนนที่แท้จริงของบุคคลเปลี่ยนแปลงไป ถ้าคะแนนที่แท้จริงในแต่ละคนเปลี่ยนแปลงไปไม่เท่ากัน ทำให้ค่าความเที่ยงลดน้อยลงด้วย ผลจากการฝึกจะมีปัญหาในกรณีที่ทิ้งระยะเวลาสั้น

ถ้าทิ้งระยะเวลามากเกินไป คะแนนที่แท้จริงของบุคคลจะเปลี่ยนแปลงไป (แล้วแต่ลักษณะเป้าหมายที่วัด) ในกรณีที่ เป็นบุคคลิกภาพ เชาวน์ปัญญา ตามทฤษฎีแล้วจะถือว่าค่อนข้างคงทนในแต่ละบุคคล แต่ถ้าทิ้งระยะเวลามากเกินไป เช่น 1 ปี ก็ทำให้คะแนนที่แท้จริงของลักษณะเหล่านี้เปลี่ยนแปลงไปด้วย ด้วยเหตุผลเหล่านี้ ต้องมีการทิ้งระยะเวลาที่เหมาะสม ระยะเวลา 2 สัปดาห์เป็นระยะเวลานิยมมากที่สุด เพราะผู้ตอบมีแนวโน้มลืมข้อสอบแล้ว และการเปลี่ยนแปลงของคะแนนที่แท้จริงยังไม่มาก จึงเหมาะสม แต่ถ้าให้ดีกว่านี้ ควรจะหาค่าความคงเส้นคงวาแบบวัดซ้ำ โดยทิ้งช่วงเวลาหลายรูปแบบ เช่น 2 3 และ 4 สัปดาห์ ดูว่าค่าความคงเส้นคงวาเปลี่ยนแปลงไปอย่างไร

ค่าความคงเส้นคงวารูปแบบนี้ต้องระวังว่าช่วงระหว่างการทำครั้งนั้น บุคคลจะไม่ได้รับผลกระทบที่ทำให้ลักษณะที่สนใจเปลี่ยนแปลงไป เช่น วัดความคงเส้นคงวาในครั้งที่ 1 เสร็จแล้ว มีผู้วิจัยพูดต่อผู้ทดสอบว่า ต้องการคัดคนที่มีความซื่อสัตย์ ด้วยเหตุนี้ทำให้การวัดในครั้งที่ 2 คะแนนของบุคคลจะตอบแบบทดสอบตามความคาดหวังของผู้วิจัย ส่งผลให้คะแนนครั้งที่ 1 และครั้งที่ 2 แตกต่างกัน อีกสิ่งหนึ่งที่ต้องระมัดระวังคือ ลักษณะทางจิตวิทยาบางอย่างเปลี่ยนแปลงได้ง่าย เช่น แรงจูงใจ บางวันผู้ทดสอบอาจต้องการเงินมาก แต่บางวันผู้ทดสอบเพียงพอแล้ว เนื่องจากคะแนนที่แท้จริงของบุคคลเปลี่ยนแปลงไป ทำให้ดูเหมือนว่าความคงเส้นคงวาจากการวัดซ้ำน้อยกว่าที่คาดหมาย แต่ก็ไม่ได้หมายความว่าแบบวัดนั้นไม่มีความคงเส้นคงวา

2) ความคงเส้นคงวาจากแบบทดสอบที่เท่าเทียมกันและจากความสอดคล้องภายใน ความคงเส้นคงวารูปแบบนี้มีหลักการอยู่ว่า ถ้าแบบทดสอบตั้งแต่สองชุดขึ้นไปมีลักษณะที่เท่าเทียมกัน คือ วัดในสิ่งๆ เดียวกัน ได้คะแนนออกมาใกล้เคียงกัน ถ้าวัดคนๆ หนึ่งด้วยแบบทดสอบทั้งสองชุดที่มีลักษณะที่เท่าเทียมกัน คะแนนที่ได้จากทั้งสองชุดต้องมีความสัมพันธ์กันสูง กล่าวคือ คนที่มีคะแนนสูงในแบบทดสอบชุดแรกต้องมีคะแนนสูงในแบบทดสอบชุดที่สอง เช่นเดียวกับคนที่คะแนนต่ำ จะต้องคะแนนต่ำทั้งสองชุด

แต่อย่างไรก็ตาม การสร้างแบบทดสอบที่เท่าเทียมกันเป็นสิ่งที่ยากลำบาก จึงมีแนวคิดขึ้นมาว่าถ้าแบ่งแบบทดสอบที่มีชุดเดี่ยวนั้นเป็นส่วนๆ ที่เท่าเทียมกัน เช่น แบบวัดจำนวน 30 ข้อ นำมาแบ่งเป็น 2 ชุด ชุดละ 15 ข้อ แล้วหาความสัมพันธ์กันแล้ว คะแนนที่ได้จากแต่ละส่วนควรจะสัมพันธ์กันสูง แต่ว่าการแบ่งเป็นส่วนๆ อาจจะไม่ดีแบบทดสอบที่เท่าเทียมกันเสมอไป จึงมีแนวคิดที่ว่าควรจะแบ่งให้ได้เท่ากับจำนวนข้อ เช่น แบบวัดจำนวน 30 ข้อ แบ่งให้เป็น 30 ชุด ชุดละ 1 ข้อ แล้วหาความสัมพันธ์ของแต่ละข้อนั้น วิธีการหาคำนี้จะใช้สูตรสัมประสิทธิ์อัลฟาของครอนบาค ค่าที่ได้นี้จะถือว่าเป็นค่าทดสอบความสอดคล้องภายในของข้อคำถามแต่ละข้อว่าไปในทางเดียวกันหรือไม่ ถ้าข้อคำถามสัมพันธ์กันมาก ค่าสัมประสิทธิ์นี้จะสูง แต่ถ้าแบบทดสอบมีข้อคำถามที่ความสัมพันธ์กันต่ำ หรือมีลักษณะทางจิตวิทยาภายในข้อคำถามมากกว่า 1 อย่าง จะส่งผลให้ค่าสัมประสิทธิ์นี้

3) ค่าความคงเส้นคงวาระหว่างผู้ประเมิน ถ้าให้ผู้ประเมินให้คะแนนลักษณะของบุคคลหนึ่ง เช่น ผลงาน บุคลิกภาพ พฤติกรรม เป็นต้น แล้ว การประเมินจะดีได้ถ้าคะแนนของผู้ประเมินทุกคนใกล้เคียงกัน สัมพันธ์กันสูง วิธีการหาอาจใช้ค่าสัมประสิทธิ์ค้ำปา หรือค่าสหสัมพันธ์แบบเพียร์สันได้ ถ้าค่าที่ได้ออกมาสูง แสดงว่าการประเมินสอดคล้องกัน

แต่อย่างไรก็ดี การให้คะแนนโดยผู้ประเมินจะมีค่าความเที่ยงสูงได้นั้น จะต้องมีภาระบุลักษณะที่ประเมินให้ชัดเจนว่าคืออะไร (มีนิยามที่ชัดเจน คนที่แตกต่างกันเข้าใจตรงกัน) และต้องระบุว่ามีลักษณะแบบใดที่เรียกว่ามาก ลักษณะใดที่เรียกว่าปานกลาง และลักษณะใดที่เรียกว่าน้อย ลักษณะเหล่านี้ต้องเขียนให้เข้าใจตรงกัน ยิ่งเป็นในรูปแบบเชิงพฤติกรรมยิ่งดี เพราะคนที่แตกต่างกัน เห็นตรงกัน ทำให้ประเมินได้เหมือนกัน นอกเหนือจากนี้ การฝึกให้คุ้นเคยกับนิยาม และข้อคำถามส่งผลให้ค่าความเที่ยงประเภทนี้เพิ่มขึ้นด้วย

## 5. การตรวจสอบความถูกต้อง (Validation)

การตรวจสอบความถูกต้องเป็นกระบวนการตรวจสอบว่าแบบทดสอบที่ได้สร้างขึ้นมาวัดในสิ่งที่ต้องการวัดหรือลักษณะเป้าหมายนั่นเอง และไม่ได้ไปวัดในสิ่งที่ไม่ต้องการวัด ในขั้นตอนนี้ต้องใช้ทฤษฎีในการตรวจสอบ กล่าวคือ หลังจากที่ใช้นิยามของลักษณะหนึ่งสร้างแบบทดสอบขึ้นมาแล้ว จะใช้แบบทดสอบนั้นในการตรวจสอบว่าตรงกับทฤษฎีที่นำมาใช้ในการสร้างนิยามหรือไม่นั่นเอง เช่น เมื่อได้นิยามเขาวนปัญญาแล้ว สร้างแบบทดสอบเขาวนปัญญาขึ้นมาแล้ว ก็นำผลคะแนนที่ได้มาทดสอบ อาจจะดูว่ามีองค์ประกอบย่อยตามทฤษฎีหรือไม่ ทฤษฎีอาจจะบอกว่าเขาวนปัญญาทั้งหมด 2 ด้านด้วยกัน คือ เขาวนปัญญาในการใช้หลักเหตุผล และเขาวนปัญญาที่ได้จากประสบการณ์ ผู้สร้างต้องนำแบบทดสอบดังกล่าวมาทดสอบว่ามีองค์ประกอบตามทฤษฎีจริงหรือไม่ หรือว่าตามทฤษฎีเขาวนปัญญาส่งผลให้ผลงานดี ดังนั้นจึงเอาคะแนนจากแบบทดสอบมาหาความสัมพันธ์กับผลงานเพื่อตรวจสอบว่าผลงานตรงตามทฤษฎีหรือไม่ ถึงตรงนี้อาจจะดูเหมือนว่าแบบสอบถามต้องสร้างให้ตรงกับทฤษฎีเท่านั้น แต่ในความเป็นจริงการสร้างแบบสอบถามก็อาจจะเป็นการตรวจสอบทฤษฎีเช่นกัน ถ้าผลออกมาว่าไม่มีความถูกต้อง ไม่ได้หมายความว่าแบบสอบถามที่สร้างขึ้นมามีความถูกต้อง

แบบสอบถามที่มีความคงเส้นคงวามาก แต่พบว่าไม่มีความตรงตามทฤษฎีก็ถือว่าเป็นแบบสอบถามที่ไม่ตรงเป้า วัดได้คงเส้นคงวาก็จริงแต่ไปวัดสิ่งอื่น ไม่ได้วัดเป้าหมายที่ต้องการ ด้วยเหตุนี้นักจิตวิทยาส่วนใหญ่จะถือว่าความตรงมีความสำคัญกว่าความคงเส้นคงวา อย่างไรก็ตามแบบสอบถามจะมีความตรงได้ ก็ต้องมีความเที่ยงเช่นเดียวกัน ยกตัวอย่างเช่น ตามทฤษฎีคะแนนที่แท้จริงของเขาวนปัญญาสัมพันธ์กับผลงาน แต่ว่าแบบสอบถามไม่ได้ฉายภาพของคะแนนที่แท้จริงมากนัก เช่น ค่าความคงเส้นคงวาเท่ากับ .30 กล่าวคือคะแนนของแต่ละบุคคลที่วัดได้เกิดจากผลของคะแนนที่แท้จริงเพียงแค่อ้อยละ 30 เท่านั้น ทำให้เมื่อไปทดสอบความตรงแล้ว ไม่พบความสัมพันธ์ เพราะคะแนนที่แท้จริงเป็นเพียงส่วนน้อย

การตรวจสอบความถูกต้องสามารถแบ่งได้อย่างง่ายเป็น 3 รูปแบบด้วยกัน ดังนี้

1) ความตรงตามเนื้อหา คือ แบบทดสอบได้เป็นตัวแทนของนิยามมากน้อยเพียงใด ความตรงตามเนื้อหาจะคล้ายๆ กับความตรงตามความเห็น แต่จริงๆ แล้วแตกต่างกัน ความตรงตามความเห็นเป็นความตรงที่ดูว่าเนื้อหาในข้อคำถามสอดคล้องกับนิยามหรือไม่ ถ้าเขียนข้อคำถามได้ดีจะถือว่ามีความตรงตามความเห็นแล้ว แต่อย่างไรก็ดี ความตรงตามเนื้อหา หมายถึงเนื้อหาของข้อคำถามที่มี สามารถวัดนิยามได้หรือไม่ ด้วยเหตุนี้ นอกจากจะต้องดูความตรงตามเห็นแล้ว จะต้องดูว่าข้อคำถามเหล่านี้ครอบคลุมนิยามทั้งหมดหรือไม่ ยกตัวอย่างเช่น ข้อสอบไล่ในวิชาคณิตศาสตร์ ควรจะมีเนื้อหาครอบคลุมทุกบทอย่างเท่าเทียมกัน ไม่ได้เน้นเนื้อหาเพียงแค่บทใดบทหนึ่ง คำว่าเท่าเทียมกันอาจจะเปรียบเทียบกับจำนวนเวลาที่สอน ความสำคัญของเนื้อหาในการสอบคัดเข้าอุดมศึกษา หรือ

เปรียบเทียบโดยใช้เกณฑ์อื่นๆ และต้องดูว่าเนื้อหาเหล่านี้ไม่ได้วัดสิ่งอื่นด้วย เช่น ถ้าภาษาที่ใช้ยากเกินไป จะทำให้มีเรื่องความยากของภาษาเจือปนกับข้อสอบคณิตศาสตร์

2) ความตรงเชิงเกณฑ์สัมพันธ์ คือ แบบทดสอบสัมพันธ์กับเกณฑ์ที่ตั้งเอาไว้มาน้อยเพียงใด เช่น ผลงาน ผลการเรียน แบบวัดมาตรฐานอื่น เป็นต้น เกณฑ์ที่ตั้งออกมาได้ต้องสอดคล้องกับวัตถุประสงค์ของการวัดดังกล่าว เช่น ถ้าใช้ทำนายผลงานในอนาคต แบบทดสอบที่ได้ต้องสัมพันธ์กับผลงาน ถ้าใช้ทำนายความซื่อสัตย์ แบบทดสอบที่ได้ต้องสัมพันธ์กับการทุจริตหรือไม่ทุจริตในอนาคต เป็นต้น ความตรงประเภทนี้สามารถแบ่งสองประเภท ประเภทแรกคือความตรงเชิงทำนาย หมายความว่าเมื่อทดสอบแบบทดสอบแล้ว ที่ระยะเวลาไว้แล้ว ไปด้วยว่าคะแนนในแบบทดสอบจะสัมพันธ์กับคะแนนของเกณฑ์ในอนาคตหรือไม่ ยกตัวอย่างเช่น ถ้าทำแบบทดสอบคัดคนเข้าอุดมศึกษา ในวิชาความถนัดทางวิศวกรรม จุดประสงค์ของแบบวัดประเภทนี้คือต้องการทำนายผลการเรียนในอนาคต ดังนั้นถ้าเกิดว่าเมื่อทดสอบความถนัดทางวิศวกรรมของนักเรียน แล้วที่ระยะเวลาไปช่วงหนึ่ง เสร็จแล้วไปดูว่าคนเหล่านี้มีผลการเรียนเป็นอย่างไร มาหาความสัมพันธ์กัน ถ้าแบบทดสอบนี้จะต้องมีความสัมพันธ์กับเกณฑ์ (ผลการเรียน) สูง ประเภทที่สองคือความตรงตามสภาพ ความตรงประเภทนี้จะนำคะแนนจากแบบทดสอบไปหาความสัมพันธ์กับเกณฑ์อื่น โดยไม่ทิ้งระยะเวลา ยกตัวอย่างเช่น บุคลิกภาพแบบมุ่งผลสัมฤทธิ์จะมีแนวโน้มทำให้ผลการทำงานดีตามทฤษฎี แต่ปัจจุบันยังไม่สามารถรอเวลาเพื่อทำความตรงเชิงทำนายได้ ก็อาจจะนำพนักงานในปัจจุบันวัดบุคลิกภาพประเภทนี้ กับผลงานที่เกิดขึ้น ถ้าสัมพันธ์กันสูงแสดงว่ามีความตรงตามสภาพดี (อย่างไรก็ตามถ้าไม่พบว่ามีความตรงรูปแบบนี้ ก็ไม่ได้หมายความว่าแบบสอบถามไม่ตรง อาจจะเป็นไปได้ว่าทฤษฎีไม่ถูกต้อง)

อย่างไรก็ตามในบางบริบท ความตรงเชิงทำนายเป็นสิ่งที่มีความสำคัญมากกว่าความตรงตามสภาพ เช่น แบบทดสอบที่ใช้ในการคัดบุคลากรเข้าทำงาน แบบทดสอบคัดบุคคลเข้าเรียนต่ออุดมศึกษา เป็นต้น เพราะจุดประสงค์ของแบบทดสอบดังกล่าวคือต้องการทำงานผลงานในอนาคต แต่ทว่าความตรงตามสภาพในบางบริบทก็จะมีเหมาะสมมากกว่า เช่น เมื่อนักจิตวิทยาต้องการตรวจสอบแบบทดสอบภาวะความกดดันในงานตามทฤษฎีพบว่าแบบทดสอบนี้ควรมีความสัมพันธ์กับความซึมเศร้า จึงนำแบบทดสอบดังกล่าวไปหาความสัมพันธ์กับแบบทดสอบที่เป็นมาตรฐานเพื่อวัดความซึมเศร้า เป็นต้น

ความตรงตามทำนายและความตรงตามสภาพจะขึ้นอยู่กับเกณฑ์ที่ใช้ เกณฑ์ที่ใช้จะต้องเหมาะสมกับเป้าหมายของการวัด เพราะว่าแบบทดสอบอาจมีความตรงกับเกณฑ์หนึ่ง แต่ไม่พบความตรงกับอีกเกณฑ์หนึ่ง ยกตัวอย่างเช่น ตามทฤษฎีบอกว่าคนที่มุ่งมั่นสู่ผลสำเร็จจะเป็นคนที่มีผลงานดี ดังนั้นแบบทดสอบวัดคนที่มีความบุคลิกภาพแบบมุ่งผลสัมฤทธิ์จะสัมพันธ์กับผลงานในการทำงานเท่านั้น แต่อาจจะไม่สัมพันธ์กับผลงานในการทำงานเป็นทีม หรือความมั่นใจช่วยผู้อื่นทำงาน หรือแบบสอบถามบางประเภทอาจจะไม่สัมพันธ์กับผลงานในช่วง 2 ปีแรก แต่อาจจะสัมพันธ์กับผลงานในระยะยาว ยกตัวอย่างเช่น ความคิดสร้างสรรค์ อาจจะไม่สัมพันธ์กับผลงานช่วง 2 ปีแรก เพราะว่าบุคคลไม่ได้รับโอกาสในการทำงานประเภทนี้ แต่จะมีผลงานที่ดีเมื่อมีโอกาส เป็นต้น

3) ความตรงตามทฤษฎี คือ คะแนนที่ได้จากแบบสอบถามสามารถไปยืนยันทฤษฎีได้ ความตรงประเภทนี้จะมีการตรวจสอบในหลายรูปแบบ เพื่อเป็นการยืนยันว่าแบบสอบถามมีรูปแบบตามทฤษฎีที่ได้กล่าวเอาไว้ แบบสอบถามที่จะมีความตรงตามทฤษฎีได้นั้น จะต้องมีความคงเส้นคงวา และความตรงรูปแบบอื่นๆ ดังที่กล่าวมาแล้วตั้งแต่ต้น เช่น ถ้าแบบสอบถามไม่มีความคงเส้นคงวาแบบความสอดคล้องภายในแล้ว แสดงว่านิยามดังกล่าวอาจจะไม่ได้วัดในสิ่งๆ เดียวกัน ทำให้ไม่ได้เป็นการยืนยันตามทฤษฎี หรือถ้าบุคลิกภาพไม่สามารถ

ทำนายผลงานได้อย่างที่ทฤษฎีกล่าวไว้ ก็อาจจะเป็นไปได้ว่าแบบสอบถามดังกล่าว ไม่มีความตรงตามทฤษฎี เป็นต้น แต่อย่างไรก็ตามกระบวนการนี้เป็นกระบวนการตรวจสอบแบบสอบถามว่ามีความตรงตามทฤษฎีหรือไม่ และเป็นกระบวนการตรวจสอบทฤษฎีตรงกับสภาพความเป็นจริงหรือไม่ด้วย

หลักการพื้นฐานในการตรวจสอบทฤษฎีคือ ตั้งบททวนทฤษฎี (ในที่นี้อาจจะเป็นสิ่งที่ผู้สร้างคิดเอาเองก็ได้) แล้วนำทฤษฎีดังกล่าวมาเขียนเป็นสมมติฐานในการทดสอบ แล้วนำแบบสอบถามไปทดสอบตามสมมติฐานดังกล่าว ยกตัวอย่างเช่น ผู้สร้างคิดว่าบุคคลที่มีผลงานดี จะต้องมีความรู้เรื่องการทำงานในระบอบราชการ ดังนั้นจึงได้สร้างข้อสอบวัดความรู้เรื่องการทำงานในระบอบราชการขึ้น ต้องการทดสอบโดยนำบุคคลที่เคยได้ 2 ชั้นในช่วง 2 ปีที่ผ่านมา มาเปรียบเทียบกับบุคคลที่ไม่ได้ ว่ามีคะแนนแบบสอบถามแตกต่างกันหรือไม่ โดยตั้งสมมติฐานว่าผู้ที่เคยได้ 2 ชั้นจะต้องมีคะแนนแบบทดสอบนี้มากกว่าผู้ที่ไม่เคย แล้วไปทดสอบเพื่อยืนยันทฤษฎีนี้ ถ้าปรากฏว่าผลออกมาไม่ได้เป็นไปตามสมมติฐาน ก็อาจจะสรุปได้ 2 ประเด็นคือ แบบทดสอบไม่ได้มาตรฐาน หรือความคิดของผู้สร้างแบบทดสอบ (ทฤษฎี) ไม่ถูกต้อง หรืออาจจะเป็นไปได้อีกนัยหนึ่งว่าการได้หรือไม่ได้ 2 ชั้น ไม่ได้เป็นตัวแสดงผลงาน

หลักการหาความตรงตามทฤษฎีแบ่งได้เป็น 3 กลุ่มอย่างง่าย ดังนี้

(1) ความตรงตามนิยามของลักษณะทางจิตวิทยาและความตรงตามโครงสร้างภายในลักษณะทางจิตวิทยา ความตรงรูปแบบนี้จะต้องมีการศึกษานิยามของลักษณะที่สนใจ แล้วมองว่าข้อกระทงสามารถเป็นตัวแทนของนิยามทั้งหมดได้หรือไม่ ซึ่งก็คือความตรงเชิงเนื้อหา นั่นเอง จากนั้นต้องดูโครงสร้างภายในของข้อคำถามต่างๆ ว่าสัมพันธ์กับข้อคำถามอื่นๆ อย่างไร ถ้านิยามบอกว่าลักษณะทางจิตวิทยานั้นไม่มีองค์ประกอบย่อย กล่าวคือทุกข้อคำถามวัดในสิ่งๆ เดียวกันแล้ว ค่าความคงเส้นคงวาแบบความสอดคล้องภายในต้องสูง และเมื่อวิเคราะห์องค์ประกอบแล้วพบว่าข้อคำถามเหล่านั้นมีเพียงองค์ประกอบเดียว แต่ถ้านิยามบอกว่ามีหลายองค์ประกอบแล้วแสดงว่าเมื่อวิเคราะห์องค์ประกอบออกมาแล้วจะต้องตรงกับนิยามที่ตั้งเอาไว้ เช่น แนวคิดหนึ่งในเรื่องความฉลาดทางเชาวน์อารมณ์ต้องมีสิ่งองค์ประกอบ คือ รู้จักตนเอง รู้จักคนอื่น ควบคุมตนเองได้ และมีปฏิสัมพันธ์ที่ดีกับผู้อื่น ผู้สร้างแบบทดสอบจะทดสอบองค์ประกอบได้ผ่านการวิเคราะห์องค์ประกอบในข้อคำถามที่ได้สร้างขึ้นตามองค์ประกอบเหล่านั้น

(2) ความตรงลู่เข้าและความตรงจำแนก แนวคิดที่ทำให้เกิดความตรงประเภทนี้คือ ลักษณะทางจิตวิทยาหนึ่งที่ต้องวัดด้วยวิธีการวัดหลายๆ วิธี ลักษณะทางจิตวิทยาก็มีหลายๆ อย่างด้วยกัน ดังนั้นถ้าวิธีการวัดที่ต่างกัน แต่วัดลักษณะทางจิตวิทยาเดียวกัน คะแนนที่ได้จากวิธีการวัดเหล่านั้นจะต้องสูง สิ่งนี้จะเรียกว่าความตรงลู่เข้า แต่ถ้าลักษณะทางจิตวิทยาที่ต่างกันแต่ใช้วิธีการวัดเดียวกัน จะต้องไม่มีความสัมพันธ์ซึ่งกันและกัน สิ่งนี้เรียกว่าความตรงจำแนก สมมติว่าผู้สร้างแบบทดสอบต้องการตรวจสอบคุณภาพของแบบวัดเชาวน์ปัญญาใหม่ ผู้สร้างอาจจะนำแบบวัดเชาวน์ปัญญาใหม่และแบบวัดเชาวน์ปัญญาที่ได้มาตรฐาน ไปวัดกับกลุ่มคนหนึ่ง แล้วหาความสัมพันธ์กัน ถ้าสูงแสดงว่ามีความตรงลู่เข้าสอดคล้องกับแบบทดสอบที่เป็นมาตรฐาน ส่วนความตรงจำแนก จะไม่นิยมตรวจสอบแบบเดี่ยวๆ จะตรวจสอบไปพร้อมๆ กับความตรงแบบลู่เข้า วิธีการตรวจสอบความตรงลู่เข้าและความตรงจำแนกที่เป็นนิยมมากที่สุดคือวิธีลักษณะหลากหลายวิธี วิธีนี้เป็นการตรวจสอบค่าสหสัมพันธ์จากเมทริกซ์สหสัมพันธ์ด้วยหลักการที่กล่าวไปข้างต้น การตรวจสอบความตรงลู่เข้าและความตรงจำแนกจะเป็นการยืนยันว่าแบบทดสอบดังกล่าววัดลักษณะทางจิตวิทยาที่เป็นเป้าหมายจริง คือ แบบสอบถามสามารถวัดได้สอดคล้องกับวิธีวัดอื่นๆ ที่วัดในลักษณะทางจิตวิทยาเดียวกัน และไม่ได้วัดสิ่งอื่น

(3) ความตรงที่มีต่อตัวแปรอื่นๆ วิธีนี้เป็นการตรวจสอบความสัมพันธ์กับตัวแปรอื่นๆ ตามทฤษฎี ก่อนที่จะตรวจสอบความตรงวิธีนี้ได้ นั่น ผู้สร้างจะต้องทบทวนวรรณกรรมถึงความเกี่ยวข้องของลักษณะทางจิตวิทยาที่ใช้ในการสร้างแบบวัด กับตัวแปรอื่นๆ ไม่ว่าตัวแปรนั้นจะเป็นลักษณะทางจิตวิทยาหรือไม่ก็ตาม บางครั้งอาจทำให้อยู่ในรูปแบบของความสัมพันธ์ว่าลักษณะทางจิตวิทยานั้นสัมพันธ์กับตัวแปรอะไร หรือทำให้ อยู่ในรูปแบบของความสัมพันธ์เชิงสาเหตุ ว่าตัวแปรใดที่เป็นเหตุของลักษณะเป้าหมาย และลักษณะเป้าหมาย ก่อให้เกิดผลอะไรบ้าง มีตัวแปรอะไรที่มากเกี่ยวข้องกับความสัมพันธ์เหล่านี้บ้าง เช่น ผู้สร้างแบบทดสอบเขาวนั ปัญหาได้ทบทวนทฤษฎีแล้วพบว่าเขาวนัปัญหาส่งผลต่อผลการเรียน แต่ว่ามีตัวแปรเรื่องฐานะมาเกี่ยวข้อง คนที่มี ฐานะปานกลางขึ้นไป ความสัมพันธ์เชิงสาเหตุของตัวแปรทั้งสองสูง แต่คนที่มีฐานะทางเศรษฐกิจต่ำ ความสัมพันธ์เชิงสาเหตุของตัวแปรทั้งสองไม่พบมากนัก จากสิ่งเหล่านี้ทำให้ผู้สร้างตั้งสมมติฐานแล้วไปทดสอบ ถ้า คะแนนที่ได้จากแบบทดสอบมีความสัมพันธ์เช่นนี้ แสดงว่าแบบทดสอบมีความตรง วิธีการทดสอบความตรงใน รูปแบบนี้อาจแบ่งได้เป็น 2 วิธีด้วยกัน ก) วิธีการทดลองจะเป็นการทดสอบว่าการสร้างให้เกิดตัวแปรใดตัวแปร หนึ่งเกิดขึ้น โดยควบคุมตัวแปรอื่นไว้ จะส่งผลต่อคะแนนในแบบวัดหรือไม่ ยกตัวอย่างเช่น แบบสอบถามเรื่อง การทำงานเป็นทีม ผู้สร้างจึงหาผู้ร่วมการทดลองมาจำนวน 20 คน แบ่งเป็นกลุ่มละ 10 คน โดยสุ่ม กลุ่มหนึ่งจะ ให้ฝึกการทำงานเป็นทีม โดยวิธีการนี้เป็นวิธีได้รับการยอมรับว่าทำให้เกิดการทำงานเป็นทีมขึ้น ส่วนอีกกลุ่มหนึ่ง ปล่อยให้ไปตามธรรมชาติ เมื่อผ่านการฝึกแล้วเปรียบเทียบกันว่าคะแนนการทำงานเป็นทีมของคนในสองกลุ่ม แตกต่างกันหรือไม่ หรืออีกนัยหนึ่งคือทำให้กลุ่ม 2 กลุ่มแตกต่างกันคะแนนที่ได้จากแบบวัด แล้วตรวจสอบว่า เกิดผลตามทฤษฎีหรือไม่ เช่น ให้กลุ่มคนที่ทำงานเป็นทีม และกลุ่มคนที่ไม่ได้ทำงานเป็นทีมไปทำงานหนึ่งที่ต้อง อาศัยทักษะนี้ แล้วเปรียบเทียบผลงานของทั้งสองกลุ่ม ข) วิธีการสำรวจ คือ ไปสำรวจลักษณะทางจิตวิทยาด้วย แบบสอบถามที่สร้างขึ้นกับตัวแปรอื่นๆ พร้อมกัน อาจจะเป็นตัวแปรทางจิตวิทยา (ต้องวัดด้วยแบบทดสอบที่เป็น มาตรฐานแล้ว) หรือลักษณะของบุคคลในโลกแห่งความเป็นจริง เช่น เปรียบเทียบพนักงานที่ผลงานสูง และ ผลงานต่ำ ว่ามีคะแนนในแบบสอบถามเรื่องการทำงานเป็นทีมต่างกันหรือไม่ หรือนำแบบสอบถามเรื่อง การ ทำงานเป็นทีมไปวัดพร้อมกับแบบสอบถามเรื่องทัศนคติต่อการทำงานคนเดียว ซึ่งต้องตรวจสอบแล้วพบว่า มี ความสัมพันธ์ทางลบ จะเห็นว่าวิธีการนี้ไม่ได้เป็นการยืนยันความสัมพันธ์เชิงสาเหตุที่ชัดเจนเช่นเดียวกับวิธีการ ทดลอง

## 6. การหาค่าเปรียบเทียบ (Norms)

อย่างที่กล่าวไปข้างต้นว่าคะแนนดิบด้วยตัวของมันเอง ไม่ได้มีความหมายในการตัดสินใจ สร้างค่า เปรียบเทียบขึ้น เพื่อให้รู้ว่าคะแนนดิบแต่ละค่ามีความหมายอย่างไรบ้าง

ก่อนอื่นต้องพิจารณาก่อนว่าแบบสอบถามของผู้สร้างใช้ในรูปแบบใด บางครั้งผู้สร้างอาจจะกำหนดว่า ถ้าผู้ทำแบบสอบถามไม่มีค่าความสามารถถึงระดับที่กำหนดไว้ ก็จะจัดให้อยู่กลุ่มหนึ่ง เช่น พวกสอบตก แต่ถ้าถึง ระดับเกณฑ์ที่กำหนดไว้ก็จัดให้อยู่อีกกลุ่มหนึ่ง เช่น สอบผ่าน การตัดสินใจแบบนี้เรียกว่าการตัดสินใจแบบจุดตัด ซึ่งจะใช้กับแบบทดสอบประเภทวัดความสามารถ แต่ถ้าผู้สร้างอาจจะต้องการดูว่าผู้ตอบแบบสอบถามแต่ละคน เมื่อเปรียบเทียบกับคนอื่นแล้ว มีลักษณะทางจิตวิทยามากน้อยอย่างไร เช่น บางคนมีบุคลิกภาพแบบกล้าแสดงออก มากกว่าคนอื่น แต่บางคนมีน้อยกว่าคนอื่น การตัดสินใจแบบนี้จะเรียกว่าการตัดสินใจแบบเปรียบเทียบ การตัดสินใจ ทั้งสองรูปแบบจะมีรายละเอียดในการหาค่าเปรียบเทียบที่แตกต่างกัน แต่หลักการเหมือนกันก็คือว่าพยายามทำให้ คะแนนดิบมีความหมายนั่นเอง

การตัดสินใจแบบจุดตัด สิ่งแรกที่ผู้สร้างต้องทำคือการนิยามจุดตัดที่ตนเองได้ตั้งเอาไว้ ยกตัวอย่างเช่น จะตัดให้นักเรียนตกเมื่อนักเรียนไม่เขียนอธิบายลักษณะของตนเอง สิ่งที่เป็นภาษาอังกฤษได้ ด้วยเหตุนี้จึงอาจจะออกข้อสอบเพื่อวัดความสามารถของนักเรียนขึ้นมา แต่อย่างไรก็ตามข้อสอบบางชุดก็ยาก ข้อสอบบางชุดก็ง่ายแตกต่างกันไป จึงต้องหาจุดตัดในข้อสอบชุดที่ง่ายและชุดที่ยากว่าจุดใดแสดงว่านักเรียนพอเขียนภาษาอังกฤษอย่างง่ายได้ ข้อสอบชุดยากอาจจะตัดที่ 30 คะแนน และข้อสอบชุดง่ายอาจจะตัดที่ 70 คะแนน ถ้าคะแนนเต็มทั้งสองชุดคือ 100 คะแนน เป็นต้น รูปแบบนี้อาจจะพบในกรณีอื่นที่ไม่ใช่การวัดความสามารถ เช่น ผู้คัดเลือกคนเข้าทำงาน อาจจะไม่ต้องการคนที่ขาดแรงจูงใจในการทำงาน ดังนั้นเมื่อสร้างแบบทดสอบวัดแรงจูงใจในการทำงานขึ้นมาแล้ว ต้องนิยามให้ได้ว่าแรงจูงใจขั้นใดที่ยอมรับได้ ขั้นใดที่เรียกว่าดี ขั้นใดที่เรียกว่าแย่ แล้วไปเปรียบเทียบว่าจุดตัดที่นิยามไปเทียบกับคะแนนเท่าไร ความหมายที่ให้กับคะแนนนั้นจะเรียกว่าค่าเปรียบเทียบ การเปรียบเทียบรูปแบบนี้ก็อาศัยการเก็บข้อมูลจากเกณฑ์อื่นๆ เข้ามาช่วยในการนิยาม ยกตัวอย่างเช่น ต้องการสร้างแบบทดสอบความเหมาะสมในงานธุรการ ก็ต้องใช้ผลงานในการเปรียบเทียบว่าผลงานดังกล่าวจะอยู่ในดัชนีขั้นใด แล้วมาใช้เป็นค่าเปรียบเทียบ

ส่วนการตัดสินใจแบบเปรียบเทียบนั้น จะต้องอาศัยเก็บกลุ่มตัวอย่างที่สนใจเป็นจำนวนมาก เพื่อหาว่าคะแนนของบุคคลที่เก็บมามีค่าสูงสุด ต่ำสุดอย่างไร เฉลี่ยแล้วมีค่าเท่าไร และจะทำให้ทราบตำแหน่งของบุคคลแต่ละคนเป็นอย่างไรบ้าง เช่น หากคะแนนของแบบทดสอบความอดทน พบว่าคนที่อดทนที่สุดมีคะแนนเท่ากับ 120 คะแนน คนที่อดทนต่ำที่สุดมีคะแนนเท่ากับ 70 คะแนน เฉลี่ยแล้วเท่ากับ 95 คะแนน พอเมื่อมีผู้มาทดสอบความอดทน สมมติว่าได้คะแนนเท่ากับ 97 คะแนน ก็จะทำให้ทราบได้ว่าเป็นผู้ที่มีความอดทนอยู่ประมาณกลางๆ เป็นต้น เทคนิคการเปรียบเทียบอาจใช้คะแนนเปอร์เซ็นต์ไทล์ คะแนนมาตรฐานช่วยได้

## 7. การทำให้เป็นมาตรฐาน (Standardization)

บางครั้งอาจจะได้ยินว่า “วันนี้อากาศร้อน เลขทำข้อสอบได้ไม่ดี” “แอร์หนาวไป คิดอะไรไม่ออก” “โชคดีคนสอบผลอนะเนี่ย เลขลอกได้” จะเห็นว่าสิ่งเหล่านี้ทำให้คะแนนเปลี่ยนแปลงไปได้ โดยไม่ได้เกี่ยวข้องกับแบบทดสอบเลย แต่เกี่ยวข้องกับกระบวนการทดสอบซึ่งส่งผลให้คะแนนที่สังเกตได้ ไม่ได้แสดงให้เห็นถึงคะแนนที่แท้จริงของบุคคลนั้น ด้วยเหตุนี้การจัดการสอบให้เป็นมาตรฐานจึงเป็นเรื่องที่สำคัญ

อากาศ การควบคุมการลอกข้อสอบ แสดงให้เห็นแล้วในตัวอย่างข้างต้น แต่ก็มีสิ่งอื่นอีกมากมายที่ต้องควบคุม ยกตัวอย่างเช่น เวลาในการสอบอาจจะส่งผลต่อความสามารถในการตอบข้อสอบ ข้อสอบชุดเก่าที่ใช้ทดสอบหลายๆ ครั้ง อาจจะมีปัญหาเรื่องข้อสอบหลุดออกไปจากการสอบ เป็นต้น สิ่งเหล่านี้เกี่ยวข้องกับกระบวนการในการสอบที่ผู้สร้างแบบทดสอบต้องคิดให้รอบคอบ แบบทดสอบบางประเภทก็มีลักษณะที่ต้องระวังเป็นลักษณะเฉพาะ เช่น ข้อสอบการฟัง จะต้องควบคุมเสียงให้ชัดเจน และดังเท่าเทียมกันทุกคน หรือการสอบแบบสัมภาษณ์ การประเมิน จะมีความผิดพลาดที่เกิดจากผู้ทดสอบที่เรียกว่าผลจากความคาดหวัง ถ้าผู้ทดสอบคาดหวังอย่างไร จะทำให้ถามคำถามที่ทำให้ผู้ตอบตอบตรงกับความต้องการของตน หรือสังเกตพฤติกรรมเฉพาะที่ตรงกับความต้องการของตน ส่งผลให้คะแนนที่ประเมินมีอคติในทางที่ดี หรืออคติในทางที่ไม่ดี เนื่องจากปัญหาในเรื่องเหล่านี้เป็นสิ่งที่ส่งผลกระทบต่อมาก ทำให้ก่อนที่ผู้สร้างจะเลือกวิธีการใดในการวัด (เช่น การสังเกตพฤติกรรม) ผู้สร้างจะต้องค้นคว้าถึงอคติที่อาจจะเกิดขึ้นในการวัด และวิธีการควบคุมอคติเหล่านั้น มิให้คะแนนที่ออกมาไม่ถูกต้อง

โดยสรุปแล้ววิธีการสร้างการวัดทางจิตวิทยาเป็นวิธีการที่ทำให้การวัดเหล่านั้นเชื่อถือได้ สามารถวัดในสิ่งที่ผู้สร้างต้องการได้ แต่ทว่าในโลกนี้ไม่มีการวัดใดทางจิตวิทยาที่ชัดเจน เชื่อถือทั้งหมด แม้กระทั่งแบบทดสอบเชาวน์ปัญญาที่เป็นมาตรฐานที่สุด ยังไม่สามารถมีค่าความเที่ยงที่สมบูรณ์ได้ ด้วยเหตุนี้ผู้สร้างจะต้องตระหนักในเรื่องเหล่านี้ โดยเฉพาะอย่างยิ่งการนำผลของการวัดเหล่านี้ไปตัดสินคน เช่น ไปตัดสินคนเข้าเรียนต่ออุดมศึกษา ไปตัดสินคนเข้าทำงาน การตัดสินด้วยเครื่องมือเหล่านี้ย่อมมีข้อผิดพลาด ดังนั้นผู้สร้างต้องไม่เชื่อเครื่องมือเพียงแค่เครื่องมือเดียว และผู้สร้างต้องไม่นำแบบวัดที่ไม่ได้มาตรฐานเข้าไปใช้ในการตัดสินคน การตัดสินผิดถือว่าเป็นการชี้นำชีวิตของบุคคลเหล่านั้นได้ ด้วยเหตุผลเหล่านี้จึงจะสอดคล้องกับหลักการแปรของจรรยาบรรณในการประเมินทางจิตวิทยาของสมาคมจิตวิทยาแห่งอเมริกาปี 2545 แทรกลงไปในตอนท้ายนี้

9 การประเมิน

9.01 พื้นฐานสำหรับการประเมิน

- (ก) นักจิตวิทยาทำให้ข้อคิดเห็นที่อยู่ในการแนะนำ รายงาน และข้อความวินิจฉัยหรือประเมิน รวมถึงการพิสูจน์ในศาล มีพื้นฐานจากข้อมูลหรือเทคนิคที่เพียงพอในการพิสูจน์ข้อค้นพบของพวกเขา (คู่มือมาตรฐาน 2.04 พื้นฐานการตัดสินใจในทางวิทยาศาสตร์และวิชาชีพ ด้วย)
- (ข) ยกเว้นที่ระบุไว้ใน 9.01ค นักจิตวิทยาให้ข้อคิดเห็นเกี่ยวกับลักษณะทางจิตวิทยาของแต่ละบุคคล เฉพาะหลังจากที่พวกเขาได้ทำการทดสอบแก่บุคคลเพียงพอที่จะสนับสนุนคำกล่าวหรือข้อสรุป ถ้าใช้ความพยายามอย่างเหมาะสมแล้ว การทดสอบไม่สามารถนำไปใช้ได้จริง นักจิตวิทยาระบุถึงความพยายามที่พวกเขาได้ทำ และผลของความพยายามเหล่านั้นแล้วทำให้กระจ่างชัดถึงผลกระทบของข้อมูลที่จำกัดต่อความเที่ยงและความตรงของข้อคิดเห็นของพวกเขา และจำกัดลักษณะหรือขอบเขตของข้อสรุปและคำแนะนำให้เหมาะสม (คู่มือมาตรฐาน 2.01 ขอบเขตของความสามารถ และ 9.06 การแปรความหมายของการประเมิน ด้วย)
- (ค) เมื่อนักจิตวิทยา ทำรายงานทบทวนการบันทึกจากการประเมิน หรือการให้คำปรึกษา หรือการดูแล และการทดสอบของแต่ละบุคคลไม่ได้จำเป็นสำหรับการให้ข้อคิดเห็น นักจิตวิทยาอธิบายถึงสิ่งเหล่านี้และแหล่งของข้อมูลที่พวกเขาใช้ในการสรุปหรือแนะนำ

9.02 การใช้การประเมิน

- (ก) นักจิตวิทยาจัดการ ปรับปรุง ให้คะแนน ตีความ หรือใช้เทคนิคการประเมิน การสัมภาษณ์ การทดสอบ หรือเครื่องมือแต่ละวิธี เพื่อจุดประสงค์ที่เหมาะสมในแง่ของการวิจัย หรือให้มีหลักฐานของประโยชน์และความเหมาะสมในการนำเทคนิคไปใช้
- (ข) นักจิตวิทยาใช้เครื่องมือประเมิน ที่ความตรงและความเที่ยงได้ระบุเอาไว้ ในการใช้สำหรับสมาชิกในประชากรที่ได้รับการทดสอบ เมื่อความตรงหรือความเที่ยงของเครื่องมือไม่ได้ระบุเอาไว้ นักจิตวิทยาอธิบายถึงจุดเด่นและข้อจำกัดของผลและการตีความของการทดสอบ
- (ค) นักจิตวิทยาใช้วิธีการประเมินให้เหมาะสมกับความชอบในการใช้ภาษา และความสามารถของแต่ละบุคคล ถ้าการใช้ภาษาอื่นไม่ได้เกี่ยวข้องกับวัตถุประสงค์ในการประเมินมากกว่า

9.03 การแจ้งให้แสดงคำยินยอมในการประเมิน

- (ก) นักจิตวิทยาได้รับการแสดงคำยินยอมสำหรับการประเมิน หรือการให้บริการวินิจฉัย จากที่ได้อธิบายไว้ในมาตรฐาน 3.10 การแจ้งให้แสดงคำยินยอม ยกเว้นเมื่อ (1) การทดสอบเป็นคำสั่งในทางกฎหมายหรือข้อบังคับของรัฐ (2) การแจ้งให้แสดงคำยินยอมได้ถูกแสดงเป็นนัยแล้ว เพราะการทดสอบถูกทำเป็นกิจวัตรในกิจกรรมการศึกษา ในสถาบัน หรือในองค์กร (เช่น เมื่อผู้ร่วมการทดลองยอมรับอย่างเต็มใจในการประเมินเมื่อสมัครเข้าทำงาน) หรือ (3) วัตถุประสงค์หนึ่งของทดสอบคือประเมินความสามารถในการตัดสินใจ การแจ้งให้แสดงคำยินยอมรวมถึงการอธิบายถึงลักษณะและวัตถุประสงค์ของการประเมิน ค่าธรรมเนียม ความเกี่ยวข้องกับกลุ่มที่สาม และขอบเขตของความลับ และควรให้โอกาสที่เพียงพอสำหรับการถามคำถามและได้รับคำตอบของลูกค้ำหรือผู้ป่วย

- (ข) นักจิตวิทยาแจ้งบุคคลที่มีความสามารถในการแสดงคำยินยอมอย่างน่าสงสัย ที่แสดงคำยินยอมหรือสำหรับผู้ทำการทดสอบเป็นคำสั่งจากกฎหมายหรือข้อบังคับของรัฐ เกี่ยวกับลักษณะและจุดมุ่งหมายของบริการการประเมินที่ได้เสนอใช้ภาษาที่ผู้ที่ได้รับการประเมินสามารถเข้าใจได้
  - (ค) นักจิตวิทยาที่ใช้บริการจากผู้ที่มีความการประเมิน ได้รับการแสดงคำยินยอมจากลูกค้าหรือผู้ป่วยที่จะใช้ผู้ที่มีความรู้รับประกันความลับของผลการทดสอบและรักษาความปลอดภัยของการวัด และรวมถึงในข้อเสนอแนะ การรายงาน และข้อความวินิจฉัยหรือประเมิน รวมถึงพิสูจน์ในศาล อภิราชเรื่องข้อจำกัดทั้งหมดของข้อมูลที่ได้รับ (คู่มือมาตรฐาน 2.05 การให้ผู้อื่นทำหน้าที่แทน, 4.01 การรับรองไว้ซึ่งความลับ, 9.01 พื้นฐานการประเมิน, 9.06 การแปรความหมายของการประเมิน และ 9.07 การประเมิน โดยผู้ที่ไม่มีคุณภาพ ด้วย)
- 9.04 การเปิดเผยข้อมูลในการทดสอบ
- (ก) ในคำศัพท์ ข้อมูลการทดสอบ หมายถึง คะแนนดิบและคะแนนที่อยู่ในมาตรวัดแบบต่อเนื่อง การตอบสนองของลูกค้าหรือผู้ป่วยต่อคำถามหรือสิ่งเร้าของแบบทดสอบ และการจับบันทึกหรือการบันทึกของนักจิตวิทยาที่เกี่ยวข้องกับคำถามและพฤติกรรมของลูกค้าหรือผู้ป่วยระหว่างการทดสอบ ส่วนของวัสดุการทดสอบที่เป็นการตอบสนองของลูกค้าหรือผู้ป่วยจัดให้รวมอยู่ในคำนิยามของข้อมูลการทดสอบด้วย จากการมอบให้ลูกค้าหรือผู้ป่วย นักจิตวิทยาให้ข้อมูลการทดสอบแก่ลูกค้าหรือผู้ป่วยหรือบุคคลอื่นที่ระบุในการมอบ นักจิตวิทยาอาจจะหยุดการมอบข้อมูลการทดสอบเพื่อที่จะปกป้องลูกค้า ผู้ป่วย หรือบุคคลอื่นจากอันตรายที่ชัดเจน หรือการนำไปใช้ในทางที่ผิด หรือการตีความข้อมูลหรือการทดสอบผิด จดจำไว้ว่ามีหลายกรณี ที่ต้องมอบข้อมูลที่เป็นความลับในสถานการณ์ที่ถูกบังคับโดยกฎหมาย (คู่มือมาตรฐาน 9.11 การรักษาความลับของแบบทดสอบ ด้วย)
  - (ข) เมื่อไม่มีการมอบให้แก่ลูกค้าหรือผู้ป่วย นักจิตวิทยาให้ข้อมูลการทดสอบเฉพาะเมื่อเป็นความต้องการของกฎหมายหรือคำสั่งจากศาล
- 9.05 การสร้างแบบทดสอบ
- นักจิตวิทยาผู้พัฒนาแบบวัดหรือเทคนิคการประเมินอื่น ใช้กระบวนการในการวัดทางจิตวิทยาที่เหมาะสมและเป็นความรู้ทางวิทยาศาสตร์หรือวิชาชีพในปัจจุบัน สำหรับการออกแบบเครื่องมือ การทำให้เป็นมาตรฐาน การทำให้ถูกต้อง การลดหรือการขจัดอคติ รวมถึงข้อเสนอแนะสำหรับการใช้
- 9.06 การแปรความหมายของการประเมิน
- เมื่อแปรความหมายผลการประเมิน รวมถึงการแปรความโดยอัตโนมัติ นักจิตวิทยาใส่ใจถึงวัตถุประสงค์ของการประเมิน เช่นเดียวกับปัจจัยในการทดสอบอันหลากหลาย ความสามารถในการทำแบบทดสอบ และลักษณะอื่นๆ ของบุคคลที่ถูกประเมิน เช่น ความแตกต่างด้านสถานการณ์ แต่ละบุคคล ภาษา และวัฒนธรรม ที่อาจจะส่งผลกระทบต่อตัดสินของนักจิตวิทยาหรือลดความถูกต้องของการแปรความหมาย พวกเขาชี้ให้เห็นถึงข้อจำกัดที่สำคัญในการแปรความหมายของพวกเขา (คู่มือมาตรฐาน 2.01 ข และ ค ขอบเขตของความสามารถ และ 3.01 การแบ่งแยกอย่างไม่ยุติธรรม ด้วย)
- 9.07 การประเมินโดยผู้ที่ไม่มีคุณภาพ
- นักจิตวิทยาไม่ส่งเสริมการใช้เทคนิคการประเมินทางจิตวิทยาที่ทำโดยผู้ที่ไม่มีคุณภาพ ยกเว้นแต่เมื่อการใช้เหล่านั้นถูกกระทำเพื่อจุดประสงค์ในการฝึกฝน โดยมีการดูแลอย่างเหมาะสม (คู่มือมาตรฐาน 2.05 การให้ผู้อื่นทำหน้าที่แทน ด้วย)
- 9.08 แบบทดสอบและผลของแบบทดสอบที่ลำสมัย
- (ก) นักจิตวิทยาไม่นำการประเมินหรือการตัดสินใจแทรกแซง หรือคำแนะนำสำหรับข้อมูล หรือผลการทดสอบ ที่ลำสมัยสำหรับวัตถุประสงค์บางอย่าง มายึดถือเป็นหลัก
  - (ข) นักจิตวิทยาไม่ยึดถือการตัดสินใจหรือข้อเสนอแนะจากการวัดและการทดสอบที่ลำสมัยและไม่เป็นประโยชน์สำหรับวัตถุประสงค์บางอย่าง
- 9.09 การให้คะแนนทดสอบและการให้บริการแปรความหมาย
- (ก) นักจิตวิทยาผู้เสนอการให้บริการประเมินหรือให้คะแนนกับนักวิชาชีพคนอื่น อธิบายวัตถุประสงค์ ค่าปกติ ความตรง ความเที่ยง และการนำกระบวนการดังกล่าวไปประยุกต์ใช้ หรือคุณสมบัติพิเศษใดๆ ที่สามารถนำไปประยุกต์ใช้ได้อย่างถูกต้อง

- (ข) นักจิตวิทยาเลือกการบริการให้คะแนนหรือแปลความหมาย (รวมถึงบริการโดยอัตโนมัติ) บนพื้นฐานของหลักฐาน ความตรงของโปรแกรมและกระบวนการ เช่นเดียวกับการพิจารณาอย่างเหมาะสมในด้านอื่นๆ (ดูมาตรฐาน 2.01ข และ ก ขอบเขตของความสามารถ ด้วย)
  - (ค) นักจิตวิทยารักษาความรับผิดชอบสำหรับการประยุกต์ใช้ การแปลความหมาย การใช้เครื่องมือ การประเมิน อย่างเหมาะสม ไม่ว่าจะพวกเขาจะให้คะแนนและตีความการทดสอบดังกล่าวด้วยตนเอง หรือใช้บริการจากผู้อื่นหรือเครื่องมืออัตโนมัติ
- 9.10 การอธิบายผลการประเมิน
- ไม่คำนึงว่าการให้คะแนนและการตีความจะทำโดยนักจิตวิทยา ลูกจ้าง ผู้ช่วย เครื่องอัตโนมัติ หรือคนอื่นที่อยู่ในบริการ นักจิตวิทยาหาทางที่เหมาะสมที่จะรับประกันการอธิบายผลการทดสอบที่ให้แต่ละบุคคล และเป็นตัวแทนของผู้ให้การทดสอบอย่างแน่ชัด ถ้าลักษณะของความสัมพันธ์ไม่กีดกันการให้คำอธิบายผลการทดสอบ (เช่น การให้คำปรึกษาแก่องค์กร บางองค์กร ลักษณะก่อนการว่าจ้าง หรือการคัดคนเพื่อความปลอดภัย หรือการประเมินทางกฎหมาย) ข้อเท็จจริงควรได้รับการอธิบายให้กระจ่างแก่บุคคลที่ได้รับการประเมินตั้งแต่แรกเริ่ม
- 9.11 การรักษาความลับของแบบทดสอบ
- ในคำศัพท์ วัสดุการทดสอบ หมายถึง คู่มือ เครื่องมือ วิธีการประดิษฐ์ปฏิบัติ และคำถามหรือสิ่งเร้าในการทดสอบ และไม่รวมถึงข้อมูลการทดสอบ ที่ได้นิยามไปแล้วในมาตรฐาน 9.04 การเปิดเผยข้อมูลในการทดสอบ นักจิตวิทยาใช้ความพยายามอย่างเหมาะสมที่จะรักษาความสมบูรณ์และความลับของวัสดุการทดสอบ และเทคนิคการประเมินอื่นๆ ให้สอดคล้องกับกฎหมายและหน้าที่ในสัญญา และในวิธีที่ปฏิบัติตามมาตรฐานจรรยาบรรณ

แปลจาก American Psychological Association. (2002). *The ethical principles of psychologists and code of conduct*. Available at [www.apa.org/ethics/code2002.html](http://www.apa.org/ethics/code2002.html)

ถึงแม้ว่าการละเมิดจรรยาบรรณเหล่านี้ในประเทศไทยจะไม่มีใครมาตรวจสอบ หรือจะไม่บดทลงโทษอันใด แต่ว่าการปฏิบัติตนให้สอดคล้องกับจรรยาบรรณเหล่านี้เป็นสิ่งจำเป็น เพื่อรักษามาตรฐานวิชาชีพของนักจิตวิทยาในภาพรวม รักษาชื่อเสียงของผู้สร้าง และเพื่อประโยชน์ของผู้ใช้บริการในการประเมินและการทดสอบทางจิตวิทยา