

หลักการวัดทางสังคมศาสตร์

ขั้นตอนที่สำคัญที่สุดอันหนึ่งในการวิจัยทางสังคมศาสตร์ คือ การวัด

ซึ่งหมายถึง กระบวนการแปรสภาพแนวคิด (Concept) ที่มีลักษณะนามธรรมให้เป็นข้อมูลทางสถิติเชิงปริมาณ หรือเชิงคุณภาพ (Qualitative or Quantitative Data) ซึ่งนักวิจัยกำหนดไว้ใช้เป็นแนวในการสำรวจลักษณะต่างๆ ของตัวแปร ผู้ที่ทำการวิจัยจะต้องอาศัยความรู้ในทางทฤษฎี และหลักเกณฑ์ในการวัดมาประกอบกัน เพื่อให้ได้เครื่องมือที่จะใช้วัดสิ่งที่ต้องการจะวัดให้มีความถูกต้อง และเชื่อถือได้ที่สุด

ปัญหาในการวัดจะยุ่งยากมากขึ้นถ้าตัวแปรมีมาก และในทางสังคมศาสตร์ ตัวแปรส่วนใหญ่จะไม่สามารถวัดได้โดยตรง ซึ่งผิดจากด้านวิทยาศาสตร์ธรรมชาติ ดังนั้น ผู้ที่ทำการวิจัยทางสังคมศาสตร์ นอกจากจะต้องมีความรู้เรื่องระเบียบวิธีการวิจัยแล้ว จะต้องมีความรู้เกี่ยวกับทฤษฎีและแนวคิดในเรื่องที่เกี่ยวข้องกับงานวิจัยของตนด้วย เพื่อช่วยในการวัด

ประโยชน์ของการวัดจะมากน้อยเพียงใด เป็นไปตามทฤษฎีหรือไม่นั้น ย่อมขึ้นอยู่กับกระบวนการทั้งหมดที่ใช้ในการวัด หากมีกระบวนการที่ดีถูกต้องตามหลักการ ก็จะทำให้มาตรวัดที่ดี แต่อย่างไรก็ตาม คุณภาพของการวัดที่ดีนั้น จะต้องประกอบด้วย 4 ตัวบ่งชี้ด้วยกันคือ 1) ความแม่นยำของการวัด (Validity) 2) ความเชื่อถือได้ของการวัด (Reliability) 3) ความไวของการวัด (Sensitivity) และ 4) การมีความหมายของการวัด (Meaningfulness) ซึ่งสามารถสรุปได้ดังนี้

ความแม่นยำของการวัด (Validity) ก็คือ สามารถวัดในสิ่งที่ต้องการวัด หรือวัดได้ตรงประเด็นที่ต้องการ เช่น

ถ้าจะวัดทัศนคติที่มีต่อพระพุทธศาสนา ก็ต้องเลือกคนที่นับถือศาสนาพุทธ เพราะถ้าไปวัดกับคนที่นับถือศาสนาอื่น คนๆ นั้นย่อมไม่รู้เรื่องราวเกี่ยวกับพระพุทธศาสนา การวัดนั้นก็จะได้ผลที่ยังตรงและแน่นอน เป็นต้น

ความเชื่อถือได้ของการวัด (Reliability) หมายความว่าเมื่อได้ทำการวัดแล้ว ผลที่ได้จากการวัดมีความสอดคล้องหรือเหมือนกันทุกครั้งหรือไม่ หากผลที่ได้จากการวัดเหมือนกันหรือสอดคล้องกัน ก็เรียกว่า มีความเชื่อถือได้ หากไม่เหมือนกันหรือไม่สอดคล้องกันก็เชื่อถือไม่ได้ เช่น

ไม้บรรทัด ถือว่าเป็นเครื่องมือที่ใช้วัด ที่มีประสิทธิภาพอย่างหนึ่ง ถ้าใช้วัดโต๊ะตัวหนึ่ง ยาว 30 นิ้ว ต่อมาอีกวันหนึ่งเอามาวัดอีก ก็จะได้ 30 นิ้วเท่าเดิม อย่างนี้ถือว่าเป็นเครื่องมือวัดที่มีความแน่นอน ตรงกันข้ามกับในการวัดทัศนคติที่คนมีต่อสิ่งใดสิ่งหนึ่ง ถ้าหากวัดครั้งแรก ผู้ตอบ ตอบว่า "พอใจ" ต่อมาเมื่อวัดอีกครั้งหนึ่ง โดยระยะเวลาและสถานการณ์ต่างๆ ใกล้เคียงกับการวัดครั้งแรก แต่ผู้ตอบคนเดียวกันกลับตอบว่า "ไม่พอใจ" ถ้าเป็นเช่นนั้น ถือว่าเครื่องมือวัดนั้นไม่แน่นอนใช้ไม่ได้

ความไวของการวัด (Sensitivity) หมายถึงความสามารถของเครื่องวัดในการจำแนกความแตกต่างระหว่างหน่วยต่าง ๆ ที่ต้องการจะศึกษาและวิจัย วิธีการวัดบางวิธีจำแนกบุคคลได้อย่างหยาบ ไม่สามารถจะวัดรายละเอียดของความแตกต่างของหน่วยที่ต้องการศึกษาวิจัยได้ วิธีการวัดนั้นก็ถือว่าขาดความไว

และการมีความหมายของการวัด (Meaningfulness) บ่อยครั้งในการวิจัยทางสังคมศาสตร์ ทั้ง ๆ ที่วิธีการที่ใช้วัดมีความถูกต้อง เชื่อถือได้ และมีความไวในการวัด แต่ให้ผลการวัดที่ไม่มีความหมาย เช่น

งานวิจัยเรื่องหนึ่งศึกษาการประกอบอาชีพของชาวชนบท แต่แบ่งอาชีพของชาวชนบทโดยยึดเฉพาะอาชีพใดอาชีพหนึ่ง การแบ่งเช่นนี้ถือว่าการแบ่งที่ไม่มีความหมายในประเด็นของการศึกษาการประกอบอาชีพของชาวชนบท ทั้งนี้เพราะชาวชนบทไทยไม่น้อยที่ประกอบอาชีพมากกว่า 1 อาชีพขึ้นไป และบุคคลที่ประกอบอาชีพเดียว กับ บุคคลที่ประกอบหลายอาชีพย่อมมีความแตกต่างกันมากมายในเรื่องอื่นๆ จะเห็นได้ว่าความผิดพลาดของการแบ่งเช่นนี้ จะส่งผลต่อการวิเคราะห์ความสัมพันธ์ของอาชีพกับปรากฏการณ์อื่นๆ ที่ใช้อาชีพเพียงอาชีพเดียวเป็นหลัก ผลการวัดที่ได้ก็จะไม่มีความหมาย

สรุปก็คือ การวัดที่ดีต้องมีความแม่นยำ มีความเชื่อถือได้ มีความไว และมีความหมาย คุณสมบัติทั้ง 4 ข้อนี้มีความสัมพันธ์ซึ่งกันและกัน คือ ถ้าการวัดไม่ถูกต้อง ผลที่ได้ก็จะไม่มีความหมาย ถ้าการวัดเชื่อถือไม่ได้ การวัดก็ไม่สามารถให้ผลที่มีความหมายได้ และถ้าไม่มีความไวการวัดก็ไม่มีมีความหมายเช่นกัน

ดังนั้น หลักสำคัญของการวัดก็คือ จะต้องวัดความแตกต่างที่เกิดขึ้นจริง (True Differences) แต่บางครั้งความแตกต่างที่ได้มานั้นอาจจะเกิดจากความคลาดเคลื่อนในการวัด (Errors in Measurement) ผู้วิจัยจำเป็นจะต้องแยกให้ออกว่า ความแตกต่างที่ได้จากการวัดนั้นเป็นความแตกต่างที่แท้จริง หรือเป็นความแตกต่างที่ได้จากความคลาดเคลื่อนในการวัด

การพิจารณาถึงความแตกต่างทั้งสองประเภทนี้ ผู้วิจัยจะต้องคำนึงถึงสาเหตุที่ทำให้เกิดความแตกต่างในลักษณะ หรือคะแนนของประชากร ซึ่งได้แก่

1. ความแตกต่างที่เป็นจริง ซึ่งได้มาจากการวัดที่ถูกต้องและแน่นอน
2. ความแตกต่างที่เป็นจริง ซึ่งได้มาจากการเกี่ยวข้องกันระหว่างลักษณะที่ต้องการจะวัด กับลักษณะอย่างอื่นของประชากร เช่น สติปัญญา การศึกษา สถานภาพทางสังคม
3. ความแตกต่างที่เกิดจากตัวบุคคล เช่น อารมณ์ สุขภาพร่างกาย ความเหนื่อยล้า สุขภาพจิต
4. ความแตกต่างที่เกิดจากสถานการณ์ เช่น การสัมภาษณ์สามีอาจจะมีความแตกต่างถ้ากรรณังอยู่ด้วย การไม่เปิดเผยชื่อจะส่งผลกระทบต่อความถูกต้องของข้อมูล
5. ความแตกต่างที่เกิดจากวิธีการรวบรวมข้อมูล เช่น ผู้สัมภาษณ์อาจจะอธิบายเพิ่มเติมจากคำถามที่ตั้งไว้ หรือถามนำ

6. ความแตกต่างเนื่องจากการเลือกสุ่มเฉพาะบางหัวข้อ เช่น การวัดทัศนคติของประชากรที่มีต่อศาสนา อาจจะเลือกเอาเฉพาะบางหัวข้อซึ่งจะทำให้เกิดความผิดพลาดขึ้นได้
7. ความแตกต่างที่เกิดจากความไม่ชัดเจนของเครื่องมือการวัด เช่น คำถามยาวเกินไป คำถามยากและซับซ้อน
8. ความแตกต่างที่เกิดจากความผิดพลาดทางเทคนิค เช่น กาเครื่องหมายผิดช่อง คำอธิบายพิมพ์ไม่ชัด
9. ความแตกต่างที่เกิดจากความผิดพลาดในการวิเคราะห์ข้อมูล เช่น รวมคะแนนพลาด การเข้าตารางแยกประเภท การคำนวณสถิติผิดพลาด เป็นต้น

ระดับการวัด

การวัด คือ การจัดระเบียบข้อมูลต่างๆ ที่ได้จากการสำรวจ ให้อยู่ในสภาพที่จะนำไปใช้ในการวิเคราะห์ ที่สามารถแบ่งออกได้เป็น 4 ระดับด้วยกันคือ ระดับกลุ่มหรือนามบัญญัติ (Nominal or Categorical) ระดับเรียงอันดับ (Ordinal Scale) ระดับช่วง (Interval Scale) และระดับอัตราส่วน (Ratio Scale) ซึ่งแต่ละระดับมีคุณสมบัติ กฎเกณฑ์ และความยากง่ายของการวัดแตกต่างกัน ความแตกต่างกันในเรื่องของระดับการวัดมีผลต่อการเลือกวิธีการทางสถิติที่จะนำมาใช้สรุป และวิเคราะห์ข้อมูล จึงจำเป็นอย่างยิ่งที่นักวิจัยจะต้องทราบล่วงหน้าว่าตนจะวิเคราะห์ข้อมูลอย่างไร และวิธีการที่ใช้วิเคราะห์นั้นต้องการข้อมูลระดับใด แต่ละระดับของการวัดนั้นมีคุณสมบัติอย่างไร ระดับการวัดทั้ง 4 ประเภทนี้มีลักษณะสำคัญดังต่อไปนี้

1) ระดับกลุ่มหรือนามบัญญัติ (Nominal or Categorical Scale)

เป็นระดับของข้อมูลที่ได้จากการวัดแบบง่ายที่สุด คือ เป็นการแบ่งแยกประชากรที่จะศึกษาออกเป็นกลุ่ม เป็นพวก โดยแต่ละกลุ่มแต่ละพวกมีความเท่าเทียมกัน สิ่งที่เราจะรู้สึกไว้เสมอก็คือ คุณสมบัติสำคัญของการวัดระดับกลุ่ม ก็คือ ความเท่าเทียมกัน (Equivalence) ดังนั้น สมาชิกในแต่ละกลุ่มจะต้องเชื่อถือได้ว่ามีความเหมือนกัน ตัวเลข หรือสัญลักษณ์ที่ใช้ระบุกลุ่มเป็นเพียงแต่ชื่อ ไม่อาจนำมาใช้ในการคำนวณทางเลขคณิตได้ เช่น

แบ่งประชากรโดยใช้เพศเป็นตัวแบ่ง คือ

1) ชาย และ 2) หญิง

แบ่งประชากรโดยใช้ภาคเป็นตัวแบ่ง คือ

1) ภาคเหนือ 2) ภาคใต้ 3) ภาคตะวันออกเฉียงเหนือ 4) ภาคตะวันออก

แบ่งประชากรโดยใช้สถานภาพสมรสเป็นตัวแบ่ง คือ

1) โสด 2) หม้าย 3) หย่าร้าง ฯลฯ

สถิติที่ใช้ในการสรุปคุณสมบัติของประชากรตามระดับการวัดนี้ ได้แก่

ความถี่ (Frequency) อัตราส่วนร้อยละ (%) ฐานนิยม (Mode) ในการทดสอบความสัมพันธ์ อาจใช้วิธีการ Chi-Square, Contingency Coefficient C, Tschuprow's T, Yule's Q, Goodman and Kruskal's tau, Lambda เป็นต้น

2) ระดับเรียงอันดับ (Ordinal Scale)

เป็นระดับข้อมูลที่กำหนดรายละเอียดของการวัดเพิ่มขึ้นจากระดับนามบัญญัติ กล่าวคือ นอกจากจะแบ่งแยกข้อมูลเป็นกลุ่มแล้ว ยังสามารถหาระดับความแตกต่างระหว่างกลุ่มได้ด้วย ซึ่งระบบการวัดแบบนี้ใช้หลักการของความมากกว่า ความน้อยกว่า เช่น

แบ่งประชากรโดยใช้ความคิดเห็น ซึ่งอาจจะมีระดับต่างๆ กัน เช่น เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่มีความเห็น ไม่เห็นด้วย ไม่เห็นด้วยอย่างยิ่ง

การแบ่งสินค้าโดยใช้คุณภาพของสินค้า คือ ดี พอใช้ เลว

จากตัวอย่างดังกล่าว เราสามารถบอกความแตกต่างของแต่ละกลุ่มข้อมูลได้ โดยสามารถจัดอันดับของข้อมูล แต่ไม่สามารถกำหนดปริมาณความน้อยกว่า หรือ มากกว่า ออกมาเป็นตัวเลขที่แน่นอนได้ เช่น เราสามารถบอกได้ว่าสินค้าที่มีคุณภาพดีนั้นย่อมจะดีกว่าสินค้าที่มีคุณภาพพอใช้ แต่ไม่สามารถบอกได้ว่าดีกว่ากันเป็นตัวเลขเท่าไร การใช้ตัวเลขหรือสัญลักษณ์ใดๆ กำหนดอันดับความมากกว่า น้อยกว่า จะไม่มีผลต่อข้อมูล เช่น การให้ร้อยตรี ติดดาว 1 ดวง ร้อยโทติดดาว 2 ดวง และร้อยเอกติดดาว 3 ดวง เราอาจจะกำหนดใหม่ให้ ร้อยเอกติดดาว 1 ดวง ร้อยโทติดดาว 2 ดวง และร้อยตรีติดดาว 3 ดวง โดยที่อันดับก่อนหลังย่อมไม่เปลี่ยนแปลง ตัวเลขที่แทนข้อมูลระดับนี้ยังไม่สามารถนำไปคำนวณทางคณิตศาสตร์ได้

สถิติที่ใช้ในการสรุปคุณสมบัติของตัวแปรตามระดับการวัดเรียงอันดับนี้ เช่น

ค่ามัธยฐาน (Median) ซึ่งไม่เปลี่ยนแปลงโดยคะแนนที่สูงกว่าหรือต่ำกว่า

สหสัมพันธ์แบบจัดอันดับ (Kendall's Tau, Spearman Rank) Gamma, dyx, dxy เป็นต้น

3) ระดับช่วง (Interval Scale)

เป็นระดับข้อมูลที่มีคุณสมบัติเพิ่มเติมจากการวัดเรียงอันดับ กล่าวคือ สามารถกำหนดปริมาณของความแตกต่างระหว่างอันดับได้ด้วย เพราะการวัดระดับนี้ หน่วยของการวัดมีลักษณะคงที่ ซึ่งเป็นมาตรฐานในการกำหนดค่าเป็นตัวเลข เช่น

ระดับอุณหภูมิ เราสามารถบอกได้ว่า อุณหภูมิ 30 องศา ร้อนกว่า 20 องศา อยู่ 10 องศา แต่การวัดระดับนี้ จุดเริ่มต้นถือว่าเป็นธรรมชาติ กล่าวคือ ไม่มีศูนย์สัมบูรณ์ (Absolute Zero) ซึ่งความหมายของศูนย์ระดับนี้ ไม่ได้แทนความหมายว่า ไม่มีค่า การวัดระดับนี้เพียงแต่ทราบระดับเปรียบเทียบเท่านั้น เช่นในการวัดอุณหภูมิ ถ้าใช้ฟาเรนไฮต์ จะเริ่มจาก 32 แต่ถ้าใช้ระบบเซลเซียสจะเริ่มจาก 0 จะเห็นได้ว่า การกำหนดจุดเริ่มต้นนั้นเป็น การกำหนดจุดเริ่มต้นตามใจชอบ ไม่มีจุดเริ่มต้นที่เป็นธรรมชาติ ถึงแม้ว่าการกำหนดจุดเริ่มต้นไม่เหมือนกัน แต่ถ้าหน่วยของการวัดมีมาตรฐานที่คงที่แล้ว การเปรียบเทียบกลับกันก็อาจทำได้ ดังตัวอย่าง

เชลเซียส	0	10	30	100
ฟาเรนไฮต์	32	50	86	212

$$\text{อัตราส่วนของเชลเซียส} \quad \frac{30 - 10}{10 - 0} = 2$$

$$\text{อัตราส่วนของฟาเรนไฮต์} \quad \frac{86 - 50}{50 - 32} = 2$$

ไม่ว่าจะใช้ระบบใด ผลต่างของการเทียบอัตราส่วนยังคงเป็น 2 แม้ว่าจุดเริ่มต้นจะไม่เท่ากัน ตัวเลขที่ได้จากการวัดระดับนี้สามารถนำมาบวกลบกันได้

การวัดชนิดนี้เป็นการวัดเชิงปริมาณ (Quantitative Scale) ที่แท้จริงสามารถที่จะนำมาใช้กับสถิติประเภทต่างๆ ที่ใช้กันอยู่ทั่วไปได้ เช่น

สถิติประเภทค่าเฉลี่ยเลขคณิต และค่าเบี่ยงเบนมาตรฐาน การวิเคราะห์ถดถอย พวกนี้อาจจะนำมาใช้กับข้อมูลประเภทนี้ได้หากเป็นไปตามข้อตกลงเบื้องต้นของสถิติเหล่านั้นๆ เช่น มีการกระจายแบบปกติ (Normal Distribution)

สิ่งที่ควรสังเกตและพึงปฏิบัติก็คือ

เมื่อใดเราสามารถจะใช้วิธีการทางสถิติแบบที่จำกัดรูปแบบการกระจาย (Parametric Statistic) ได้ ควรจะใช้วิธีนั้นทันที เพราะการใช้สถิติแบบที่ไม่จำกัดการกระจาย (Non-Parametric Statistic) ซึ่งต้องการข้อมูลในระดับที่ต่ำกว่า เช่น ข้อมูลระดับกลุ่มหรือนามบัญญัติ ทำให้ต้องมีการยุบรวมรายละเอียดข้อมูลให้เป็นกลุ่ม) จะทำให้เสียรายละเอียดข้อมูลโดยไม่จำเป็น นอกจากนั้นประสิทธิภาพในการทดสอบของสถิติประเภทที่จำกัดรูปแบบการกระจายยังสูงกว่าแบบที่ไม่จำกัดการกระจายอีกด้วย ในจำนวนตัวอย่างที่เท่าๆ กัน

4) ระดับอัตราส่วน (Ratio Scale)

เป็นระดับข้อมูลที่ดีกว่ามีความสมบูรณ์ที่สุด และเป็นระดับการวัดสูงสุด มีจุดเริ่มต้นตามธรรมชาติ คือ มีศูนย์แท้ ที่หมายความว่าถึงการไม่มีค่า เช่น น้ำหนัก ส่วนสูง อายุ ฯลฯ ข้อมูลบางชนิดไม่สามารถวัดได้ถึงระดับนี้ เช่น ข้อมูลทางด้านพฤติกรรม ทักษะคติ ข้อมูลระดับนี้สามารถนำไปคำนวณคณิตศาสตร์ได้ทุกประเภท

หมายเหตุ

ข้อมูลในระดับต่างๆ ที่กล่าวมานั้นอาจจะจัดแบ่งได้เป็น 2 ลักษณะใหญ่ คือ

1) ข้อมูลเชิงคุณภาพ (Qualitative)

2) ข้อมูลเชิงปริมาณ (Quantitative)

ข้อมูลเชิงคุณภาพ คือ ข้อมูลที่ไม่สามารถนำไปคำนวณทางคณิตศาสตร์ได้ ซึ่งก็คือ ข้อมูลระดับกลุ่มหรือนามบัญญัติ (Nominal Scale) และเรียงอันดับ (Ordinal Scale) นั้นเอง

ส่วนข้อมูลเชิงปริมาณ คือ ข้อมูลระดับช่วง (Interval Scale) และระดับอัตราส่วน (Ratio Scale)

มาตรวัด และวิธีการสร้างมาตรวัด

ลักษณะทั่วไปของมาตรวัด

มาตรวัด คือ ชุดของสัญลักษณ์หรือตัวเลขที่สร้างขึ้นมาจากลักษณะที่สามารถนำไปใช้วัดบุคคลหรือสิ่งที่ต้องการวัด โดยมีการกำหนดไว้แน่นอนว่าคุณสมบัติใดควรจะได้คะแนนเท่าไร

มาตรวัด มีองค์ประกอบสำคัญ 3 ประการคือ

- 1) ปรากฏการณ์
- 2) ตัวเลขหรือคะแนน
- 3) วิธีการให้คะแนนกับปรากฏการณ์นั้นๆ

เมื่อนักวิจัยต้องการสร้างมาตรวัด ก็จะมุ่งความสนใจไปที่เหตุการณ์หรือปรากฏการณ์ที่สามารถสังเกตการณ์ได้ แล้วกำหนดว่าต้องการที่จะวัดอะไรหรือสิ่งใด จากนั้นจึงกำหนดเกณฑ์หรือกติกาในการให้คะแนนกับปรากฏการณ์ที่จะไปสังเกตการณ์ สุดท้ายคือ การวัดหรือให้คะแนนกับสิ่งนั้นๆ

การสร้างมาตรวัดมีหลายวิธี แต่ไม่ว่าจะใช้วิธีใดก็ตาม ควรคำนึงถึงความแม่นยำหรือความเชื่อถือได้ของมาตรวัดนั้น ปัจจัยที่อาจจะกระทบต่อความเชื่อถือได้สามารถพิจารณาได้ใน 3 ประเด็น คือ จำนวนตัวแปรในมาตรวัด ผู้ให้ข้อมูล และข้อคำถามเกี่ยวกับสิ่งที่ต้องการทราบ โดยมีรายละเอียดคือ

1. จำนวนตัวแปรในมาตรวัด

ในการวัดสิ่งใดสิ่งหนึ่ง อาจใช้ตัวแปรเพียงตัวเดียววัดสิ่งนั้น หรืออาจจะใช้ตัวแปรหลายๆ ตัวร่วมกันวัดในสิ่งเดียวกันก็ได้ นั่นคือ

1) มาตรวัดที่ใช้ตัวแปรเดียว

คือมาตรวัดที่ใช้ในการวัดเฉพาะด้าน เช่น การประเมินผลการปฏิบัติงานในด้านผลงาน อาจวัดจากคุณภาพหรือปริมาณของงาน เป็นต้น ซึ่งบางครั้ง การใช้มาตรวัดแบบนี้ อาจให้คนอื่นทำการประเมินหรือวัด เช่น

ผู้บังคับบัญชา หรือหัวหน้างานที่เกี่ยวข้อง แต่ในบางสถานการณ์ ผู้ประเมินอาจจะเป็นเจ้าตัวเองก็ได้ เช่น การวัดความพึงพอใจกับคุณภาพชีวิตของตัวเอง ผู้ตอบแบบสอบถามอาจจะเป็นผู้ประเมินว่า พึงพอใจมากน้อยเพียงใด เป็นต้น

2) มาตรการที่ใช้ตัวแปรหลายตัว

คือมาตรการที่สร้างขึ้นมาโดยรวมคะแนนจากข้อความหลายๆ ข้อเข้าด้วยกัน แล้วสรุปรวมเป็นคะแนนเดียว เช่น

จากตัวอย่างการประเมินผลการปฏิบัติงานข้างต้น นอกจากจะประเมินด้านผลงานแล้ว อาจจะประเมินคุณลักษณะในการปฏิบัติงานซึ่งประกอบด้วยความรู้ความสามารถในการปฏิบัติงาน การรักษาวินัย ความอดุสาหะ และความรับผิดชอบ และเมื่อประเมินแล้วจะรวมคะแนนในทุกข้อเพื่อสรุปจากคะแนนรวมว่า ผู้ที่ได้รับการประเมินอยู่ในระดับใด

2. ผู้ให้ข้อมูล

ในการตอบแบบสอบถามที่ใช้เป็นมาตรวัดนั้น ข้อมูลที่ได้อาจมาจากเจ้าตัวเอง หรือจากการประเมินของผู้ที่เกี่ยวข้องก็ได้ ซึ่งการจะเลือกผู้ใดเป็นผู้ให้ข้อมูลขึ้นอยู่กับเนื้อหาสาระของเรื่องหรือสิ่งที่ต้องการวัดเป็นสำคัญ ซึ่งอาจพิจารณาทั้ง 2 กรณีได้ดังนี้

1) การประเมินโดยตนเอง (Self - evaluation)

วิธีการนี้ผู้ให้ข้อมูลคือเจ้าตัวเอง ซึ่งในหลายกรณีจะเป็นการให้ข้อมูลที่ถูกต้องที่สุด เพราะเป็นการสมเหตุสมผลที่ว่าตนต้องรู้จักตนมากกว่าคนอื่น เช่น ข้อมูลเกี่ยวกับภูมิหลังทางครอบครัว การศึกษา ทัศนคติ

หรือการประเมินความสามารถทางกายภาพบางอย่าง เช่น การยกสิ่งของ การทำงานภายในกำหนดระยะเวลาหรือเงื่อนไขบางประการ เป็นต้น

อย่างไรก็ตาม ในหลายกรณีการประเมินตนเองก็อาจเบี่ยงเบนได้ เพราะบางครั้ง "การรู้จักตนเอง" นั้น ก็อาจจะไม่ตรงกับที่เป็นอยู่จริง หรือรู้จริงแต่ไม่ยอมบอก หรืออยากบอกแต่มีความลำเอียงเข้าข้างตนเอง อย่างใดอย่างหนึ่ง

กระนั้นก็ตาม ในการเก็บข้อมูลจากแบบสอบถามโดยทั่วไป มักให้ผู้ตอบแบบสอบถามเป็นผู้ให้ข้อมูลเกี่ยวกับตัวเองมากที่สุด ฉะนั้น นักวิจัยจึงควรพึงระวังถึงข้อจำกัดเหล่านี้เป็นพิเศษด้วย

2) การประเมินตนโดยบุคคลอื่น

วิธีการนี้ผู้ให้ข้อมูลคือผู้ที่เกี่ยวข้อง และสามารถตอบคำถามเกี่ยวข้องกับบุคคลที่เป็นกลุ่มตัวอย่างที่ต้องการวัดได้ เช่น ในกรณีการประเมินผลงานหรือความสามารถในการทำงานของเจ้าหน้าที่โดยผู้บังคับบัญชา หรือในกรณีศึกษาการบริหารงานของผู้บังคับบัญชา โดยเก็บข้อมูลจากผู้ใต้บังคับบัญชา เป็นต้น

วิธีการนี้สามารถใช้แก้ปัญหาการไม่สามารถประเมินตนเองได้ถูกต้อง หรือการเกิดอคติของผู้ที่เราต้องการวัดได้ในบางกรณี แต่ไม่ใช่จะสามารถใช้วิธีนี้ได้เสมอไป ดังได้กล่าวไว้ข้างต้นไว้แล้วว่า การจะเลือกกว่าควรเก็บข้อมูลจากใคร หรือใครควรเป็นผู้ให้ข้อมูลนั้น ขึ้นอยู่กับสาระของแบบสอบถามเป็นสำคัญ ซึ่งในบางกรณีอาจใช้ทั้ง 2 วิธีร่วมกันก็ได้

3. ข้อคำถามเกี่ยวกับสิ่งที่ต้องการทราบ

ในการสร้างแบบสอบถามเพื่อวัดสิ่งใดสิ่งหนึ่งนั้น นักวิจัยอาจทำได้ 2 ลักษณะคือ

1) การวัดโดยถามตรงไป

เป็นการสร้างข้อคำถามที่ถามเกี่ยวกับคุณสมบัติ หรือสิ่งที่ต้องการวัดเป็นการเฉพาะเจาะจงลงไป เช่น

เมื่อต้องการวัดว่า บุคคลมีความผูกพันกับงานหรือไม่ ก็อาจสร้างคำถามว่า "ท่านรู้สึกผูกพันกับงานหรือไม่"

หรือในกรณีที่ต้องการวัดความรับผิดชอบ ก็อาจใช้คำถามตรงไปตรงมาว่า "ท่านมีความรับผิดชอบในหน้าที่การงาน มากน้อยเพียงใด"

การสร้างข้อคำถามในลักษณะนี้เป็นการให้ผู้ตอบใช้ดุลพินิจหรือ วิจารณ์ญาติความ และประเมินในเรื่องนั้นๆ ซึ่งในบางเรื่องที่ไม่กำกวมก็อาจใช้ได้

แต่ในบางเรื่องที่มีการตีความได้กว้าง เช่น ความรับผิดชอบ หรือความผูกพันในงานดังกล่าวนั้น การวัดโดยตรงก็อาจได้คำตอบที่คลาดเคลื่อนไปตามการตีความของผู้ตอบได้

2) การวัดโดยถามทางอ้อม

เป็นการสร้างข้อคำถามโดยคำนึงถึงมิติต่างๆ ที่เกี่ยวข้องกับสิ่งที่ต้องการวัด และสร้างคำถามให้ครอบคลุมมากที่สุด เพื่อหลีกเลี่ยงการตีความที่อาจไม่ตรงกันระหว่างนักวิจัยกับผู้ตอบแบบสอบถามหรือกล่าวอีกนัยหนึ่งก็คือ การหาเครื่องชี้วัดมาช่วยประเมินสิ่งที่ต้องการวัดนั่นเอง เช่น จากกรณีตัวอย่างของการวัดโดยตรง เรื่องความรับผิดชอบ นักวิจัยอาจกำหนดสิ่งบ่งชี้เพื่อสร้างคำถามทางอ้อมเป็น 3 ลักษณะ คือ

- การส่งงานตรงเวลา
- การไม่มึนงานทำค้าง และ
- ผลงานไม่เคยผิดพลาด

เพื่อวัดว่าบุคคลนั้นๆ มีความรับผิดชอบหรือไม่ แทนการถามตรงๆ ว่า บุคคลนั้นๆ มีความรับผิดชอบหรือไม่ เป็นต้น

วิธีการสร้างมาตรวัด

1. การสร้างมาตรวัดแบบให้คะแนน หรือให้ประเมิน (Rating Scale)

เอกลักษณ์ของมาตรวัดแบบ Rating Scale คือ

การวัดที่กำหนดกรอบหรือแบ่งระดับของเกณฑ์ในการวัด เพื่อบอกให้ทราบว่า ผู้ที่ถูกวัดอยู่ ณ จุดใด เช่น

กำหนดเกณฑ์การวัดออกเป็น 9 ระดับ ตั้งแต่น้อยที่สุด (คะแนนเท่ากับ 1) ไปจนถึงมากที่สุด (คะแนนเท่ากับ 9)

การสร้างมาตรวัดแบบให้คะแนนดังกล่าวนี้ สามารถแบ่งออกเป็น 3 ชนิดด้วยกันคือ

ชนิดแรก : Graphic Rating Scales

เป็นการสร้างมาตรวัดที่แสดงกรอบของสิ่งที่เราจะวัดเป็นแผนภาพ โดยมีคำอธิบายประกอบว่า ในแต่ละจุดของแผนภาพนั้นหมายถึงอะไร จากนั้นให้ผู้ตอบแบบสอบถามเป็นผู้ที่เลือกว่า เขาจะอยู่ในระดับใด เช่น การวัดความพึงพอใจในชีวิต อาจแสดงเป็นรูปคนที่แสดงอารมณ์ต่างๆ หรือในกรณีที่ต้องการความพึงพอใจของประชาชนต่อการแก้ไขปัญหาเศรษฐกิจของรัฐบาลปัจจุบัน อาจแบ่งมาตรวัดออกเป็น 7 ระดับ ซึ่งผู้ให้ข้อมูลในเรื่องหนึ่งๆ จะต้องเลือกภาพ หรือระบุหมายเลขที่ใกล้เคียงที่สุดกับความรู้สึกของตนเกี่ยวกับเรื่องนั้นๆ

การสร้างมาตรวัดแบบ Graphic Rating นี้ มีข้อดีคือ

ง่ายต่อการใช้ แต่ถ้าจะใช้ให้ได้ผลจริงๆ แล้ว ควรจะมีการระวังในบางเรื่อง เช่น ชุดข้อความที่อยู่ปลายสุดของทั้ง 2 ข้างนั้น ไม่ควรจะเป็นข้อความที่สุดขั้วเกินไป เพราะคนจะไม่เลือกตอบ แต่จะมาเลือกตอบในช่วงตรงกลาง ซึ่งจะทำให้แยกความแตกต่างไม่ได้ ฉะนั้น การกำหนดเกณฑ์หรือข้อความ และการอธิบายข้อความแต่ละช่วงของมาตรวัด ก็ควรสอดคล้องกับคะแนนที่ให้อยู่ด้วย

ชนิดที่สอง : Itemized Rating Scales

บางท่านเรียกว่า "Specific Category Scales" หรือบางท่านเรียกว่า "Numerical Scales"

มาตรวัดแบบให้คะแนนในลักษณะนี้ประกอบด้วยปฏิริยาของผู้ตอบต่อชุดข้อความที่กำหนดให้ เช่น เห็นด้วยหรือไม่อย่างไร ส่วนใหญ่จะแบ่งเป็น 5 - 7 ช่วง เช่น

ความเห็นต่อการฝึกอบรมเพื่อเลื่อนตำแหน่ง ว่าเป็นสิ่งที่จำเป็นในระบบราชการไทย ซึ่งผู้ตอบอาจจะเลือกคำตอบตามที่กำหนดไว้ว่า เห็นด้วยอย่างยิ่ง เห็นด้วย ไม่มีความเห็น ไม่เห็นด้วย หรือไม่เห็นด้วยอย่างยิ่ง

อนึ่ง ในการวัดทั้งชนิดที่หนึ่งและชนิดที่สองนี้ เป็นการประเมินข้อความหนึ่งๆ หรือบุคคลหนึ่งๆ โดยไม่มีการเปรียบเทียบกับบุคคลอื่นแต่อย่างใด

ชนิดที่สาม : Comparative Rating Scales

เป็นการให้คะแนนเชิงเปรียบเทียบระหว่างสิ่งที่ถูกวัด หรือบุคคลที่ถูกวัด หรือสิ่งอื่นหรือบุคคลอื่น เช่น

ในแบบฟอร์มการประเมินนักศึกษาเพื่อศึกษาต่อในระดับสูงนั้น จะเป็นการเปรียบเทียบว่า เมื่อเทียบกับเพื่อนร่วมกลุ่มของเขาแล้ว นักศึกษานั้นจัดอยู่ในลำดับใด

หรือในการประเมินผลการทำงานของพนักงาน อาจเป็นการให้คะแนนความสามารถในการปฏิบัติงานเปรียบเทียบกับบุคคลอื่นในหน่วยงานว่า บุคคลที่ได้รับการประเมินนั้นอยู่ในระดับใด เช่น ใน 10 เปอร์เซนต์แรกของพนักงานดีเด่น เป็นต้น

หรืออาจเป็นการเรียงลำดับความสำคัญหรือความสามารถของบุคคลทุกคนในกลุ่ม ตั้งแต่สูงสุดลงมาถึงต่ำสุด ซึ่งการใช้มาตรวัดเช่นนี้ จะสามารถกระทำได้อีกต่อเมื่อผู้ประเมินรู้จักทั้งกลุ่มเป็นอย่างดี

ปัญหาที่สำคัญก็คือ บางครั้งความรู้ในเรื่องที่จะประเมินต่อกลุ่มนั้นมีไม่เพียงพอที่จะบอกได้ โดยเฉพาะการจัดเรียงลำดับของทั้งกลุ่ม จึงทำให้การใช้ Comparative Rating Scales ค่อนข้างเป็นไปได้ยาก และจะใช้ได้เฉพาะกับกลุ่มที่มีขนาดเล็กเท่านั้น โดยจะเป็นประโยชน์ก็ต่อเมื่อต้องการทราบอันดับจริงๆ ว่า ใครมาลำดับที่ 1-2 หรือ 3

หรือในอีกกรณีหนึ่ง อาจทำได้เมื่อมีเกณฑ์การวัดหรือทดสอบที่มีมาตรฐานเชื่อถือได้ เช่น การสอบภาษาอังกฤษ TOEFL (Test of English as Foreign Language)

ทั้งนี้มาตรวัดทั้ง 3 ชนิดที่กล่าวมาข้างต้น สามารถนำมาใช้ได้ทั้งกรณีที่เป็น การประเมินบุคคลอื่น หรือเป็นการประเมินตนเองของบุคคลหนึ่งๆ

สิ่งที่พึงระวังในการสร้างมาตรวัดแบบให้คะแนน

เนื่องจากการใช้มาตรวัดแบบให้คะแนน (Rating Scales) จริงๆ แล้วก็คือการ ใช้วิจารณ์ญาติในการประเมิน เพราะฉะนั้น โอกาสที่จะมีความลำเอียง หรือที่เรียกว่า "อคติ" (Bias) เข้ามานั้น เป็นไปได้ค่อนข้างมาก ความลำเอียงในที่นี้อาจแบ่งได้ 3 ประเภทคือ

ประเภทแรก คือ

ความลำเอียงที่มีแนวโน้มจะให้คะแนนระดับเดียวกับบุคคลใดบุคคลหนึ่งในทุกๆ มิติ (Halo Effect) เช่น ถ้าเชื่อว่าบุคคลที่เป็นพนักงานขายชั้นแข็ง ย่อมเป็นพนักงานที่ซื่อตรง และมีคุณธรรมด้วย ดังนี้ ถ้าเราให้คะแนนเขาในด้านหนึ่งๆ สูง ก็จะทำให้คะแนนในด้านอื่น สูงตามไปด้วย ในกรณีนี้ขึ้นอยู่กับความเชื่อของเราว่า คุณลักษณะหนึ่งๆ นั้น จะสัมพันธ์หรือครอบคลุมถึงคุณลักษณะอื่นๆ อย่างไรบ้าง เช่น นาย ก. เป็นบุคคลที่มีบุคลิกขี้อาย ก็มีแนวโน้มว่า นาย ก. จะปรับตัวยากไปด้วย เมื่อเป็นเช่นนี้ แม้ว่าจะมีความพยายามวัดในหลายมิติ แต่ก็อาจคลาดเคลื่อนได้ ซึ่งทำให้บางครั้งผลการวัดดูเหมือนแต่ละมิติมีความสัมพันธ์ซึ่งกันและกัน แต่จริงๆ แล้วกลับไม่ใช่

ประเภทที่สอง คือ

ความลำเอียงที่จะเกิดขึ้นคือความใจกว้างของผู้ประเมิน (Generosity Error) เป็นการให้คะแนนแก่ผู้ถูกประเมินสูงกว่าที่เขาควรจะได้รับ โดยเฉพาะกับคนที่ผู้ประเมินมีความชอบพอเป็นทุนเดิมอยู่แล้ว ซึ่งในที่นี้หมายความว่า ผู้ประเมินมักจะประเมินบุคคลที่ตนชอบให้ดีเกินควร และประเมินผู้ที่ไม่ชอบในลักษณะปานกลาง

ประเภทที่สาม คือ

เป็นการให้คะแนนโดยเปรียบเทียบผู้อื่นกับตัวผู้ประเมินเอง หรือใช้ตนเองเป็นมาตรฐาน ซึ่งเป็นการประเมินบุคคลอื่นในทางตรงข้ามกับเรา เช่น ถ้าเราเป็นคนค่อนข้างเจ้า

ทำอย่างไรจึงจะป้องกันไม่ให้เกิดความลำเอียง

จะเห็นว่า ความลำเอียงเกิดขึ้นได้ในหลายลักษณะ และมีผลกระทบอย่างมากต่อการวัดแต่ละครั้ง ซึ่งจำเป็นอย่างยิ่งที่ผู้วิจัยจะต้องป้องกันมิให้เกิดความลำเอียง หรือให้เกิดน้อยที่สุดเท่าที่จะเป็นไปได้ ซึ่งอาจจะทำได้ดังต่อไปนี้

- 1) ผู้ประเมินจะต้องตระหนักถึงปัญหาและระมัดระวังในเรื่องนี้เป็นอย่างมาก
- 2) การใช้ถ้อยคำในการจำแนกคะแนนต่างๆ ไม่ควรเป็นการสุดโต่งหรือสุดขั้วจนเกินไป เช่น

ไม่ควรใช้คำว่า ตลอดเวลา น้อยที่สุด มากที่สุด เพราะคำว่า "ที่สุด" คนมักจะไม่เลือก เช่น ถ้าให้เลือกระหว่างข้าพเจ้าพอใจงานที่ทำอยู่ กับข้าพเจ้าพอใจที่สุดกับงานที่ทำอยู่ คนจะเลือกคำตอบข้อแรกมากกว่า ในขณะที่คำตอบหลังอาจจะไม่มีผู้เลือกเลย

หรือในปัญหาของอคติที่เกิดจากการให้คะแนนดีเกินควร ก็ควรหาถ้อยคำที่เป็นกลางมากกว่าถ้อยคำที่มีความหมายบวกหรือลบในตัวของมันเอง เช่น กรณีการใช้ข้อความ "ชอบประจบผู้ใหญ่" กับ "ให้ความเอาใจใส่กับความต้องการของผู้ใหญ่ และพยายามสนองความต้องการเหล่านั้น" มีการตีความหมายแตกต่างกันอย่างมาก แม้ว่าต่างก็เป็นความหมายที่พยายามอธิบายพฤติกรรมที่มีลักษณะคล้ายๆ กัน

- 3) การแก้ไขความลำเอียงที่มีแนวโน้มจะให้คะแนนระดับเดียวกับบุคคลใดบุคคลหนึ่งในทุกๆ มิติ (Halo Effect)

3.1) ให้ประเมินบุคคลหนึ่งๆ โดยเก็บข้อมูลคราวละด้าน เพื่อที่จะได้ลิ้มว่า ครั้งก่อนประเมินไว้อย่างไร และผู้ประเมินไม่ควรทราบด้วยว่ากำลังประเมินใคร หรือ

3.2) ให้มีผู้ประเมินหลายคน โดยแยกประเมินคนละมิติ อาจจะโดยการให้นาย ก. ประเมินมิติที่ 1 นาย ข. ประเมินมิติที่ 3 เช่น กรณีการประเมินการปฏิบัติงานของพนักงานคนหนึ่ง อาจให้หัวหน้างาน 3 คนร่วมกันประเมิน โดยหัวหน้างานคนแรกประเมินผลงาน คนที่สองประเมินความมีมนุษยสัมพันธ์ และคนที่สามประเมินความรับผิดชอบ เป็นต้น

อย่างไรก็ตาม มาตรวัดที่ดีจะต้องมีความแม่นยำ และมีความเชื่อถือได้ ทั้งนี้ผู้อ่านควรศึกษาเรื่องดังกล่าวนี้ควบคู่กันไป เพราะเป็นเรื่องสำคัญอย่างยิ่งในการสร้างมาตรวัด

2. การสร้างมาตรวัดจากชุดข้อความถาม (Questionnaires that form scale)

แบ่งออกได้ 3 ลักษณะ คือ 1) Summated Rating Scales 2) Equal

Appearing Scales และ 3) Cumulative Scales

Summated Rating Scales

คือ ชุดข้อความที่ผู้ประเมินต้องประเมิน โดยตอบว่าเห็นด้วยหรือไม่ มากน้อยเพียงใด คะแนนที่ได้ของผู้ตอบแต่ละรายจะเป็นผลรวมของทุกข้อ ซึ่งอาจเป็นคะแนนดิบ หรืออาจหารด้วยจำนวนข้อ หรือตามเกณฑ์ที่กำหนดขึ้น

อาจจะแบ่งคำตอบในมาตรวัดประเภทนี้ ออกเป็น 4 ช่วง เช่น เห็นด้วยอย่างยิ่ง (คะแนนเท่ากับ 4) เห็นด้วย (คะแนนเท่ากับ 3) ไม่เห็นด้วย (คะแนนเท่ากับ 2) ไม่เห็นด้วยอย่างยิ่ง (คะแนนเท่ากับ 1) โดยไม่อนุโลมให้ผู้ตอบไม่แน่ใจ ไม่มีความเห็น หรือรู้สึกเฉยๆ

มาตรวัดแบบ Summated Rating Scales ที่ใช้กันแพร่หลายที่สุด ก็คือ มาตรวัดที่สร้างโดยเรนซิส ลิเคิร์ท (Rensis Likert) ซึ่งเรียกชื่อตามผู้คิดค้นว่า มาตรวัดแบบลิเคิร์ท (Likert - type Scale)

การสร้างมาตรวัดแบบลิเคิร์ทประกอบด้วยขั้นตอนต่อไปนี้

1) นักวิจัยจะรวบรวมข้อมูลเกี่ยวกับสิ่งที่ต้องการจะวัดได้ในจำนวนข้อที่มากพอสมควร โดยที่ชุดข้อความเหล่านั้นจะต้องไม่เป็นข้อความกลางๆ คือ ควรแสดงชัดเจนไปในด้านบวกหรือด้านลบ เพื่อให้ผู้ตอบจะได้มีปฏิกิริยาในเรื่องนั้นๆ ได้

ที่สำคัญอีกประการหนึ่งคือ ชุดข้อความเหล่านั้นควรต้องมีทั้งข้อความด้านบวก คือเชิงสนับสนุน และข้อความด้านลบ คือเชิงปฏิเสธหรือต่อต้านสลับกันไป เพื่อให้ผู้ตอบจะได้มีโอกาสแสดงทั้งความเห็นด้วยและไม่เห็นด้วยทั้ง 2 อย่างพร้อมๆ กัน ทั้งนี้เพื่อป้องกันอคติ (Bias) ซึ่งอาจเกิดขึ้นได้ถ้าข้อความชวนให้ผู้ตอบตอบในทางเดียวกันหมด เช่น "เห็นด้วยกับทุกข้อ" หรือ "ไม่เห็นด้วยกับทุกข้อ"

2) นำชุดข้อความนี้ไปทดสอบกับกลุ่มตัวอย่างที่มีลักษณะคล้ายคลึงกับกลุ่มตัวอย่างที่จะศึกษาวิจัย โดยให้แต่ละคนในกลุ่มตัวอย่างนี้ตอบว่า เห็นด้วยหรือไม่เห็นด้วยกับข้อความแต่ละข้อ

3) ให้คะแนนกับข้อความต่างๆ โดยที่ถ้าเห็นด้วยอย่างยิ่งก็ได้รับคะแนนสูงสุด ส่วนข้อความที่ไม่เห็นด้วยอย่างยิ่งก็ได้รับคะแนนต่ำสุด ส่วนการกำหนดคะแนนสูงสุดจะเริ่มจาก 4 3 2 1 0 หรือ 5 4 3 2 1 นั้นไม่สำคัญ แต่ควรกำหนดให้คะแนนเป็นไปในทิศทางใดทิศทางหนึ่ง

และการตัดสินใจว่าข้อความหนึ่งๆ ควรกำหนดคะแนนให้ในทิศทางใด ควรดูที่เนื้อหาและถ้อยคำของการใช้คำถามนั้นๆ เช่น

"ฉันไม่คิดจะลาออกจากการงาน" กับ "ฉันจะลาออกจากการงานถ้าสามารถหางานอื่นได้" การตอบว่าเห็นด้วยกับข้อความที่ 1 และข้อความที่ 2 จะเป็นที่ยืนยันว่าไม่ได้เห็นด้วยในสิ่งเดียวกัน ดังนั้น ถ้าเรากำลังวัดความผูกพันต่อองค์กร เราอาจจะให้คะแนนแก่คนที่เห็นด้วยกับข้อแรกเท่ากับ 5 ในขณะที่ให้คะแนนข้อที่สองเท่ากับ 1 เป็นต้น

4) คำนวณคะแนนรวมของแต่ละบุคคล ซึ่งได้จากการบวกคะแนนของข้อความทั้งหมดที่ได้ตอบไป

5) ขั้นสุดท้าย นักวิจัยจะวิเคราะห์ว่าข้อความใดมีอำนาจจำแนก และข้อความใดควรจะคงไว้ในมาตรวัด โดยพิจารณาจากอำนาจจำแนกและความเชื่อถือได้ ดังวิธีการที่จะเสนอในส่วนที่เกี่ยวข้องต่อไป หากไม่ผ่านขั้นตอนนี้ จะยังไม่ถือว่าเป็นมาตรวัดแบบลิเคิร์ท ตัวอย่างของคำถามซึ่งมีฐานการเรียงอันดับแบบมาตรวัดลิเคิร์ท

ตัวอย่างของคำถามซึ่งมีฐานการเรียงอันดับแบบมาตรวัดลิเคิร์ต

คุณภาพ (Quality)				
เยี่ยมยอด (Excellent)	ดี (Good)	ปานกลาง (Fair)	ไม่ดี (Poor)	
ดีมาก (Very good)	ค่อนข้างดี (Fairly good)	ปานกลาง (Neither good or bad)	ค่อนข้างไม่ดี (Not very good)	ไม่ดีเลย (Not good at all)

ความสำคัญ (Importance)				
สำคัญมาก (Very important)	ค่อนข้างสำคัญ (Fairly important)	ปานกลาง (Neutral)	ค่อนข้างไม่สำคัญ (Not so important)	ไม่สำคัญ (Not at all important)

ความน่าสนใจ (Interest)		
น่าสนใจอย่าง มาก (Very interested)	น่าสนใจบ้าง (Somewhat interested)	ไม่น่าสนใจ (Not very interested)

ความพึงพอใจ (Satisfaction)				
พึงพอใจอย่าง มาก (Completely satisfied)	พึงพอใจบ้าง (Somewhat satisfied)	เฉยๆ (Neither satisfied nor dissatisfied)	ไม่ค่อยพึงพอใจ (Completely dissatisfied)	ไม่พึงพอใจอย่าง มาก (Somewhat dissatisfied)
พอใจมาก (Very satisfied)	พอใจบ้าง (Quite satisfied)		พอใจเล็กน้อย (Somewhat satisfied)	ไม่พอใจ (Not at all satisfied)

ความถี่ (Frequency)				
ตลอดเวลา (All of the time)	บ่อยครั้งมาก (Very often)	บ่อยครั้ง (Often)	เป็นบางครั้ง (Sometimes)	นานๆ ครั้ง (Hardly ever)
บ่อยครั้งมาก (Very often)	บ่อยครั้ง (Often)	บางครั้ง (Sometimes)	นานๆ ครั้ง (Rarely)	ไม่เคย (Never)
ตลอดเวลา (All of the time)	ใช้เวลาส่วนใหญ่ (Most of the time)		ใช้เวลาบางส่วน (Some of the time)	เพิ่งครั้งแรก (Just now and then)

ความเป็นจริง (Truth)			
เป็นจริงอย่าง มาก (Very true)	เป็นจริงบ้าง (Somewhat true)	ไม่เป็นจริงอย่าง มาก (Not very true)	ไม่เป็นจริง ทั้งหมด (Not at all true)
ใช่แน่นอน (Definitely yes)	บางครั้งใช่ (Most of the time)	บางครั้งไม่ใช่ (Probably no)	ไม่ใช่ทั้งหมด (Definitely no)

ความเป็นเอกลักษณ์ (Uniqueness)				
แตกต่างกันอย่าง มาก (Very different)	แตกต่างกัน บ้าง (Somewhat different)	แตกต่างกัน เล็กน้อย (Slightly different)	ไม่แตกต่างกัน ทั้งหมด (Not at all different)	
มีความเป็น เอกลักษณ์อย่าง เต็มที่ Extremely unique)	มีความเป็น เอกลักษณ์มาก (Very unique)	มีความเป็น เอกลักษณ์ บ้าง (Somewhat unique)	มีความเป็น เอกลักษณ์เพียง เล็กน้อย (Slightly unique)	ไม่มีความเป็น เอกลักษณ์ (Not at all unique)

ข้อสังเกตประการหนึ่งของคะแนนที่ได้จากมาตรวัดแบบลิเคิร์ทนี้ จะเป็น
คะแนนที่มีระดับการวัดเรียงอันดับ (Ordinal Scale) ซึ่งนับว่าเป็นจุดอ่อนของมาตรวัดแบบลิ
เคิร์ท

Equal Appearing Scales

เป็นมาตรวัดที่พยายามจะวัดในระดับการวัดแบบช่วง ซึ่งครอบคลุมถึงมาตรวัดประเภทอื่นๆ เช่น Paired Comparison Scale ของ Thurstone และ Successive Intervals ของ Saffir ในที่นี้จะพูดถึงเฉพาะมาตรวัดของเทอร์สโตน ซึ่งกล่าวได้ว่ามีชื่อเสียงมากที่สุดในการวัดแบบนี้ การสร้างมาตรวัดแบบเทอร์สโตนประกอบด้วยขั้นตอนต่อไปนี้

1) นักวิจัยจะรวบรวมข้อความที่คิดว่าเกี่ยวข้องกับทัศนคติที่กำลังจะวัด จำนวนมากกว่า 100 ข้อขึ้นไป (200 - 300 ข้อ) 2) ทำการทดสอบชุดข้อความนั้นกับกลุ่มผู้รู้หรือผู้เชี่ยวชาญ ระหว่าง 50 - 300 คน โดยขอให้ผู้เชี่ยวชาญแต่ละคนจำแนกข้อความออกเป็น 11 กลุ่มด้วยกัน กลุ่มแรกประกอบด้วยชุดข้อความที่ผู้เชี่ยวชาญคิดว่าสอดคล้องหรือเห็นด้วยที่สุด ลดหลั่นไปจนถึงกลุ่มที่เห็นด้วยน้อยที่สุด โดยให้กลุ่มที่ 6 เป็นจุดกึ่งกลาง

3) ข้อความหนึ่งๆ จะได้รับคะแนนโดยถือค่าเฉลี่ยหรือมัธยฐานเลขคณิต (Mean) ที่ได้รับจากผู้เชี่ยวชาญต่างๆ (ในกรณีมาตรวัดแบบอื่นที่ไม่ใช่ของเทอร์สโตน จะใช้จากค่ามัธยฐาน หรือ Median) ถ้าข้อความใดมีการกระจายมาก หรือผู้เชี่ยวชาญมีความเห็นแตกต่างกันมากก็จะต้องตัดออกไป เพราะถือว่าไม่สามารถที่จะวัดสิ่งนั้นๆ ได้ และเนื่องจากข้อความแต่ละข้อจะได้คะแนนเท่าไร ขึ้นอยู่กับวิจารณ์ของผู้เชี่ยวชาญ ดังนั้น จึงจำเป็นที่จะต้องใช้กลุ่มของผู้เชี่ยวชาญมากกว่า 50 คนขึ้นไป

4) ขั้นตอนต่อไปคือ การเก็บข้อมูล ซึ่งในขั้นตอนนี้ ผู้กรอกแบบสอบถามจะถูกถามอย่างใดอย่างหนึ่งต่อไปนี้

(ก) ให้เลือกเฉพาะข้อความที่เขาเห็นด้วย หรือ

(ข) ให้เลือกเฉพาะข้อความ 2 - 3 ข้อ ที่เขาเห็นด้วยมากที่สุด

การจัดเรียงข้อในแบบสอบถามนี้ จะต้องสลับข้อไม่ให้ผู้ตอบรู้ว่า แต่ละข้อได้รับคะแนนเท่าใด ซึ่งในเชิงทฤษฎี ข้อความที่เขาจะเลือกขึ้นมาควรจะเป็นข้อความที่มีน้ำหนักอยู่ใกล้เคียงกัน เช่น ถ้าคะแนนรวมเท่ากับ 24 เขาอาจจะเลือกข้อความที่มีค่า 8, 9, และ 7 ตามลำดับ ไม่ใช่เลือกจากข้อความที่มีค่า 5, 11, และ 8 ซึ่งภายหลังจากคำนวณค่าเฉลี่ยของคะแนนแต่ละบุคคลแล้ว นำคะแนนเฉลี่ยนั้นมาเป็นคะแนนของบุคคล เพื่อสะท้อนถึงจุดยืนของเขาถึงสิ่งๆ นั้น ในการตีความว่าเห็นด้วยหรือไม่ เพียงไร

อนึ่งมาตรวัดแบบเทอร์สโตนมีข้อเสียบางประการ คือ

ประการแรก

สิ้นเปลืองเวลามาก ทั้งในการสร้างข้อความและการหากลุ่มผู้เชี่ยวชาญเพื่อเป็นกลุ่มทดสอบมาตรวัดนั้น

ประการที่สอง

ความเห็นของผู้เชี่ยวชาญอาจแตกต่างกัน แม้ว่าจะมีผลคะแนนรวมเท่ากัน เช่น ถ้าผู้เชี่ยวชาญคนหนึ่งเลือกข้อความที่มีน้ำหนักคะแนนเท่ากับ 8, 8 และ 7 ทำให้ได้คะแนน

ประการที่สาม

เนื่องจากมาตรวัดแบบเทอร์สโตน มีคะแนนประจำแต่ละข้อโดยยึดความเห็นของผู้เชี่ยวชาญเป็นสำคัญ จึงอาจเกิดอคติขึ้นได้ ถ้าการคัดเลือกผู้เชี่ยวชาญไม่สัมพันธ์กับเรื่องที่ต้องการวัด เช่น ถ้าต้องการสร้างมาตรวัดเพื่อวิจัยเกี่ยวกับรัฐวิสาหกิจ แต่ใช้ความเห็นจากผู้เชี่ยวชาญซึ่งเป็นอาจารย์มหาวิทยาลัยเป็นส่วนมาก คะแนนที่ได้ในแต่ละข้อความอาจไม่สัมพันธ์กับเรื่องที่จะวัดก็เป็นได้ กระนั้นก็ตาม จากการทบทวนวรรณกรรมในเรื่องนี้ พบข้อสรุปว่าสามารถสร้างมาตรวัดแบบ Equal Appearing Scales ได้

อย่างไรก็ดี เมื่อเปรียบเทียบระหว่างมาตรวัดแบบลิเคิร์ตกับเทอร์สโตนแล้ว จะเห็นได้ว่า แม้ทั้งสองแบบจะใช้เป็นมาตรวัดทัศนคติเช่นเดียวกัน แต่มาตรวัดแบบลิเคิร์ตถือว่าสร้างได้ง่ายกว่า และมีความเชื่อถือได้มากกว่าในกรณีที่มาตรวัดมีความยาวเท่าเทียมกัน ทั้งนี้สืบเนื่องมาจากมาตรวัดแบบลิเคิร์ตมีการจำแนกกระหว่างการตอบออกได้ประมาณ 5 ประเภท แต่มาตรวัดแบบเทอร์สโตนจะเลือกได้เฉพาะเห็นด้วยหรือไม่เห็นด้วยเท่านั้น

Cumulative Scales หรือ Guttman Scales

มาตรวัดประเภทนี้เป็นมาตรวัดในมิติเดียว ประกอบด้วยชุดข้อความที่ผู้ตอบลงความเห็นว่าย เห็นด้วยหรือไม่เห็นด้วย โดยชุดข้อความมีไม่มากข้อ อาจจะเพียง 4 - 5 ข้อขึ้นไป แต่ชุดข้อความนั้นจะต้องมีความสัมพันธ์กันในลักษณะต่อเนื่อง คือ คนที่เห็นด้วยกับข้อที่ 2 ก็จะต้องเห็นด้วยกับข้อที่ 1 และคนที่เห็นด้วยกับข้อที่ 3 ก็จะต้องเห็นด้วยกับข้อที่ 1 และ 2 เป็นต้น

ตัวอย่างเช่น หากการยอมรับอำนาจหน้าที่ของผู้บังคับบัญชาขึ้นอยู่กับการยอมรับของผู้ใต้บังคับบัญชา เราอาจทดลองวัดการยอมรับอำนาจหน้าที่โดยการวัดระดับการยอมรับ และปฏิบัติตามคำสั่งของผู้บังคับบัญชานั้นๆ จากผู้ใต้บังคับบัญชา ซึ่งอาจกำหนดขอบเขตการยอมรับคำสั่งที่ผู้ใต้บังคับบัญชาจะปฏิบัติใน 4 ระดับ คือ

- 1) เมื่อคำสั่งนั้นเป็นเรื่องเกี่ยวกับงานในหน้าที่และเกิดขึ้นในเวลางาน
- 2) เมื่อคำสั่งนั้นเป็นเรื่องเกี่ยวกับงานในหน้าที่แต่เกิดขึ้นนอกเวลางาน
- 3) เมื่อคำสั่งนั้นไม่ใช่งานในหน้าที่แต่เกิดขึ้นในเวลางาน
- 4) เมื่อคำสั่งนั้นไม่ใช่งานในหน้าที่และเกิดขึ้นนอกเวลางาน

การเลือกตอบว่าเห็นด้วยหรือไม่ในระดับข้างต้นนี้ จะบอกให้ทราบถึงระดับการยอมรับอำนาจหน้าที่ของผู้บังคับบัญชาจากผู้ใต้บังคับบัญชาว่า จะยอมให้ผู้บังคับบัญชามีอำนาจเหนือตนเพียงไร หรือผู้ใต้บังคับบัญชาจะยอมรับอำนาจและปฏิบัติงานให้แก่ไหนเพียงไร ซึ่งอาจนำคำตอบนั้นมาเขียนวิเคราะห์ให้รูปตารางแบบ Scalogram Method ได้ดังนี้

จะเห็นได้ว่า ถ้าผู้ตอบเลือกตอบข้อ 4 คือ เห็นด้วยที่จะปฏิบัติตามคำสั่งของผู้บังคับบัญชา เมื่อคำสั่งนั้นไม่ใช้งานในหน้าที่และเกิดขึ้นนอกเวลา ก็ย่อมเป็นที่แน่นอนว่า เขาจะเลือกตอบว่าเห็นด้วยกับข้อ 1, 2 และ 3 นั่นคือ ถ้าเป็นคำสั่งของผู้บังคับบัญชาแล้ว ไม่ว่าจะอยู่ในสถานการณ์ใด เขาก็ยินดีที่จะปฏิบัติตามคำสั่งเสมอ แสดงถึงการยอมรับผู้บังคับบัญชาในระดับที่สูงมาก

ในขณะที่ถ้าผู้ตอบเลือกข้อ 2 ก็เป็นเพียงว่าเขายินดีที่จะปฏิบัติตามคำสั่ง ถ้างานนั้นเป็นงานในหน้าที่ความรับผิดชอบ ไม่ว่าจะในหรือนอกเวลางาน แต่ไม่ยินดีที่จะปฏิบัติตามคำสั่ง ถ้างานนั้นๆ ไม่ใช้งานในหน้าที่รับผิดชอบของเขา

จากลักษณะดังกล่าว เราจะสร้าง Scalogram เพื่อทดสอบว่าทัศนคติหรือสิ่งที่กำลังศึกษาอยู่นั้นมีมิติเดียวหรือไม่ เพราะถ้ามีมิติเดียวก็จะสามารถวิเคราะห์ได้ในรูปตารางแบบ Scalogram ได้ ซึ่งหมายความว่า เมื่อรู้คะแนนของบุคคลหนึ่ง ก็จะได้รู้ได้ทันทีว่าบุคคลนั้น เลือกข้อความใดในมาตรวัดนั้นบ้าง

การวิเคราะห์ด้วยตาราง Scalogram จึงเป็นความพยายามที่จะกำหนดให้แต่ละรายการของข้อความในมาตรวัดนั้นๆ ว่าควรจะได้คะแนนอย่างไร และมีการเรียงลำดับความสำคัญและเชื่อมโยงระหว่างข้ออย่างไร ทั้งเพื่อช่วยให้การสร้างมาตรวัดแบบกัทแมนมีความเป็นไปได้มากขึ้น

ซึ่งการทดสอบว่ามาตรวัดแบบกัทแมนที่สร้างขึ้นเป็นมาตรวัดในอุดมคติและยอมรับได้หรือไม่ สามารถดูได้จากกลุ่มตัวอย่างขนาดใหญ่ที่ผลการวัดอยู่ในรูปของตารางแบบ Scalogram และมีความสอดคล้องหรือค่าสัมประสิทธิ์ (Reproducibility Coefficient) เท่ากับ 0.90

ข้อเสียของมาตรวัดแบบกัทแมน

เป็นเพียงเทคนิคที่จะทดสอบว่าชุดข้อความนั้นคือมาตรวัดในมิติเดียวกันหรือไม่ แต่ไม่ได้ให้คำแนะนำว่าจะหาหรือสร้างชุดข้อความนั้นด้วยวิธีใด

นอกจากนี้ ได้มีการตั้งข้อสังเกตเกี่ยวกับมาตรวัดมิติเดียว (Unidimensional) ไว้ดังนี้

1) มาตรวัดประเภทนี้ ไม่สามารถใช้เป็นฐานในการวัดทัศนคติต่อปรากฏการณ์ที่มีความสลับซับซ้อน หรือไม่อาจใช้เป็นฐานในการพยากรณ์ หรือคาดการณ์เกี่ยวกับพฤติกรรมของบุคคลได้ เพราะคนเราอาจมีความรู้สึกได้หลายๆ ด้าน ซึ่งการถามเพียงด้านเดียว อาจจะทำให้ได้ภาพที่ไม่ชัดเจนเท่ากับการวัดในหลายมิติ (Multi-Dimensional)

2) การกำหนดว่าสิ่งใดมีเพียงมิติเดียว สิ่งใดมีหลายมิติ ยังเป็นสิ่งที่ไม่ชัดเจน เพราะว่าเมื่อขอความหมายมากขึ้น ก็ยังมีได้หลายมิติมากขึ้น

และสิ่งที่เป็นมิติเดียวในช่วงเวลาหนึ่ง อาจกลายเป็นหลายมิติได้ในอีกสถานการณ์หนึ่ง ขึ้นอยู่กับเวลา สถานที่และกลุ่มบุคคล

เช่น การวัดความมั่งคั่งในทรัพย์สินที่มีอยู่ในชนบทด้วยมาตรวัดแบบกัทแมน ซึ่งกำหนดรายการให้เลือกระหว่าง บ้าน โทรทัศน์ รถยนต์ แล้วผู้ตอบเลือกตอบว่ารถยนต์ก็มีความเป็นไปได้มากกว่าผู้นั้นจะมีบ้านและโทรทัศน์แล้ว แต่ถ้านำแบบสอบถามเดียวกันมาวัดคนในกรุงเทพมหานคร การเลือกตอบว่าเขามีรถยนต์ ย่อมหมายถึงหมายความว่า บุคคลนั้นมีบ้านเป็นของตนเองด้วย เป็นต้น

การวัดทัศนคติ

มีผู้ให้ความหมายของ "ทัศนคติ" ไว้ดังนี้ คือ

พจนานุกรมทางการศึกษา (Dictionary of Education) ให้คำจำกัดความว่า "ทัศนคติ" หมายถึง แนวโน้มหรือท่าทีที่มีต่อสิ่งหนึ่งหรือสถานการณ์หนึ่ง โดยปกติจะมีความรู้สึกและอารมณ์เกี่ยวข้องอยู่ในความเห็นนั้น ทัศนคติสังเกตไม่ได้ แต่จะอนุมานได้จากพฤติกรรมทั้งที่เป็นวาจา และท่าทาง

เทอร์สโตน (Therstone) ให้ความหมายว่า "ทัศนคติ" เป็นผลรวมทั้งหมดของความรู้สึก อคติ ความคิด ความกลัว ต่อบางสิ่งบางอย่าง การแสดงออกทางด้านการพูด เช่น ความคิดเห็น ดังนั้นเราวัดทัศนคติได้โดยวัดความคิดเห็นของบุคคลที่มีต่อสิ่งเร้าต่างๆ

อลพอร์ต (Allport) กล่าวว่า "ทัศนคติ" เป็นสภาพความพร้อมของสมองและประสาท ที่จัดไว้เป็นระเบียบโดยอาศัยประสบการณ์เข้าช่วย ซึ่งจะมีผลโดยตรงต่อการตอบสนองของบุคคลที่มีต่อสิ่งต่างๆ หรือสถานการณ์ต่างๆ ทั้งหมดที่เกี่ยวข้องกับบุคคลนั้น

อนาสตาซี (Anastasi) กล่าวว่า "ทัศนคติ" เป็น ความโน้มเอียงที่จะมีปฏิกิริยาตอบสนองสิ่งเร้าที่กำหนดให้เป็นพวกๆ ในทางชอบหรือไม่ชอบ

กิลฟอร์ด (Gilford) ให้ความหมายว่า "ทัศนคติ" เป็นความโน้มเอียงของจิตใจในทางที่ชอบ เห็นด้วย สนับสนุน หรือ ไม่ชอบ ไม่เห็นด้วย ไม่สนับสนุนต่อการกระทำหรือต่อสิ่งใดสิ่งหนึ่งในสังคม

เชดคัทท์ โฆวาลินส์ ให้ความหมายว่า "ทัศนคติ" คือความรู้สึกของบุคคลที่มีต่อสิ่งต่างๆ อันเป็นผลเนื่องมาจากการเรียนรู้แลประสบการณ์ และเป็นตัวกระตุ้นให้บุคคลแสดงพฤติกรรมหรือแนวโน้มที่จะตอบสนองต่อสิ่งเร้านั้นๆ ไปในทิศทางใดทิศทางหนึ่ง อาจเป็นไปในทางสนับสนุนหรือคัดค้านก็ได้ ทั้งนี้ขึ้นอยู่กับกรอบการเรียนรู้ระเบียบของสังคม (Socialization) ทัศนคติจึงเป็นสิ่งที่ปลูกฝังและ พัฒนาได้ และเป็นสิ่งที่กำหนดให้บุคคลประพฤติปฏิบัติ หรือตอบสนองในการที่จะชอบหรือไม่ชอบต่อวัตถุ บุคคล สถานการณ์ หรือแม้แต่ความคิดเห็นต่างๆ

พิพัฒน์ ชารุณนทกร อธิบายความหมายของทัศนคติไว้ในบทความทางวิชาการเรื่อง "ทัศนคติกับการประยุกต์ใช้ทางการตลาด" อย่างน่าสนใจว่า หมายถึง ความโน้มเอียงที่จะตอบสนองในสิ่งที่ชอบหรือไม่ชอบต่อสิ่งเร้าที่กำหนดให้ คำว่า "สิ่งเร้า" ที่ใช้ในความหมายของ

ประการแรก คือ เป็นสิ่งที่เรียนรู้ได้ซึ่งจะเป็นการเรียนรู้จากประสบการณ์ทางตรงที่ประสบกับตนเอง หรือประสบการณ์ทางอ้อมที่ได้รับข่าวสารข้อมูลมา ซึ่งล้วนแล้วแต่ทำให้เกิดการเรียนรู้และทำให้เกิดทัศนคติได้ในที่สุด

ประการที่สอง คือ ความสม่ำเสมอ ทัศนคติมีความสม่ำเสมอ แต่ไม่ได้หมายความว่ามีความคงทนถาวร เพราะว่าทัศนคติเป็นสิ่งที่เรียนรู้ได้ ดังนั้นเราจึงสามารถที่จะเปลี่ยนแปลงทัศนคติได้ และถึงแม้ว่าบุคคลนั้นจะมีได้เปลี่ยนทัศนคติ แต่ก็ไม่ได้หมายความว่าเขาจะต้องทำตามหรือแสดงพฤติกรรมทัศนคติเสมอไป

ประการสุดท้าย คือ อิทธิพลของสถานการณ์จะเป็นสิ่งที่ทำให้เกิดทัศนคติ บางอย่างก็อาจจะเหมือนหรือแตกต่างไปจากทัศนคติเดิมหรืออาจจะทำให้บุคคลแสดงพฤติกรรมที่แตกต่างไปจากทัศนคติที่มีอยู่

จากการให้ความหมายของนักวิชาการในข้างต้น สามารถสรุปได้ว่า "ทัศนคติ" เป็นพฤติกรรมอย่างหนึ่งที่อยู่ภายในจิตใจของบุคคล เกิดขึ้นจากการรับรู้และประสบการณ์ต่างๆ อยู่ในสภาพที่พร้อมจะตอบสนองต่อสิ่งเร้าต่างๆ ในลักษณะที่สนับสนุนหรือไม่สนับสนุน ทัศนคติเป็นสิ่งที่สังเกตไม่ได้ วัด

โดยตรงไม่ได้ แต่วัดได้จากความคิดเห็นของบุคคลที่มีต่อสิ่งเร้าต่างๆ ซึ่งนักจิตวิทยาได้แยกตัวแปรที่เกี่ยวข้องไว้ 3 ประเภทคือ

1) ตัวแปรที่วัดทัศนคติเกี่ยวกับการรับรู้หรือความเชื่อ (Cognitive Variables) เป็นการวัดการรับรู้หรือความเชื่อหรือความคิดว่าสิ่งๆ หนึ่งเป็นอย่างไร เช่น การประเมินความขยันของพนักงาน หรือการประเมินว่าผู้นำแบบใดจึงจะเกิดประสิทธิภาพสูงสุดของหน่วยงาน เป็นต้น

2) ตัวแปรที่วัดทัศนคติเกี่ยวกับความสมัครใจที่จะประกอบพฤติกรรม (Behavioral Variables) เป็นการวัดความสมัครใจที่จะประกอบพฤติกรรมหนึ่งๆ หรือยินยอมหรือสนับสนุนให้เกิดพฤติกรรมนั้นๆ เช่น ถ้ากล่าวว่า "ควรจะให้บุคคลที่ทุจริตต่อหน้าที่ออกจากงาน" การจะตอบว่าเห็นด้วยหรือไม่เห็นด้วยนั้นขึ้นอยู่กับว่า ความเชื่อหรือความคิดของเราที่มีต่อการกระทำเพื่อ "ไล่ออก" นั้น เห็นว่าเป็นการสมควรกับพฤติกรรมของบุคคลที่ทุจริตหรือไม่

3) ตัวแปรที่วัดทัศนคติเกี่ยวกับความรู้สึกส่วนตัว (Affective Variables) เป็นเรื่องของความรู้สึกชอบไม่ชอบอย่างไร พอใจหรือไม่พอใจอย่างไร ซึ่งกล่าวได้ว่าเป็นเรื่องความรู้สึกส่วนตัว ต่างจิต

ต่างใจ โดยแท้

ดังนั้น ในการสร้างมาตรวัดทัศนคติ ส่วนใหญ่จะให้ความสนใจรวมทุกข้อในมาตรวัดโดยไม่แยกดูเฉพาะข้อใดข้อหนึ่ง ซึ่งการเลือกข้อความเพื่อบรรจุไว้ในมาตรวัดนี้ อาศัยหลักเกณฑ์ 2 ประการด้วยกัน คือ

ประการแรก : ข้อความหนึ่งจะต้องเกี่ยวข้องกับหรือเชื่อมโยงกับคุณสมบัติที่จะวัด และนักวิจัยต้องพร้อมที่จะอธิบายว่าเกี่ยวข้องกับสัมพันธ์กันอย่างไร

ประการที่สอง : ข้อความนั้นๆ ควรจะสามารถจำแนกหรือแยกความแตกต่างในแต่ละระดับได้ เช่น ข้อความที่ทุกคนเห็นด้วยทั้งหมด หรือไม่เห็นด้วยทั้งหมดนั้น เป็นข้อความที่ไม่ดี เพราะไม่สามารถแยกความแตกต่างในสิ่งที่ทำการวัดได้ และจะทำให้การวัดนั้นๆ ไม่มีผลใดๆ เกิดขึ้นตามมา

ข้อเสนอแนะในการเขียนข้อความวัดทัศนคติ

1) ควรเป็นข้อความที่สั้นไม่ยุ่งยาก ไม่เป็นคำถามที่ใช้คำปฏิเสธซ้ำ เช่น "ท่านเห็นด้วยหรือไม่ว่าไม่มีผู้ใดที่ไม่เห็นด้วยกับนโยบายส่งเสริมเสรีภาพทางด้านการรับรู้ข่าวสารของประชาชน"

2) ไม่ควรเป็นข้อความที่ถามถึงสองสิ่งพร้อมกัน เช่น "ท่านอ่านหนังสือพิมพ์ หรือฟังวิทยุเป็นประจำหรือไม่" เพราะจะเป็นการลำบากต่อการตีความหมายคำตอบที่ว่า ตอบรับหรือปฏิเสธส่วนใดของคำถาม ถ้าต้องการจะถามควรแยกถามทีละข้อ

3) ควรหลีกเลี่ยงที่จะนำข้อความหรือสุภาษิตที่เป็นที่รู้จักกันดีมาถาม เพราะผู้ตอบมักจะให้คำตอบทันทีโดยไม่คิดให้ลึกซึ้ง ยกเว้นแต่ว่าคำตอบในลักษณะนั้นเป็นสิ่งที่เราต้องการ

4) ข้อความทัศนคติที่ดีควรเป็นข้อความสดที่ทำให้ผู้ตอบคิด และต้องหาจุดยืนเองโดยทั่วไปแล้ว มาตรฐานวัดทัศนคติที่นิยมใช้กันแพร่หลายก็คือ มาตรฐานที่ยกมาเป็นตัวอย่างในตอนที่แล้ว ทั้งมาตรฐานแบบลิเคิร์ท เทอร์สโตน และกัทแมน อย่างไรก็ตาม นอกเหนือจากมาตรฐานทั้ง 3 แบบดังกล่าว ยังมีมาตรฐานวัดอื่นๆ อีกมากมาย อย่างไรก็ตาม ไม่ว่าจะเป็นการสร้างมาตรฐานวัดแบบใดก็ตาม สิ่งสำคัญที่มองข้ามไปไม่ได้ก็คือ การพิจารณาความเชื่อถือได้ และความแม่นยำของมาตรฐานวัดซึ่งจะได้กล่าวถึงในตอนต่อไป

ความแม่นยำของการวัด(Validity)

ถ้ามาตรฐานวัดหรือเครื่องชี้วัดหนึ่งๆ สามารถวัดสิ่งที่ต้องการวัดได้ ก็กล่าวได้ว่ามาตรฐานวัดหรือเครื่องชี้วัดนั้นๆ มีความแม่นยำ (Valid) เช่น

เราต้องการทดสอบความรู้ความเข้าใจที่ประชาชนมีต่อปัญหาสังคม เราก็ควรจะวัดความรู้ความเข้าใจของประชาชนกลุ่มนั้นในเรื่องเกี่ยวกับปัญหาสังคม

โดยผู้วัดจะต้องเข้าใจเสียก่อนว่าปัญหาสังคมนั้นคืออะไร มีอะไรบ้างที่ควรจะนำมาทดสอบความรู้ความสามารถของประชาชน และวัดความรู้ดังกล่าวได้อย่างไร

ตัวอย่างของการวัดที่ขาดความแม่นยำในเรื่องปัญหาสังคม คือการออกข้อสอบปัญหาสังคมแล้วให้ผู้ตอบตอบเป็นภาษาอังกฤษ ผลการวัดดังกล่าวไม่ได้บ่งชี้ว่าผู้ตอบมีความรู้ความเข้าใจในปัญหาสังคม แต่เป็นการวัดความสามารถในการเขียนภาษาอังกฤษมากกว่า

สมมติว่ามีการออกข้อสอบเป็นภาษาไทย และตอบเป็นภาษาไทย แต่การออกข้อสอบไม่ถูกต้อง เพราะคำถามไม่ได้เป็นตัวแทนของเรื่องราวที่เป็นความรู้ความเข้าใจปัญหา

ความผิดพลาด (Error) ของการวัด

ในการวัดใดๆ มักจะมีความผิดพลาดเกิดขึ้นเสมอไม่มากก็น้อย ความผิดพลาด (Error) ที่เกิดขึ้นนี้มีอยู่ 2 ประเภทคือ ความผิดพลาดสุ่ม (Random Error) และความผิดพลาดไม่สุ่ม (Nonrandom Error)

1) ความผิดพลาดสุ่ม (Random Error)

คือ ความผิดพลาดที่เกิดขึ้นไม่สม่ำเสมอ ไม่มีระบบ ไม่สามารถคาดคะเนทิศทางได้ เช่น การยิงปืน ความผิดพลาดสุ่มคือการกระจายของลูกปืนไปรอบๆ เป้าที่ต้องการจะยิงอย่างไม่มีระบบ อาจจะสูง อาจจะต่ำ อาจจะไปทางด้านซ้าย อาจจะไปทางด้านขวา ปืนที่ให้ผลการยิงดังกล่าวเรียกว่า ขาดความเชื่อถือได้ แต่ถ้าปืนดังกล่าวให้ผลการยิงตรงเป้า ก็เรียกได้ว่ามีความเชื่อถือได้มาก

ในการวิจัย ไม่ว่าจะเป็นในทางสังคมศาสตร์และวิทยาศาสตร์ จะมีความผิดพลาดสุ่มเกิดขึ้นได้เสมอ โดยเฉพาะอย่างยิ่งในงานวิจัยด้านสังคม ความผิดพลาดอาจเกิดขึ้นจากตัวคำถามในแบบสอบถาม ผู้สัมภาษณ์ ผู้ให้สัมภาษณ์ การลงรหัสข้อมูล การ key ข้อมูล เพื่อบันทึกลงฮาร์ดดิสก์ ผู้วิจัยจึงควรพยายามควบคุมมิให้ความผิดพลาดสุ่มเกิดขึ้นมากจนข้อมูลขาดความเชื่อถือได้

2) ความผิดพลาดไม่สุ่ม (Nonrandom Error)

เป็นความผิดพลาดที่เกิดขึ้นอย่างเป็นระบบ มีทิศทางที่แน่นอน เช่น ในการยิงปืน หากมีบางสิ่งบางอย่างเข้ามารบกวนจนทำให้ผลการยิงเบี่ยงเบนไปจากเป้า เราเรียกความเบี่ยงเบนดังกล่าวนี้ว่า ความผิดพลาดอย่างมีระบบ เพราะเป็นความผิดพลาดที่เกิดขึ้นเนื่องจากมีปัจจัยอย่างอื่นที่ไม่ใช่ปืน (เครื่องมือวัด) เข้ามาเกี่ยวข้อง ดังนั้น หากเราสามารถลดความผิดพลาดที่มีรบกวนสถานการณ์นั้นๆ ให้น้อยลงได้ การยิงปืนก็จะแม่นยำมากขึ้น

ความผิดพลาดที่สามารถป้องกันได้ หรือความผิดพลาดไม่สุ่มนี้ ถือว่าเป็นหัวใจสำคัญของการศึกษาความแม่นยำของมาตรวัด เนื่องจากในสถานการณ์ของการวัดสิ่งใดสิ่งหนึ่งในการวิจัยนั้น

ผู้วิจัยต้องแน่ใจว่าได้ป้องกันมิให้มีสิ่งอื่นใดมารบกวนผลของการวัดอย่างรอบคอบที่สุด

การวัดโดยทั่วไปต้องอาศัยแนวคิด (Concept) เกี่ยวกับสิ่งที่เราต้องการวัด การระบุคุณสมบัติ และการหาตัวบ่งชี้คุณสมบัติ เราจะต้องแน่ใจว่าตัวบ่งชี้ที่กำหนด สามารถชี้ให้เห็นถึงคุณสมบัติของสิ่ง (แนวคิด) ที่เราต้องการวัด และไม่ได้เป็นตัวแทนหรือตัวบ่งชี้ของสิ่งอื่น

โดยสรุป ความแม่นยำของการวัด เป็นเรื่องที่ต้องอาศัยความคิด (หรือทฤษฎี) มาช่วยตัดสินว่าเราวัดได้ถูกต้องหรือไม่ ความแม่นยำของการวัดพิสูจน์ได้จากความสอดคล้องกันระหว่างตัวบ่งชี้ ว่าได้วัดสิ่งที่ต้องการจะวัดหรือไม่ หรือว่าไปวัดเอาสิ่งอื่นที่เข้ามารบกวนสถานการณ์ของการวัดจนทำให้เครื่องมือวัดผิดพลาดไป

ชนิดของความแม่นยำ

เนื่องจากความผิดพลาดไม่สุ่ม (Nonrandom Error) ที่เกิดขึ้นกับการวัด มีหลายประเภท ความแม่นยำของการวัดซึ่งขึ้นอยู่กับความผิดพลาดไม่สุ่มจึงมีหลายประเภท เช่นเดียวกัน

เคอร์ลิงเจอร์ (Kerlinger) ได้จำแนกความแม่นยำออกเป็น

1) ความแม่นยำตามเนื้อหา (Content Validity)

2) ความแม่นยำตามความสัมพันธ์กับเกณฑ์ (Criterion-related Validity) ซึ่งอาจแบ่งย่อยออกเป็น ความแม่นยำตามความสอดคล้องกับเกณฑ์ในปัจจุบัน (Concurrent Validity) กับความแม่นยำพยากรณ์ (Predictive Validity) และ

3) ความแม่นยำตามมโนทัศน์ (Construct Validity)

ความแม่นยำตามเนื้อหา (Content Validity)

แนวคิดหลักของความแม่นยำตามเนื้อหา คือ จะถือว่ามาตรวัดหนึ่ง ๆ มีความแม่นยำ ถ้า

(ก) ข้อความในมาตรวัดนั้น ๆ เป็นข้อความที่ตรงเป้าหรือตรงประเด็นกับสิ่งที่ต้องการวัด และ

(ข) ข้อความในมาตรวัดนั้น ๆ ครอบคลุมปรากฏการณ์ที่ต้องการวัดอย่างกว้างขวางเพียงพอ

พิจารณาตามข้อ (ก) การที่ข้อความในมาตรวัดหนึ่ง ๆ เป็นข้อความที่ตรงเป้า หรือตรงประเด็นกับสิ่งที่ต้องการวัดหรือไม่เพียงใดนั้น เรียกว่า เป็นการพิจารณาความแม่นยำจากรูปลักษณะของมาตรวัด (Face Validity) เช่น

สมมติว่า ต้องการวัด "ความผูกพันต่อหน่วยงาน" ข้อความที่กำหนดขึ้นมาเพื่อบรรจุในมาตรวัดนี้ อาจประกอบด้วยข้อความหลายๆ ข้อด้วยกัน อาทิ

"ฉันรักหน่วยงานของฉัน"

"ฉันไม่คิดจะเปลี่ยนงาน"

"แม้ว่ามีงานอื่นที่ดีกว่า ฉันก็ยังอยากจะทำงานที่เดิม"

"เพื่อนร่วมงานของฉันทำงานกันอย่างสามัคคีเสมือนฉันท์" ฯลฯ

ซึ่งการพิจารณาความแม่นยำจากรูปลักษณะของมาตรวัด ก็ต้องพิจารณาเนื้อหาของข้อความแต่ละข้อว่าตรงประเด็นหรือไม่ ในตัวอย่างที่ให้ไว้ ข้อความ 3 ข้อแรกน่าจะตรงประเด็น ส่วนข้อความสุดท้ายยังอาจต้องถกเถียงกันต่อไปว่า สมควรจะบรรจุในมาตรวัดหรือไม่

พิจารณาตามข้อ (ข) เป็นการพิจารณาว่าเนื้อหาของมาตรวัดหนึ่ง ๆ มีลักษณะครอบคลุมเรื่องที่ต้องการวัดไว้ทั้งหมดหรือไม่ ซึ่งกระทำได้โดยการตรวจสอบเนื้อหาของมาตรวัด (Content Validity) นั้นเอง เช่น

ในกรณีการวัดความผูกพันต่อหน่วยงาน ควรต้องตรวจสอบดูว่า เนื้อหาครอบคลุม
สิ่งต่างๆ ที่บ่งชี้ถึงความผูกพันต่อหน่วยงานหรือไม่เพียงใด เป็นต้นว่า การยินดีที่จะทำงาน
ล่วงเวลา การปกป้องหน่วยงานจากการวิพากษ์วิจารณ์ในทางลบ ความรู้สึกภาคภูมิใจใน
หน่วยงาน ฯลฯ

ในทางปฏิบัติ การที่จะให้ได้มาซึ่งความแม่นยำตรงตามเนื้อหา นั้น นักวิจัยควรต้อง
กำหนดทุกมิติ หรือทุกแง่มุมที่เกี่ยวข้อง จากนั้นจึงพยายามที่จะสร้างเครื่องชี้วัดให้ได้ในทุกๆ
มิติ ซึ่งในบางครั้ง ถ้าหากสิ่งที่ต้องการวัดมีความซับซ้อนในตัวเองมาก (เช่น การวัด
"ประสิทธิผล" ของหน่วยงาน เป็นต้น) นักวิจัยอาจแยกมิติบางมิติออกมาเป็นมาตรวัดต่างหากก็
ได้ ทั้งนี้เพื่อความสะดวกในการสร้างมาตรวัด

การสร้างมาตรวัดที่มีความแม่นยำตรงตามเนื้อหา นั้น ควรพิจารณาความแม่นยำ
ก่อนที่จะดำเนินการเก็บข้อมูล ทั้งนี้ผู้วิจัยจะต้องมีความแตกฉานในเนื้อหาที่จะวิจัย เพื่อให้
สามารถเข้าใจได้ชัดเจน และเมื่อมีความต้องการตรวจสอบอีกครั้ง ก็อาจนำรายการที่เลือกเฟ้น
แล้วนั้นไปให้ผู้เชี่ยวชาญหรือผู้รู้พิจารณาให้คำปรึกษาอีกครั้งหนึ่ง ก่อนลงมือเก็บรวบรวมข้อมูล
หรือกล่าวให้ชัดเจนก็คือ ความแม่นยำตรงตามเนื้อหาต้องพยายามสร้างก่อนออกแบบสอบถาม
และก่อนการเก็บข้อมูล

อนึ่ง การพิจารณาความแม่นยำตรงของมาตรวัดตามเนื้อหาโดยผู้ทรงคุณวุฒิหรือ
ผู้เชี่ยวชาญนั้น ซัชแมน (Suchman) เรียกว่า Consensual Validity หรือ Experts' Judgement
of Content

ความแม่นยำตรงตามความสัมพันธ์กับเกณฑ์ (Criterion-related Validity)

ความแม่นยำตรงตามความสัมพันธ์กับเกณฑ์ หมายถึง การที่มาตรวัดหนึ่งๆ ที่สร้าง
ขึ้นมาใหม่นั้น สอดคล้องหรือสัมพันธ์กับเกณฑ์อื่นที่มีอยู่เดิม (และที่ถือว่าถูกต้องแล้ว) มากน้อย
เพียงใด เช่น

การวัดความรู้ภาษาอังกฤษเพื่อไปเรียนต่อในต่างประเทศ มักใช้การทดสอบ
TOEFL โดยถือว่าข้อสอบ TOEFL เป็นเกณฑ์มาตรฐานสากล สมมติว่าต่อมาหน่วยงานราชการ
ซึ่งทำหน้าที่จัดสอบความรู้ภาษาอังกฤษให้แก่ข้าราชการที่ชิงทุนไปศึกษาต่อต่างประเทศ ได้
ตัดสินใจจะพัฒนาชุดข้อสอบภาษาอังกฤษขึ้นเอง วิธีหนึ่งที่จะทดสอบความแม่นยำตรงของชุด
ข้อสอบที่สร้างขึ้นใหม่นี้ ก็โดยการเปรียบเทียบว่าผลของการสอบในครั้งหนึ่งๆ สอดคล้องกับผล
การสอบ TOEFL มากน้อยเพียงใด โดยให้กลุ่มตัวอย่างกลุ่มเดียวกันสอบข้อสอบทั้งสองชุดแล้ว
เทียบผลระหว่างกัน

เกณฑ์ที่ใช้วัดความแม่นยำตรงดังกล่าวนี้อาจแบ่งได้ 2 ชนิดคือ

- 1) ปรากฏในปัจจุบันเรียกว่า "ความแม่นยำตรงตามความสอดคล้องกับเกณฑ์ใน
ปัจจุบัน" (Concurrent Validity) ดังตัวอย่างการทดสอบความรู้ภาษาอังกฤษที่เพิ่งอ้าง
- 2) เป็นเรื่องที่เกิดขึ้นในอนาคต ซึ่งเรียกว่า "ความแม่นยำตรงพยากรณ์" (Predictive)
ซึ่งจะทดสอบได้นั้น ก็ต้องรอให้เหตุการณ์ในอนาคตเกิดขึ้นเสียก่อน เช่น

จากตัวอย่างการทดสอบความรู้ภาษาอังกฤษข้างต้น เกณฑ์ที่อาจกำหนดขึ้นเพื่อทดสอบความแม่นยำคือ คะแนนการศึกษาในมหาวิทยาลัยต่างประเทศที่สอนเป็นภาษาอังกฤษ (โดยอาจยกเว้นวิชาที่เน้นความรู้ทางด้านคณิตศาสตร์) กล่าวคือ การทดสอบภาษาอังกฤษชุดหนึ่งๆ จะถือว่ามีความแม่นยำพยากรณ์ต่อเมื่อสามารถนำผลการทดสอบไปเป็นฐานในการพยากรณ์แม่นยำได้ว่า ใครจะได้คะแนนการเรียนอย่างไรบ้าง

อาจสังเกตได้ว่า ความแม่นยำพยากรณ์ที่ดี ความแม่นยำตามความสอดคล้องกับเกณฑ์ในปัจจุบันก็ดี ต่างตั้งอยู่บนฐานของความสัมพันธ์กับเกณฑ์ที่กำหนดไว้ก่อนแต่เบื้องต้น จะต่างกันก็เพียงเงื่อนไขของเวลาเท่านั้น

ดังนั้นจะเห็นได้ว่า สิ่งที่สำคัญในการพิจารณาความแม่นยำตามความสัมพันธ์กับเกณฑ์ก็คือ จะต้องมีความสัมพันธ์ที่มีความแม่นยำและเชื่อถือได้ก่อนแล้ว เพื่อเป็นบรรทัดฐานหรือมาตรฐานที่ใช้เพื่อเทียบเคียงความแม่นยำของมาตรวัดที่สร้างขึ้นใหม่

ความแม่นยำตามมโนทัศน์ (Construct Validity)

การสร้างมาตรวัด บ่อยครั้งก็เพื่อมุ่งวัดปรากฏการณ์ที่เป็นมโนทัศน์ (Construct) ที่มีลักษณะเป็นนามธรรม เช่น

กำลังขวัญในการทำงาน ความผูกพันต่อหน่วยงาน ความขัดแย้งภายในองค์กร ลักษณะการเป็นผู้นำ ความซับซ้อนของโครงสร้างองค์กร ประสิทธิภาพในการทำงาน ฯลฯ เป็นต้น

ซึ่งมโนทัศน์ดังกล่าวไม่สามารถวัดได้ด้วยเครื่องชี้วัดใดเครื่องชี้วัดหนึ่งโดยเฉพาะแท้ที่จริงแล้ว มาตรวัดเหล่านี้มีวัตถุประสงค์เพื่อวัดมโนทัศน์ในภาพรวม ซึ่งค่อนข้างจะเป็นนามธรรม ในขณะที่เครื่องชี้วัดต่างๆ ที่บรรจุอยู่ในมาตรวัดแต่ละมาตรวัด คงจะมีลักษณะเป็นรูปธรรมพอสมควร การพิจารณาความแม่นยำของมาตรวัดมโนทัศน์นั้น เรียกตรงตัวว่า ความแม่นยำตามมโนทัศน์

นักวิชาการที่บุกเบิกเสนอเรื่องความแม่นยำตามมโนทัศน์นี้ได้แก่ ครอนบาคและมิลล์ (Cronbach and Meehl) โดยได้เสนอว่า การนิยามหรือการให้คำจำกัดความเกี่ยวกับสิ่งที่เป็นมโนทัศน์นั้น ส่วนหนึ่งมักจะประกอบด้วยข้อความที่อธิบายถึงความสัมพันธ์ระหว่างมโนทัศน์กับตัวแปรอื่นๆ ที่เกี่ยวข้อง

ดังนั้น ในการพิจารณาถึงความแม่นยำตามมโนทัศน์ จึงเป็นการเหมาะสมที่จะตั้งคำถาม เช่น

ถ้าทราบคะแนนจากการวัดตามมโนทัศน์หนึ่งๆ แล้ว จะสามารถคาดการณ์หรือพยากรณ์คะแนนหรือค่าของตัวแปรอื่นๆ ที่มีความสัมพันธ์กับมโนทัศน์นั้นๆ ได้หรือไม่ มองอีกแง่หนึ่งคือ เท่ากับเป็นการตั้งสมมติฐานไว้ล่วงหน้าว่า รูปแบบความสัมพันธ์น่าจะเป็นไปในรูปแบบใด การทดสอบสมมติฐานนี้จึงเป็นการทดสอบความแม่นยำของมาตรวัด ซึ่งเป็นที่น่าสังเกตว่า การระบุความสัมพันธ์ระหว่างมโนทัศน์หนึ่งๆ กับตัวแปรอื่นๆ ที่เกี่ยวข้องนั้น เป็น

สมมติฐานเกี่ยวกับความสัมพันธ์ระหว่างมโนทัศน์กับตัวแปรอื่นๆ ที่เกี่ยวข้องนั้น อาจได้มาจากการอนุมานหลายชั้นก็ได้

ที่สำคัญก็คือ สมมติฐานที่ตั้งไว้ควรมีความหลากหลาย โดยครอบคลุมจำนวนตัวแปรที่เกี่ยวข้องสัมพันธ์กับมโนทัศน์ให้มากที่สุดเท่าที่จะเป็นไปได้ โดยในการทดสอบความแม่นยำตรงตามมโนทัศน์นั้น จะต้องยืนยันสมมติฐานที่ตั้งไว้ และถ้าหากผลออกมาในเชิง ปฏิเสธ สมมติฐานแม้แต่สมมติฐานเดียว ก็ส่งผลไปในทางลบต่อความแม่นยำตรงของมาตรวัด

ยกตัวอย่างเช่น

สมมติว่าได้สร้างมาตรวัดความซับซ้อนขององค์การขึ้นมา แล้วทดสอบความแม่นยำตรงตามมโนทัศน์โดยพิจารณาความสัมพันธ์ระหว่าง "ความซับซ้อน" กับ "ขนาดขององค์การ" และ "ความเป็นทางการ" ซึ่งตามแนวคิดเชิงทฤษฎี ความซับซ้อนควรมีความสัมพันธ์เชิงบวกกับตัวแปรทั้งสองตัวหลัง

หลังจากวิเคราะห์ข้อมูลแล้ว หากพบว่ามีความสัมพันธ์เชิงบวกระหว่างความซับซ้อนกับขนาดขององค์การ และความเป็นทางการ ก็ถือว่ายืนยันความแม่นยำตรงตามมโนทัศน์ของมาตรวัด แต่ในทางกลับกัน หากผลการวิเคราะห์ปฏิเสธความสัมพันธ์ระหว่างความซับซ้อนกับขนาดขององค์การ และความเป็นทางการ ก็คงสรุปได้ว่า มาตรวัดความซับซ้อนขาดความแม่นยำตรงตามมโนทัศน์

แคมเบลและฟิสก์ (Campbell and Fiske) ได้เสนอเพิ่มเติมว่า เพื่อความรอบคอบก่อนที่จะดำเนินการทดสอบความแม่นยำตรงตามมโนทัศน์ดังที่ได้กล่าวแล้วข้างต้น น่าจะมีการทดสอบเบื้องต้นเสียก่อนใน 2 ประเด็นด้วยกัน โดยแคมเบลและฟิสก์ได้เสนอว่า ควรจะตรวจสอบหลักฐาน 2 ประเภทก่อนที่จะพิจารณาความสัมพันธ์ระหว่างมาตรวัดตามมโนทัศน์หนึ่งๆ กับตัวแปรอื่นๆ ที่เกี่ยวข้อง การตรวจสอบดังกล่าวได้แก่

ประการแรก ตรวจสอบว่ามาตรวัดอื่นๆ ที่วัดในสิ่งเดียวกัน จะให้ผลที่สอดคล้องต้องกันเพียงใด เรียกว่า Convergent Validity เช่น จากตัวอย่างการสร้างมาตรวัดความซับซ้อนขององค์การ ควรสร้างมาตรวัดเพิ่มขึ้นมาอีกเพื่อวัดความซับซ้อน แล้วจึงทดสอบความสอดคล้องระหว่างมาตรวัดต่างๆ ที่วัดความซับซ้อนขององค์การ

ประการที่สอง ตรวจสอบว่ามโนทัศน์ที่วัดด้วยมาตรวัดที่สร้างขึ้นมานั้น สามารถแยกออกมาได้จากมโนทัศน์ที่มีความหมายใกล้เคียงกัน เรียกว่า Discriminant Validity หรือความแม่นยำตรงในการจำแนกมโนทัศน์

กล่าวคือ ต้องให้แน่ใจได้ว่า มาตรวัดหนึ่งๆ สามารถวัดมโนทัศน์ที่ต้องการวัดแต่เพียงมโนทัศน์นั้นๆ เท่านั้น มิใช่วัดครอบคลุมไปถึงสิ่งอื่นด้วย เช่น ในตัวอย่างการวัดความ

หนึ่ง เทคนิคในการพิจารณาความแม่นยำในการจำแนกมโนทัศน์มีหลายวิธีการด้วยกัน แต่ในปัจจุบันนิยมใช้ เทคนิคการวิเคราะห์ปัจจัย (Factor Analysis) ซึ่งเป็นเทคนิคทางสถิติที่ช่วยแยกมิติต่าง ๆ ออกจากกันได้ เช่น ในกรณีเรื่องความซับซ้อนขององค์การ อาจเก็บข้อมูลทั้งทางด้านความซับซ้อน และทางด้านการรวมอำนาจ จากนั้นจึงใช้การวิเคราะห์ปัจจัย เพื่อชี้ให้เห็นว่า ทั้งสองนี้คือมิติเดียวกันหรือไม่ ซึ่งถ้าผลการวิเคราะห์พบว่า ความซับซ้อนและการรวมอำนาจไม่สามารถแยกออกจากกันเป็น 2 มิติได้ มาตรการวัดความซับซ้อนนั้นก็ต้องถือว่าขาดความแม่นยำในการจำแนกมโนทัศน์ และถ้าเป็นเช่นนั้น นักวิจัยจะต้องรีบแก้ไขในจุดนี้ก่อนที่จะพิจารณาความแม่นยำตามมโนทัศน์ต่อไป

ความเชื่อถือได้ของการวัด(Reliability)

คือ เมื่อมีการวัดหลายครั้ง หรือมีเครื่องมือที่ใช้วัดหลายชนิดที่คล้ายกัน และการวัดแต่ละครั้งเป็นอิสระต่อกัน เมื่อใช้วัดในสิ่งเดียวกัน ถ้าการวัดมีความแน่นอน ก็จะได้ผลออกมาเหมือนกัน

ผลคะแนนที่ได้จากมาตรวัดหนึ่ง ๆ โดยปกตินอกเหนือจากจะเป็นการสะท้อนคุณสมบัติที่เราต้องการวัดแล้ว ยังจะสะท้อนอิทธิพลอื่นๆ ที่แฝงเข้ามาด้วย อิทธิพลของปัจจัยอื่นๆ ที่แฝงเข้ามานี้ แยกออกได้เป็น 2 ประเภทด้วยกันคือ

1) แนวโน้มที่ผู้ตอบคนหนึ่ง ๆ จะมีความลำเอียง (Bias) ไปในทิศทางใดทิศทางหนึ่ง เช่น ประเมินตนเองดีเกินสมควร หรือมีแนวโน้มที่จะตอบคำถามไปในทิศทางใดทิศทางหนึ่ง อันถือได้ว่าเป็นความแตกต่างคงเส้นคงวา ซึ่งการทดสอบความแม่นยำตามมโนทัศน์ได้พยายามแยกอิทธิพลเหล่านี้ออกไปแล้ว

2) ปัจจัยอื่นๆ ที่เข้ามาเกี่ยวข้องโดยบังเอิญ และเกิดขึ้นเป็นครั้งคราว (Random)

การประเมินความเชื่อถือได้ของมาตรวัดหนึ่ง ๆ ก็คือประเมินว่าปัจจัยที่เข้ามาเกี่ยวข้องโดยบังเอิญนั้น มีอิทธิพลแค่ไหนนั่นเอง หรือกล่าวอีกนัยหนึ่งคือ การทดสอบความเชื่อถือได้เป็นการวัดระดับความคลาดเคลื่อนที่เกิดขึ้นโดยบังเอิญ ยิ่งความคลาดเคลื่อนที่เกิดขึ้นโดยบังเอิญมีน้อย ก็ยิ่งมีความเชื่อถือได้มาก แต่ถ้าความคลาดเคลื่อนมีมากก็เชื่อถือได้น้อย เช่น

การทดสอบเชาว์ ถ้าหากทดสอบครั้งแรกปรากฏว่าบุคคลๆ หนึ่งได้คะแนนสูงมาก แต่เมื่อทดสอบอีกครั้งภายในระยะเวลาอันใกล้เคียงกัน ปรากฏว่าบุคคลเดียวกันนั้นได้คะแนนต่ำ สถานการณ์เช่นนี้ย่อมสันนิษฐานได้ว่ามาตรวัดขาดความคงเส้นคงวา คือขาดความเชื่อถือได้

เครื่องมือหนึ่ง ๆ หรือเครื่องวัดหนึ่ง ๆ ที่มีความแม่นยำจะต้องมีความเชื่อถือได้ด้วย แต่มาตรวัดที่มีความเชื่อถือได้ไม่จำเป็นต้องแม่นยำเสมอไป เพราะความเชื่อถือได้เป็นเรื่องของความคงเส้นคงวา ส่วนความแม่นยำเป็นเรื่องของความตรงประเด็น (แต่จะแม่นยำได้

หลักในการทดสอบความเชื่อถือได้

ความเชื่อถือได้นั้น ไม่ว่าจะพิจารณาในแง่มุมใดก็ตาม สิ่งที่เหมาะสมในภาพรวม ก็คือการประเมินความคงเส้นคงวา หรือความสอดคล้องกันของคำตอบที่ได้รับ ซึ่งในปัจจุบัน อาจจะแยกมิติของความสอดคล้องหรือคงเส้นคงวานี้ได้ 3 ประการคือ

1) เสถียรภาพ (Stability) คือ คือผลที่ได้ในแต่ละครั้งของการวัดด้วยมาตรวัดเดิมน่าจะคล้ายคลึงกัน

2) การทดแทนซึ่งกันและกันได้ (Equivalence) คือ คำตอบจากการใช้มาตรวัดที่ต่างกัน หรือการใช้มาตรวัดอันเดิม แต่วัดกับต่างบุคคล น่าจะคล้ายคลึงกัน

3) ความเป็นอันหนึ่งอันเดียวกัน (Homogeneity) คือ คำตอบที่บุคคลหนึ่ง ๆ ให้ต่อชุดข้อความที่สมมติว่าจะเป็นการวัดมโนทัศน์หนึ่ง ๆ ควรมีความสอดคล้องเป็นอันหนึ่งอันเดียวกัน

ความมั่นคงหรือเสถียรภาพ (Stability)

เสถียรภาพของมาตรวัดหนึ่ง ๆ จะประเมินได้จากการวัดที่สอดคล้องกัน เมื่อมีการวัดซ้ำแล้วซ้ำอีก โดยมีสิ่งที่ต้องระวังคือ

- ต้องสามารถจำแนกได้ระหว่างความไม่สอดคล้องอันเนื่องมาจากเครื่องมือวัดขาดความเชื่อถือได้

เช่น ชั่งน้ำหนักบุคคล ครั้งแรกชั่งได้ 56 กิโลกรัม แต่อีก 5 นาทีต่อมากลับชั่งได้ 60 กิโลกรัม ความเปลี่ยนแปลงลักษณะนี้น่าจะแสดงถึงปัญหาในตัวเครื่องชั่งนั่นเอง

- ความไม่สอดคล้องเนื่องจากสิ่งที่วัดนั้นมีการเปลี่ยนแปลงโดยธรรมชาติในตัวเอง

เช่น วัดทัศนคติที่ผู้ได้บังคับบัญชาที่มีต่อผู้บังคับบัญชา ซึ่งถ้าทำการวัดเข้า-ป่าย อาจได้คำตอบที่แตกต่างกัน ทั้งที่มาตรวัดนั้นไม่ได้บกพร่อง ขึ้นอยู่กับว่าช่วงเวลาระหว่างเข้า-ป่ายนั้นได้มีสถานการณ์อะไรแทรกซ้อนเข้ามา อาทิ เกิดความขัดแย้งเรื่องงานที่ทำ ประกาศข่าวสองชั้นแล้วรู้ว่าปีนั้นไม่ได้รับการเสนอสองชั้น ฯลฯ เป็นต้น

อย่างไรก็ตาม ไม่ว่าจะโดยกรณีใด ก็ควรดูที่ความคงเส้นคงวาของมาตรวัดนั้น ๆ ซึ่งถ้าทำการวัดโดยสังเกตการณ์ ก็อาจจะทำการสังเกตการณ์ซ้ำซากหลายครั้ง เพื่อดูว่าคำตอบเหมือนกันหรือไม่ แต่ถ้าหากวัดโดยการเก็บข้อมูลจากแบบสอบถามหรือแบบสัมภาษณ์ ก็ต้องทำการถามซ้ำกับชุดคำถามบางข้อ (Test-retest Procedure) เพื่อดูว่าคำตอบตรงกันหรือไม่ ทั้งนี้ต้องคำนึงถึงปฏิกิริยาของผู้ตอบด้วย กล่าวคือ อาจมีปัญหว่าผู้ตอบอาจหย่อนความสนใจ หรือมีฉะนั้นก็อาจเพิ่มความสนใจในรอบที่ 2 หรืออาจจำคำตอบจากครั้งแรกมาตอบครั้งที่สองก็ได้

ความทดแทนซึ่งกันและกันได้ (Equivalence)

การทดแทนซึ่งกันและกันได้ หมายถึงความสอดคล้องของคำตอบที่ได้ หรือได้คำตอบเดียวกันจากการวัดบุคคลหรือสิ่งเดียวกัน โดยผู้วัดที่ต่างกัน เช่น

การที่มีแบบสอบถาม 2 ชุดในลักษณะเดียวกัน แล้วให้คนสองคนถามคนๆ เดียวกัน คำตอบที่ได้ก็น่าจะเหมือนกัน

หรือถ้าใช้เทคนิคการสังเกตการณ์ที่ถูกต้องในการสังเกตสิ่งๆ เดียวกันแล้ว ผลที่ได้ก็น่าจะเหมือนกัน

หรือในอีกความหมายหนึ่งก็คือ เมื่อใช้เครื่องวัดที่ต่างกันแล้ว ได้คำตอบจากการวัดสิ่งที่เหมือนกันนั้นเหมือนกัน เช่น การวัดผลการสอบภาษาอังกฤษ (TOEFL) ซึ่งมีมาตรฐานวัดหลายมาตรฐานวัดที่ให้ผลของการวัดเทียบเท่ากัน

อย่างไรก็ดี การทดสอบความเชื่อถือได้โดยการสร้างมาตรวัดที่มีฟอร์มต่างกันแต่ให้คำตอบเดียวกันนี้ เป็นเรื่องยากและใช้เวลานานมากในการพัฒนาเครื่องมือ และในการเก็บข้อมูลซ้ำโดยใช้เครื่องมือดังกล่าวก็มีจุดอ่อนต่อความเปลี่ยนแปลงของผลการวัด เช่นเดียวกับจุดอ่อนที่กล่าวถึงในมิติของเสถียรภาพ

ความเป็นอันหนึ่งอันเดียวกัน (Homogeneity)

บางครั้งสิ่งที่ต้องการวัดนั้นไม่สามารถจะวัดได้ด้วยเครื่องชี้วัดเพียงตัวเดียว แต่จะต้องอาศัยเครื่องชี้วัดหลายๆ ตัวรวมกัน เพื่อให้ได้ภาพรวมของสิ่งนั้น ซึ่งการรวมเครื่องชี้วัดหลายๆ ตัวเข้าด้วยกันนี้ เรียกว่าเป็นการสร้างมาตรวัด

และในการพิจารณาความเชื่อถือได้ของมาตรวัดหนึ่งๆ เกณฑ์ที่ใช้เป็นเกณฑ์หลักประการหนึ่งได้แก่ การพิจารณาความสอดคล้องกันระหว่างชุดของข้อความหรือเครื่องชี้วัดที่บรรจุอยู่ในมาตรวัดนั้นๆ (Internal Consistency)

แต่เดิมการวัดความเป็นอันหนึ่งอันเดียวกัน โดยดูความสอดคล้องหรือความสัมพันธ์นั้น ใช้การแบ่งแยกข้อความในมาตรวัดออกเป็น 2 ส่วน แล้วหาความสัมพันธ์ระหว่าง 2 ส่วนนี้ โดยอาศัยฐานคติว่า แต่ละข้างมีความเท่าเทียมกัน (Equivalent Form) เรียกว่า เป็นการทดสอบแบบ Split-half Correlation ซึ่งคำตอบที่ได้มาก็คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างชุดข้อความครึ่งหนึ่งกับชุดข้อความอีกครึ่งหนึ่งของมาตรวัด

เช่น ถ้ามาตรวัดมี 40 ข้อก็จะแบ่งครึ่ง คือครึ่งละ 20 ข้อ แล้วจึงคำนวณหาความสัมพันธ์ระหว่างทั้ง 2 ชุด แต่ในความเป็นจริง มาตรวัดนั้นมีถึง 40 ข้อ มิใช่มีเพียง 20 ข้อ จึงต้องดำเนินการปรับค่าสัมประสิทธิ์สหสัมพันธ์ให้สะท้อนจำนวนข้อที่แท้จริง ซึ่งนิยมใช้สูตรที่เรียกว่า สูตรสปีร์แมน-บราวน์ (Spearman-Brown Correlation for Split-half Reliability)

ต่อมาในระยะหลัง มีแนวความคิดขึ้นมากว่า วิธีการทดสอบหาความเชื่อถือได้ของทุกข้อความในมาตรวัดหนึ่งๆ ควรจะเป็นการทดสอบหาความสัมพันธ์ระหว่างการจัดแบ่งครึ่งข้อความโดยวิธีการสุ่ม (Random Halves) มากกว่าระหว่างการแบ่งครึ่งข้อความเท่ากันเช่นที่กล่าวมาในย่อหน้าที่แล้ว

ดังนั้นจึงได้มีการคิดค้นสูตรใหม่ขึ้น และได้มีการใช้กันแพร่หลายมาจนถึงทุกวันนี้ นั่นคือ สูตรของคูเดอร์-ริชาร์ดสัน (Kuder-Richardson Formula) และสูตรสัมประสิทธิ์อัลฟาของครอนบาค (Cronbach's Alpha)

ซึ่งสูตรทั้งสองนี้มีคุณสมบัติประการหนึ่งเป็นการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์โดยเฉลี่ย (Average Split-half Correlation) ในเมื่อมีการแบ่งรายการในมาตรวัดออกเป็น 2 ส่วนในทุกๆ ทางที่เป็นไปได้

สูตรของคูเดอร์-ริชาร์ดสัน (Kuder-Richardson Formula) ใช้ในกรณีที่ข้อความจัดอยู่ในลักษณะของ Dichotomy คือ การให้คะแนนเป็น 0 หรือ 1 เท่านั้น โดยใช้สูตรดังนี้

ตัวอย่างการคำนวณตามสูตร K-R 20

สูตรสัมประสิทธิ์อัลฟาของครอนบาค (Cronbach's Alpha) ใช้ในกรณีที่ระดับการวัดของข้อความแต่ละข้อเป็นการวัดจัดอันดับ การทดสอบความเชื่อถือได้อาจใช้สูตรคำนวณค่าสัมประสิทธิ์อัลฟา (Coefficient Alpha) ดังนี้

หลังจากหาค่าสัมประสิทธิ์ความเชื่อถือได้ (Reliability Coefficient) แล้ว นักวิจัยควรต้องพิจารณาต่อไปว่า ค่าอัลฟา (Alpha) ที่ได้มานั้นสูงเพียงพอหรือไม่ เกี่ยวกับเรื่องนี้ Jum C. Nunally เสนอว่า เกณฑ์ที่น่าจะยอมรับได้คือ 0.50

ตัวอย่าง

จากการวิจัยเรื่อง "ประสิทธิผลของการประชาสัมพันธ์เผยแพร่ความรู้ความเข้าใจเกี่ยวกับงานประกันสังคมในสถาบันการศึกษาด้านวิชาชีพของรัฐ"

เก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างที่เป็นนักศึกษาสถาบันเทคโนโลยีราชมงคล 7 แห่ง ในกรุงเทพมหานครได้แก่ วิทยาเขตเทคนิคกรุงเทพ วิทยาเขตชุมพรเขตอุดมศักดิ์ วิทยาเขตพัฒนการพระนคร วิทยาเขตบพิตรพิมุข จักรวรรดิ วิทยาเขตบพิตรพิมุข มหาเมฆ วิทยาเขตโชติเวช และวิทยาเขตอุเทนถวาย จำนวน 350 ราย

เพื่อให้มั่นใจในมาตรวัดที่ใช้ จึงวิเคราะห์หาค่าความเชื่อถือได้ จากค่าสัมประสิทธิ์อัลฟา (Alpha Coefficient) โดยใช้เกณฑ์ของค่าอัลฟาไม่ต่ำกว่า 0.50 ขึ้นไป ตามเกณฑ์ของ Nunally ผลการวิเคราะห์สรุปได้ดังนี้

โดยสรุปพบว่าค่าอัลฟาทุกค่าที่คำนวณได้สูงกว่า 0.50 ขึ้นไป แสดงว่าค่าความเชื่อถือได้ของมาตรวัดนี้มีศักยภาพเพียงพอที่จะนำไปใช้ได้

วิธีเพิ่มความเชื่อถือได้ของมาตรวัด

เมื่อจะสร้างมาตรวัดใหม่ขึ้นมา เป็นไปได้อย่างยิ่งว่ามาตรวัดนั้นอาจมีความเชื่อถือได้ต่ำกว่าเกณฑ์ที่ตั้งเป้าไว้ แต่ก็มีใช้เรื่องที่นักวิจัยต้องเสียใจหรือกังวล เพราะหาวิธีการพัฒนาเครื่องมือขึ้นมาใหม่ได้ แม้ว่าจะมีเงื่อนไขเรื่องเวลาซึ่งต้องเพิ่มขึ้น...

ดังนั้น ในทางปฏิบัติ นักวิจัยจึงควรกันเวลาส่วนหนึ่งของการวิจัยเพื่อพัฒนาเครื่องมือขึ้นมาก่อน โดยเก็บข้อมูลจากกลุ่มตัวอย่างที่มีลักษณะใกล้เคียงกับกลุ่มที่จะไปศึกษาจริงภายหลัง นักวิจัยคงต้องทำใจไว้ด้วยว่า อาจจะต้องพัฒนาสองถึงสามรอบก่อนจะได้เครื่องมือที่ภาคภูมิใจ

เทคนิคหรือวิธีการเพิ่มความเชื่อถือได้ของมาตรวัดนั้น เป็นที่ทราบกันดีว่า ในเชิงสถิติ ระดับความเชื่อถือได้ ส่วนหนึ่งขึ้นอยู่กับข้อความเป็นองค์ประกอบของมาตรวัด โดยที่ข้อความยังมีจำนวนมากเท่าใด มาตรวัดนั้นก็จะมีความเชื่อถือได้เพิ่มขึ้นด้วย

วิธีการที่มักใช้กันเพื่อเพิ่มระดับความเชื่อถือได้ของมาตรวัดที่ประกอบด้วยข้อความหลายข้อความก็คือ

วิธีแรก : เพิ่มจำนวนข้อความเข้าไปในมาตรวัด

Nunnally กล่าวว่า ถ้าหากทราบค่าสัมประสิทธิ์ความเชื่อถือได้ของมาตรวัดหนึ่งๆ ซึ่งประกอบด้วยข้อความจำนวนหนึ่งแล้ว ถ้าหากมีการเพิ่มจำนวนข้อความเข้าไปในอัตราส่วน k เท่าตัว ความเชื่อถือได้จะเพิ่มขึ้นดังต่อไปนี้

ตัวอย่างเช่น

สมมติว่ามาตรวัดตัวหนึ่งประกอบด้วยข้อความ 20 ข้อ ค่าสัมประสิทธิ์ความเชื่อถือได้เท่ากับ .70 และถ้าเพิ่มข้อความเข้าไปอีก 40 ข้อความ โดยมีเงื่อนไขว่า ข้อความใหม่เหล่านี้จะต้องตรงประเด็นกับสาระสำคัญของมาตรวัด ผลที่ได้คือ มาตรวัดใหม่ประกอบด้วยข้อความ 60 ข้อความ ซึ่งนับว่ามีความยาวมากกว่ามาตรวัดเก่าถึง 3 เท่าตัว ($k=3$) และค่าสัมประสิทธิ์ความเชื่อถือได้ที่คาดว่าจะได้ (Estimated Reliability) ก็คือ

ในทำนองเดียวกัน สูตรในสมการนี้ สามารถนำไปใช้ในการคำนวณค่าความเชื่อถือได้ในกรณีที่ต้องการลดจำนวนข้อความในมาตรวัดหนึ่ง

ดังเช่นต้องการคัดเลือกข้อความจากมาตรวัดชุดเดิม 19 ข้อจาก 50 ข้อ (เท่ากับ .38 เท่าของ 50 ข้อ) การคำนวณความเชื่อถือได้ของมาตรวัดใหม่คาดว่าจะเท่ากับประมาณ

ทั้งนี้เพราะสัมประสิทธิ์ความเชื่อถือได้ของมาตรวัดข้างต้น ซึ่งประกอบด้วยข้อความ 50 ข้อ มีค่า $r = .95$

ดังนั้นเมื่อตัดรายการข้อความลดลงเหลือเพียง 19 ข้อ จึงคาดว่า ค่าสัมประสิทธิ์ความเชื่อถือได้คงลดลงเหลือประมาณ $r = .87$ ซึ่งนับว่ายังอยู่ในเกณฑ์ใช้ได้ และข้อดีในแง่ปฏิบัติก็คือ เป็นการประหยัดเวลาและค่าใช้จ่ายในการเก็บและวิเคราะห์ข้อมูลด้วย

อย่างไรก็ตาม ค่าสัมประสิทธิ์ในสมการหลังสุดนี้ เป็นเพียงค่าที่คาดว่าจะได้ (Estimated Reliability) เมื่อดำเนินการรวบรวมข้อมูลจากแบบสอบถามเสร็จ จะต้องคำนวณหา ค่าสัมประสิทธิ์ความเชื่อถือได้ที่แท้จริง เพื่อพิสูจน์ให้ได้ว่าค่าที่คำนวณได้ครั้งนี้ยังใกล้เคียงกับ ค่าแรกที่เราได้

วิธีที่สอง : การตัดข้อความที่ไม่สอดคล้องออก

วิธีนี้โดยปกติเป็นสิ่งที่ต้องทำอยู่แล้ว เพราะในมาตรวัดหนึ่ง ๆ นั้น หากพบว่ามีบาง ข้อที่ไม่สอดคล้องกันก็ควรตัดข้อความเหล่านั้นออกไป จากนั้นจึงจะคำนวณหาความเชื่อถือ ได้จากข้อความที่เหลือและแน่ใจว่าสามารถใช้วัดได้ การตัดข้อความออกนี้อาจทำได้หลายวิธี ด้วยกัน

ในที่นี้จะเสนอวิธีที่สำคัญและแพร่หลายในปัจจุบัน นั่นคือ การวิเคราะห์รายข้อ (Item Analysis) ซึ่งอาจแบ่งออกได้เป็น 2 วิธีย่อยคือ

1. วิธีแรก เป็นการวิเคราะห์แต่ละข้อ โดยดูความสัมพันธ์ของแต่ละข้อกับมาตรวัด ในส่วนรวม เช่น ถ้ามาตรวัดมี 20 ข้อความ ก็ได้ค่าสัมประสิทธิ์ออกมา 20 ค่า ในทางปฏิบัติเรา จะตัดข้อความที่มีค่าของ p มากกว่า 0.5 ออกไป แล้วคำนวณหาความเชื่อถือได้ใหม่ ซึ่งใน ปัจจุบันเครื่องคอมพิวเตอร์จะช่วยอำนวยความสะดวกในการวิเคราะห์ได้เป็นอย่างดี

2. วิธีที่สอง ในการวิเคราะห์รายข้อ (Item Analysis) ใช้การหาค่าอำนาจจำแนก ของแต่ละข้อความ (Discriminant Value) ซึ่งเป็นการทดสอบว่าข้อความหนึ่ง ๆ สามารถจำแนก ออกได้ระหว่างผู้ที่ได้คะแนนสูง กับ ผู้ที่ได้คะแนนต่ำ ตามมาตรวัดที่ข้อความนั้น ๆ บรรลุอยู่ หรือไม่เพียงใด วิธีการทดสอบมีขั้นตอนดังต่อไปนี้คือ

ภายหลังจากเก็บข้อมูลแล้ว นักวิจัยจะ

1) รวมคะแนนของทุก ๆ ข้อ ที่บรรลุอยู่ในมาตรวัดหนึ่ง ๆ เพื่อหาคะแนนรวม ของแต่ละคน (Total Score) ตามมาตรวัดนั้น ๆ

2) เรียงลำดับคะแนนรวมของทุกคนจากสูงไปต่ำ

3) คัดเลือกเฉพาะคนที่มีความคะแนนสูงสุด 27 เปอร์เซนต์แรก และคนที่มีความคะแนน ต่ำสุด 27 เปอร์เซนต์เช่นกัน (บางตำราอาจใช้เพียง 25 เปอร์เซนต์ ของคะแนนสูงสุดต่ำสุด)

4) วิเคราะห์หาอำนาจจำแนกของแต่ละข้อความจากกลุ่มที่คัดมาทั้ง 2 กลุ่ม โดยใช้วิธีการเปรียบเทียบความแตกต่างของค่าเฉลี่ยระหว่างกลุ่ม (สูงสุด/ต่ำสุด) ว่าแตกต่างกัน อย่างมีนัยสำคัญทางสถิติหรือไม่ เพราะถ้าข้อความหนึ่ง ๆ มีอำนาจจำแนก ก็จะต้องสามารถแยก ให้เห็นความแตกต่างระหว่างกลุ่มได้

โดยวิธีนี้ เมื่อพบว่าข้อความข้อใดไม่มีอำนาจจำแนก (มีค่าเฉลี่ยระหว่าง กลุ่มสูงและกลุ่มต่ำไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ) ก็ให้ตัดข้อความนั้นออกไปเสีย เพื่อให้ทุกข้อที่คงไว้มีอำนาจจำแนกสูง ซึ่งจะมีส่วนช่วยให้มาตรวัดนั้นมีความเชื่อถือได้มากขึ้น ด้วย

วิธีที่สาม : การใช้ถ้อยคำในการเขียน

ส่วนหนึ่งของความเชื่อถือได้ของมาตรวัดจะขึ้นอยู่กับการใช้ถ้อยคำในการเขียนข้อความในมาตรวัดแต่ละข้อ ฉะนั้น จึงต้องพิจารณาด้วยว่าทุกข้อความเขียนได้ชัดเจน รัดกุม หรือไม่เพียงใด เพราะถ้าเป็นการเขียนที่กำกวมแล้ว ผู้ที่ถูกวัดอาจจะไม่เข้าใจ ซึ่งอาจส่งผลให้ได้คำตอบที่ไม่ถูกต้อง และเชื่อถือไม่ได้นั่นเอง

เอกสารอ้างอิง

- กาญจนา มณีแสง. หลักการวิจัยเบื้องต้นทางพฤติกรรมศาสตร์และสังคมศาสตร์. กรุงเทพฯ
มหานคร: โอเดียนสโตร์. 2522.
- จิรพรรณ กาญจนจิตรา. ระเบียบวิธีการวิจัยขั้นสูงทางสังคมวิทยาและมานุษยวิทยา. กรุงเทพฯ
มหานคร: ฝ่ายตำราและอุปกรณ์การศึกษา มหาวิทยาลัยรามคำแหง. 2528.
- เชิดศักดิ์ โฉวาสินธุ์. การวิจัยพฤติกรรมศาสตร์และสังคมศาสตร์. กรุงเทพมหานคร:
โอเดียนสโตร์. 2523.
- พิพัฒน์ ชารุณนทกร. "ทัศนคติกับการประยุกต์ใช้ทางการตลาด." วารสารร่วมพักษ์ 14
(มิถุนายน-ตุลาคม 2538) : 73-79.
- ศิริชัย พงษ์วิชัย. การวิเคราะห์ข้อมูลทางสถิติด้วยคอมพิวเตอร์. กรุงเทพมหานคร: สำนักพิมพ์
จุฬาลงกรณ์มหาวิทยาลัย. 2535.
- ศิริวรรณ เสรีรัตน์ และคณะ. การวิจัยการตลาด Marketing Research ฉบับสมบูรณ์. กรุงเทพฯ
มหานคร: A.N. การพิมพ์. 2540.
- สุจิตรา บุญรัตพันธุ์. ระเบียบวิธีวิจัยทางรัฐประศาสนศาสตร์. กรุงเทพมหานคร: โครงการ
ส่งเสริมตำราและเอกสารวิชาการ มูลนิธิ 30 ปี คณะรัฐประศาสนศาสตร์
สถาบันบัณฑิตพัฒนบริหารศาสตร์. 2534.
- สุชาติ ประสิทธิ์รัฐสินธุ์ . เทคนิคการวิเคราะห์ตัวแปรหลายตัวสำหรับการวิจัยทางสังคมศาสตร์
และพฤติกรรมศาสตร์. กรุงเทพมหานคร: สถาบันบัณฑิตพัฒนบริหารศาสตร์.
2537.
- สุพรรณิ มังคะลี. การวิจัยด้านการประชาสัมพันธ์. กรุงเทพมหานคร: ฝ่ายตำราและอุปกรณ์
การศึกษา มหาวิทยาลัยรามคำแหง: 2529.