



วรรณคดีที่เกี่ยวข้อง

การวิจัยนี้เป็นการศึกษาเกี่ยวกับค่าความเที่ยงของการวัดโดยใช้แบบสอบถามชนิดเลือกตอบ เมื่อกำหนดเทคนิคการให้คะแนนที่ต่างกัน นับเป็นเรื่องหนึ่งที่ยังไม่มีใครทำมาก่อนในประเทศไทย ดังนั้นโดยส่วนใหญ่แล้ว จึงจำเป็นต้องอาศัยความรู้ และข้อมูลจากตำราหรือรายงานการวิจัยที่ทำในต่างประเทศ การพิจารณาขอเขียนจากตำรา หรือรายงานการวิจัยซึ่งเกี่ยวข้องกับเรื่องนี้ ผู้วิจัยได้จำกัดขอบเขตของเทคนิคการให้คะแนนเป็นสองแบบด้วยกัน คือ แบบแaggerการเคา และไมแaggerการเคา แบบที่ไม่แaggerการเคา คือการคิดคะแนนโดยนับจากจำนวนข้อที่ตอบถูก เมื่อกำหนดน้ำหนักคะแนน หนึ่งให้กับข้อที่ตอบถูก และศูนย์ให้กับข้อที่ตอบผิดหรือข้อเว้น ซึ่งเป็นวิธีที่ใช้กันโดยทั่วไป ส่วนแบบแaggerการเคานั้นได้จำแนกออกเป็นสามวิธีคือ วิธีลดคะแนนของข้อที่ตอบผิด เพิ่มคะแนนให้กับข้อที่ตอบถูก และวิธีทดสอบความมั่นใจ จากข้อเขียนในตำราและรายงานการวิจัยเท่าที่ค้นคว้าได้ทำให้ทราบว่า มีการศึกษาและวิจัยเกี่ยวกับเทคนิคการให้คะแนนเพื่อแaggerการเคากันอย่างกว้างขวางและเป็นเวลานาน ตั้งแต่ต้นศตวรรษที่ 20 ซึ่งปรากฏผลการวิจัยที่สนับสนุนเทคนิคการให้คะแนนเพื่อแaggerการเคาหลายฉบับ แต่ขณะเดียวกันก็มีข้อเขียน และผลงานวิจัยบางฉบับไม่สนับสนุนการให้คะแนนเพื่อแaggerการเคา เพราะวิธีการที่ยุ่งยาก และให้ผลไม่แตกต่างกับการไม่แaggerการเคา จึงยังไม่มีข้อสรุปแน่นอนว่าเทคนิคการให้คะแนนวิธีใดดีที่สุด ผลการวิจัยครั้งนี้อาจเป็นส่วนหนึ่งที่จะช่วยท่านที่ใช้แบบสอบถามชนิดเลือกตอบตัดสินใจว่า การสอบครั้งต่อไปนั้น ท่านควรใช้เทคนิคการให้คะแนนวิธีใด

โดยทั่วไปแล้ว เราถือว่าข้อสอบแต่ละข้อในแบบสอบถามชนิดเลือกตอบมีความเป็นตัวแทนของคุณสมบัติที่แบบสอบถามนั้นต้องการวัดเท่า ๆ กัน จึงกำหนดน้ำหนักคะแนนหนึ่งให้กับข้อที่ตอบถูก ส่วนข้อที่ตอบผิด หรือไม่ตอบกำหนดคะแนนเป็นศูนย์ การหาคะแนนรวมสำหรับผู้เข้าสอบแต่ละคน จึงนับจากจำนวนข้อที่ตอบถูก ดังสมการ

$$X = R \quad (1)$$

เมื่อ X แทนคะแนนดิบของแบบทดสอบฉบับนั้น
 R แทนจำนวนข้อที่ตอบถูก

วิธีการให้คะแนนแบบนี้ ฮาโรลด์ กัลลิกเซน ¹(Harold Gulliksen) ก็ยอมรับในความสะดวกและเหมาะสม เขากล่าวว่า การใช้แบบสอบในปัจจุบันมักจะกำหนดค่าตั้งชี้แจงใหญ่ เขาสอบตอบข้อสอบทุกข้อติดต่อกันไป จึงถือได้ว่าจำนวนข้อที่ผู้เขาสอบเว้นข้ามไปจะเป็นศูนย์ หรือมีน้อยจนไม่จำเป็นต้องนำมาพิจารณา และถึงแม้ว่าจำนวนข้อที่เว้นกับจำนวนข้อที่ยังทำไม่เสร็จจะมีมาก การนับคะแนนจากจำนวนข้อที่ตอบถูกก็น่าจะเป็นวิธีที่ใช้โดยผู้เฉลย เขาสอบทุกคนอ่านข้อสอบและพยายามแก้ปัญหาให้โจทย์ก่อนจะทำเครื่องหมายลงในกระดาษคำตอบ ทั้งนี้เพราะโดยทั่วไปแล้วคนที่มีความรู้ในเนื้อวิชาย่อมจะแก้ปัญหาได้ถูกต้องและรวดเร็วกว่าผู้ที่ไม่มีความรู้ แต่ถึงกระนั้นการใช้แบบสอบประเภทนี้ก็มีโอกาสเป็นไปได้ที่ผู้ที่ไม่มีความรู้ สามารถจะทำเครื่องหมายได้ถูกโดยบังเอิญ ผลจากการตอบถูกโดยบังเอิญนี้อาจจะไม่ร้ายแรง และสามารถมองข้ามไปได้โดยผู้เขาสอบทุกคนทำเครื่องหมายลงในข้อสอบทุกข้อโดยไม่เว้นไว้เลย แต่บางครั้งอาจมีผู้เขาสอบเพียงบางคนเท่านั้นสังเกตว่า เหลือเวลาอีกสองนาทีก่อนหมดเวลา จึงรีบทำเครื่องหมายลงในกระดาษคำตอบกับข้อสอบอีกยี่สิบ หรือสามสิบข้อ ที่เหลือซึ่งไม่มีเวลาพอที่จะอ่านเพื่อจะได้ประโยชน์จากโอกาสที่จะตอบถูกโดยบังเอิญนี้ เมื่อคิดคะแนนโดยนับจำนวนข้อที่ตอบถูกแล้ว ภายในเวลาสองนาทีก่อนที่เหลือ เขาก็ย่อมมีโอกาสทำคะแนนได้สูงกว่าคนที่พยายามอ่านโจทย์แล้วทำข้อสอบไปที่ละข้อ

กัลลิกเซน ได้เสนอวิธีการง่าย ๆ วิธีหนึ่งที่จะสำรวจดูว่า ในการสอบแต่ละครั้งจะมีผู้ตอบเพราะการเดามากน้อยเพียงไร วิธีนี้ทำได้โดยการเขียนกราฟ กำหนดแกนนอน (Absissa)

¹ Harold Gulliksen, The Theory of Mental Tests, (New York : John Wiley and Sons, 1958), p. 247.

แทนจำนวนข้อทั้งหมดที่ผู้เข้าสอบทำ (R+W) แกนตั้ง (Ordinate) แทนจำนวนข้อที่ตอบถูก (R) สมการเส้นตรง $y=x$ จะแสดงทางเดินของคะแนนที่สมบูรณ์ขณะที่ยังทำข้อสอบไปมากข้อขึ้น และเส้นตรง $y=(1/C)x$ เมื่อ C คือจำนวนตัวเลือกในแต่ละข้อก็จะชี้ให้เห็นทางเดินของคะแนนที่ได้ เพราะการตอบถูกโดยบังเอิญ กัลลิกเซนเรียกเส้นตรงนี้ว่า ชานส์ไลน์ (Chance line) ถ้าแบบสอบมีห้าตัวเลือก สมการเส้นตรง $y=(1/5)x$ ก็จะต้องแสดงถึงทางเดินของคะแนนที่ได้จากการตอบถูกโดยบังเอิญนี้ เมื่อนำคะแนนของแต่ละคนมากำหนดพิกัดลงในกราฟ ถ้ามีจุดใดที่แสดงจำนวนข้อที่ตอบถูกสูงแต่อยู่ใกล้เส้น ชานส์ไลน์ ก็แสดงว่าคะแนนจำนวนนั้นได้มาจากผู้เข้าสอบที่เดาข้อสอบหลายข้อที่เกี่ยวข้อง

ผลโคผลเสียของการตอบข้อสอบโดยการเดานี้ เฟรดเคอริก บี เดวิส (Federick B. Davis) ได้แสดงตัวอย่างที่เป็นไปได้ให้เห็นชัดเจนขึ้นว่า ถ้าผู้เข้าสอบสองคนทำแบบสอบชนิดถูก - ผิด ซึ่งเป็นแบบสอบใช้ความเร็ว มีทั้งหมดหนึ่งร้อยข้อ ในเวลาสองนาทีก่อนจะหมดเวลา แต่ละคนทำได้ถึงข้อหกสิบเท่ากัน และปรากฏว่ามีจำนวนข้อที่ตอบถูกคนละห้าสิบข้อเท่ากัน ถ้าคนแรกใช้เวลาสองนาทีก่อนที่เหลือทำเครื่องหมายลงในกระดาษคำตอบกับข้อสอบอีกสี่สิบข้อที่เหลือโดยไม่ได้อ่านข้อสอบเลย เขาก็มีโอกาสจะตอบถูกถึงยี่สิบข้อ ดังนั้นคะแนนของเขาเมื่อนำมาจากจำนวนข้อที่ถูกจึงเป็น 70 ส่วนผู้เข้าสอบอีกคนหนึ่งใช้เวลาสองนาทีก่อนที่เหลือทำข้อสอบไปที่ละข้อได้อีกสองข้อทั้งหมดเวลา ถ้าเขาทำถูกทั้งสองข้อ คะแนนก็จะ เป็น 52 แต่จากเหตุการณ์ดังกล่าว เราทราบว่าผู้เข้าสอบทั้งสองคนมีความสามารถใกล้เคียงกัน แต่เพราะการทำเครื่องหมายโดยการเดา คนแรกจึงได้คะแนนสูงกว่ามาก วิธีตอบข้อสอบของคนแรกได้ทำลายความตรงของแบบสอบที่ตองการจะวัดความเร็ว และความแม่นยำในการคำนวณไปโดยสิ้นเชิง¹

¹ Davis, Educational Measurement and their Interpretation, pp.75-78.

ถ้าเราไม่สามารถหลีกเลี่ยงผลที่จะเกิดขึ้นเหล่านี้ได้ ก็จะเป็นข้อบกพร่องอย่างร้ายแรงของแบบจำลองประเภทนี้ อย่างไรก็ตาม วิศวกรกล่าวว่า เราสามารถจะป้องกันได้โดยวิธีการให้คะแนนเพื่อจำกัดโอกาสที่จะตอบถูกโดยบังเอิญออกไป วัตถุประสงค์ของการให้คะแนนแบบนี้เพื่อจะประมาณจำนวนข้อที่ผู้เข้าสอบแต่ละคนตอบถูกเพราะมีความรู้ หรือทักษะพอที่จะชี้ตัวเลือกที่ถูกได้ หรืออย่างน้อยที่สุดก็พอจะรู้ถึงความไม่ถูกของตัวลวงต่าง ๆ ได้ กล่าวอีกอย่างหนึ่งก็คือเพื่อลดคะแนนซึ่งเกิดจากโอกาสที่ผู้เข้าสอบจะตอบถูกโดยการเดาให้เป็นศูนย์ การให้คะแนนแบบนี้ วิศวกรกล่าวว่า สามารถทำได้ภายใต้ข้อตกลงเบื้องต้นว่า พฤติกรรมที่อยู่เบื้องหลังการทำเครื่องหมายลงในกระดาษคำตอบของผู้เข้าสอบ เป็นดังนี้

1. มีความรู้ และทักษะพอเพียงที่จะชี้ตัวเลือกที่ถูกได้ หรือ
2. เดาจากตัวเลือกต่าง ๆ ที่มีในแต่ละข้อ

ดังนั้นข้อที่ตอบผิดทุกข้อ จึงถือว่าเป็นผลเนื่องมาจากการเดา

จากข้อตกลงเบื้องต้นดังกล่าวนี้ จำนวนข้อที่ผู้เข้าสอบตอบถูก ก็จะแสดงได้โดยสมการ

$$R = R_k + R_g \quad (2)$$

เมื่อ R_k แทน จำนวนข้อที่ตอบถูกเพราะมีความรู้
 R_g แทน จำนวนข้อที่ตอบถูกเพราะการเดา

เราต้องการทราบค่า R_k จึงเขียนสมการ (2) ใหม่ในรูป

$$R_k = R - R_g \quad (3)$$

จากกระดาษคำตอบของผู้เข้าสอบ เราสามารถนับจำนวนข้อที่เขาตอบถูก (R) ได้และขณะเดียวกันเราก็สามารถนับจำนวนข้อที่เขาตอบผิด (W) ได้เช่นกัน จำนวนข้อที่ผู้เข้าสอบทำเครื่องหมายคำตอบโดยการเดาจึงเท่ากับผลรวมของจำนวนข้อที่ตอบถูกกับตอบผิด ($R+W$) หักออกเสียด้วย

จำนวนข้อที่ตอบถูกเพราะมีความรู้ในเนื้อหาวิชา $(R+W)-R_k$ และโอกาสที่ข้อซึ่งตอบโดยการเดานี้จะเป็นข้อถูก จะมีค่าสูงสุด คือ $(R+W)-R_k/C$ เมื่อ C คือจำนวนตัวเลือกในแต่ละข้อ จึงเขียนเป็นสมการ

ได้ว่า $R_g = \frac{1}{C} [(R+W)-R_k]$

แทนค่า R_g ในสมการ (3) $R_k = R - \frac{1}{C} [(R+W)-R_k]$

$= R - \frac{R}{C} - \frac{W}{C} + \frac{R_k}{C}$

$\frac{R_k - R_k}{C} = R - \frac{R}{C} - \frac{W}{C}$

$[\frac{C-1}{C}] R_k = [\frac{C-1}{C}] R - \frac{W}{C}$

$\frac{C}{C-1}$ คูณตลอด $R_k = \frac{R-W}{C-1}$

แต่ R_k เป็นตัวแทนของการประมาณจำนวนข้อ ที่ผู้เข้าสอบตอบถูกเพราะมีความรู้ ซึ่งก็คือคะแนนที่แก้โอกาสที่จะตอบถูกโดยการเดานั้นเอง ถ้าให้ X_c แทนคะแนนนี้ สูตรของคะแนนเพื่อแก้การเดาจึงสามารถเขียนในรูป

$X_c = \frac{R-W}{C-1}$ (4)

ในกรณีที่ $C = 2$ ดังเช่นแบบสอบถาม ถูก - ผิด สูตรของคะแนนแก้การเดาก็คือ

$X_c = R - W$

เมื่อนำสูตรนี้ไปใช้กับการทำข้อสอบของผู้เข้าสอบสองคนที่กล่าวมาข้างต้น คะแนนที่แก้การเดาแล้วจะเป็นดังนี้

คนแรก $X_{c_1} = 70 - 30 = 40$

คนที่สอง $X_{c_2} = 52 - 10 = 42$

คะแนนที่แกการเคาแลวนี้สอดคล้องกับเหตุการณ์ที่เราทราบว่า ผู้เข้าสอบทั้งคู่ทำคะแนนได้เท่ากันในเวลาสองนาทีก่อนจะหมดเวลา แต่คนแรกใช้วิธีเคาโดยไม่ได้อ่านข้อสอบเลยจึงสามารถทำได้ถึงข้อสุดท้าย ส่วนอีกคนหนึ่งพยายามอ่านโจทย์ และทำข้อสอบต่อไปจนหมดเวลา เมื่อใช้เทคนิคให้คะแนนเพื่อแกการเคา จึงไม่ทำให้คนแรกได้คะแนนสูงกว่าคนที่สอง

อย่างไรก็ตามถึงแม้ โรเบิร์ต แอล อีเบล¹ (Robert L. Ebel) จะไม่คัดค้านเทคนิคการให้คะแนนแบบนี้ของเควิส แต่เขาก็กล่าวว่ววิธีการให้คะแนนแบบนี้เคยถูกวิจารณ์ว่าผิด ตั้งแต่ข้อตกลงเบื้องต้นแล้ว นั่นคือข้อตกลงที่ว่า ข้อซึ่งตอบผิดทุกข้อเป็นผลเนื่องมาจากการเคา เขาจึงเสนอวิธีการให้คะแนนอีกแบบหนึ่ง แทนที่จะปรับโทษจากข้อที่เคา เขากลับเพิ่มคะแนนให้กับข้อที่ไม่เคา โดยมีข้อตกลงเบื้องต้นว่า ข้อที่ผู้ตอบเว้นไว้นั้น ถ้าตอบมาโดยการเคา เขาก็มีโอกาสหนึ่งในจำนวนตัวเลือกที่จะตอบถูก ทั้งนี้สูตรในการแกการเคาก็คือ

$$X_c = R + \frac{O}{C} \quad (5)$$

เมื่อ 0 แทนจำนวนข้อที่เว้น

อีเบลเชื่อว่า วิธีนี้จะขจัดผลประโยชน์ของผู้เข้าสอบ ที่จะได้จากการตอบถูกโดยบังเอิญออกไปได้เช่นเดียวกัน การให้คะแนนโดยพิจารณาจำนวนข้อที่ผู้เข้าสอบเว้นไว้ ซึ่งรวมทั้งข้อที่เว้นข้ามไป และข้อที่ยังทำไม่เสร็จนี้ ผู้วิจัยเองเห็นว่า เป็นประเด็นหนึ่งที่น่าสนใจ และกลัลดิเซนเองก็ได้ให้ข้อคิดไว้ว่า "มีโอกาสนั้นเป็นไปได้ที่คะแนนจากข้อที่เว้น หรือข้อที่ยังทำไม่เสร็จจะมีประโยชน์ เพราะคะแนนดังกล่าวอาจเป็นเครื่องชี้ให้เห็นถึงลักษณะนิสัย ตลอดจนบุคลิกภาพอื่น ๆ ของผู้เข้าสอบได้ ผู้เข้าสอบบางคนอาจแสดงให้เห็นถึงความมั่นคงในการที่จะตอบข้อสอบทุกข้อถึงแม้จะผิด ขณะที่บางคนอาจจะลังเลไม่กล้าตอบ"²

¹ Ebel, Measuring Educational Achievement, pp. 224-225.

² Gulliksen, Theory of Mental Test, pp. 245-246.



ถึงกระนั้นเทคนิคการให้คะแนนที่เดวิสเสนอไว้ก็ไม่น่าจะถูกปฏิเสธ เพราะเขาเองก็เห็นว่าเหมาะสมกับแบบสอบใช้ความเร็ว วัดความรู้วิชาพื้นฐานเช่นเลขคณิต สำหรับแบบสอบใช้ความสามารถ ซึ่งวัดความสามารถทางสมองที่ซับซ้อนกว่านี้ พฤติกรรมภายในอันเป็นพื้นฐานในการตอบข้อสอบย่อมแตกต่างกันออกไปอีก แต่การตอบแบบสอบชนิดเลือกตอบนี้ ผู้ตอบมีพฤติกรรมภายในอะไรเป็นเกณฑ์ในการตัดสินใจว่า ตัวเลือกข้อใดเป็นข้อถูก เรื่องนี้มันเป็นเรื่องที่น่าสนใจและเป็นเรื่องหนึ่งที่ยังไม่มีรายละเอียด หรือข้อมูลให้ศึกษาอย่างพอเพียง เดวิสกล่าวว่า จากการศึกษาพฤติกรรมของผู้เข้าสอบขณะกำลังทำแบบสอบชนิดเลือกตอบ และการวิเคราะห์โดยตรรกวิทยาถึงสถานการณ์ที่ผู้เข้าสอบกำลังเผชิญ แสดงให้เห็นว่า คำตอบแทบทั้งหมด มีพื้นฐานการตอบมาจากพฤติกรรมภายในข้อใดข้อหนึ่ง หรือมากกว่า ทั้งต่อไปนี้

1. มีความรู้พอเพียง (Sufficient knowledge) ที่จะชี้ตัวเลือกที่ถูกได้
2. มีความรู้เพียงบางส่วน (partial knowledge) ซึ่งทำให้กำจัดตัวลวงตัวใดตัวหนึ่งหรือมากกว่าออกไป แล้วจึงเอาจากตัวเลือกที่เหลือ
3. เอาสุ่มจากตัวเลือกทั้งหมด หลังจากที่ได้อ่านข้อสอบโดยตลอดแล้ว
4. มีความรู้ที่ผิดในบางส่วน (partial misinformation) ซึ่งทำให้กำจัดตัวเลือกตัวหนึ่งตัวใด หรือมากกว่ารวมทั้งตัวเลือกซึ่งเป็นข้อถูกออกไปด้วย แล้วจึงเอาจากตัวเลือกที่เหลือ
5. เกิดจากความรู้ที่ผิด (Sufficient misinformation) ทำให้เลือกข้อผิดมาโดยเข้าใจว่าเป็นคำตอบถูก
6. เอาจากลักษณะที่ผิดปกติกบางอย่าง เช่นตัวเลือกที่มีความยาวมาก การใช้ถ้อยคำจำนวนครั้งที่ผู้เข้าสอบเองได้ทำเครื่องหมายแต่ละตัวเลือกมาแล้ว หรือลักษณะของตำแหน่งตัวเลือกที่ปรากฏบนกระดาษคำตอบ
7. เอาจากตัวเลือกทั้งหมดโดยไม่ได้อ่านข้อสอบเลย

ขอบเขตซึ่งพฤติกรรมเหล่านี้จะถูกนำมาใช้ ย่อมแตกต่างกันไป แล้วแต่ผู้เข้าสอบแต่ละคนและแบบสอบแต่ละฉบับ ซึ่งขึ้นอยู่กับลักษณะของเนื้อหาวิชา ระดับความยากของข้อสอบ วัตถุประสงค์ของการใช้แบบสอบ คำสั่งชี้แจง และวิธีให้คะแนน แต่การสอบโดยทั่วไปนั้นผู้เข้าสอบมักจะมีพฤติกรรมตาม

ข้อหนึ่ง, สอง, สาม และเจ็ด พฤติกรรมข้อสามคือการเคาะสุ่มจากตัวเลือกทั้งหมด หลังจากที่ได้อ่านข้อสอบแล้วมักไม่ค่อยเกิดขึ้น และข้อเจ็ดคือการเคาะจากตัวเลือกทั้งหมดโดยที่ยังไม่ได้อ่านข้อสอบ จะเกิดขึ้นบ่อยในกรณีที่เป็นแบบสอบใช้ความเร็ว ซึ่งไม่มีเวลาพอจะอ่านข้อสอบทุกข้อ โดยปกติ ถ้าผู้สอบได้อ่านและพิจารณาข้อสอบแล้ว หากจะมีการเคาะก็มักเป็นการเคาะจากตัวเลือกเพียงไม่กี่ข้อ แทนที่จะเคาะจากตัวเลือกทั้งหมด การได้ความรู้ซึ่งอาจจะถูกหรือผิดมากก็ตามจะทำให้เขาคัดตัวเลือกบางข้อออกไปจากการพิจารณา เมื่อเผชิญกับการทบทวนข้อผิดพลาดที่เลือกระหว่างคำตอบที่เหลือ เขาอาจใช้วิธีโยนเหรียญ หรือใช้พฤติกรรมในข้อหก เพื่อช่วยในการตัดสินใจ แต่แบบสอบที่สร้างมาอย่างนี้นั้น ลักษณะต่าง ๆ ในข้อหกจะถูกกำจัดโดยการจัดเตรียมข้อสอบมาอย่างระมัดระวัง ไม่ให้มีความสัมพันธ์กับความถูก หรือผิดของคำตอบเลย

การให้คำสั่งชี้แจง และวิธีให้คะแนน ก็มีอิทธิพลต่อพฤติกรรมภายในอันเป็นพื้นฐานในการตอบข้อสอบเช่นเดียวกัน ถ้าผู้เขาสอบทราบว่าคะแนนของการสอบจะคิดจากจำนวนข้อที่ตอบถูกหักด้วยสัดส่วนของจำนวนข้อที่ตอบผิด ก็หมายความว่า การเคาะสุ่มจากตัวเลือกทั้งหมดแทบจะไม่ได้ประโยชน์อะไร และยังอาจต้องเสียคะแนนไปอีก เขาก็จะหลีกเลี่ยงพฤติกรรมในข้อสาม และเจ็ด แต่ถ้าวางว่ามีโอกาสจะได้คะแนนจากคำตอบคำถาม ถึงแม้ไม่แน่ใจว่าจะถูก เขาก็จะรู้สึกกล้าที่จะใช้พฤติกรรมข้อสองและสี่ ถึงกระนั้นก็ยังมิใช่ผู้เขาสอบบางคนที่ยอมรับเพียง กล้าได้กล้าเสีย คำสั่งชี้แจงจึงไม่สามารถป้องกันการเคาะของผู้เขาสอบทุกคนได้เท่ากัน

ทัตสันย์ สุขสมบูรณ์ ได้ทำการวิจัยเกี่ยวกับการให้คำสั่งชี้แจงในการสอบและพบว่า "การให้คำสั่งชี้แจง ที่แตกต่างกัน คือ แบบให้รางวัล ลงโทษ แนะให้เคาะ และแบบไม่เสนอแนะ ไม่ทำให้จำนวนข้อที่ตอบถูกแตกต่างกัน แต่ข้อตอบผิดและเว้นว่างต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01"¹

002159

¹ ทัตสันย์ สุขสมบูรณ์, "ผลของคำสั่งชี้แจงในการให้คะแนน และระยะเวลาในการตอบที่ต่างกันต่อการตอบข้อสอบแบบเลือกตอบ" (วิทยานิพนธ์ปริญญาโทมหาบัณฑิต แผนกวิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2518, หน้า 60.

งานวิจัยเรื่องนี้ได้ทำกับกลุ่มตัวอย่างที่เป็นนักเรียนไทย เป็นงานวิจัยเรื่องหนึ่งที่สนับสนุนว่า การให้ คำสั่งชี้แจงและวิธีให้คะแนนมีผลต่อพฤติกรรมภายในซึ่ง เป็นพื้นฐานในการทดสอบข้อสอบ

คำสั่งชี้แจง ที่แนะนำและสนับสนุนให้ผู้เข้าสอบพยายามตอบข้อสอบทุกข้อ ถึงแม้ไม่แน่ใจว่า จะถูก ก็ควรเหตุผลดังที่ ดี เจ ครอนแบค¹ (Lee J. Cronbach) กล่าวไว้ว่าแนวโน้มของการ เคาหรือไม่เคานั้น ถูกกำหนดโดยองค์ประกอบทางบุคลิกภาพ (Personality factors) และไม่สามารถที่จะขจัดให้หมดไปได้โดยคำสั่งชี้แจงให้ผู้เข้าสอบเคา หรือกำหนดโทษในการปรับเมื่อมีการ เคา เขาเชื่อว่าวิธีเดียวที่จะขจัดความแปรปรวนของการเคาได้ก็คือ ให้คำสั่งชี้แจงให้ผู้เข้าสอบ ตอบคำถามทุกข้อ เมื่อทำเช่นนี้จะไม่มีการโยกย้ายเป็นพิเศษ และไม่มี ความจำเป็นที่จะต้องให้ คะแนนแบบแกลการเคาด้วย การที่สั่ง เช่นนี้ขึ้นเพื่อจะควบคุมพฤติกรรมภายในซึ่ง เป็นพื้นฐานในการทดสอบ ข้อสอบของผู้เข้าสอบทุกคนให้เป็นอย่างเดียวกัน แต่ผู้เข้าสอบบางคนก็ไม่สามารถจะทำตามได้ เพราะ เป็นการฝืนลักษณะนิสัย นับเป็นเรื่องลำบากที่จะให้เขาทำเครื่องหมายลงในคำตอบโดยที่เขาไม่ทราบ หรือยังไม่ได้อ่านข้อสอบ เควสมีความเห็นคัดค้านคำสั่งชี้แจงชนิดนี้ด้วยเหตุผลสำคัญที่ว่า จะทำให้นักเรียน นักการศึกษา และคนทั่วไปเสื่อมศรัทธาต่อแบบสอบประเภทนี้โดยคิดว่าเป็นแบบสอบที่สามารถ เคาได้

สมการ (4) เป็นวิธีการประมาณจำนวนข้อสอบที่ผู้ตอบรู้จริง โดยมีข้อตกลงเบื้องต้นว่าผู้ เข้าสอบตอบข้อสอบโดยมีพฤติกรรมดังข้อหนึ่ง, สาม และเจ็ด แต่ในการสอบโดยทั่วไปผู้เข้าสอบยอม มีพฤติกรรมดังข้อสอง, สี่, ห้า และหกด้วยเช่นกัน ดังนั้นการไขสูตรให้คะแนนเพื่อแกลการเคาดังที่ เควส เสนอไว้ในสมการ (4) จะสมควรนำมาใช้หรือไม่ ปัญหานี้เควสตอบว่า ถ้าผู้เข้าสอบมี พฤติกรรมดังข้อสอง คือมีความรู้เพียงบางส่วน ทำให้กำจัดกังวลตัวใดตัวหนึ่งหรือมากกว่าออกไป แล้วจึง เคาจากตัวเลือกที่เหลือ จะทำให้การประมาณดังกล่าวสูงกว่าที่ควรจะเป็น (Over estimate)

¹ Lee J. Cronbach, Essential of Psychological Testing, 2d ed.

(New York : Harper & Row, 1960), p. 51.

ข้อถูก การทดสอบความมั่นใจมีวิธีการแตกต่างกันหลายวิธี และถึงแม้จะมีประวัติที่ย้อนหลังไปยังต้นศตวรรษที่ 20 นี้ แต่ก็เริ่มมีการใช้ในฐานะเป็นวิธีในการเพิ่มข้อมูลให้กับ การทดสอบ และลดผลที่จะเกิดจากการเดาของแบบสอบประเภทถูก - ผิด¹

ในปีคริสต์ศักราช 1932 เฮฟเนอร์ (Hevner) ได้รายงานผลการวิจัยว่า การทดสอบความมั่นใจทำให้ค่าความเที่ยงสูงขึ้น² เครื่องมือที่เขาใช้เป็นแบบสอบชนิดถูก - ผิด และกำหนดระดับความมั่นใจไว้สามระดับ โดยกำหนดน้ำหนักคะแนนสาม ให้กับข้อที่ตอบด้วยความมั่นใจสูง ข้อที่มีความมั่นใจปานกลางให้สอง และข้อที่ไม่มั่นใจให้หนึ่ง เฮฟเนอร์พบว่า คะแนนที่คิดโดยวิธีถ่วงน้ำหนักของข้อที่ถูก ทำให้ค่าความเที่ยงที่คิดโดยวิธี สปลิต ฮาล์ฟ (Split half) สูงกว่าคะแนนที่นับจากจำนวนข้อถูก

การศึกษาเกี่ยวกับเรื่องนี้ของโซเดอร์ควิสต์³ (Soderquist) เมื่อปีคริสต์ศักราช 1936 ไม่มีวิธีการอะไรที่ใหม่สำหรับการทดสอบความมั่นใจ แต่ต่างกันตรงที่วิธีกำหนดน้ำหนักคะแนนซึ่งต่างไปจากของเฮฟเนอร์ โดยที่ผู้เข้าสอบจะได้คะแนนสำหรับข้อถูกเป็น 1, 3, 2 หรือหนึ่งตามระดับความมั่นใจ และข้อผิดจะกำหนดน้ำหนักเป็นสองเท่า คือ 2, 6, 4 หรือ สอง แต่คะแนนรวมนั้นศึกษาเปรียบเทียบกันสองวิธี คือ ถัดจากคะแนนที่ถ่วงน้ำหนักของข้อถูก หักเสียด้วยคะแนนที่ถ่วงน้ำหนัก

¹ Gary J. Echternacht, "The Use of Confidence Testing in Objective Tests," Review of Educational Research, 42 (1972) : 217.

² K.A.Hevner, "A Method of Correcting for Guessing in True-False Test and Empirical Evidence in Support of it," Journal of Social Psychology, 3 (1932) : 359-362.

³ H.O.Soderquist, "A New Method of Weighting Scores in A True-False Test," Journal of Educational Research, 30 (1936) : 290-292.

ของข้อผิด และคะแนนจากข้อถูก หักด้วยคะแนนของข้อผิด โดยไม่มีการดัดน้ำหนักเมื่อหาค่าความเที่ยง โดยคำนวณตามวิธีเดียวกับที่เฮฟเนอร์ใช้ เขาพบว่าคะแนนดัดน้ำหนักของข้อถูกหักด้วยคะแนนดัดน้ำหนักของข้อผิดให้ค่าความเที่ยงที่สูงกว่า ซึ่งเป็นข้อค้นพบที่ต่างไปจากของเฮฟเนอร์ ทั้งนี้อาจจะเป็นเพราะกลุ่มตัวอย่างของไซเคอควิสท์ ได้รับคำสั่งชี้แจงอธิบายวิธีให้คะแนนขณะที่วิธีให้คะแนนนี้เป็นความลับสำหรับกลุ่มตัวอย่างของเฮฟเนอร์

ปีคริสต์ศักราช 1938 สไวน์ฟอร์ด¹ (Swineford) ได้ศึกษาเกี่ยวกับตัวแปรทางบุคลิกภาพ โดยใช้แบบสอบถามชนิดถูก - ผิด และให้คำสั่งชี้แจงเช่นเดียวกับที่ไซเคอควิสท์ใช้ เขาสามารถหาคะแนนซึ่งเขาเองเรียกว่า เกมมิง สคอร์ (gambling score) จากการทดสอบความมั่นใจนี้และต่อมาในคริสต์ศักราช 1941 เขาก็ได้ใช้แบบสอบถามอีก สี่ ฉบับศึกษาเกี่ยวกับแนวโน้มที่จะเสี่ยง ได้ขอคนพบว่าเด็กชายมีแนวโน้มที่จะเคาน์ในเนื้อวิชาที่ไม่คุ้นเคยมากกว่าเด็กหญิง และในปีเดียวกันนี้เอง กริทเพน และ จอห์นสัน² (Gritten and Johnson) ก็ได้ใช้วิธีการทดสอบความมั่นใจกับงานวิจัยของเขา ถึงแม้จะมีวัตถุประสงค์ที่จะศึกษาระดับความมั่นใจในการตอบข้อสอบเมื่อใช้คำสั่งชี้แจงแบบให้เคา และไม่ให้เคา แต่ประเด็นสำคัญของการทดสอบความมั่นใจที่เขาทั้งสองทำก็คือ เขาใช้แบบสอบถามเลือกตอบ ซึ่งต่างจากที่คนอื่นเคยทำมาแล้ว และระดับความมั่นใจที่ใช้ ก็ได้แบ่งละเอียดออกเป็นห้าระดับด้วยกัน

หลังจากนี้ ความสนใจในเรื่องการทดสอบความมั่นใจ ก็ดูเหมือนจะชะงักลงจนกระทั่งปีคริสต์ศักราช 1953 เดรสเซลและชมิท (Dressel and Schmid) ได้ทำการวิจัยโดยใช้แบบสอบถาม

¹ F.Swineford, "Measurement of a Personality Trait," Journal of Educational Psychology, 29 (1938) : 295-300

² F.Gritten and D.M.Johnson, "Individual Differences in Judging Multiple-Choice Questions," Journal of Educational Psychology, 32 (1941) : 423-430.

ต่างกัน และมีวิธีให้คะแนนที่ต่างกัน เพื่อจะปรับปรุงอำนาจจำแนกของแบบสอบชนิดเลือกตอบ แบบสอบชนิดเลือกตอบที่เขาใช้มี ๓ แบบด้วยกัน ¹ คือ

1. แบบเลือกตอบเสรี (Free-choice test) เป็นแบบสอบชนิดเลือกตอบหาตัวเลือกที่ผู้เข้าสอบจะเลือกคำตอบมากเท่าไรก็ได้

2. แบบแสดงระดับความมั่นใจ (A degree of certainty test) เป็นแบบที่ผู้เข้าสอบจะต้องแสดงระดับความมั่นใจในการตอบ ซึ่งแบ่งเป็น ๓ ระดับ แต่คำถามแต่ละข้อให้ตอบได้เพียงคำตอบเดียว

3. แบบมีคำตอบถูกมากกว่าหนึ่ง (A multiple-answer test) แบบสอบชนิดนี้จะมีคำตอบถูกอยู่หลายคำตอบ ผู้เข้าสอบจะต้องทำเครื่องหมายทุกข้อที่ถูก

4. แบบมีคำตอบถูกสองคำตอบ (A two-answer test) ผู้เข้าสอบจะเลือกตอบได้เพียงสองคำตอบเท่านั้น

ในแต่ละกรณีผู้เข้าสอบจะต้องเข้าใจถึงชนิดของแบบสอบที่ตนทำ และเพราะเหตุว่าแบบสอบสองแบบแรกเท่านั้นที่เกี่ยวข้องกับการทดสอบความมั่นใจ จึงขอแสดงวิธีการให้คะแนนเฉพาะสองแบบแรก ดังตารางที่ 1 และตารางที่ 2 ตามลำดับ

¹ P.L.Dressel, and J.Schmid, "Some Modification of the Multiple-Choice Item," Educational and Psychological Measurement, 13(1953):574-595.

ตารางที่ 1 วิธีการให้คะแนนของแบบสอบถามเลือกตอบเสรีตามวิธีของเกรสเชด

จำนวนคำตอบที่ทำเครื่องหมาย	คะแนน	
	ผู้สอบทำเครื่องหมายถูกต้อง	ผู้สอบไม่ได้ทำเครื่องหมายถูกต้อง
1	4	-1
2	3	-2
3	2	-3
4	1	-4
5	0	

ตารางที่ 2 วิธีการให้คะแนนของแบบสอบถามชนิดทดสอบความมั่นใจตามวิธีของเกรสเชด

ระดับความมั่นใจ	คะแนน	
	ผู้สอบทำเครื่องหมายถูกต้อง	ผู้สอบไม่ได้ทำเครื่องหมายถูกต้อง
มั่นใจมาก	4	-4
มั่นใจปานกลาง	3	-3
เคา แต่มีเหตุผลในการเคา	2	-2
ไม่มั่นใจ	1	-1

เครสเชด และซมิต พบว่าการทดสอบของนักเรียนเก่ง ซึ่งพิจารณาจากคะแนนการสอบ จะแตกต่างกับนักเรียนประเภท ปานกลาง และอ่อนอย่างมีนัยสำคัญ เมื่อใช้แบบสอบชนิดเลือกตอบเสรี คือกลุ่มเก่งจะทำเครื่องหมายคำตอบน้อยกว่ากลุ่มอ่อน ทั้งข้อสอบที่มีระดับความยาก (Difficulty level) สูง ปานกลาง และต่ำ ในทางตรงกันข้ามแบบสอบชนิดทดสอบความมั่นใจไม่ได้จำแนก นักเรียน เก่ง ปานกลาง และอ่อนได้ก็เท่าวิธีแรก เพราะเมื่อพิจารณาการทดสอบแล้วพบว่า ผู้เข้าสอบตอบข้อสอบที่มีความยากปานกลาง และยากมาก ด้วยระดับความมั่นใจอย่างเดียวกัน เครสเชด และซมิตได้สรุปไว้ว่า องค์ประกอบของความมั่นใจที่วัดได้โดยการใช่แบบสอบชนิดเลือกตอบเสรีนั้น แตกต่างกับที่วัดโดยแบบใหญ่ เขาสอบแสดงระดับความมั่นใจ

ปีคริสต์ศักราช 1965 ปรากฏข้อเขียนของอีเบลภายใต้หัวข้อเรื่องการทดสอบความมั่นใจใน คำว่าที่เกี่ยวกับการวัดทางจิตวิทยา¹ อีเบลได้บรรยายถึงวิธีการที่เพิ่มเติมเติมจากของไซเคอร์ควิสต์ โดย ใช้แบบสอบชนิดถูก - ผิด ที่มีวิธีการให้คะแนนละเอียดกว่า แรงจูงใจของอีเบลที่เขียนเรื่องนี้เช่นเดียวกับผู้บุกเบิก วิธีการทดสอบความมั่นใจคนอื่นในระยะแรก คือต้องการลดความคลาดเคลื่อนของการสอบที่เนื่องมาจากการเดา ข้อสอบที่เขาใช้จัดอยู่ในประเภทถูก - ผิด ซึ่งแต่ละข้อจะมีข้อความ (Statement) ใหญ่ เขาสอบประเมินความถูก หรือ ผิด ของข้อความนั้น คำตอบซึ่งแสดงการประเมินข้อความมี ทำ ระดับ และมีวิธีให้คะแนนดังตารางที่ 3

1

Ebel, Measuring Educational Achievement, pp. 130-135.

ตารางที่ 3 การให้คะแนนการทดสอบความมั่นใจตามวิธีของอีเบล

หมายเลขคำตอบ	ความหมาย	คะแนน		
		ถูก	ผิด	เว้น
1	ขอความนั้นน่าจะถูกต้อง	2	-2	
2	ขอความนั้นอาจจะถูกต้อง	1	0	
3	ไม่มีความรู้พอที่จะตอบ			.5
4	ขอความนั้นอาจจะผิด	1	0	
5	ขอความนั้นน่าจะผิด	2	-2	

อีเบล กำหนดหมายเลข 1 ถึง 5 ให้กับคำตอบที่แสดงการประเมินความถูกต้อง หรือ ผิดของข้อความในข้อกระทง และกำหนดคะแนน 2 ให้กับคำตอบที่แสดงความมั่นใจ เมื่อคำตอบนั้นถูกต้อง แต่ถ้ามืดก็
จะปรับเป็น -2 สำหรับคำตอบซึ่งแสดงความไม่มั่นใจ กำหนดน้ำหนักคะแนนให้เพียง 1 เมื่อคำตอบนั้น
ถูกต้อง และไม่ให้คะแนนเลยเมื่อคำตอบนั้นผิด แต่ถ้ามืดเลือกคำตอบหมายเลข 3 คือ ปฏิเสธที่จะ
ประเมินความถูกต้อง หรือผิดของข้อความในข้อกระทง เขากำหนดคะแนนให้ .5

ในไม่เคียวกันนี้ อีเบล ได้รายงานค่าความเที่ยงจากงานวิจัยของเขาซึ่งตีพิมพ์ในวารสาร
ฉบับหนึ่ง¹ ว่าเมื่อใช้วิธีให้คะแนนดังกล่าวกับการทดสอบนักเรียนทั้งหมดสามชั้นเรียน ปรากฏว่าให้
ค่าความเที่ยงที่สูงขึ้นทั้งสามชั้นเรียน

งานวิจัยชิ้นหลังสุดที่เกี่ยวกับการทดสอบความมั่นใจ เท่าที่พบในขณะนี้ดูเหมือนจะเป็นของ

1

R.L.Ebel, "Confidence Weighting and Test Reliability," Journal of Educational Measurement, 2 (1965) : 49-57.

ชิว - ซุง เวิน¹ (Shih-Sung Wen) ซึ่งตีพิมพ์ในปีคริสต์ศักราช 1975 ศึกษาความสัมพันธ์ระหว่างคะแนนที่ได้จากผลรวมของจำนวนข้อที่ตอบถูก กับคะแนนการทดสอบความมั่นใจ พบว่ามีสหสัมพันธ์กันสูงถึง .97 ซึ่งในทางปฏิบัติแล้ว คะแนนที่ได้จากทั้งสองวิธีสามารถนำมาจัดอันดับ (rank order) ได้เหมือนกัน แต่เขาก็ให้ข้อคิดว่า ถึงแม้สหสัมพันธ์ระหว่างคะแนนที่ได้จากสองวิธีนี้จะสูง แต่ก็ไม่ได้เป็นข้ออ้างที่จะปฏิเสธวิธีการทดสอบความมั่นใจ ถึงอย่างไรการทดสอบความมั่นใจก็สามารถให้รายละเอียด หรือข้อมูลในการตอบข้อสอบที่มากกว่า

การให้คะแนนตามสมการ (4) และ (5) ตลอดจนการทดสอบความมั่นใจ เป็นเทคนิคการให้คะแนนที่มีวัตถุประสงค์จะจัดโอกาสให้ผู้เข้าสอบจะได้คะแนนเพราะการเดา ซึ่งโดยทฤษฎีแล้ว เทคนิคการให้คะแนนแบบแก้การเดานี้ ย่อมจะทำให้ค่าความตรง และความเที่ยงสูงขึ้น ขณะเดียวกันก็ลดความคลาดเคลื่อนในการวัดลง ซึ่งก็ปรากฏผลงานของนักวิจัยหลายท่านสนับสนุนทฤษฎีนี้ แต่การศึกษาเพื่อเปรียบเทียบค่าความเที่ยงของการวัด เมื่อใช้เทคนิคการให้คะแนนตามสมการ (4) และ (5) ซึ่งอาศัยหลักของความน่าจะเป็น กับการทดสอบความมั่นใจซึ่งอาศัยค่ารายงานของผู้เข้าสอบเอง ดูเหมือนจะเป็นของใหม่ที่ยังไม่เคยมีใครศึกษามาก่อน

¹ Shih-Sung Wen, "The Relationship between Verbal Meaning Test Scores and Degree of Confidence in Item Responses," Journal of Educational Measurement, 12 (Fall 1975) : 197-199.