หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2
วรรณคดีที่เกี่ยวข้อง

ในบทนี้จะได้กล่าวถึงประเด็นที่มีส่วนเกี่ยวข้องกับการวิจัย 3 ประการคือ

1. ผลงานวิจัยที่เกี่ยวกับวิธีการตอบและการให้คะแนนแบบสอบเลือกตอบ เพื่อแสดงให้เห็นถึงแนวคิดและประเด็นปัญหาในความต้องการพัฒนาการสูตรการให้คะแนนเกี่ยวกับความรู้บางส่วนของผู้ตอบ
2. วิธีการพัฒนาสูตรการให้คะแนนของอาร์โนลด์และแอมดาน ซึ่งใช้เป็นวิธีการที่นำมาประยุกต์ในการสร้างสูตรการให้คะแนนสำหรับการวิจัยนี้
3. การศึกษาคุณภาพของสูตรการให้คะแนนด้านความตรงเชิงทฤษฎีและวิธีการวิเคราะห์

ในตอนต่อไปนี้เกี่ยวข้องกับรายละเอียดของแต่ละประเด็น ดังนี้

1. <u>ผลงานวิจัยที่เกี่ยวกับวิธีการตอบและการให้คะแนนในแบบสอบแบบเลือกตอบ</u>

แบบสอบแบบเลือกตอบที่ใช้กันอยู่มีรูปแบบการตอบที่กำหนดให้ผู้ตอบเลือกตอบตัวเลือกที่ถูกต้องมากที่สุดเพียงหนึ่งตัว ส่วนวิธีการให้คะแนนก็พิจารณาจากคำตอบว่าผู้ตอบเลือกตัวเลือกได้ถูกต้องหรือไม่ ถ้าเลือกถูก จะให้ 1 คะแนน และถ้าตอบผิดจะได้ 0 คะแนน วิธีการตอบและการให้คะแนนในแบบสอบเลือกตอบที่ใช้กันอยู่นี้ก่อให้เกิดปัญหาในเชิงทฤษฎีที่สำคัญ 2 ประการ คือ

ปัญหาแรก คือ ปริมาณที่ได้จากการวัดไม่ได้มาซึ่งสารสนเทศที่พอเพียงที่จะจำแนกผู้ตอบในระดับต่างๆของความรู้ ระหว่างผู้ที่รู้จริง ผู้ที่รู้บางส่วน และผู้ที่ไม่รู้ ออกจากกัน

ปัญหาที่สอง คือ ปัญหาการตอบแบบเดาสุ่มซึ่งเป็นแหล่งความคลาดเคลื่อนในการวัดอันเป็นสาเหตุทำให้ความตรงและความเที่ยงของการวัดลดลง

การศึกษาเพื่อแก้ปัญหาทั้ง 2 ประการ ประกอบด้วยการศึกษากระบวนการตอบแบบสอบเลือกตอบเพื่อให้ได้มาซึ่งปริมาณสารสนเทศสูงสุดในการวัดเป็นการศึกษาที่เกี่ยวกับวิธีการตอบ และ การให้คะแนน (the response/scoring method) โดยยอมให้ผู้ตอบได้แสดงความรู้บางส่วน และ การศึกษาที่เกี่ยวกับการให้คะแนนในลักษณะต่างๆเพื่อแสดงระดับความรู้ของผู้ตอบได้ตรงกับความสามาถของเขาส่วนการแก้ปัญหาการเดาสุ่มของผู้ตอบงานส่วนใหญ่เกี่ยวข้องกับการพัฒนาสูตรการให้คะแนนเพื่อแก้การเดาของผู้ตอบ งานวิจัยในประเด็นดังกล่าว จะเป็นกรอบในการแก้ปัญหาประเด็นการวัดทั้ง 2 สำหรับงานวิจัยนี้ต่อไป

1.1 การศึกษาเกี่ยวกับความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบ

การศึกษาเกี่ยวกับความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบมีวิธีการศึกษาแตกต่างกันในงานวิจัยที่เกี่ยวกับวิธีการตอบและการให้คะแนน (the response/scoring method) ของแบบสอบเลือกตอบทั้งนี้ขึ้นอยู่กับการนิยามความรู้บางส่วนของผู้วิจัย ซึ่งมีอยู่ด้วยกันหลายลักษณะ ดังนี้

ก. การที่ผู้ตอบแบบสอบเลือกตอบสามารถตัดตัวลวงบางตัวออกได้อย่างถูกต้องแม้ว่าจะไม่ทราบว่าคำตอบที่ถูกต้องคืออะไร (Coombs, Milholland,and Womer, 1956)

ข. การที่ผู้ตอบแสดงโอกาสที่สูงกว่าในการเลือกตัวเลือกที่ถูกหลังจากเลือกตัวเลือกที่ผิดเป็นคำตอบแรกมาแล้ว

ค. การที่ผู้ตอบแสดงความมั่นใจในระดับที่สูงกว่าในคำตอบที่ถูกแทนที่จะมีความมั่นใจต่ำในคำตอบที่ถูก (Hutchinsun, 1982)

ง. การที่ผู้ตอบแสดงความน่าจะเป็นเชิงอัตนัย (Subjective probability) ในการแสดงต่อตัวเลือกทุกตัวว่าจะเป็นคำตอบที่ถูกต้อง (de Finetti, 1965)

ผลงานวิจัยที่ศึกษาเกี่ยวกับความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบในนิยามต่างๆ พอสรุปได้ ดังนี้

1.1.1 ผลงานวิจัยที่เกี่ยวกับความรู้บางส่วนในความหมายที่ผู้ตอบสามารถตัดตัวลวงออกไปได้บ้างอย่างถูกต้องแม้จะไม่ทราบว่าคำตอบที่ถูกต้องคืออะไร ได้แก่ งานที่เกี่ยวกับวิธีการตอบและการให้คะแนนแบบคูมบ์ วิธีการตอบและการให้คะแนนแบบอาร์โนลด์ วิธีการตอบและการให้คะแนนแบบครอส วิธีการให้คะแนนแบบแฟรรี่ย์ วิธีการเลือกชุดย่อยของคำตอบ ซึ่งมีรายละเอียดของแต่ละวิธี และผลงานวิจัยที่ให้การสนับสนุนวิธีต่างๆ คือ

<u>วิธีการตอบและการให้คะแนนแบบคูมบ์</u> (the Coombs response/scoring method หรือ CBS) Coombs ได้จำแนกปริมาณความสามาถของผู้ตอบแบบสอบ จากสารสนเทศในการวัดออกเป็น 4 ประเภทคือ ประเภทที่รู้จริง (complete information) คือผู้ที่สามารถตอบได้อย่างถูกต้องว่าคำตอบคืออะไร และ สามารถจำแนกตัวเลือกได้ว่าตัวเลือกใดเป็นคำตอบตัวเลือกใดเป็นตัวลวง ประเภทที่สองรู้ผิด (misinformation) เป็นผู้ที่เข้าใจว่าตัวเลือกที่ถูกเป็นตัวลวงและไม่มีพื้นฐานในการจัดประเภทตัวเลือกปรากฏเลยว่าเป็นคำตอบหรือเป็นตัวลวง ประเภทที่สามรู้บางส่วน (partail information) เป็นผู้ที่ไม่สามารถตอบได้ว่า ตัวเลือกใดเป็นคำตอบที่ถูกต้องแต่สามารถตัดตัวลวงบางตัวออกไปได้แต่ไม่ใช่ทั้งหมด และประเภทที่สี่ รู้ผิดบางส่วน (partail

misinformation) เป็นผู้ที่คิดว่าตัวเลือกที่เป็นคำตอบเป็นตัวลวง และเลือกหาคำตอบจากตัวลวงที่เหลืออยู่แต่ก็มีตัวลวงบางตัวที่สามารถตัดออกไปได้บ้าง (Coombs, and others, 1956)

วิธีการตอบและการให้คะแนนแบบคูมบ์ ด้วยรูปแบบเลือกตอบแบบเดิมที่มีตัวเลือกที่เป็นคำตอบถูกเพียงตัวเดียว จะให้ผู้ตอบตัดตัวลวงที่ไม่ใช่คำตอบที่เขาแน่ใจออกไปมากที่สุด การให้คะแนนผู้ตอบจะได้รับ 1 คะแนนในแต่ละตัวลวงที่ตัดออกได้อย่างถูกต้องและในกรณีที่ผู้ตอบมีความรู้ผิดตัดตัวเลือกที่ถูกออกไปโดยเข้าใจว่าเป็นตัวลวง เขาจะถูกลงโทษโดยถูกหักคะแนนออกไปเท่ากับจำนวนตัวลวงในข้อสอบนั้น หรือเท่ากับจำนวนตัวเลือกลบด้วยหนึ่ง จากผลการวิจัยของ Coombs, and others (1956) พบว่าการให้คะแนนแบบคูมบ์จะทำให้ความเที่ยงของแบบสอบเพิ่มขึ้น เมื่อเทียบกับระบบคะแนน 0 หรือ 1 แบบเดิม และจากงานของ Bradbard and Green (1986) พบว่าวิธีการของคูมบ์ช่วยลดการเดาของผู้ตอบและทำให้ผู้ตอบแบบสอบมีความระมัดระวังในการตอบมากขึ้น ความเที่ยงด้วยระบบการให้คะแนนแบบคูมบ์จะให้ผลเหนือกว่าวิธีเดิม

<u>วิธีการตอบและการให้คะแนนแบบอาร์โนลด์</u> (The Arnold Scoring/Response Method หรือ AA) วิธีนี้ใช้กับข้อสอบแบบเลือกตอบรูปแบบเดิมวิธีการตอบเช่นเดียวกับวิธีการของคูมบ์ โดยให้ผู้ตอบกาตัวเลือกที่เขาแน่ใจว่าเป็นตัวลวง หรือ คำตอบที่ผิดออกไป แต่วิธีนี้มีระบบการให้คะแนนที่ต่างออกไป โดยจะให้ 1 คะแนนสำหรับทุกตัวลวงที่ถูกกาออกไปอย่างถูกต้องแต่เมื่อใดที่ผู้ตอบกาตัวเลือกที่ถูกออกไปโดยเข้าใจว่าเป็นตัวลวงจะได้ $-1/(k-1)$ คะแนน (เมื่อ k คือจำนวนตัวเลือกในแต่ละข้อ) แม้ว่าเขาจะตัดตัวลวงตัวอื่นออกไปได้ถูกต้องบ้างแล้วก็ตาม ระบบการให้คะแนนของอาร์โนลด์นี้วางอยู่บนพื้นฐานของการให้คะแนนเฉลี่ยรายข้อ (expected item score) เท่ากับ 0 เมื่อผู้ตอบเดาสุ่มจากผลการวิจัยของ Arnold and Arnold (1970) พบว่าวิธีการให้คะแนนความรู้บางส่วนของเขาในกรณีที่ผู้ตอบสามารถตัดตัวลวงบางตัวออกได้อย่างถูกต้อง และไม่ให้คะแนนเมื่อรู้ผิดหรือเดาสุ่ม พบว่าลำดับที่ของผู้ตอบไม่เปลี่ยนแปลงเมื่อเทียบกับวิธีการให้คะแนนแบบเดิม แต่วิธีนี้สามารถจำแนกลำดับที่ผู้ตอบในตอนกลางๆออกได้ดี ซึ่งมักจะรวมกันเป็นกลุ่มในการให้คะแนนแบบเดิม

<u>วิธีการตอบและการให้คะแนนแบบครอส</u> (The Cross Response/Scoring Method หรือ CRS) วิธีการนี้การตอบให้ผู้ตอบเลือกชุดย่อยของคำตอบที่เป็นตัวลวงเช่นเดียวกับวิธีของคูมบ์ และอาร์โนลด์ แต่วิธีการให้คะแนนเปลี่ยนไป โดยวิธีการให้คะแนนของครอสจะให้คะแนนรายข้อเฉลี่ยต่ำสุดเมื่อผู้ตอบเลือกตัวเลือกที่เป็นคำตอบโดยเข้าใจว่าเป็นตัวลวงและเลือกตัวลวงอีก c-2 ตัว ว่าเป็นชุดของตัวเลือกที่ผิด ซึ่งผิดกับการให้คะแนนแบบคูมบ์ที่คะแนนรายข้อเฉลี่ยต่ำสุดจะเกิด จากการเลือกตัวเลือกที่ถูกว่าเป็นตัวลวงแม้เพียงคำตอบเดียว การให้คะแนนแบบครอสผู้ตอบจะได้ 2 คะแนนในแต่ละตัวลวงที่เขาตัดออกได้อย่างถูกต้อง และจะได้ 1 คะแนนสำหรับแต่ละตัวเลือกที่ไม่ได้กา

ว่าเป็นตัวลวง เมื่อคำตอบถูกยังไม่ได้ถูกกาแต่เมื่อใดที่ผู้ตอบกาตัวเลือกที่ถูกด้วย misinformation เขาจะได้ตัวเลือกละ 1 คะแนน สำหรับแต่ละตัวลวงที่ไม่ได้กา สำหรับในกรณีที่ข้อสอบมี 4 ตัวเลือก คะแนนที่ยอมให้เกิดขึ้นได้ มีค่าตั้งแต่ 1 คะแนน ถึง 7 คะแนน และ คะแนนต่ำสุดจะเกิดจากการที่ผู้ตอบกาตัว.ลือกที่ถูกต้องหลังจากการกาตัวลวงมาแล้ว 2 ตัว Frary (1982) เปรียบเทียบวิธีการให้คะแนนแบบครอสกับวิธีการให้คะแนนอื่นๆพบว่าคะแนนแบบครอสให้ค่าความเที่ยงต่ำกว่าวิธีการให้คะแนนของอาร์โนลด์และคูมบ์ส่วนความตรงตามเกณฑ์ไม่แตกต่างกัน

<u>วิธีการให้คะแนนแบบแฟรรี่ย์</u> (The proposed new method of Frary หรือ PNM) วิธีนี้เสนอโดย แฟรรี่ย์เขาต้องการจะปรับวิธีการให้คะแนนของครอสใหม่จากช่วงคะแนน 1 ถึง 7 เป็น -1 ถึง 1 คะแนน เพราะวิธีของครอสในกรณีที่ไม่มีการเลือกตัวเลือกใดเลยผู้ตอบจะได้คะแนนเท่ากับจำนวนตัวเลือกในข้อสอบ แฟรรี่ย์เสนอว่าเขาควรได้ 0 คะแนน และใช้ระบบการแปลงคะแนนเชิงเส้นจาาคะแนนของครอสใหม่ Frary (1980) ได้วิเคราะห์ผลของการให้คะแนนแก่ความรู้บางส่วน การลงโทษผู้รู้ผิด และการเดาจากคะแนนเฉลี่รายข้อของวิธีการให้คะแนนแบบต่างๆ ให้ข้อเสนอว่าวิธี PNM เป็นวิธีการให้คะแนนวิธีหนึ่งที่เหมาะกับการให้คะแนนความรู้บางส่วนเช่นเดียวกับวิธีการให้คะแนนของคูมบ์และอาร์โนลด์ ที่จะช่วยให้ความเที่ยงของการวัดสูงขึ้น

<u>วิธีการเลือกชุดย่อยของคำตอบ</u> (A Subset Selection Technique) วิธีนี้จะให้ผู้ตอบเลือกชุดย่อยของคำตอบที่เขาเห็นว่าน่าจะมีตัวเลือกที่เป็นคำตอบถูกรวมอยู่ด้วย การให้คะแนนขึ้นอยู่กับจำนวนตัวเลือกที่มีอยู่ในชุดย่อยของคำตอบ ยิ่งน้อยเท่าใดคะแนนจะได้เพิ่มขึ้น ในกรณีที่ตัวเลือกที่ถูกต้องรวมอยู่ในชุดย่อยของคำตอบนั้นด้วย แต่ถ้าชุดย่อยนั้นไม่มีคำตอบที่ถูกต้องอยู่ด้วยคะแนนจะถูกหักออกไปเท่ากับขนาดของจำนวนตัวเลือกที่ผู้ตอบเลือก ในกรณีที่ผู้ตอบไม่เลือกตัวเลือกใดเลยจะได้ 0 คะแนนวิธีการนี้ Gibbon, Olkin, and Sobel (1979) ได้ศึกษาพบว่าเป็นวิธีที่ช่วยลดการเดาสุ่มของผู้ตอบลง และยังยอมให้ผู้ตอบได้แสดงความรู้บางส่วนของผู้ตอบด้วย

1.1.2 ผลงานวิจัยที่เกี่ยวกับความรู้บางส่วนในความหมายที่ผู้ตอบสามารถแสดงโอกาสที่สูงกว่าการเลือกตัวเลือกที่ถูก หลังจากเลือกตัวเลือกที่ผิดเป็นคำตอบแรกมาแล้ว วิธีการนี้คือวิธีการตอบจนกว่าจะพบข้อที่ถูก (The answer-until correct method หรือ AUC) วิธีนี้จะให้ผู้ตอบเลือกคำตอบที่เขาแน่ใจว่าเป็นคำตอบที่ถูกต้องทีละตัวเลือกจนกว่าจะพบข้อที่ถูกจริงๆ การกาคำตอบจะกาลงในกระดาษคำตอบพิเศษที่ให้ผลย้อนกลับทันทีต่อผู้ตอบ คะแนนที่ได้รับจะเท่ากับจำนวนตัวเลือกที่เหลืออยู่หลังจากที่คำตอบถูกได้ถูกเปิดเผยแล้ว วิธีนี้ได้รับการสนับสนุนว่าช่วยส่งเสริมการเรียนรู้ของผู้ตอบ ระบบการให้คะแนนจะแก้ไขการเดาสุ่มของผู้ตอบได้ และช่วยเพิ่ม

ความเที่ยงของแบบสอบ แต่การให้ตอบและการให้คะแนนด้วยวิธีนี้ได้รับการคัดค้านจากระบบการให้คะแนนของครอส เพราะครอสเสนอว่าการที่ผู้ตอบเลือกตอบตัวเลือกที่เขาเห็นว่าถูกไปจนพบการเฉลย ไม่ใช่เป็นวิธีวัดความเข้าใจจริงๆของผู้ตอบที่เชื่อว่าตัวเลือกใดเป็นตัวลวงจริงๆเพราะเขาได้รับการเฉลยคำตอบเสียก่อน ดังนั้นหากไม่มีการเฉลยคำตอบและให้ผู้ตอบเลือกต่อไป ก็จะแสดงถึงระดับความรู้ของเขาจริงๆการให้คะแนนแบบครอสดังที่กล่าวมาจึงเป็นระบบการให้คะแนนที่เพิ่มขึ้นจากวิธีการตอบจนกว่าจะถูกนี้ (Cross, quoted in Frary 1980)

1.1.3 ผลงานวิจัยที่เกี่ยวกับความรู้บางส่วนรู้ความหมายที่ผู้ตอบสามารถแสดงความมั่นใจในระดับที่สูงกว่าสำหรับตัวเลือกที่เป็นคำตอบ แทนที่จะมีความมั่นในต่ำในตัวเลือกที่ถูกวิธีการนี้ได้แก่วิธีการที่ให้ผู้ตอบได้ตอบพร้อมแสดงความมั่นใจต่อคำตอบที่ตอบนั้น (confident testing) วิธีการนี้นอกจากจะให้ผู้ตอบเลือกคำตอบที่ถูกแล้วยังให้ผู้ตอบได้ตอบว่าเขามีความมั่นใจในระดับใดในคำตอบของเขาว่าจะเป็นคำตอบที่ถูกต้อง การให้คะแนนจะมากหรือน้อยขึ้นอยู่กับระดับความมั่นใจที่ผู้ตอบได้แสดงเมื่อเขาเลือกคำตอบที่ถูก และการให้คะแนนจะลดลงเมื่อเขาแสดงความมั่นใจสูงในคำตอบที่จริง ๆเป็นคำตอบที่ผิด วิธีการดังกล่าวนี้นำมาใช้เพื่อควบคุมการเดาในแบบสอบถูกผิด จากการทบทวนผลงานวิจัยที่เกี่ยวข้องกับการสอบความมั่นใจของผู้ตอบ Echternacht (1972) พบว่าคะแนนที่ผู้ตอบได้รับขึ้นอยู่กับบุคลิกภาพของผู้ตอบว่าจะเสี่ยงที่จะแสดงความมั่นใจในการตอบของเขาสักเพียงใด ซึ่งทำให้ผลการวัดมิได้วัดในสิ่งที่ต้องการวัดจริง แม้ว่าผลการวิจัยจะแสดงว่าวิธีการนี้ทำให้เกิดความเที่ยงของแบบสอบสูงขึ้นกว่าการตอบ และ การให้คะแนนแบบ 0,1 แต่วิธีนี้ก็ต้องการเวลาในการตอบ และเวลาในการให้คะแนนมากขึ้นกว่าวิธีการแบบเดิม

1.1.4 ผลงานวิจัยที่เกี่ยวกับความรู้บางส่วนในความหมายที่ผู้ตอบได้แสดงความน่าจะเป็นเชิงอัตนัย (subjective probability) ต่อตัวเลือกทุกตัวว่าเป็นคำตอบที่ถูกต้อง ได้แก่การใช้ความน่าจะเป็นส่วนบุคคล (personal probability approaches) วิธีการนี้จะให้ผู้ตอบกำหนดความน่าจะเป็นกับตัวเลือกทุกตัวในข้อสอบแต่ละข้อว่ามีความน่าจะเป็นที่จะเป็นคำตอบที่ถูกสักเพียงใดโดยให้ผลรวมของความน่าจะเป็นในข้อหนึ่งไม่เกิน 1ในวิธีการของ de finetti (1965) ได้กำหนดดาว 5 ดวง ให้กับผู้เข้าสอบที่จะแสดงความน่าจะเป็นในการตอบถูกแต่ละตัวเลือก ซึ่งดาวแต่ละดวงจะแทนค่าความน่าจะเป็น 0.2 เมื่อมี 5 ตัวเลือก ผู้ตอบจะได้รับคำชี้แจงให้แจกแจงดาวระหว่างตัวเลือกที่เขาพิจารณาว่าเป็นตัวเลือกที่ถูกต้องด้วยความน่าจะเป็นตามจำนวนดาวที่ให้ วิธีการนี้มีข้อตกลงเบื้องต้นหลายประการ แต่ที่สำคัญก็คือผู้ตอบต้องเข้าใจต่อวิธีการตอบและการให้คะแนนอย่างแจ่มชัดเพื่อที่เขาจะได้ใช้กลวิธีในการที่จะได้มาซึ่งคำตอบที่ได้คะแนนสูงสุด และมีข้อวิจารณ์ว่าวิธีการของ de Finetti นี้เสนออยู่ในรูปของทฤษฎีที่ไม่ได้มีการพิจารณาถึงองค์ประกอบในทางปฏิบัติโดยเฉพาะองค์ประกอบทางจิตวิทยา

ด้วยความไม่แน่ใจว่าผู้ตอบข้อสอบจะเข้าใจอย่างเต็มที่ในผลของการแสดงความน่าจะเป็น ด้วยคำตอบของเขาต่อระดับความถูกต้องของตัวเลือกที่เขาพิจารณา นอกจากนี้วิธีการนี้ก็เป็นวิธีการที่มีความซับซ้อนในการตอบและการให้คะแนนมาก

1.1.5 นอกจากวิธีการตอบและการให้คะแนนที่กล่าวมาในการศึกษาความรู้บางส่วนของผู้ตอบ ยังมีวิธีการที่ให้ผู้ตอบตอบทุกตัวเลือก (multiple true-false method) วิธีนี้จะให้ผู้ตอบตอบทุกตัวเลือกว่าเป็นคำตอบของคำถามข้อนั้นหรือไม่ โดยการตอบว่าถูกหรือผิดในแต่ละตัวเลือก ปัญหาของวิธีการนี้อยู่ที่วิธีการให้คะแนน ซึ่งมีความแตกต่างกันหลายวิธี เช่น ผู้ตอบต้องตอบทุกตัวเลือกได้ถูกต้องจึงจะได้คะแนน 1 คะแนน ถ้าผิดไปจากนี้จะไม่ได้คะแนน ซึ่งวิธีการให้คะแนนเช่นนี้ไม่ได้มีการพิจารณาความรู้บางส่วนของผู้ตอบเลย ทำให้สารสนเทศที่มีผู้ตอบบางคนที่สามารถตอบในบางตัวเลือกได้ถูกต้องหายไป อีกวิธีหนึ่งก็เป็นการพิจารณาให้คะแนนทุกตัวเลือกที่ผู้ตอบ ตอบถูกโดยถือเสมือนว่าการตอบแต่ละตัวเลือกก็คือข้อสอบถูกผิด 1 ข้อ แต่วิธีการนี้ได้รับการศึกษาและวิจารณ์ถึงความไม่เหมาะสมที่จะกระทำให้ตัวเลือกแต่ละตัวคือข้อสอบ 1 ข้อ อย่างเป็นอิสระจากกัน เพราะจากการศึกษาพบว่าความสัมพันธ์ของตัวเลือกนั้น ภายในข้อเดียวกันสูงมากกว่าความสัมพันธ์ในระหว่างข้อที่แสดงให้เห็นถึงความไม่เป็นอิสระระหว่างตัวเลือกในข้อเดียวกันที่พอจะแยกให้คะแนนออกเป็นข้อๆได้ (Albanese and Sablers, 1988) แต่อย่างไรก็ตามเมื่อเปรียบเทียบวิธีการให้คะแนนที่ผู้ตอบต้องตอบทุกตัวเลือกถูกจึงจะได้คะแนนกับการให้คะแนนที่แสดงความรู้บางส่วนของผู้ตอบโดยให้คะแนนถูกผิดในแต่ละตัวเลือกพบว่า วิธีการหลังที่ให้คะแนนความรู้บางส่วนแก่ผู้ตอบมีประสิทธิภาพเชิงสัมพัทธิ์ (relative efficiency) เหนือกว่าวิธีที่ไม่ให้ความรู้บางส่วนแก่ผู้ตอบ ตั้งแต่ผู้ตอบที่มีระดับความสามารถต่ำจนถึงสูง ในขณะที่การให้คะแนนแบบไม่ให้ความรู้บางส่วนมีประสิทธิภาพเชิงสัมพัทธิ์สูงกว่าวิธีอื่นๆ ในกรณีความสามารถของผู้ตอบสูงมากเท่านั้น (จักรกฤษณ์ สำราญใจ, 2531)

จากการวิเคราะห์วิธีการตอบและการให้คะแนนโดยยอมให้ผู้ตอบได้แสดงความรู้บางส่วนและให้คะแนนแก่ผู้ตอบตามความรู้ที่เขาแสดงออกมานี้ พบว่าวิธีการให้ผู้ตอบได้แสดงความมั่นใจในคำตอบที่ตอบ ยังมีจุดอ่อนเพราะคะแนนที่ได้แสดงลักษณะทางบุคลิกภาพของผู้ตอบ เช่นการกล้าเสี่ยง แทนที่จะเป็นความสามารถของผู้ตอบที่ต้องการวัดจริงๆ ในส่วนวิธีการที่ให้ผู้ตอบแสดงความน่าจะเป็นส่วนบุคคลต่อตัวเลือกว่าจะเป็นคำตอบถูกหรือไม่ ยังมีปัญหาอยู่ว่าผู้ตอบมีความเข้าใจเกี่ยวกับงานที่เขาจะต้องตอบข้อสอบโดยการแจกแจงค่าความน่าจะเป็นให้กับตัวเลือกแต่ละตัวได้ตรงกับที่เขามีความรู้หรือไม่ แนวทางในการศึกษาความรู้บางส่วนของผู้ตอบส่วนใหญ่จึงอยู่ที่การที่ผู้ตอบสามารถเลือกชุดย่อยของตัวเลือกได้ว่าควรที่จะเป็นตัวเลือกถูก หรือ ควรที่จะเป็นตัวลวงโดยที่เขาสามารถจัดตัวเลือกบางตัวเข้าใน

กลุ่มหนึ่งกลุ่มใด ซึ่งเป็นวิธีการตามโมเดลที่สอง จึงควรจะเป็นวิธีการที่เหมาะกับการตอบแบบสอบเลือกตอบเพื่อแสดงความรู้บางส่วนของผู้ตอบแทน วิธีการตอบและการให้คะแนนที่เปลี่ยนแปลงจากการเลือกตัวเลือกที่ถูกต้องเพียงตัวเดียว (single response) และจากการให้คะแนนในระบบ 0,1มาเป็นการตอบที่ตอบชุดของตัวเลือกที่เห็นว่าถูกหรือชุดที่เห็นว่าผิด (multiple response) หรือว่าจะเป็นการตอบจนกว่าจะพบข้อถูกก็ตาม จะมีการให้คะแนนที่เกิดขึ้นมากไปหรือน้อยไปเมื่อเปรียบเทียบในแต่ละวิธีการ การศึกษาของ Frary (1980) ในการเปรียบเทียบคะแนนเฉลี่ยรายข้อของวิธีการตอบและการให้คะแนนแบบต่างๆให้ข้อสังเกตว่าแต่ละวิธีมีการลงโทษกรณีของ misinformation และให้รางวัลในกรณีของ partail knowledge ที่กำหนดขึ้นตามอำเภอใจแต่ Frary เขาได้ให้ข้อสนับสนุนว่าวิธีการให้คะแนนแบบอาร์โนลด์ (A) ผู้ตอบจะได้รับการคาดหวังให้รายงานตำแหน่งจริงของความรู้ (true state of knowledge) ซึ่งวิธีการให้คะแนนจะมีการควบคุมเมื่อมีการเดาเกิดขึ้นอย่างยุติธรรมโดยให้ค่าเฉลี่ยรายข้อที่เกิดขึ้นจากการเดาสุ่มมีค่าเป็นศูนย์ วิธีการนี้จึงเป็นการลดความแปรปรวนของการเดาซึ่งเป็นแหล่งความคลาดเคลื่อนลง จากผลการประเมินของ Frary แนะนำว่าวิธีการให้คะแนนของอาร์โนลด์ให้ผลเหนือกว่าวิธีอื่น แต่อย่างไรก็ตามจากการเปรียบเทียบวิธีการตอบและการให้คะแนนแบบต่างๆนี้กับวิธีแบบเดิม พบว่าวิธีการเหล่านี้ให้ค่าความเที่ยงของแบบสอบเพิ่มขึ้นเล็กน้อยแต่ความตรงเชิงทำนายที่ได้จากผลการวิจัยไม่ได้แตกต่างไปจากวิธีการแบบเดิม ซึ่ง Frary ได้ให้พิจารณาเกี่ยวกับเกณฑ์ที่เลือกใช้ในการหาความตรงว่าเป็นเกณฑ์ที่เหมาะสมหรือไม่

1.2 <u>การศึกษาเกี่ยวกับการให้คะแนนในแบบสอบแบบเลือกตอบ</u>

การให้คะแนนตามทฤษฎีการตอบข้อสอบ (Item Response Theory หรือ IRT) Lord (1980 ได้เสนอสูตรการให้น้ำหนักคะแนนที่เหมาะสมในการตอบข้อสอบรายข้อ ในแต่ละระดับความสามารถ (θ) ของผู้ตอบไว้ดังนี้

$$w_i(\theta) = p_i'(\theta)/p_i(\theta)q_i(\theta)$$

เมื่อ $w_i(\theta)$ คือ น้ำหนักที่เหมาะสมของข้อสอบข้อที่ i ณ ผู้สอบที่มีความสามารถ θ

$p_i(\theta)$ คือ ค่าความน่าจะเป็นที่ผู้ตอบข้อสอบข้อ i ที่ความสามารถ θ สามารถตอบได้ถูกต้อง ซึ่งสามารถแสดงได้ด้วยฟังก์ชั่นโลจิสติคตามโมเดลพารามิเตอร์ 1 หรือ 2 หรือ 3 ตัว

$q_i(\theta)$ คือ $1-p(\theta)$

$p_i'(\theta)$ คือ first derivative ของฟังก์ชั่นโลจิสติค

Lord (1980) สามารถพิสูจน์ได้ว่าเมื่อกำหนดน้ำหนักคะแนนที่เหมาะสมรายข้อตามระดับความสามาถของผู้ตอบแล้วและรวมคะแนนทุกข้อเข้าด้วยกัน จะทำให้ได้สารสนเทศสูงสุดจากระบบการให้คะแนนนี้ในทางปฏิบัติเราไม่สามารถที่จะทราบค่าความสามารถ (θ) ของแต่ละคนได้ ทำให้เราไม่ทราบค่า $w_i(\theta)$ จริงๆ แต่เราสามารถประมาณค่าความสามารถ (θ)ได้ เมื่อกำหนดให้ $p_i(\theta) = c + (1-c)/[1+e^{-1.7a(\theta-b)}]$ ซึ่งเป็นฟังก์ชั่นโลจิสติคแบบพารามิเตอร์ 3 ตัว จะได้

$$w_i(\theta) = [Da_i(p_i-c_i)/p_i(1-c_i)]$$
$$= [Da_i/(1+c_i e^{-DL}) \quad \text{เมื่อ } L_i = a_i(\theta-b_i)$$

สังเกตว่า c=0 ในกรณีของพารามิเตอร์ 2 ตัว ค่า $w_i(\theta) \rightarrow Da_i$ ซึ่งก็คืออำนาจจำแนกของข้อสอบนั่นเอง ด้วยเพราะว่า D คือค่าคงที่ ที่ 1.7 $w_i(\theta) \rightarrow Da_i$ และในกรณีของพารามิเตอร์หนึ่งตัว ค่าอำนาจจำแนกของข้อสอบเท่ากันทุกข้อ น้ำหนักที่เหมาะสมก็คือ 1 ซึ่งก็คือการรวมคะแนนรายข้อเข้าด้วยกันในวิธีการแบบเดิมโดยไม่มีการถ่วงน้ำหนักนั่นเอง ในกรณีที่ไม่คิดถึงการเดาของการตอบข้อสอบ (c=0) Lord พิสูจน์ให้เห็นว่าค่าน้ำหนักคะแนนที่เหมาะสมรายข้อ $w_i(\theta)$ สามารถประมาณด้วยค่าอำนาจจำแนกของข้อสอบ (a_i) ซึ่งไม่ขึ้นอยู่กับค่าความสามารถ (θ) ของผู้ตอบและได้เสนอการรวมคะแนนเชิงประกอบที่ได้ถ่วงน้ำหนักที่เหมาะสมรายข้อ คือ $S = \Sigma_i a_i u_i$ ในกรณีของ two-parameter logistic function ซึ่ง $w_i(\theta)=a_i$ คืออำนาจจำแนกของข้อสอบโดยไม่ขึ้นอยู่กับค่าความสามารถของผู้ตอบนั่นเอง

จากการศึกษาวิธีการให้น้ำหนักคะแนนที่เหมาะสมด้วย $w_i=a_i$ ที่นำมาใช้ในแบบสอบเชิงถ้อยคำ(SCAT II 2 A)ด้วย information curve พบว่าคะแนน $\Sigma_i a_i u_i$ มีประสิทธิภาพเหมาะสมกับระดับความสามารถสูงๆแต่มีประสิทธิภาพน้อยลงกว่าคะแนนรวมข้อถูกที่ระดับความสามารถต่ำๆ ซึ่งสิ่งนี้เป็นผลที่ได้รับจากแบบสอบเลือกตอบที่กำหนดให้ $w_i=a_i$ ในกรณีที่ไม่คิดถึงค่าการเดา (Lord, 1980) ในเรื่องค่าความเที่ยงของแบบสอบเมื่อเปรียบการให้คะแนนที่มีการถ่วงน้ำหนักรายข้อที่เหมาะสมกับการให้คะแนนแบบรวมข้อถูกแบบเดิมพบว่าจะให้ค่าความเที่ยงสูงขึ้น (นภา กาญจนกิจโสภณ, 2528 และ สุพัฒน์ สุกมลสันต์, 2528) อย่างไรก็ตามการกำหนดน้ำหนักคะแนนที่เหมาะสมรายข้อนี้ยังวางอยู่บนพื้นฐานของวิธีการตอบและการให้คะแนนเฉพาะข้อที่ตอบถูกแบบเดิมซึ่งไม่ได้มีการพิจารณาให้ความรู้บางส่วนแก่ผู้ตอบ ด้วยระบบของวิธีการให้คะแนนคำตอบถูก 1 คะแนน คำตอบผิด 0 คะแนน ซึ่งเป็นระบบการให้คะแนนแบบ dichotomous scoring ใน Binary Model เป็นระบบการให้คะแนนที่ให้ความสนใจแต่การเลือกตัวเลือกที่ถูกเท่านั้น ไม่สนใจคำตอบอื่นๆของผู้ตอบ มีผลการวิจัยยืนยันว่าการใช้คำตอบผิดของผู้ตอบมาเป็นข้อมูลในการประมาณค่า

ความสามารถของผู้ตอบด้วยจะทำให้ได้ค่าประมาณความสามารถที่แม่นยำขึ้น (Levine and Drasgow, 1983) ผลการวิจัยนี้นำไปสู่การใช้ Polychotomous test model ว่าเป็นเทคนิคการวัดที่เหมาะสมและมีอำนาจมากกว่าการประมาณความสามารถจากคำตอบถูกเพียงอย่างเดียวในโมเดลการให้คะแนนดังกล่าวเสนอระบบการให้คะแนนที่จำแนกคำตอบของผู้ตอบมากกว่า 2 ประเภท คือแทนที่จะมีประเภทคำตอบถูก และ ผิด จะจัดคำตอบผิดออกเป็นประเภทต่าง ๆ มากกว่าที่จะจัดรวมเข้าไว้เป็นคำตอบผิดเหมือนกัน ถ้ามีระดับของความไม่ถูกต้องในระดับต่างๆของตัวเลือกหรือถ้าผู้ตอบมีความสามารถต่างกันได้แสดงรูปแบบที่แตกต่างกันในการเลือกตัวผิด Polychotomous test model ก็ควรจะมีประโยชน์ในการประมาณความสามารถ (θ) ของผู้ตอบ (Hulin, Drasgow, and Parsons, 1983) โมเดลดังกล่าวนี้ได้รับการสนับสนุนว่าเป็นโมเดลที่เหมาะสมในแบบสอบเลือกตอบ (Thissen and Steinberg, 1984) แนวทางในการได้มาซึ่งสารสนเทศที่แสดงความสามารถของผู้สอบระดับต่างๆในประเด็นนี้จึงอยู่ที่ระบบการให้คะแนนรายข้อที่เหมาะสมที่จะแสดงระดับของความสามารถของผู้ตอบได้ตรงกับความสามารถจริงๆของเขา

จากการศึกษางานวิจัยที่เกี่ยวกับการให้น้ำหนักคะแนนรายข้อหรือแบบสอบย่อย Wang และ Stanley (1968) ได้ให้ข้อสรุปว่าน้ำหนักที่กำหนดให้กับแต่ละข้อนั้นมีผลต่อความแปรปรวนของแบบสอบทั้งฉบับไม่มากไปกว่าเป็นเพียงน้ำหนักตามตัวเลขที่กำหนดขึ้นโดยผู้วิจัยเท่านั้น แท้ที่จริงความแปรปรวนของการวัดเชิงประกอบ (Composite measure) ขึ้นอยู่กับ (1) น้ำหนัก ตามตัวเลขที่กำหนดขึ้น (2) ความแปรปรวนของข้อกระทงแต่ละข้อ (3) สหสัมพันธ์ระหว่างข้อกระทงหนึ่งกับข้อกระทงอื่น ๆใน การวัดเชิงประกอบนั้น และ (4) ความแปรปรวนของข้อกระทงข้ออื่นๆ ในการวัดเชิงประกอบนั้น จากงานวิจัยของ Stanley และ Wang (1970) แสดงว่าการให้น้ำหนักคะแนนรายข้อที่แตกต่างกันจะเพิ่มค่าความเที่ยงและความตรงเพียงเล็กน้อย ซึ่งผลที่ได้ไม่เพียงพอที่จะคุ้มกับการลงทุนที่จะกำหนดน้ำหนักรายข้อหรือในแต่ละแบบสอบย่อยให้แตกต่างกันเมื่อรวมคะแนนทั้งฉบับ แต่อย่างไรก็ตาม Stanley และ Wang ได้เสนอแนะว่าแม้การกำหนดน้ำหนักต่อข้อกระทง (item weighting) จะให้ผลในการเพิ่มความเที่ยงและความตรงอย่างไม่มีนัยสำคัญต่อการวัดเชิงประกอบ การให้น้ำหนักที่แตกต่างกันต่อการตอบภายในข้อกระทง (reponse within items) ก็ควรเป็นสิ่งที่ได้รับการศึกษามากขึ้น เพราะเขาเชื่อว่าการกำหนดน้ำหนักให้กับตัวเลือก (option weighting) ในการตอบอาจจะทำให้ประสิทธิภาพแบบสอบเพิ่มขึ้นเพราะน้ำหนักแต่ละตัวเลือกเป็นการพิจารณาความแปรปรวนจากการตอบที่ไม่ถูกต้องด้วยซึ่งเป็นการพิจารณาความรู้บางส่วนของผู้ตอบจากงานวิจัยของ Waters (1980) ได้เปรียบเทียบวิธีการให้คะแนนโดยการกำหนดน้ำหนักให้กับตัวเลือกด้วยค่าสหสัมพันธ์แบบ point biserial (r_{pb}) ระหว่างตัวเลือกกับคะแนนสอบทั้งหมด และ การให้น้ำหนักตัวเลือกด้วยสัมประสิทธิ์ของ Brogden (r_{B}) เปรียบกับการให้คะแนนแบบเดิมคือการรวมข้อถูก ผลการศึกษาให้ข้อสนับสนุนว่าการให้คะแนนด้วยน้ำหนักตามตัวเลือกที่เขาเลือกในแต่ละข้อกระทงเป็นสิ่งที่ช่วยขยายสารสนเทศของข้อมูลเพื่อนำไปใช้เป็นพื้นฐานในการตัดสินใจได้ชัดเจนขึ้น ซึ่งข้อมูลบางส่วนได้สูญเสียไปโดยผ่านกระบวนการให้คะแนนแบบเดิม

1.3 การศึกษาเกี่ยวกับการแก้ปัญหาการเดาของการตอบแบบสอบเลือกตอบ

แหล่งสำคัญของความคลาดเคลื่อนของการวัดด้วยแบบสอบแบบเลือกตอบ ก็คือ การเดา "การเดา" เป็นเทอมที่ชี้ถึงจำนวน (number) ของพฤติกรรมที่ปรากฏเมื่อผู้สอบคนหนึ่งตอบข้อสอบข้อหนึ่งด้วยความมั่นใจน้อยกว่าความมั่นใจอย่างสมบรณ์ (Abu-Sayf, 1979) วิธีการแก้ปัญหาการเดาที่ใช้กันอยู่ ก็คือการใช้สูตรการให้คะแนนเพื่อแก้การเดาของผู้ตอบและสูตรการให้คะแนนที่ นิยมใช้มี 2 สูตรคือ

สูตรแรกนี้เป็นสูตรที่พยายามที่จะประเมินระดับจริงของความรู้ของผู้ตอบแบบสอบเลือกตอบ โดยการกำจัดคะแนนการตอบถูกของผู้ตอบที่เป็นผลมาจากการเดาสุ่มออก สูตรการให้คะแนนนี้คือ

$$S = R - W/(k-1)$$

เมื่อ S คือ คะแนนที่ประมาณความสามารถจริงของผู้ตอบ

R คือ จำนวนคำตอบผิด

k คือ จำนวนตัวเลือกในแต่ละข้อ

มีงานวิจัยจำนวนมากที่ใช้สูตรการให้คะแนนนี้เมื่อพูดถึงสูตรการให้คะแนนผู้วิจัยส่วนใหญ่ก็จะนึกถึงสูตรการให้คะแนนสูตรนี้ งานวิจัยจำนวนมาก แสดงถึงความตรงของคะแนนที่ได้จากคะแนนสูตรนี้และพบว่าคะแนนสูตรนี้ช่วยให้คุณสมบัติของแบบสอบดีขึ้น (Jaradat and Tollefson, 1988) และจากการศึกษาของ Lyerly (1951) พบว่าสูตรการให้คะแนนได้ให้ค่าประมาณที่ใกล้เคียงคะแนนจริงของผู้สอบแบบสอบ เมื่อการใช้สูตรการให้คะแนนเป็นไปตามข้อตกลงเบื้องต้นที่กำหนดในกรณีที่ผู้ตอบไม่รู้คำตอบและมีการเดาสุ่ม สูตรการให้คะแนนที่เป็นที่นิยมใช้อีกสูตรหนึ่งไม่ใช่สูตรแก้การเดาแต่พยายามจะลดการเดาคำตอบของผู้สอบลงด้วยการให้รางวัลผู้ตอบที่ไม่ยอมเดาสุ่ม แต่จะละข้อนั้นไว้เมื่อเขาไม่มีสารสนเทศเพียงพอที่จะตอบหรือมีความมั่นใจไม่พอว่าจะเลือกคำตอบได้ถูกต้อง สูตรการให้คะแนนนี้คือ

$$S = R + u/k$$

เมื่อ u คือ จำนวนข้อที่ผู้ตอบละไม่ตอบ

คะแนนสูตรที่ให้รางวัลการละคำตอบ (omission reward) แก่ผู้ตอบที่ไม่เดาคำตอบนี้ได้รับการ

สนับสนุนถึงผลกระทบทางจิตวิทยา ในการที่จะลดพฤติกรรมการเดาสุ่มลงด้วยการให้รางวัลแทนที่จะเป็นการลงโทษดังสูตรแรก (Traub, Hambleton, and Singh, 1969 quoted in Abu-Sayf, 1979) สูตรการให้คะแนนนี้จะให้ค่าประมาณที่ตรงกับคะแนนจริงเมื่อผู้ตอบ ตอบเมื่อเขาแน่ใจว่าเขามีสารสนเทศเกี่ยวกับคำตอบนั้นพอเพียง และละคำตอบเมื่อเขาไม่มีความรู้ที่จะตอบ เมื่อตัวเลือกทุกตัวมีความดึงดูดเท่ากัน ซึ่งข้อตกลงของสูตรการให้คะแนนนี้เหมือนกับข้อตกลงของสูตรการให้คะแนนแบบเดิมที่ใช้กันอยู่ คือการให้คะแนน 0 และ 1 ซึ่ง Lord (1975) ได้พิสูจน์ว่าสูตรการละคำตอบนี้มีความสัมพันธ์กับสูตรการให้คะแนนแบบเดิมอย่างสมบูรณ์ Gulliksen (1950) ได้แนะนำสูตรการให้คะแนนที่รวมเอาสูตรการให้คะแนนข้างต้นมาไว้ด้วยกันคือ

$$S = R - W/c + u/d$$

เมื่อ c และ d เป็นค่าคงที่ที่กำหนดขึ้นจากผลการสอบโดยให้ค่า c น้อยกว่า d แต่อย่างไรก็ตามสูตรนี้พัฒนาขึ้นอย่างไม่มีหลักฐานเชิงประจักษ์และแนวทางทางทฤษฎีสนับสนุน และไม่เป็นที่นิยมแพร่หลายนัก (Abu-Sayf, 1979)

สูตรการให้คะแนนทั้งหมดที่กล่าวมาพัฒนาขึ้นบนพื้นฐานของการแก้การเดาสุ่มในแบบสอบความเร็ว (speed test) (Gulliksen, 1950) ซึ่งได้มีการพัฒนาเพิ่มขึ้นจากงานของ Abu-Sayf (1979) เขาได้พัฒนาขึ้นมาจากพื้นฐานที่ผู้ตอบที่มีความสามารถเท่ากันแต่ผู้สอบคนหนึ่งตอบช้าอีกคนหนึ่งตอบเร็วทำให้คนที่ตอบข้อสอบช้าทำข้อสอบไม่ทันและมีข้อที่ละไม่ตอบ จึงไม่ได้คะแนนทั้งที่ความจริงเขาควรได้คะแนนเท่ากัน สูตรที่เขาพัฒนาขึ้นเป็นดังนี้

$$S = R - W/(k-1) + aU$$

โดยที่ a คือ คะแนนเฉลี่ยรายข้อสำหรับผู้สอบที่ตอบข้อสอบทั้งหมดซึ่งเป็นค่าแก้สำหรับผู้ที่ละไม่ตอบ ส่วนผู้สอบที่ตอบทุกข้อเทอม aU ก็จะหายไปคงเหลือแต่สูตรการให้คะแนนเพื่อแก้การเดาที่ใช้กันอยู่เดิม

จะพบว่าสูตรการให้คะแนนแบบสอบเลือกตอบที่พัฒนาขึ้นไม่มีการพิจารณาความรู้บางส่วนของผู้ตอบและจะวางอยู่บนพื้นฐานของแบบสอบความเร็ว ซึ่งผู้สอบจะเดาสุ่มเมื่อไม่มีเวลาพิจารณาข้อสอบแต่ละข้ออย่างพอเพียง ดังนั้นการนำสูตรการให้คะแนนดังกล่าวมาใช้กับแบบสอบเลือกตอบที่เป็น power test โดยที่ผู้ตอบมีเวลา พิจารณาคำตอบอย่างพอเพียงจึงไม่เหมาะสม อีกประการหนึ่ง การใช้สูตรการให้คะแนนจำเป็นที่จะต้องชี้แจงแก่ผู้สอบอย่างเหมาะสมเพื่อที่เขาจะได้สามารถ

ใช้ยุทธวิธีที่เหมาะสมในการที่จะให้ได้คะแนนมากที่สุดในผู้สอบแต่ละคน การให้คำชี้แจงโดยห้ามผู้ตอบไม่ให้เดาจะให้ผลกับผู้ตอบแต่ละคนไม่เหมือนกันสำหรับผู้ที่ยอมเสี่ยงที่จะเดามีแนวโน้มว่าจะได้คะแนนมากกว่าผู้ที่เชื่อตามคำชี้แจงและเป็นการลงโทษผู้ที่ละคำตอบ (Slakter, 1968) และจากการศึกษาวรรณคดีที่เกี่ยวกับการแก้การเดาของผู้ตอบแบบสอบเลือกตอบ Diamond และ Evans (1973) พบว่าการแก้ปัญหาการเดาของผู้ตอบแบบสอบเลือกตอบด้วยการชี้แจงห้ามไม่ให้เดาหรือให้เดาด้วยการให้สูตรการให้คะแนนเพื่อแก้การเดาหรือด้วยการให้รางวัลผู้ละคำตอบจะไม่ทำให้ค่าความเที่ยงของแบบสอบเพิ่มขึ้นอย่างน่าพอใจนัก และ โดยเฉพาะการไม่เพิ่มในค่าความตรงของแบบสอบอย่างพอใจ

สูตรการให้คะแนนที่กล่าวมาแต่ต้น ไม่ได้มีการพิจารณาถึงความรู้บางส่วนของผู้ตอบในกรณีที่เขาสามารถตัดตัวลวงบางตัวออกไปได้อย่างถูกต้อง การให้คะแนนวางอยู่บนพื้นฐานของการเดาสุ่มเท่านั้น Hamdan (1979) ได้ศึกษาถึงวิธีการให้คะแนนด้วยสูตรการให้คะแนนเพื่อประมาณความสามารถจริงของผู้ตอบแบบสอบเลือกตอบ โดยใช้ค่าแก้การเดา 2 ตัว คือคะแนนที่ปราศจากการเดาสุ่มกับคะแนนที่ปราศจากการเดาอย่างมีการศึกษา โดยเขาได้ใช้ทฤษฎีความน่าจะเป็น (probability theory) เพื่อเลือกตัวแก้ที่เหมาะสมของการเดาโดยผ่านการศึกษาการแจกแจงของพารามิเตอร์เพื่อที่จะนำไปสู่การแก้การเดาอย่างสุ่มและการเดาอย่างมีการศึกษาที่แตกต่างกัน ด้วยค่าประมาณ maximum likelihood ของพารามิเตอร์ที่เกี่ยวข้อง จะให้ตัวประมาณค่าการแก้การเดาที่สอดคล้องกับการตอบของผู้สอบ สมหวัง พิธิยานุวัฒน์ และ เพ็ญศิริ ด่านชนะ (2524) ได้นำแนวคิดวิธีการให้คะแนนของแฮมดานมาใช้เป็นกรอบในการวิจัยเพื่อศึกษาถึงรูปแบบในการเดาของผู้ตอบแบบสอบเลือกตอบซึ่งเป็นนักศึกษาปริญญาโทที่เรียนวิชาสถิติและมีพื้นความรู้ทางคณิตศาสตร์แตกต่างกัน ได้พบหลักฐานว่ากลุ่มตัวอย่างที่มีพื้นฐานความรู้ทางคณิตศาสตร์และเรียนวิจัยโดยตรง จะใช้รูปแบบการเดาสุ่มและการเดาแบบผสมน้อยกว่าการเดาอย่างมีความรู้ ส่วนกลุ่มตัวอย่างที่ไม่มีพื้นฐานความรู้ทางคณิตศาสตร์จะใช้วิธีการเดาแบบสุ่มมากกว่าการเดาแบบผสมและการเดาอย่างมีความรู้ งานวิจัยของ สมหวัง และ คณะให้ข้อสังเกตและยืนยันหลักฐานที่ว่า ในการตอบแบบสอบแบบเลือกตอบนั้นผู้ตอบไม่ได้ใช้การเดาสุ่มเพียงอย่างเดียวผู้ตอบยังใช้ความรู้บางส่วนในการตอบคำถามด้วย ดังนั้นสูตรการให้คะแนนที่เหมาะสมจึงควรพิจารณาความรู้บางส่วนในการตอบของผู้ตอบ ซึ่งเมื่อพิจารณาสูตรการให้คะแนนของแฮมดานสามารถหาค่าตัวแก้การเดาแบบสุ่มและการเดาอย่างมีการศึกษาได้ขึ้นอยู่กับลักษณะการตอบของผู้ตอบว่าจะสามารถมีความรู้บางส่วนในการตัดตัวลวงออกได้เท่าใด เมื่อนำปรับให้เข้ากับแนวทางการให้คะแนนเพื่อพิจารณาความรู้บางส่วนของผู้ตอบ สูตรการให้คะแนนแบบแฮมดานจึงควรได้รับความสนใจที่ได้ทำการศึกษาต่อไป

จากการศึกษางานวิจัยที่เกี่ยวข้องกับวิธีการตอบและการให้คะแนนแบบสอบเลือกตอบที่กล่าวมามีความพยายามที่จะแก้ปัญหาการวัดที่เกิดจากความคลาดเคลื่อนในการเดาและการได้มาซึ่ง

สารสนเทศเกี่ยวกับความสามารถของผู้ตอบได้ตรงที่สุด ได้พบแนวทางที่จะแก้ปัญหาการวัดจากแบบสอบเลือกตอบ ดังนี้

ก. การได้มาซึ่งปริมาณสารสนเทศสูงสุดในการวัด ด้วยการให้ผู้ตอบเลือกคำตอบถูกเพียงคำตอบเดียวในแต่ละข้อ และให้คะแนนเฉพาะข้อที่ผู้ตอบ ตอบได้ถูกต้องทำให้ขาดสารสนเทศเกี่ยวกับความรู้บางส่วนของผู้ตอบ ดังนั้นในการศึกษาเพื่อที่จะให้ทราบถึงระดับความรู้ของผู้ตอบระดับต่างๆจึงต้องเปลี่ยนจากการให้ผู้ตอบตอบคำตอบเดียว (single response) เป็นการตอบหลายอย่างมากขึ้น (multiple response)

ข. ประเด็นการใช้สูตรการให้คะแนนเพื่อแก้การเดาจะใช้ได้ผลหรือไม่ขึ้นอยู่กับว่าพฤติกรรมของการตอบแบบสอบ (test-taking) ต้องแสดงถึงความสามารถที่แท้จริงของผู้ตอบ ดังนั้นการตอบทุกข้อด้วยวิธีการแบบเดิมทำให้เกิดข้อโต้แย้ง 2 ประการ ประการแรก ในเชิงเทคนิค แหล่งความคลาดเคลื่อนได้ถูกกระทำให้เพิ่มขึ้นจากการเดาสุ่ม ประการที่สอง ในเชิงการศึกษา ซึ่งควรจะยึดหลักการที่ว่าการกระตุ้นให้นักเรียนเดาไม่ใช่สิ่งที่ต้องการในการให้การศึกษา แต่กระบวนการให้คำชี้แจงในการสอบควรกระตุ้นให้นักเรียนได้ใช้ความคิดในการแก้ปัญหามากกว่า (Hutchinsun, 1982) และให้นักเรียนได้แสดงการตอบตามความสามารถที่แท้จริงของเขา

ค. การใช้สูตรการให้คะแนนที่พัฒนาขึ้น ส่วนมากเป็นสูตรการให้คะแนน สำหรับแบบสอบความเร็วโดยมีข้อตกลงเบื้องต้นที่เกี่ยวกับการที่ผู้ตอบไม่รู้คำตอบจะมีการเดาแบบสุ่มในระหว่างตัวเลือกที่ให้มาโดยให้ตัวเลือกทุกตัวมีโอกาสในการได้รับเลือกเท่ากัน ส่วนองค์ประกอบที่เกี่ยวกับความรู้บางส่วนของผู้ตอบ องค์ประกอบที่เกี่ยวกับความแตกต่างในระดับของความเต็มใจที่จะเดาคำตอบของผู้ตอบ หรือการละคำตอบ องค์ประกอบที่เกี่ยวกับการขาดความเป็นอิสระระหว่างตัวเลือกภายในข้อ องค์ประกอบเหล่านี้ทำให้ความตรงของโมเดลทางคณิตศาสตร์ที่เอามาใช้ในการสร้างสูตรการให้คะแนนลดลง แต่อย่างไรก็ตามตราบใดที่ผลขององค์ประกอบเหล่านี้ยังไม่สามารถที่จะสังเกตได้จากคะแนนดิบที่สังเกตมาได้ จึงเป็นความหวังที่จะแสวงหาสูตรการให้คะแนนที่เหมาะสมต่อไป (Lyerly, 1951) ซึ่งเป็นประเด็นปัญหาสำหรับการวิจัยครั้งนี้

2. <u>วิธีการให้คะแนนของอาร์โนลด์และแฮมดาน</u>

ในตอนต่อไปนี้จะได้กล่าวถึงวิธีการให้คะแนนของอาร์โนลด์และวิธีการให้คะแนนของแฮมดาน ซึ่งทั้งสองวิธีเป็นวิธีที่ผู้วิจัยได้นำมาประยุกต์เพื่อเสนอเป็นสูตรการให้คะแนนที่ใช้ในการวิจัยนี้ ดังนี้

2.1 วิธีการให้คะแนนแบบอาร์โนลด์

Arnold and Arnold (1970) ได้เสนอวิธีการในแบบสอบเลือกตอบซึ่งยอมให้คะแนนเกี่ยวกับความรู้บางส่วนและควบคุมคะแนนที่เพิ่มขึ้นด้วยโอกาสของการเดาเป็นศูนย์ กระบวนการในการพัฒนาการให้คะแนนใช้ทฤษฎีเกมเบื้องต้น (game theory) ในการให้คะแนนโดยสมมติคะแนนของผู้ตอบสำหรับผลที่เป็นไปได้ ดังนี้ สำหรับการให้คำชี้แจงให้ผู้เข้าสอบตัดตัวเลือกที่เป็นคำตอบที่ไม่ถูกเท่าที่จะเป็นไปได้ โดยสมมติคะแนนที่เป็นผลลัพธ์จากการตอบของผู้เข้าสอบ ดังนี้

คะแนน C_0 ถ้าผู้เข้าสอบไม่สามารถตัดตัวลวงใด ๆ ได้เลย

คะแนน C_1 ถ้าผู้เข้าสอบตัดตัวลวงออกได้ 1 ตัว

คะแนน C_2 ถ้าผู้เข้าสอบตัดตัวลวงออกได้ 3 ตัว

คะแนน $-p$ ถ้าผู้สอบตัดตัวถูกออกเพราะเข้าใจว่าเป็นตัวลวง

ในกรณีที่ถ้าข้อกระทงมีตัวเลือก k ตัว จะมีจำนวนตัวลวงทั้งหมด k-1 ตัวเมื่อมีคำตอบที่ถูกต้องเพียงคำตอบเดียว และเมื่อ d แทนจำนวนตัวลวงที่ถูกตัดออไปได้โดยผู้เข้าสอบค่าของ d จะมีค่าตั้งแต่ 0,1,2,...,k-1 ในกรณีที่ผู้สอบไม่สามารถตัดตัวลวง k-1 ตัวใด ๆ ออกได้แล้ว เขาก็มีทางเลือกที่จะตัดตัวเลือกออกอย่างสุ่ม d ครั้ง สำหรับคะแนนที่เป็นไปได้ C ในขณะเดียวกันผู้สอบที่ต้องเสี่ยงที่จะเลือกตัวเลือกที่ถูกออกไปเป็นตัวลวงด้วยการลงโทษ -p

ด้วยคะแนนที่ได้รับการคาดหวังเกี่ยวกับคำตอบของผู้เข้าสอบจะถูกกำหนดโดย

$$\left(\begin{array}{c}\text{คะแนนที่ได้ถ้าชนะเกม}\\\text{คือคำตอบถูก}\end{array}\right)\cdot\left(\begin{array}{c}\text{โอกาสของการ}\\\text{ชนะเกม}\end{array}\right) - \left(\begin{array}{c}\text{คะแนนที่รับ}\\\text{การลงโทษ}\end{array}\right)\cdot\left(\begin{array}{c}\text{โอกาสในการแพ้เกม}\\\text{เมื่อแพ้เกม}\end{array}\right)$$

จากค่าที่ถูกกำหนดแต่ต้นคะแนนที่ถูกคาดหวังของผู้เข้าสอบที่ตัดตัวลวง d ตัวออกไปได้จะเป็น

$$(C_d)[(k-1)/d] - (p)(d/k) \quad \text{.......................(1)}$$

แต่คะแนนที่ได้เพิ่มจากการเดาสุ่มในกรณีที่ (1) นี้เป็นไปตามข้อตกลงเบื้องต้นที่ศึกษานี้แล้วว่าจะให้เป็นศูนย์ ซึ่งเป็นคะแนนที่ไม่สามารถตัดตัวลวงใดๆออกไปได้เลยและปฏิเสธการเดาของผู้สอบเมื่อไม่สามารถตัดตัวลวงใดๆออกได้โดยให้คะแนนที่เพิ่มขึ้นจากการเดาสุ่มเป็นศูนย์ ดังนั้นค่าคะแนนเฉลี่ยรายข้อที่คาดหวังใน (1) จะเป็นศูนย์ ดังนั้นเมื่อ (1) มีค่าเท่ากับศูนย์ จะได้

$$C_d = (p)(d/(k-d)) \quad \text{........................(2)}$$

C_d คือคะแนนที่ยุติธรรมในกรณีที่คะแนนที่เพิ่มขึ้นจากการเดาเป็นศูนย์และมีการให้คะแนนเมื่อผู้สอบมีความรู้บางส่วนที่สามารถตัดตัวลวงบางตัวออกไปได้ และคำถามที่มีตัวเลือก k ตัวมี k-1 ตัวลวง และ -p คะแนน เป็นการลงโทษในกรณีที่ผู้ตอบจัดตัวเลือกที่ถูกเข้าไว้เป็นตัวลวง

ถ้าเราพิจารณากรณีพิเศษอื่นๆซึ่งคะแนนความรู้บางส่วนไม่ได้รับการพิจารณาแล้วการเลือกเพียงคำตอบที่ดีที่สุดก็เท่ากับการตัดตัวลวงออกไป k-1 ตัว นั่นคือ d = k-1 คะแนน C_d ใน (2) จะเป็น

$$C_{k-1} = p(k-1)$$

เมื่อเป็นการตัดคำตอบที่ถูกต้องออกไปทั้งหมดคะแนนแบบเดิมประเพณีนิยมที่กำหนดคือ $C_{(k-1)}$ = 1 นั่นคือ p=1/(k-1) นั่นเอง ข้อพิจารณาดังกล่าวก็คือการแก้การเดาโดยการใช้สูตรการให้คะแนนแบบเดิม คือ S = R - W/(k-1) นั่นเอง สูตรการให้คะแนนดังกล่าวก็คือสูตรการแก้การเดาสุ่มแบบดั้งเดิมที่กล่าวมาแล้วแต่ต้นที่ไม่ได้มีการพิจารณาความรู้บางส่วนของผู้ตอบ แต่เป็นการพิจารณาเมื่อผู้ตอบมีการเดาแบบสุ่มนั่นเอง

สำหรับวิธีการให้คะแนนแบบยุติธรรมในกรณีของอาร์โนลด์นี้ ในกรณีของ 4 ตัวเลือก การลงโทษอย่างยุติธรรมเมื่อตัวเลือกที่เป็นคำตอบถูกกาออกไปว่าเป็นตัวลวงคือ -1/3 การให้คะแนนอย่างยุติธรรมในกรณีที่ผู้สอบไม่สามารถตัดตัวลวงใดๆออกได้เลย (C_0) จะเป็น 0 คะแนน และในกรณีที่กาตัวลวงออกได้ถูกต้อง 1 ตัว (C_1) คะแนนที่ได้จะเป็น 1/9 กรณีที่กาตัวลวงออกได้ถูกต้อง 2 ตัว (C_2) คะแนนจะเป็น 1/3 และในกรณีที่กาตัวลวงออกถูกทั้งหมด 3 ตัว คะแนนที่ได้รับ (C_3) จะเป็น 1 คะแนน

การศึกษาคุณภาพของสูตรการให้คะแนนของอาร์โนลด์

Arnold and Arnold (1970) ได้เปรียบเทียบสูตรการให้คะแนนที่สร้างขึ้นโดยการนำไปใช้ให้คะแนนในการสอบวิชาสถิติเบื้องต้นกับนักเรียนจำนวน 25 คน และเพื่อให้มีการเปรียบเทียบกับวิธีการให้คะแนนอื่นๆ เขาได้กำหนดคำชี้แจงในการสอบ ดังนี้

'คำถามแต่ละข้อในการสอบนี้มี 4 ตัวเลือก โดยมีคำตอบที่ถูกต้องเพียงคำตอบเดียว ให้ท่านกาตัวเลือกที่เป็นตัวลวงมากเท่าที่ท่านจะสามารถกาออกได้ ถ้าท่านกาตัวเลือกที่เป็นตัวลวงออกได้มากเท่าใด ท่านจะได้คะแนนมากเท่านั้น แต่อย่างไรก็ตามท่านจะถูกลงโทษในกรณีที่ท่านกาตัวเลือกที่เป็นคำตอบออกไปด้วย คะแนนที่ได้และการลงโทษในแต่ละข้อจะถูกคำนวณเพื่อว่าท่านจะได้คะแนนเฉลี่ยรายข้อ เป็น 0 เมื่อท่านเดาคำตอบในแต่ละข้อให้ท่านวงกลมรอบตัวเลือกที่ท่านรู้สึกว่าน่าจะเป็นคำตอบมากที่สุด ในกรณีที่ท่านสามารถกาตัวลวงออกทุกตัว ท่านต้องวงคำตอบที่เหลือด้วย "

คำชี้แจงตอนสุดท้ายที่ให้วงคำตอบในแต่ละข้อเป็นสิ่งที่ต้องการ เพื่อที่จะสามารถนำมาให้คะแนนเพื่อเปรียบเทียบกับวิธีการให้คะแนนวิธีอื่นๆได้ วิธีการให้คะแนนเพื่อนำมาเปรียบเทียบกันได้ ได้จากการตรวจข้อที่ผู้ตอบเลือกวงเพื่อแสดงว่าตัวเลือกที่วงนั้นเป็นคำตอบ ซึ่งเลือกวงได้เพียงตัวเลือกเดียว

ในแต่ละข้อและคำตอบนี้สามารถให้คะแนนแบบที่เคยให้ได้ แบบสอบที่ใช้ประกอบด้วยคำถาม 9 ข้อ แต่ละคำถามมีตัวเลือกที่ถูกต้องเพียง 1 ข้อ และตัวลวง 3 ตัว โดยให้คะแนนแต่ละข้อ สูงสุด 9 คะแนน สำหรับคำถาม 8 ข้อ แรก และคะแนน สูงสุด 12 คะแนน ในข้อที่เหลือ ในข้อที่ให้คะแนนสูงสุด 9 คะแนน คะแนนรายข้อจะประกอบด้วย คะแนนลงโทษผู้รู้ผิด -3 คะแนน คะแนนที่ไม่สามารถตัดตัวลวงใดๆออกได้เลย 0 คะแนน คะแนนที่ตัดตัวลวงออกได้ 1ตัว 1 คะแนน คะแนนที่ตัดตัวลวงออกได้ 2 ตัว 3 คะแนน และ คะแนนที่ตัดตัวลวงออกได้ 3 ตัว 9 คะแนน ในกรณีที่คะแนนสูงสุดรายข้อคือ 12 คะแนน คะแนนลงโทษผู้รู้ผิดจะเป็น -4 คะแนน คะแนนที่ไม่สามารถตัดตัวลวงใดๆออกได้เลย 0 คะแนน คะแนนที่สามารถตัดตัวลวงออกได้ 1 ตัว 4/3 คะแนน คะแนนที่ตัดตัวลวงออกได้ 2 ตัว 4 คะแนน และคะแนนที่ตัดตัวลวงออกได้ 3 ตัว 12 คะแนน วิธีการให้คะแนนที่นำมาเปรียบเทียบกับคะแนนความรู้บางส่วนของอาร์โนลด์นี้ ประกอบด้วยวิธีการให้คะแนน 4 วิธี

วิธีที่ 1 ให้คะแนนเต็มในแต่ละข้อสำหรับข้อที่วงได้ถูกต้อง โดยไม่มีการลงโทษเมื่อเดาคำตอบ

วิธีที่ 2 ให้คะแนนเต็มในแต่ละข้อที่วงคำตอบได้ถูกต้องแต่จะลงโทษ -3 คะแนนในข้อที่คะแนนเต็ม 9 คะแนน และ -4 คะแนนในข้อที่คะแนนเต็ม 12 คะแนน สำหรับแต่ละข้อที่คำตอบถูกวง ซึ่งวิธีนี้เป็นวิธีการให้คะแนนด้วยสูตรการแก้การเดาแบบเดิมที่กำหนดให้ $S=R-W/(k-1)$ เมื่อ R คือจำนวนคำตอบที่ถูก W คือ คำตอบที่ผิด และ k คือจำนวนตัวเลือกในแต่ละคำถาม

วิธีที่ 3 กำหนดให้ $S=\lambda\{(kN_c-N)/n_c(k-1)\}$ เมื่อ $\lambda=9$ หรือ 12 ในแต่ละคำถาม $k=4$ และ N_c คือจำนวนคำตอบที่ถูกต้องจากผู้ตอบในชั้นทั้งหมด และ N คือจำนวนผู้สอบที่พยายามตอบคำถามข้อนั้น วิธีนี้เป็นวิธีการให้คะแนนวิธีหนึ่งซึ่งทำให้ mean square error ลดลง ในกรณีที่คะแนนจริงของผู้สอบเป็น 1 ผู้สอบจะรู้คำตอบที่ถูกต้องและมีคะแนนจริงเป็น 0 ถ้าผู้สอบไม่รู้คำตอบที่ถูกต้อง

วิธีที่ 4 เหมือนกับวิธีที่ 3 ยกเว้นแต่ว่าจะจัดกลุ่มผู้สอบตามพื้นฐานของคะแนนดิบเป็น 3 กลุ่มคือกลุ่มที่ได้คะแนนสูงมี 25 เปอร์เซนต์ กลุ่มกลางมี 50 เปอร์เซนต์ และกลุ่มต่ำมี 25 เปอร์เซนต์ คะแนนในแต่ละตำแหน่งของผู้สอบจะคำนวณแยกในแต่ละกลุ่ม

จากคำตอบของผู้ตอบ 25 คน ได้นำมาให้คะแนนความรู้บางส่วนตามวิธีของอาร์โนลด์และวิธีที่นำมาเปรียบเทียบอีก 4 วิธี จัดอันดับผู้ตอบจากคะแนนสูงไปต่ำในวิธีการให้คะแนนแต่ละวิธี ตำแหน่งที่ซ้ำกันจะนำมาเฉลี่ยกัน เปรียบเทียบตำแหน่งตามอันดับของการให้คะแนนแต่ละวิธี พบว่าผู้ตอบที่ได้คะแนนสูงในวิธีการให้คะแนนวิธีหนึ่งจะได้คะแนนสูงในอีกวิธีหนึ่ง เช่นเดียวกับผู้ตอบที่ได้คะแนนต่ำในวิธีการให้คะแนนวิธีหนึ่งจะได้คะแนนต่ำในอีกวิธีหนึ่งด้วย แต่พบว่าคะแนนความรู้บางส่วนที่ให้ด้วยวิธีของอาร์โนลด์จะจำแนกตำแหน่งตอนกลางซึ่งมักมีตำแหน่งซ้ำๆกันในการให้คะแนนวิธีอื่นๆ ออกได้ชัดเจนกว่า ซึ่งอาร์โนลด์ได้ให้เหตุผลไว้ 2 ประการคือเนื่องจากองค์ประกอบของการเดาสุ่มที่ถูกกำจัดออกไปด้วย

จึงเป็นสาเหตุให้มีการกระจายสูงในช่วงคะแนนตอนนี้ ประการที่สองเป็นเพราะคะแนนมีแนวโน้มจะรวมอยู่ที่ค่าเฉลี่ยระหว่างตำแหน่งในตอนกลางโดยทั่วไปน้อยกว่าตำแหน่งในตอนปลาย จึงเป็นสาเหตุทำให้มีการกระจายสูงในตำแหน่งของคะแนนตอนกลาง อาร์โนลด์ให้คำแนะนำต่อสูตรการให้คะแนนของเขาว่า เหมาะกับการให้คะแนนความรู้บางส่วนในการตอบแบบสอบเลือกตอบที่คะแนนความรู้บางส่วนได้รับการนิยามว่าเป็นจำนวนตัวลวงในข้อสอบแต่ละข้อที่ผู้ตอบสามารถตัดออกได้แม้ว่าจะไม่ทราบคำตอบ คะแนนรายข้อเฉลี่ยที่ได้จะมีค่าเป็น 0 ในกรณีที่ผู้ตอบเดาคำตอบ ในกรณีที่ผู้ใช้สูตรการให้คะแนนต้องการให้ผู้ตอบท้อถอยต่อการเดาคำตอบผู้ใช้สามารถปรับการขยายการลงโทษผู้รู้ผิดได้ตามความเหมาะสม และคำนวณค่าคะแนนความรู้บางส่วนได้จากสูตรการให้คะแนนนั้น แต่ถ้ากระบวนการให้คะแนนไม่ยอมให้มีการพิจารณาให้คะแนนความรู้บางส่วน สูตรการให้คะแนนแบบอาร์โนลด์จะเหมือนกับสูตรการแก้การเดาแบบเดิม (CG) คือ S=R-W/(k-1) วิธีการให้คะแนนความรู้บางส่วนนี้มีข้อได้เปรียบในการปฏิบัติคือเป็นวิธีการที่ง่ายในการให้คะแนนและคำชี้แจงง่ายที่จะเข้าใจ สังเกตได้ว่าวิธีนี้ช่วยให้ได้ข้อมูลเพิ่มขึ้นในการสอบด้วยแบบสอบเลือกตอบ และพบว่าวิธีนี้ผู้ตอบข้อสอบรู้สึกพอใจในกระบวนการให้คะแนนนี้

Frary (1980) ได้ศึกษาวิเคราะห์ผลของความรู้ผิด (misinformation) ความรู้บางส่วน (Partial information) และการเดา ที่มีต่อคะแนนเฉลี่ยรายข้อในแบบสอบเลือกตอบด้วยการวิเคราะห์ถึงคะแนนเฉลี่ยรายข้อ (expected item score) จากวิธีการให้คะแนน 6 วิธี ซึ่งประกอบด้วยวิธีการให้คะแนนแบบประเพณีนิยม (NR) วิธีการแก้การเดาแบบเดิม (CG) การให้คะแนนแบบอาร์โนลด์ (A) การตอบจนกว่าจะถูก (AUC) การให้คะแนนแบบคูมบ์ (CBS) และวิธีการที่ Frary เสนอใหม่ (FNM) จากการวิเคราะห์คะแนนเฉลี่ยรายข้อของอาร์โนลด์ ซึ่งให้คะแนนแก่จำนวนตัวลวงที่ผู้ตอบสามารถตัดออกได้อย่างถูกต้อง และควบคุมในกรณีที่ผู้ตอบรู้ผิดโดยกาคำตอบว่าเป็นตัวลวงด้วยการลดคะแนนลง -1/(k-1) คะแนน ซึ่งเป็นการให้คะแนนและการลงโทษที่เทียบเท่ากับ NR และ CG แต่สูตรการให้คะแนนแบบอาร์โนลด์จะได้เปรียบในการที่คะแนนรายข้อที่กำหนดจะเท่ากับคะแนนเฉลี่ยรายข้อในกรณีที่ผู้ตอบไม่เดาสุ่มอย่างสมบูรณ์เกิดขึ้น ซึ่งจะช่วยลดความแปรปรวนของคะแนนรวมลง และช่วยส่งเสริมให้ความเที่ยงของการวัดสูงขึ้น แต่ถ้ามีการเดาเกิดขึ้น การใช้สูตรการให้คะแนนของอาร์โนลด์จะไม่มีการลงโทษผู้สอบเพราะคะแนนที่เกิดจากโอกาสในการเดาสุ่มจะเป็นศูนย์ ทำให้คะแนนเฉลี่ยรายข้อเท่ากับคะแนนเดิมที่ผู้ตอบมีความรู้บางส่วนในการตัดตัวลวงบางตัวออกได้จริงก่อนที่จะใช้การเดาแบบสุ่ม เมื่อเปรียบเทียบคะแนนแบบ A NR และ CG จะพบว่ามีการลงโทษผู้รู้ผิดในรูปแบบเดียวกันและมีค่าคะแนนลงโทษเพียงเล็กน้อย นอกจากนี้คะแนนทั้ง 3 วิธีนี้ จะให้คะแนนเฉลี่ยรายข้อเช่นเดียวกัน (ให้คะแนนผลลัพธ์ที่สามารถแปลงไปสู่เสกลเดียวกัน) และการเดาไม่มีผลต่อคะแนนเฉลี่ยรายข้อ แต่การใช้วิธีการให้คะแนนทั้ง 3 นี้จะแตกต่างกันในปริมาณของการกระจายของคะแนนรวมซึ่งเกี่ยวกับโอกาสที่เป็นผลมาจากการเดา NR ไม่เหมาะสมที่สุดในกรณีนี้ ส่วน CG ผู้ตอบจะได้รับการกระตุ้นให้รายงานตำแหน่งความรู้ที่แสดงออก (actaul state of knowledge) ในกรณีที่อ้างอิงว่าเมื่อตอบผิดเขา

จะไม่มีความรู้โดยสิ้นเชิงซึ่งทำให้แหล่งความแปรปรวนในการเดาสุ่มลดลงในการให้คะแนนนี้ ส่วนการให้คะแนนแบบอาร์โนลด์ ผู้สอบได้รับการคาดหวังให้รายงานตำแหน่งจริงของความรู้ (true state of knowledge) ของเขาในทุกข้อ จึงลดความแปรปรวนในการเดาลง ซึ่งผลลัพธ์นี้ควรแนะนำได้ว่าวิธีการให้คะแนนแบบอาร์โนลด์เหนือกว่า NR และ CG อย่างไรก็ตามจากงานวิจัยที่ใช้ CG ในการให้คะแนนให้ข้อแนะนำว่าผู้สอบไม่สามารถหรือไม่อาจแสดงระดับจริงของความรู้ของเขาได้ (Cross and Frary, 1977) เพราะผู้ตอบยากที่จะหลีกเลี่ยงที่จะไม่แสดงว่ารู้ผิดในการเลือกคำตอบเข้าเป็นตัวลวง จึงทำให้วิธีการให้คะแนนทั้ง 3 วิธีนี้ใกล้เคียงกัน ทางเลือกจึงขึ้นอยู่กับความเหมาะสมในการลงโทษผู้รู้ผิดด้วยคะแนนการลงโทษจำนวนน้อยๆ (โดยการพิจารณาจากประเด็นของความตรงที่จะศึกษา) และขึ้นอยู่กับความสามารถของผู้สอบที่จะรายงานระดับจริงของความรู้ของเขา จากการวิเคราะห์ของ Frary (1980) นี้ให้ข้อสรุปว่า

(1) ในกรณีที่ผู้ตอบถูกคาดหวังให้แสดงตำแหน่งจริงของความรู้ของเขาในข้อสอบรายข้อ วิธี CBS PNM และ A เป็นวิธีการที่ควรนำมาใช้มากที่สุด ส่วนวิธี CG AUC และ NR เป็นวิธีที่ควรใช้น้อยที่สุด การเลือกวิธีใดต้องสอดคล้องกับคุณลักษณะของผู้ตอบ เช่นแรงจูงใจ และความสามาถในการเข้าใจคำชี้แจง ซึ่งจะนำไปสู่การเพิ่มความเที่ยงในการวัด ในการใช้วิธี CBS A และ PNM

(2) คะแนนความรู้บางส่วนวิธี NR CG และ A ให้คะแนนในส่วนนี้น้อยกว่าเมื่อเทียบกับวิธี AUC และ CBS จะมีการให้คะแนนความรู้บางส่วนมากกว่า การเลือกใช้วิธีใดคำถามนี้เกี่ยวกับความตรงตามเกณฑ์และธรรมชาติของเกณฑ์ที่ใช้เป็นพื้นฐาน แต่อย่างไรก็ตามก็เกี่ยวข้องกับเนื้อหาของแบบสอบที่ใช้ทำนายด้วย ซึ่งต้องพิจารณาจากการวิเคราะห์ด้วยหลักเหตุผลด้วยว่าควรจะให้คะแนนที่ให้คะแนนแก่ความรู้บางส่วนมากหรือน้อย การวิเคราะห์ที่ต้องการจะกำหนดถึงวิธีการตอบและการให้คะแนนที่ดีที่สุดสำหรับเกณฑ์เฉพาะนั้นๆ

(3) การรู้ผิด (misinformation) มีความเกี่ยวข้องกับความตรง เนื่องจากวิธีการตอบและการให้คะแนนต่างๆ มีการลงโทษผู้รู้ผิดในระดับแตกต่างกัน งานการวิจัยเชิงประจักษ์จะเปิดเผยให้ทราบถึงความแตกต่างของความตรงตามวิธีให้คะแนนต่างๆ ที่เอามาใช้

Frary (1980) ได้สรุปว่าข้อสรุปดังกล่าวได้แสดงถึงความสำคัญของวิธีการตอบและการให้คะแนนที่จะกำหนดขึ้นในกระบวนการพัฒนาแบบสอบ คุณลักษณะนี้ได้ปรากฏบ่อยครั้งว่าถูกละเลยไปโดยสิ้นเชิงในกระบวนการพัฒนาแบบสอบ และ หลักฐานที่ได้วิเคราะห์มาแสดงถึงความจริงว่าวิธี NR และ CG ได้นำมาใช้บ่อยครั้งในงานการทดลองส่วนวิธี AUC ได้นำมาใช้บ้างแต่ดูเหมือนว่ายังต้องการวิธีการปรับอุปกรณ์ที่เป็นกระดาษคำตอบให้เหมาะสม ส่วนวิธี CBS และ A สามาถนำมาใช้ด้วยกระดาษคำตอบแบบเดิมที่สามารถปรับให้เหมาะสมขึ้นอีกเพียงเล็กน้อย วิธีการให้คะแนนนี้กำลังเป็นที่ต้องการของผู้ตอบ และโชคร้ายที่งานวิจัยในอดีตที่เกี่ยวกับความตรงที่ปรับปรุงขึ้นจากวิธีการตอบและการให้คะแนนอื่นๆ นอกจาก NR ยังไม่ได้ผลมากนักในอดีต ซึ่งเป็นสิ่งที่ควรแสวงหาคำตอบต่อไป

Frary (1980) ได้ใช้วิธีการจำลองแบบ (simulation) ในการศึกษาเปรียบเทียบความเที่ยง และความตรงของคะแนนจากวิธีการตอบและการให้คะแนน 6 วิธี ซึ่งประกอบด้วยวิธีการให้คะแนนแบบแระเพณีนิยม (NR) วิธีการแก้การเดาแบบเดิม (CG) วิธีการให้คะแนนแบบคูมบ์ (CBS) วิธีการให้คะแนนของอาร์โนลด์ (A) วิธีการตอบจนกว่าจะถูก (AUC) วิธีการของครอส (CRS) สำหรับข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ ซึ่งจำลองคำตอบขึ้นจากผู้ตอบ 120 คน ชุดของคะแนนได้สร้างขึ้นเพื่อสะท้อนให้เห็นถึง 5 ระดับของการมีอยู่ของการรู้ผิด โดยเชื่อว่าคำตอบคือตัวลวง และ 5 ระดับของการเดาซึ่งละเมิดต่อคำชี้แจงการตอบซึ่งห้ามไม่ให้เดา เกณฑ์ถูกจำลองขึ้นด้วยการตอบแบบคะแนนรวมข้อถูก ซึ่งมี 5 ระดับของการรู้ผิด และ 4 ระดับของความสัมพันธ์ของคะแนนจริงกับตัวทำนาย กระบวนการทั้งหมดถูกระทำซ้ำๆ ด้วยการแจกแจงปกติ และความคลาดเคลื่อนสุ่มในระดับข้อสอบ กระบวนการนี้ให้คะแนน 260 ชุดของคะแนน 5 ตัว ระหว่างตัวทำนาย และเกณฑ์ซึ่งถูกตรวจสอบโดยการกำหนดผลที่ต่างกันในเรื่องของความเที่ยงและความตรงในแต่ละลักษณะของการตอบและการให้คะแนน ผลการวิจัยพบว่าวิธีการให้คะแนนที่ยอมให้มีการตอบหลายคำตอบ (multiple response) ในข้อสอบข้อหนึ่งๆ จะช่วยให้ได้ค่าประมาณความเที่ยงสูงกว่าวิธี NR และ CG ซึ่งพบว่าวิธีการให้คะแนนแบบ A และแบบ CBS มีค่าความเที่ยงแบบความสอดคล้องภายในสูงกว่าวิธีอื่นๆ ซึ่งมีแนวโน้มจะทำให้ค่าสัมประสิทธิ์ความตรงสูงขึ้น แต่ให้ค่าความแตกต่างของความตรงจากวิธีการให้คะแนนต่างๆไม่แตกต่างกันอย่างมีนัยสำคัญ

จะเห็นว่าจากการศึกษาคุณภาพของสูตรการให้คะแนนแบบอาร์โนลด์ เป็นวิธีการให้คะแนนที่มีการพิจารณาให้คะแนนความรู้บางส่วนแก่ผู้ตอบและในขณะเดียวกันก็มีวิธีการแก้การเดาซึ่งพัฒนาขึ้นมาจากโมเดลทางคณิตศาสตร์ที่ให้คะแนนที่เพิ่มขึ้นจากการเดาสุ่มเป็นศูนย์ และสามารถปรับการลงโทษผู้รู้ผิดได้ตามความเหมาะสมกับความมุ่งหมายของผู้สอบเพื่อทำให้ผู้ตอบท้อถอยต่อการเดาคำตอบ เมื่อเปรียบเทียบคะแนนที่ได้จากวิธีการให้คะแนนแบบอาร์โนลด์กับวิธีการให้คะแนนแบบอื่นๆ จะสามารถทำให้ผู้ตอบในตำแหน่งตอนกลางมีการกระจายตำแหน่งกันได้ชัดเจนกว่า (Arnold and Arnold, 1970) การให้คะแนนแบบอาร์โนลด์ช่วยทำให้คะแนนที่ได้จากการสอบมีความเที่ยงสูงขึ้น เมื่อเทียบกับการให้คะแนนแบบประเพณีนิยม ถึงแม้ว่าจะยังไม่สามารถสรุปได้ชัดเจนถึงการปรับปรุงขึ้นในเรื่องความตรงตามเกณฑ์ (Frary, 1982) ด้วยความเหมาะสมของวิธีการให้คะแนนแบบอาร์โนลด์นี้จึงควรที่จะนำมาใช้เป็นกรอบในการพัฒนาสูตรการให้คะแนนในการวิจัยนี้ต่อไป

2.1 วิธีการให้คะแนนแบบแฮมดาน

วิธีการให้คะแนนด้วยสูตรการให้คะแนนที่ประมาณความรู้จริงของผู้สอบ เสนอโดย Hamdan (1979) เขาได้ใช้ทฤษฎีความน่าจะเป็น (probability theory) เพื่อเลือกตัวแก้ที่เหมาะสมของการเดา โดยผ่านการศึกษาการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่เกี่ยวข้อง ซึ่งสามารถทำให้

แปลงค่าที่เหมาะสมของการแจกแจงของพารามิเตอร์เพื่อนำไปสู่การแก้การเดาอย่างสุ่ม และ การเดาอย่างมีความรู้บางส่วนที่แตกต่างกัน maximum likelihood estimator ของพารามิเตอร์ที่เกี่ยวข้องจะให้ตัวประมาณการแก้การเดาที่สอดคล้องกัน แอมดานได้แปลงการแจกแจงความน่าจะเป็นร่วม (joint probability distribution) ของตัวแปรสุ่ม (x) ที่วัดได้ (ซึ่งหมายถึงจำนวนคำตอบที่ถูกต้องที่เป็นคำตอบ) กับ ตัวแปรสุ่ม (x_1) ที่วัดไม่ได้ (จำนวนคำตอบที่แสดงความรู้จริงของนักเรียน) โดยการแจกแจงความน่าจะเป็นร่วมระหว่างตัวแปรสุ่ม X และ X_1 นี้ ได้รวมพารามิเตอร์ 2 ตัวคือ p_1 และ p_2 ซึ่ง p_1 คือสัดส่วนของเนื้อหาวิชาที่รู้จริงโดยนักเรียน และ p_2 คือโอกาสของความน่าจะเป็นของการเดาถูก maximum likelihood estimator ของ $E(x_1)$ ที่ได้มาจากการกำหนด p_2 ทั้ง 2 กรณี ที่จะได้รูปแบบของการให้คะแนนการสอบในแต่ละประเด็นต่อไป

ต่อไปนี้จะแสดงวิธีการที่แอมดานได้เสนอขั้นตอนในการประมาณค่าต่างๆเพื่อให้ได้มาซึ่งสูตรการให้คะแนน เมื่อกำหนดสัญญลักษณ์ต่างๆเพื่อแสดง the joint probability distribution ของ x และ x_1 ดังนี้

- x คือ จำนวนคำตอบที่ถูกทั้งหมด
- x_1 คือ จำนวนคำตอบที่รู้จริงโดยนักเรียน
- x_2 คือ จำนวนการเดาถูก
- p_1 คือ สัดส่วนของเนื้อหาที่รู้โดยนักเรียน
- p_2 คือ ความน่าจะเป็นของการเดาถูก
- k คือ จำนวนข้อในข้อสอบ

เมื่อให้ $X_1=x_1$ (x_1 คือคำตอบที่รู้จริง) นักเรียนจะเดาคำถามที่เหลืออีก $k-x_1$ ด้วยความน่าจะเป็นในการเดาถูก p_2 ดังนั้น the conditional distribution ของ x_2 เมื่อกำหนด $X_1 = x_1$ เป็น binomial with probability function

$$p(X_2 = x_2 \mid X_1 = x_1) = \binom{K-X}{X} p_2^{x_2} q_2^{k-x_1-x_2} \quad , \; p_2 + q_2 = 1$$

$$x=0,1,2,\ldots,k-x_1 \quad \ldots\ldots\ldots\ldots\ldots(3)$$

แต่ marginal distribution ของ x_1 เป็น binomial distribution ด้วยพารามิเตอร์ k และ p_1 นั่นคือ

$$p_1(x_1) = p(X_1 = x_1) = \binom{K}{X_1} p_1^{x_1} q_1^{k-x_1} \quad \text{เมื่อ } p_1+q_1 = 1$$

$$x=0,1,2,\ldots,k \quad \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

117337082

ดังนั้น the joint probability function ของ x และ x_1 คือ

$$p(x,x_1) = p(X=x, X_1=x_1)$$
$$= p(X_1=x_1)\, p(x_1+x_2 = x \mid X_1=x_1)$$
$$= p(X_1=x_1)\, p(X_2=x-x_1 \mid X_1=x_1)$$
$$= \binom{k}{x_1} p_1^{x_1} q_1^{k-x_1} \binom{k-x_1}{x-x_1} p_2^{x-x_1} q_2^{k-x}$$
$$= \binom{k}{x}\binom{x}{x_1} p_1^{x} q^{k-x_1} p_2^{x-x_1} q_2^{k-x}$$

เมื่อ $x_1 = 0,1,\ldots,x;\ x=0,1,\ldots,k$ ....(5)

the marginal probability probability function ของ x คือ

$$p(x) = p(X=x) = \sum_{x_1=0}^{x} P(x,x_1)$$
$$= \binom{k}{x}(p_2/q_2)^{x}(q_1 q_2)^{k} \sum (p_1/p_1 q_1)^{x}$$
$$= \binom{k}{x}(p_2/q_2)^{x}(q_1 q_2)^{k}[1+p_1(p_2+q_1)]^{x}$$
$$= \binom{k}{x}(p_1+p_2 q_1)^{x}(q_1 q_2)^{k-x},$$

$x=0,1,\ldots,k$...(6)

ดังนั้น x จึงเป็น binomial ที่มีพารามิเตอร์ k และ $p_1+p_2 q_1$ ในท้ายที่สุด conditional probability function x_1 ที่กำหนดให้ X=x คือ ratio ของ (5) และ (6) ซึ่งทำให้ได้

$$p(x_1 \mid x) = \binom{x}{x_1} p^{x} (1-p)^{x-x_1} \quad , x_1=0,1,\ldots,x \quad \text{......(7)}$$

ซึ่ง $p = p_1/(p_1+p_2 q_1)$ ....(8)

การเดาสุ่มและการเดาอย่างมีความรู้บางส่วน

เราได้เลือกตัวเลือก 2 ตัวของพารามิเตอร์ p_2 ตัวเลือกแรกอยู่บนพื้นฐานของการเดาสุ่ม ซึ่งทำให้ได้ค่า $p_2(r)=1/n$ และ สอดคล้องกับข้อตกลงเบื้องต้นที่ว่านักเรียนไม่ได้ใช้ความรู้ใด ๆ เกี่ยวกับเนื้อ.เรื่องที่ตอบในกระบวนการเดา อีกตัวอย่างหนึ่งสำหรับตัวเลือกที่ 2 ของพารามิเตอร์

p_2 อยู่บนพื้นฐานของข้อตกลงที่ว่านักเรียนจะนำความรู้ของเขามาตัด m ตัวเลือกออกจาก n ตัวเลือก โดยนักเรียนจะเดา n-m คำตอบที่เหลือที่เหมาเอาว่ารวมตัวถูกเข้าไว้ด้วย n จึงเป็นสัดส่วนกับ p_t ซึ่งเป็นสัดส่วนคำตอบที่รู้จริงโดยนักเรียน และได้ค่าเฉลี่ย $m=np_t$ ซึ่งนำไปสู่การเดาอย่างมีความรู้บางส่วน $p_2(e) = 1/(n-np_t) = 1/nq_t$ ซึ่ง $nq_t > 1$ หรือ $p<(n-1)/n$ Maximum likelihood estimator of p_t และคะแนนที่จะประมาณได้สุดท้าย

สูตร (6) แสดงว่าค่าเฉลี่ยของตัวแปรสุ่มที่วัดได้ภายใต้การเดาสุ่มและการเดาอย่างมีความรู้บางส่วน คือ ค่าต่อไปนี้

$$E_r(x) = k(p_t+q_t/n) = k/n[1+(n-1)p_t] \quad \text{.....................(9)}$$

$$E_e(x) = k(p_t+1/n) = k/n(1+np_t) \quad \text{.......................(10)}$$

ถ้ากำหนด x_o เป็นผลรวมคะแนนสังเกตทั้งหมดจากข้อที่ถูกวัดได้จริง maximum likelihood estimator ของ p_t จะเป็นดังนี้

$$\hat{p}_t(r) = (nx_o - k) / [k(n-1)] \quad \text{.......................(11)}$$

$$\hat{p}_t(e) = (nx_o - k) / (kn) \quad \text{.......................(12)}$$

ดังนั้นจาก (9) และ (10) ค่าเฉลี่ยของ x_t ที่ได้จากการเดาสุ่มหรือการเดาอย่างมีการศึกษาคือ $E(x_t) - kp_t$ ซึ่งสามารถได้มาจากสูตร ที่ (11) และ (12)

$$E_r(x_t) = (nX_o - k)/(n-1) = x_o - (k-x_o)/(n-1) \quad \text{............(13)}$$

$$E_e(x_t) = (nx_o - k)/n = X_o - (k/n) \quad \text{..............(14)}$$

ค่าประมาณคะแนนความรู้จริงจากสูตรที่ (13) และ (14) คือค่าประมาณความรู้จริงที่ปราศจากการเดาสุ่ม และ การเดาอย่างมีความรู้บางส่วนตามลำดับ แฮมคานได้อภิปรายว่าค่าประมาณคะแนนจริงจากการเดาอย่างมีความรู้บางส่วนในสูตรที่ (14) ไม่เหมาะสม เพราะตัวแก้คะแนนที่สังเกตได้มีเพียงรูปแบบเดียวคือ k/n ซึ่งไม่ขึ้นอยู่กับค่า x_o ซึ่งเป็นคะแนนที่สังเกตหรือวัดได้ ค่าแก้ดังกล่าวจึงดูไม่สมเหตุผลเพราะนักเรียนสองคนที่มี $p_t=0$ และ $p_t=1/2$ จะได้ค่าแก้ที่เท่ากัน เหตุผลก็เพราะการไปเลือกจำนวนตัวลวงที่ผู้สอบตัดออกไปโดยมีความรู้บางส่วน ($m = np_t$) ซึ่งเป็นค่าคงที่ในสัดส่วนที่ขึ้นกับค่า n ซึ่งก็คือจำนวนตัวเลือกในแต่ละข้อนั่นเอง ถ้าเลือก $m = cp_t$ และกำหนดให้ $0<c<n$ แล้ว $p_2 = 1/n-cp_t$ เมื่อ $(n-cp_t)>1$ จาก (6) จะได้

$$E_{*}(x) = k(p_i + p_2 q_i) = k[p_i + (1-p_i)/(n-cp_i)] \quad \ldots\ldots\ldots\ldots(15)$$

เมื่อกำหนด $E_{*}(x) = X_o$ ใน (15) จะได้สมการ ควอดราตริค p_i ดังนี้

$$kc\hat{p}_i^2 - (cX_o - k + nk)\hat{p}_i + (nX_o - k) = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots(16)$$

ในกรณีสุดโต่งคือการเดาอย่างสมบรณ์ (c=0) แทนใน 16 จะได้ผลลัพธ์ $p_i(r)$ ที่กำหนดโดยสูตร (11) และได้ตัวประมาณค่าความรู้จริงจากการเดาสุ่มในสูตร (13) และ ในกรณีสุดโต่งอีกกรณีหนึ่ง (c=r) แทนในสมการที่ (16) จะได้ $p_i(e)$ ในสูตร (12) ที่ให้ตัวประมาณค่าความรู้จริงที่ไม่เหมาะสมในสูตร (14) ดังนั้นค่า c จึงควรมีค่าอยู่ระหว่าง 0 และ n เพื่อที่จะได้ค่าประมาณ p_i ซึ่งให้ค่าแก้การเดาซึ่งขึ้นอยู่กับค่า X_o แฮมดานสรุปว่าค่า c นี้ขึ้นอยู่กับความสามารถของนักเรียนที่จะใช้ความรู้บางส่วนของเขา ในเนื้อหาวิชาตัดตัวเลือกบางตัวที่เป็นตัวลวงออกไป และให้ข้อสังเกตว่าค่า c ที่มาแทนในสูตร (16) จะเป็นค่าเฉลี่ยที่เท่ากันทุกคำถามในการสอบแต่แตกต่างกันไปในผู้สอบแต่ละคน

เมื่อพิจารณากรณีตัวอย่างที่จำนวนข้อสอบทั้งหมด (k) มี 10 ข้อ จำนวนตัวเลือกในแต่ละข้อ (n) มี 4 ตัวเลือก และกรณีที่ผู้ตอบมีความรู้บางส่วนตัดตัวลวงออกได้ (C) 2 ตัว สูตรการให้คะแนนสูตรที่ 16 จะกลายเป็น $10\hat{p}_i^2 - (X_o+15)\hat{p}_i + (2X_o - 5) = 0$ ซึ่งราก $p_i = (X_o + 15 \pm\sqrt{X_o^2 - 50X_o + 425})/20$ เพราะว่า X_o มีค่าระหว่าง 0 ถึง 10 รากที่สองจึงต้องน้อยกว่า 5 เครื่องหมายบวกหน้ารากที่สองจึงตัดทิ้งไปคงใช้แต่เครื่องหมายลบ ดังนั้นในกรณีที่ X_o=8 เราจะได้ p_i= .687 $E(x_i)=k\hat{p}_i$=6.87 ผลลัพธ์นี้เพื่อเปรียบเทียบกับผลที่ได้จากสมการที่ 11 และ 12 จะได้ $E_r(X_i)$=7.33, $E_{*}(X_i)$=5.5 สำหรับในกรณีที่ X_o=5 จะได้ $\hat{p}_i$=.293 $E(X_i)$=2.93 ในขณะที่ $E_r(x_i)$=3.33 และ $E_{*}(X_i)$=2.5

จากตัวอย่างข้างต้นพบว่าแนวความคิดในการสร้างสูตรการให้คะแนนของแฮมดานได้มีการพิจารณาเลือกตัวแก้การเดาที่มีความรู้บางส่วนโดยการตัดตัวลวงบางตัวออกได้ แต่คะแนนที่ได้จากสูตรของแฮมดานเป็นการเลือกตัวแก้ที่ทำให้ไม่มีการพิจารณาให้คะแนนแก่การตอบด้วยการเดา ไม่ว่าจะเป็นการเดาสุ่มหรือการเดาอย่างมีความรู้บางส่วน ซึ่งพบว่าคะแนนจากสูตรการให้คะแนนของแฮมดานในกรณีการเดาอย่างมีความรู้บางส่วนจะได้คะแนนน้อยกว่าการเดาอย่างสุ่ม ซึ่งเมื่อนำมาพิจารณาว่าการที่ผู้สามารถตัดตัวลวงบางตัวออกได้อย่างถูกต้องก็เป็นความสามารถที่ควรได้รับการพิจารณาให้คะแนนเช่นเดียวกัน ดังนั้นจึงสามารถนำสูตรการให้คะแนนของแฮมดานมาประยุกต์ใช้กับวิธีการให้คะแนนที่มีการพิจารณาให้คะแนนความรู้บางส่วนของนักเรียนได้ต่อไป ซึ่งวิธีการของแฮมดานนี้ผู้วิจัยได้นำมาใช้เป็นกรอบในการพิจารณาในการให้คะแนนในการวิจัยนี้

การศึกษาคุณภาพของสูตรการให้คะแนนของแอมดานยังไม่มีงานวิจัยที่นำสูตรการให้คะแนนนี้ไปศึกษาเปรียบเทียบกับสูตรการให้คะแนนอื่นๆ คงมีแต่งานวิจัยของสมหวังและเพ็ญศิริ (2524) ได้นำสูตรการให้คะแนนนี้มาใช้เพื่อศึกษาพฤติกรรมการเดาสุ่มและการเดาอย่างมีความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบ ในกลุ่มที่มีพื้นฐานความรู้ทางคณิตศาสตร์แตกต่างกัน โดยศึกษารูปแบบการเดาของผู้ที่มีระดับความรู้พื้นฐานแตกต่างกันในวิชาสถิติสำหรับนักศึกษาปริญญาโทชั้นปีที่ 1 จำนวน 88 คน คะแนนความรู้บางส่วน (c) ที่ผู้วิจัยกำหนดมีค่า c=0 ถือว่าผู้ตอบตอบแบบเดาสุ่ม c=3 ถือว่าผู้ตอบตอบแบบมีความรู้บางส่วน และ c=ระดับเกรดเฉลี่ยตามพื้นฐานการเรียนคณิตศาสตร์ในภาคเรียนที่ผ่านมาใช้รูปแบบการเดาแบบผสม ข้อค้นพบจากการวิจัยคือ กลุ่มตัวอย่างที่มีพื้นฐานความรู้ทางคณิตศาสตร์ใช้รูปแบบการเดาสุ่มและการเดาแบบผสม ให้ค่าคะแนนคาดคะเนใกล้เคียงกว่ารูปแบบการเดาอย่างมีความรู้บางส่วน ส่วนในกลุ่มที่ไม่มีความรู้พื้นฐานทางคณิตศาสตร์ปรากฏว่ารูปแบบการสุ่มให้ค่าคะแนนคาดคะเนใกล้เคียงกว่าการเดาแบบอื่นๆ ข้อสรุปและอภิปรายที่ได้จากการวิจัยนี้ ผู้วิจัยเสนอว่าได้พบหลักฐานว่ากลุ่มตัวอย่างที่มีพื้นฐานความรู้คณิตศาสตร์และเรียนวิจัยโดยตรงจะใช้การเดาสุ่มและการเดาแบบผสมมากกว่าการเดาอย่างมีความรู้บางส่วน ซึ่งปฏิเสธสมมติฐานที่ผู้วิจัยตั้งไว้ ซึ่งผู้วิจัยให้เหตุผลว่าเป็นเพราะข้อสอบยากเกินไป ส่วนกลุ่มตัวอย่างที่ไม่มีพื้นความรู้คณิตศาสตร์ใช้วิธีการเดาสุ่มมากกว่าการเดาแบบผสมและการเดาอย่างมีความรู้บางส่วน ซึ่งทำให้ผลการศึกษาไม่เหมาะสมกับรูปแบบการเดาที่ผู้วิจัยกำหนดแต่แรก ซึ่งผู้วิจัยให้เหตุผลว่า ขึ้นอยู่กับค่าคะแนนความรู้บางส่วน (c) ที่กำหนดตามระดับเกรดในภาคเรียนที่แล้ว ซึ่งนักศึกษาได้เกรดเฉพาะ A และ B เท่านั้น จึงทำให้การเดาแบบผสมและการเดาอย่างมีความรู้บางส่วนใกล้เคียงกัน ผู้วิจัยให้ข้อสังเกตว่าถ้าทราบคะแนนดิบ แล้วใช้เป็นเกณฑ์ในการกำหนดว่าคะแนนความรู้บางส่วน ซึ่งเป็นดัชนีที่แสดงความสามารถของผู้ตอบจากการเดาสุ่มจนถึงการเดาอย่างมีความรู้บางส่วน รูปแบบการเดาแบบผสมน่าจะมีความเหมาะสมดังเช่นผลการวิจัยในกลุ่มตัวอย่างที่มีพื้นความรู้ทางคณิตศาสตร์

จากการศึกษาวิธีการสร้างสูตรการให้คะแนนของแอมดาน คุณสมบัติของคะแนนที่ได้จากสูตรนี้และผลงานวิจัยของสมหวัง และเพ็ญศิริ (2524) พบว่าการนำสูตรการให้คะแนนของแอมดานมาใช้ ผู้วิจัยต้องกำหนดคะแนนความรู้บางส่วน (c) ของผู้ตอบให้เหมาะสม ซึ่งงานวิจัยของสมหวังและเพ็ญศิริได้ใช้เกรดเฉลี่ยเดิมจากภาคเรียนที่แล้ว ส่วนแอมดานเสนอว่าค่า c คือค่าคงที่ที่แตกต่างกันในแต่ละคนและมีระดับเท่ากันในทุกข้อ ปัญหาสำคัญในการใช้สูตรการให้คะแนนของแอมดานจึงอยู่ที่การประมาณค่าคะแนนความรู้บางส่วน (c) ของผู้ตอบอย่างไร จึงจะใกล้เคียงกับความสามารถที่แท้จริงของเขา ซึ่งหากค่า c นี้ได้มาจากข้อมูลเชิงประจักษ์จากคำตอบที่ผู้ตอบสามารถตัดตัวลวงบางตัวออกได้จริง จะได้ค่าประมาณความรู้บางส่วนของผู้ตอบได้เหมาะสม ซึ่งในกรณีนี้จะได้นำมาใช้ในกรณีที่จะนำสูตรการให้คะแนนของแอมดานมาประยุกต์ใช้ในการวิจัยนี้ต่อไป

3. การศึกษาคุณภาพของสูตรการให้คะแนนในด้านความตรงเชิงทฤษฎีและวิธีการวิเคราะห์

จากการตรวจสอบวรรณคดีที่เกี่ยวข้องกับแบบสอบเลือกตอบ มีงานวิจัยจำนวนมากที่พยายามแก้ปัญหาการเดาที่เกิดจากการตอบแบบสอบเลือกตอบ ด้วยการพัฒนาสูตรการให้คะแนน และการพัฒนาวิธีการตอบและการให้คะแนนแบบต่าง ๆ โดยเฉพาะงานวิจัยที่เกี่ยวกับการให้ผู้ตอบได้แสดงความรู้บางส่วนในการตอบเพื่อจำแนกความสามารถของผู้ตอบในระดับต่าง ๆ ทั้งนี้ก็เพื่อให้คะแนนได้แสดงถึงระดับความสามารถที่แท้จริงของผู้ตอบได้ใกล้เคียงที่สุด การศึกษาถึงสูตรการให้คะแนนที่ให้คะแนนความรู้บางส่วนแก่ผู้ตอบ ผู้วิจัยใช้วิธีตรวจสอบคุณภาพของสูตรการให้คะแนนด้วยการอ้างอิงจากการเทียบกับสูตรการให้คะแนนแบบประเพณีนิยมและสูตรการให้คะแนนอื่นๆ ในด้านความเที่ยง และความตรงของคะแนนจากแบบสอบชนิดต่างๆที่นำมาใช้วัดกับกลุ่มตัวอย่างต่างๆ โดยอ้างอิงว่าสูตรการให้คะแนนที่มีคุณภาพดีกว่าย่อมช่วยส่งเสริมให้ผลของคะแนนที่วัดด้วยแบบสอบนั้นๆมีความเที่ยง และความตรงสูงขึ้น งานวิจัยที่ผ่านมาให้ข้อสรุปถึงการเพิ่มความเที่ยงของคะแนนจากแบบสอบที่ให้คะแนนแก่ความรู้บางส่วนของผู้ตอบ แต่ยังไม่มีหลักฐานเพียงพอที่จะสรุปได้ในเรื่องความตรงที่เพิ่มขึ้นในเรื่องความตรงตามเกณฑ์ ซึ่งเป็นประเด็นที่ควรจะได้ศึกษาต่อไปในวรรณคดีของการให้คะแนนแก่แบบสอบเลือกตอบ อย่างไรก็ตาม Lord (1975) ได้ให้แนวความคิดว่า ในการเปรียบเทียบวิธีการให้คะแนนแบบต่างๆจากแบบสอบเลือกตอบ คำถามสำคัญ ที่จะต้องตอบเกี่ยวกับวิธีการให้คะแนนก็คือว่าวิธีการให้คะแนนวิธีหนึ่งให้คะแนนที่เป็นตัวแทนของสิ่งที่ต้องการวัดได้แม่นยำกว่าอีกวิธีหนึ่งหรือไม่ และสิ่งที่ต้องการวัดนั้นเป็นสิ่งเดียวกัน หรือ ลักษณะ (trait) เดียวกันไม่ใช่สิ่งอื่น คำถามดังกล่าวเป็นสิ่งที่ผู้ศึกษาถึงวิธีการให้คะแนนแบบต่างๆควรจะแสวงหาคำตอบ ซึ่งเป็นการแสวงหาคำตอบเกี่ยวกับความตรงเชิงทฤษฎีของคะแนนนั่นเอง ในตอนต่อไปนี้จะได้กล่าวถึงวิธีการศึกษาความตรงเชิงทฤษฎีของสูตรการให้คะแนนซึ่งอ้างอิงจากผลการให้คะแนนจากแบบสอบที่วัดคุณลักษณะหรือความสามารถอย่างอย่างใดอย่างหนึ่ง ซึ่งมีโครงสร้างการนิยามคุณลักษณะหรือความสามารถนั้นตามทฤษฎีที่จะวัดคุณลักษณะนั้นอย่างชัดเจน หากคะแนนที่ให้ด้วยวิธีใดแสดงความสัมพันธ์กับคุณลักษณะที่ต้องการวัดได้สูง ย่อมสามารถอ้างอิงได้ถึงคุณภาพในด้านความตรงเชิงทฤษฎีของสูตรการให้คะแนนนั้น

3.1 วิธีการศึกษาความตรงเชิงทฤษฎีของสูตรการให้คะแนน

ในการศึกษาความตรงเชิงทฤษฎีของการวัด Campbell และ Fiske (1959) ได้ให้แนวทางในการวิเคราะห์ไว้ว่า วิธีการวัดหลายวิธีจะลู่เข้า (converge) กับลักษณะเดียวกัน แต่วิธีวัดต้องแยกออก (discriminant) ระหว่างลักษณะที่ต่างกันด้วย โดยที่ถ้าชนิดของพฤติกรรมที่สังเกตหรือวัดได้เป็นชนิดเดียวกันได้ถูกนำมาใช้เพื่อเป็นพื้นฐานในการอ้างอิงไปถึงระดับต่างๆของลักษณะที่มีอยู่นั้น การวัดที่ต่างกันก็ควรจะให้ผลเช่นนี้ด้วย Campbell และ Fiske ได้เสนอให้ตรวจสอบ

แนวคิดดังกล่าวโดยการนำวิธีต่าง ๆ กัน ที่วัดลักษณะต่าง ๆ กันมาพิจารณาร่วมกันในเมตริคซ์ลักษณะพหุ-วิธีพหุ (Multitrait-Multimethod Matrix or MTMM Matrix) ซึ่งเป็นเมตริคซ์ที่แสดงความสัมพันธ์ระหว่างลักษณะต่างๆที่วัด ด้วยวิธีการต่างๆกัน มากกว่า 1 วิธี ทั้งนี้เพราะกระบวนการหาความตรงต้องควรพิจารณาบนพื้นฐานของวิธีการดังต่อไปนี้

ก. กระบวนการหาความตรง (validation) เน้นการลู่เข้าเพื่อให้เกิดความมั่นใจในการศึกษาต้องใช้กระบวนการวัดที่เป็นอิสระจากกันมากที่สุดเพื่อตรวจสอบความตรงแบบลู่เข้า ซึ่งเป็นกระบวนการที่แตกต่างจากวิธีหาความเที่ยงคือต้องการวิธีวัดที่เหมือนกันมากที่สุด หรือใช้เครื่องมือชุดเดียวกัน

ข. ในการวัดลักษณะ (trait) ที่เป็นคุณลักษณะทางทฤษฎี และไม่สามารถสังเกตได้โดยตรง การพิจารณาถึงความตรงตามทฤษฎี ในกระบวนการหาความตรงแบบแยกออก (discriminant validity) ถือเป็นกระบวนการที่ต้องการการตรวจสอบเช่นเดียวกับการหาความตรงแบบลู่เข้า

ค. แบบสอบแต่ละชุดที่ใช้ในจุดมุ่งหมายของการวัดหนึ่งเป็นหน่วยที่เป็นผลรวมมาจากลักษณะและวิธีวัด (trait-method unit) เพราะความแปรปรวนที่เป็นระบบ (systematic variance) ระหว่างคะแนนสอบสามารถที่จะเกี่ยวกับผลที่มาจากคุณลักษณะของวิธีวัด เช่นเดียวกับผลที่มาจากเนื้อหาของ trait นั้นด้วยเช่นเดียวกับที่ Cronbach (1946 อ้างถึงใน Campbell and Fiske, 1599) กล่าวว่า " สิ่งที่แบบสอบวัดเป็นสิ่งที่ถูกกำหนดโดยเนื้อหาของข้อสอบทั้งหลาย แต่คะแนนสุดท้ายที่ได้มา เป็นส่วนผสมของผลที่มาจากเนื้อหาของข้อสอบและผลที่มาจากรูปแบบ (form) ของข้อสอบที่เอามาใช้ด้วย"

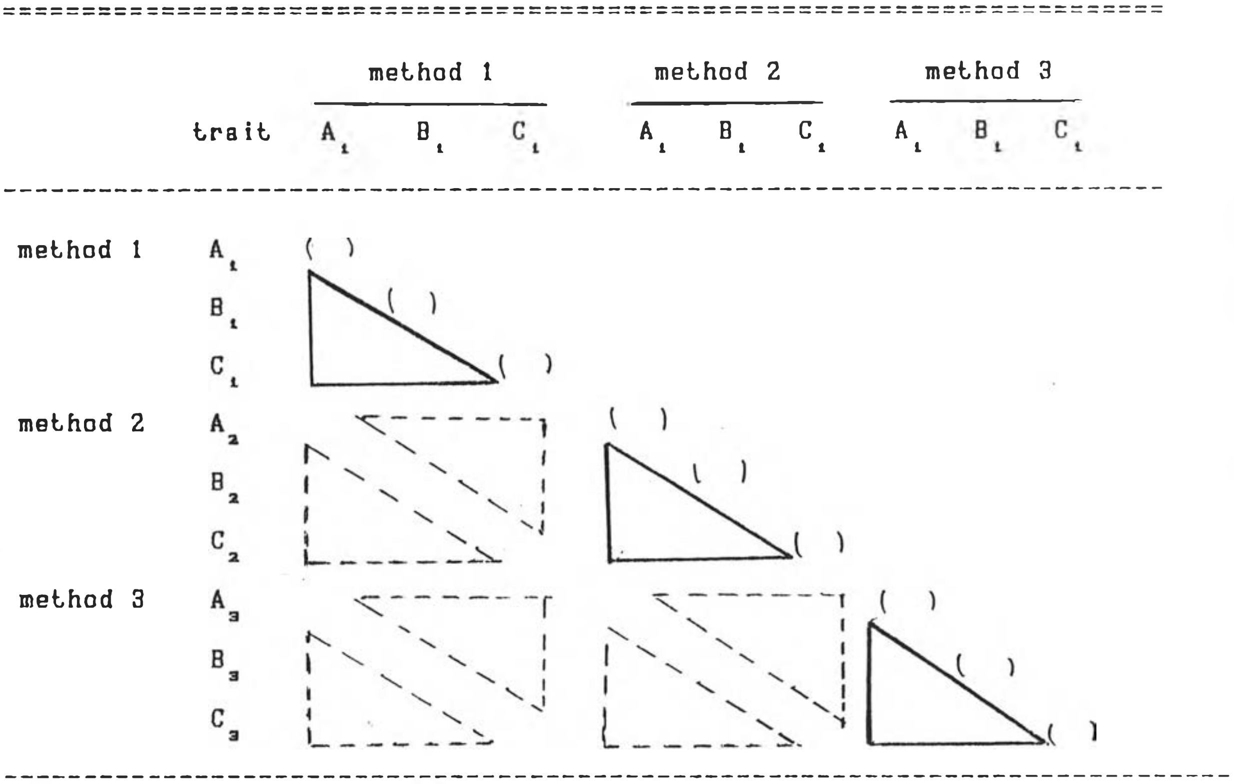
ง. เพื่อตรวจสอบเกี่ยวกับความตรงแบบแยกออก และการประมาณค่าความแปรปรวนที่เกี่ยวกับคุณลักษณะ (trait variance) และความแปรปรวนที่เกี่ยวกับวิธี (method variance) กระบวนการหาความตรงจึงต้องใช้วิธีการวัดมากกว่า 1 วิธี และคุณลักษณะมากกว่า 1 คุณลักษณะ ดังนั้นการศึกษาโดยผ่านเมตริคซ์ลักษณะพหุ-วิธีพหุ จะช่วยทำให้ประสบผลในการหาความตรงนี้ได้ Campbell และ Fiske ได้ยกตัวอย่างเมตริคซ์ที่มี ลักษณะ แตกต่างกัน 3 ลักษณะ และวิธีการวัดที่แตกต่างกัน 3 วิธี ดังตารางที่ 1 ซึ่งมีส่วนประกอบต่างๆของเมตริคซ์ ดังนี้

ความเที่ยง สามารถกำหนดได้จากค่าตามแนวเส้นทะแยงมุมใหญ่ของเมตริคซ์ทั้งหมด สามารถแสดงได้จาก monotrait-monomethod value (MTMM values) ซึ่งแสดงค่าบนเส้นทะแยงมุมของสามเหลี่ยมกรอบทึบ ส่วนค่าต่าง ๆ ในสามเหลี่ยมกรอบทึบจะเป็นค่าที่เรียกว่า heterotrait-monomethod values (HTMM values)

ความตรง สามารถกำหนดได้จากค่าตามเส้นทะแยงมุมในกรอบประซึ่งค่าความตรงนี้สามารถแสดงได้ด้วยค่าที่เรียกว่า monotrait-heteromethod values (MTHM values)

ในส่วนของสามเหลี่ยมทึบเรียกว่า monotrait-monomethod triangles และส่วนของสามเหลี่ยมประเรียกว่า heterotrait-heteromethod triangles

ตารางที่ 1 แสดงเมตริคซ์ลักษณะพหุ-วิธีพหุ ซึ่งมีลักษณะ 3 ลักษณะและวิธี 3 วิธี



ข้อสังเกต ค่าความตรงคือค่าตามเส้นทะแยงมุมระหว่างสามเหลี่ยมประ ส่วนค่าความเที่ยงคือค่าตามเส้นทะแยงมุมเหนือสามเหลี่ยมกรอบทึบในวงเล็บ

3.2 <u>วิธีการวิเคราะห์เมตริคซ์ลักษณะพหุ-วิธีพหุเพื่อศึกษาความตรงเชิงทฤษฎีของสูตรการให้คะแนน</u>

Campbell และ Fiske (1959) ได้แนะนำเกณฑ์ 4 เกณฑ์จากการตรวจสอบเมตริคซ์ลักษณะพหุ-วิธีพหุ ในการประเมิน trait-method combinations คือ

1) สัมประสิทธิ์ความตรง (validity coeffifients) คือสหสัมพันธ์ระหว่างวิธีการวัดที่ต่างกันใน trait เดียวกัน ควรจะมีนัยสำคัญและมีค่าสูงพอที่จะสนับสนุนในการวิจัยต่อไป ข้อกำหนดนี้เป็นหลักฐานของความตรงแบบลู่เข้า (convergent validity)

2) ค่าความตรงควรสูงกว่าค่าสหสัมพันธ์ที่อยู่ในแถว และ คอลัมภ์เดียวกัน ซึ่งอยู่ใน heterotrait-heteromethod triangles นั่นคือค่าความตรงสำหรับตัวแปรควรที่จะสูงกว่าสหสัมพันธ์ที่ได้มาระหว่างตัวแปรนั้น และ ตัวแปรอื่นๆที่ไม่ใช่ trait เดียวกัน และ ไม่ใช่วิธีเดียวกัน

3) ค่าความตรงจะมีค่าสูงกว่าค่าสหสัมพันธ์ในสามเหลี่ยมที่วัด Trait ต่างกันด้วยวิธีเดียวกัน (heterotrait-monomethod triangles)

4) แบบแผนความสัมพันธ์ระหว่าง trait ควรเหมือนกัน โดยสังเกตจากค่าที่แสดงใน heterotrait triangles ทั้งใน monomethod และ heteromethod Blockes เกณฑ์ 3 ข้อสุดท้ายนี้เป็นหลักฐานที่แสดงถึง Discriminant validity

สำหรับเกณฑ์ทั้ง 4 ของ Campbell และ Fiske ในการวิเคราะห์ MTMM Matrix มีผู้วิจัยหลายคนได้ชี้ถึงปัญหาในการใช้เกณฑ์ดังกล่าว คือ

1) Campbell และ Fiske ไม่ได้แสดงถึงระดับของเกณฑ์ในเชิงปริมาณว่าเท่าไรจึงจะถึงเกณฑ์การตัดสินค่าความตรงต้องอาศัยลักษณะเชิงคุณภาพเป็นเกณฑ์

2) ประเด็นที่มีความสำคัญต่อผู้วิจัยหลายคนอยู่ที่ความต้องการที่จะแยก method variance ออกจาก random error ซึ่ง Campbell และ Fiske ชี้ว่าการประเมินค่าสหสัมพันธ์ต่างๆ ใน MTMM ต้องสนใจค่าความเที่ยงของวิธีการวัดแต่ละวิธีด้วย ถ้าวิธีวัดหนึ่งมีความเที่ยงต่ำมากแล้วจะทำให้ method variance ในวิธีการวัดในเมทริคซ์ที่เหลือสูงขึ้น (inflate) ซึ่ง Jackson (1969) และ Althausu, and Herberlein (1970 อ้างใน Scmitt and Stults, 1986) ได้แนะนำให้มีการแก้ attenuation เกี่ยวกับความไม่เที่ยงของวิธีวัดใน MTMM Matrix

3) ปัญหาข้อนี้เกี่ยวกับเกณฑ์ของ Campbell และ Fiske ไม่สมบูรณ์ นักวิจัยหลายคน (Althausu, and Heberlein, 1970; Alwin, 1974; Kalleberge ,and Kluegel, 1975 Krause, 1977 อ้างใน Schmitt,and Stults, 1986) ชี้ถึงเกณฑ์ของ Campbell และ Fiske ว่ามีข้อตกลงเบื้องต้นที่ซ่อนอยู่ว่าไม่มีความสัมพันธ์ระหว่างตัวประกอบลักษณะและตัวประกอบวิธี (นั่นคือทุกๆ ลักษณะ ได้รับอิทธิพลเท่าๆกันจาก ตัวประกอบวิธี) และตัวประกอบวิธีไม่มีความสัมพันธ์กัน

จากการศึกษาถึงวิธีการในการวิเคราะห์ MTMM Matrix เพื่อแก้ปัญหาเกี่ยวกับเกณฑ์ของ Campbell และ Fiske Schmitt, and Stults (1986) ได้สรุปว่าวิธีการวิเคราะห์ความแปรปรวน และ วิธี Confirmatory factor analysis เป็นวิธีการที่ใช้กันมากที่สุด และ สามารถตอบคำถามที่ผู้วิจัยต้องการศึกษาได้ชัดเจน ซึ่งจะเสนอในรายละเอียดของแต่ละวิธี ดังนี้

3.2.1 <u>การวิเคราะห์ความแปรปรวนของเมทริคซ์ลักษณะพหุ-วิธีพหุ</u> (Analysis of Variance of MTMM Matrix)

วิธีการวิเคราะห์ความแปรปรวนสำหรับ MTMM Matrix เป็นวิธีการที่เอามาใช้กันมากที่สุดวิธีหนึ่ง วิธีนี้เสนอโดย Guilford (1954) และ พัฒนาต่อมาโดย Stanley (1961) Boruch, and Wolinns (1970) และ Borouch, Larkiing, and Mackinney (1979) (อ้างใน Schmitt, and Stults, 1986) สำหรับโมเดลการวิเคราะห์ความแปรปรวนนี้ ตัวแปรที่สังเกตได้แต่ละตัวเป็นการรวมกันขององค์ประกอบ 4 ตัว คือ

1) ตัวประกอบทั่วไป (A general factor) ซึ่งเป็นการวัดคนทั้งหมดในทุกลักษณะด้วยวิธีการทุกวิธีที่ศึกษา

2) ตัวประกอบลักษณะ (A trait dimension) ซึ่งเป็นการวัดที่บ่งชี้ว่า คนอยู่ในตำแหน่งที่เด่น หรือด้อยตามตัวประกอบทั่วไป (general factor) นั้น

3) ตัวประกอบวิธี (A method factor) ซึ่งเกี่ยวกับการวัดที่ว่าวิธีการวัดหนึ่ง ๆ ได้ให้คะแนนสูงหรือต่ำกว่าในทุก trait กับคนใดคนหนึ่ง

4) ความคลาดเคลื่อนสุ่ม (random error)

โมเดลการวิเคราะห์ความแปรปรวนได้นำมาใช้ด้วยข้อตกลงว่า ตัวประกอบลักษณะและตัวประกอบวิธีเป็นอิสระจากกัน และเป็นอิสระจากตัวประกอบทั่วไปทำให้ไม่ต้องมีการประมาณ trait intercorrelation, method intercorrelation หรือ trait-method intercorrelation Stanley (1969) ได้กำหนดวิธีการวิเคราะห์โดยคำนวณค่าเฉลี่ยของรูปแบบของสหสัมพันธ์ 3 ชนิดใน MTMM Matrix คือ

r_a เป็นค่าเฉลี่ยของสหสัมพันธ์ในกลุ่มของตัวแปรที่เกี่ยวกับลักษณะต่างกันวัดด้วยวิธีการต่างกันเรียกว่า heterotrait-heteromethod values

r_m เป็นค่าเฉลี่ยของสหสัมพันธ์ระหว่างตัวแปรที่เกี่ยวกับ ลักษณะต่างกันที่วัดด้วยวิธีเดียวกันเรียกว่า Heterotrait-monomethod values

r_t เป็นค่าเฉลี่ยของสหสัมพันธ์ระหว่างตัวแปรที่เกี่ยวกับลักษณะเดียวกันแต่วัดด้วยวิธีการที่ต่างกันเรียกว่า monotrait-monomethod values

Stanley ได้แสดงว่าค่าเหล่านี้เกี่ยวกับโมเดลการวิเคราะห์ความแปรปรวนแบบ 3 ทาง คือ Person X traits X Methods ตัวประกอบทั่วไปก็เปรียบเหมือนกับ person effect ซึ่งแสดงถึงความตรงแบบลู่เข้า ส่วนผลของตัวประกอบลักษณะก็เปรียบเหมือน person X method interaction ซึ่งแสดงถึง halo effect ของวิธีการวัดที่นำมาศึกษาไม่วิธีการใดก็วิธีการหนึ่งหรือทั้งหมด แผนการวิเคราะห์ ข้อมูลจาก MTMM Matrix แสดงในตารางที่ 2

วิธีการวิเคราะห์ความแปรปรวนใน MTMM Matrix ได้มีผู้วิจัยนำมาใช้กันมากที่สุดวิธีการหนึ่ง ด้วยข้อตกลงที่ค่อนข้างจำกัดของโมเดลทำให้ได้เพียงค่าประมาณรวมๆ trait/method/person/ variance ซึ่งไม่สามารถประเมินถึง trait-method unit แต่ละอันได้ ซึ่งบางครั้งวัตถุประสงค์ของการวิจัยต้องการประเมินค่าเหล่านี้ด้วย สำหรับโมเดลที่ยอมให้มีการประเมิน trait-method unit ก็คือ Confirmatory factor analytic model ซึ่งจะได้กล่าวในรายละเอียดต่อไป

ตารางที่ 2 สูตรการคำนวณสำหรับการวิเคราะห์ความแปรปรวนในเมทริคซ์ลักษณะพหุ-วิธีพหุ

แหล่งความแปรปรวน	SS	df	ค่าประมาณองค์ประกอบความแปรปรวน
SUBJECT	$N_{nm}(\bar{r}_t)$	N-1	$(MS_S - MS_{STM})/nm$
SUBJECT X TRAITS	$N_{nm}(\bar{r}_{SM} - \bar{r}_T)$	(N-1)(n-1)	$(MS_{ST} - MS_{STM})/m$
SUBJECT X METHODS	$N_{NM}(\bar{r}_{ST} - \bar{r}_T)$	(N-1)(m-1)	$(MS_{SM} - MS_{STM})/n$
SUBJECT X TRAITS X METHODS	$N_{nm}(1 - \bar{r}_{SM} - \bar{r}_{ST} + \bar{r}_T)$	(N-1)(n-1)(m-1)	MS_{STM}

$\bar{r}_T$ คือ สหสัมพันธ์เฉลี่ยของเมทริคซ์เต็มรูปที่มีค่า 1 อยู่ที่เส้นทะแยงมุมหลัก

$\bar{r}_{SM}$ คือ ความตรงเฉลี่ย คำนวณจาก (2(ผลรวมความตรงตามเส้นทะแยงมุม)+nm)/nm^2

$\bar{r}_{ST}$ คือ (2(ผลรวมของค่าสหสัมพันธ์ในสามเหลี่ยม monomethod-heterotrait)+nm)/nm^2

N คือ จำนวน SUBJECT n คือ จำนวน trait และ m คือจำนวน method

3.2.2 <u>Confirmatory Factor Analysis</u> (CFA)

โมเดลที่ใช้ในการวิเคราะห์ตัวประกอบ หรือ ตัวแปรแฝงซึ่งผู้วิจัยสนใจจะวัดแต่ไม่สามารถสังเกต หรือ วัดได้โดยตรงมีอยู่ 2 โมเดล คือ Exploratory factor model และ Confirmatory factor model ใน Exploratory factor model มีคำถามที่ผู้วิจัยค้นหาคำตอบเกี่ยวกับจำนวนของตัวประกอบจากชุดของตัวแปรที่สังเกตได้ ซึ่งผู้วิจัยตั้งข้อตกลงเบื้องต้นว่า

ก. common factor ทุกตัวสัมพันธ์กัน (หรือบางรูปแบบ common factor ไม่สัมพันธ์กัน)

ข. ตัวแปรที่สังเกตได้ทุกตัวถูกกระทบโดยตรงจาก common factor ทุกตัว

ค. unique factor ไม่สัมพันธ์กับ unique factor ของตัวแปรตัวอื่นๆ

ง. ตัวแปรที่สังเกตได้ทุกตัวถูกกระทบโดย unique factor และ

จ. ตัวแปรแฝง (ξ) ทุกตัวไม่สัมพันธ์กับ unique (δ) ทุกตัว

ข้อตกลงที่จำกัดของ Exploratory factor model นี้สามารถลดหย่อนได้ด้วย Confirmatory factor model และยอมให้มีการทดสอบทางสถิติว่าข้อมูลที่เก็บมาได้ยืนยันถึง

โมเดลที่สร้างขึ้นในทางทฤษฎีหรือไม่ ในตอนต่อไปนี้จะได้บรรยายถึงขั้นตอนในการวิเคราะห์ CFM และการนำ CFM ไปใช้ในการวิเคราะห์ MTMM Matrix เพื่อตอบปัญหาในการวิจัยเกี่ยวกับความตรงเชิงทฤษฎีของคะแนนที่ให้กับแบบสอบเลือกตอบ

ขั้นตอนในการวิเคราะห์ Confirmatory factor analysis ประกอบด้วยขั้นตอนต่างๆ ดังนี้

1) Specification

2) Identification

3) Estimation

4) Assessment of fit

การวิเคราะห์แต่ละขั้นตอนมีรายละเอียดดังนี้

1) Specification เป็นการกำหนดรูปแบบอย่างเป็นทางการเกี่ยวกับ จำนวนของ common factor, จำนวนของ obseved variables, ความแปรปรวน และความแปรปรวนร่วมระหว่าง common factor, ความสัมพันธ์ระหว่าง observed variable และ latent factor, ความสัมพันธ์ระหว่าง unique factor และ observed variables, ความแปรปรวนและความแปรปรวนร่วมระหว่าง unique factor

factor analysis พยายามที่จะอธิบายถึงความแปรปรวนและความแปรปรวนร่วมในชุดของตัวแปรที่สังเกตได้ในรูปของตัวประกอบที่สังเกตไม่ได้โดยตรง ซึ่งมีจำนวนน้อยกว่า โดยตัวแปรที่สังเกตได้แต่ละตัวจะถูกมองว่าเป็นฟังชั่นเชิงเส้นของตัวประกอบตัวหนึ่ง หรือ มากกว่าหนึ่ง ซึ่งตัวประกอบเหล่านี้มีอยู่ 2 ชนิด คือ ตัวประกอบร่วม (common factor) ซึ่งเป็นตัวที่กระทบต่อตัวแปรที่สังเกตมากกว่า 1 ตัว ส่วนชนิดที่สองเป็นตัวประกอบเดี่ยว (unique factor หรือ residual factor) เป็นตัวประกอบที่กระทบโดยตรงกับตัวแปรที่สังเกตได้เพียงหนึ่งตัวเท่านั้น ความสัมพันธ์เชิงเส้นระหว่างตัวแปรที่ สังเกตได้ และ ตัวประกอบสามารถแสดงได้ดังนี้

$$X = \Lambda \xi + \delta \qquad \text{...............} (17)$$

เมื่อ X คือ เวคเตอร์ของตัวแปรที่สังเกตได้ขนาด q X 1

ξ คือ เวคเตอร์ของ common factor ขนาด s X 1

Λ คือ เมตริคซ์ของ factor loadings ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปรที่สังเกตได้ (x) กับตัวประกอบหรือตัวแปรแฝง (ξ) ขนาด q X s ซึ่งกำหนดให้ q > s

δ คือ เวคเตอร์ของ unique factor ขนาด q x 1

เนื่องจาก factor equation ในสมการที่ (1) ไม่สามารถจะประมาณค่าที่ไม่สามารถสังเกตได้โดยตรง เช่นสมการถดถอยธรรมดา จึงจำเป็นต้องมีการตรวจสอบจากโครงสร้างของความแปรปรวนร่วม (structure of covariance)ระหว่างตัวแปรที่สังเกตได้ ดังนี้

$$\Sigma = \Delta \Phi \Delta' + \Theta \qquad \ldots\ldots\ldots\ldots\ldots\ldots (18)$$

สมการนี้เรียกว่า covariance equation ซึ่งเป็นโมเดลในการวิเคราะห์สำหรับ confirmatory factor analysis

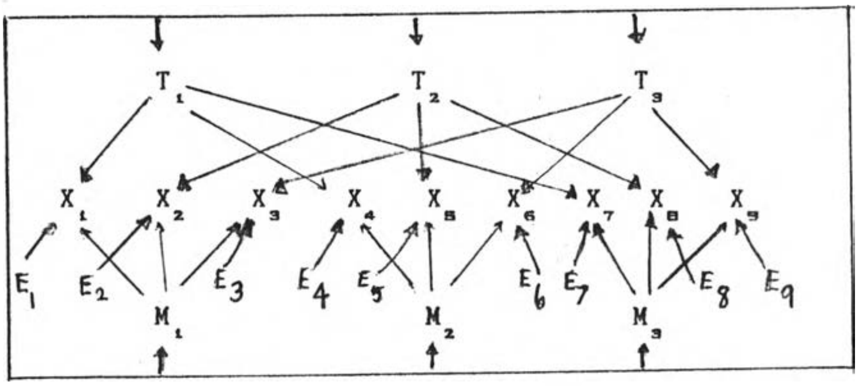
สำหรับเมตริคซ์ความแปรปรวนร่วม (Σ) นิยามว่าเป็น $\Sigma = E(xx')$ ซึ่งมีลักษณะเป็น symmetric matrix มีขนาด q x q สมาชิกตัวที่ i และ j ของ Σ เขียนได้ว่าเป็น σ_{ij} ซึ่งเป็นค่าประชากรของความแปรปรวนร่วมระหว่าง X_i และ X_j ซึ่งนิยามได้ว่า $\sigma_{ij} = E(x_i x_j)$ ถ้า x ถูกกระทำให้เป็นมาตรฐานมีความแปรปรวนเป็นหนึ่ง $E(x_i x_j)$ จะเป็นการแสดงแสดงความสัมพันธ์ระหว่าง x_i และ x_j และ Σ จะเป็นเมตริคซ์สหสัมพันธ์

ความแปรปรวนร่วมระหว่าง common factor จะใส่อยู่ในเมตริคซ์ Φ ซึ่งมีลักษณะเป็น symmetric matrix ขนาด s x s สมาชิกในเมตริคซ์ Φ คือ Φ สามารถที่จะแสดงถึงค่าความแปรปรวนร่วมระหว่างตัวแปรแฝง ξ_i และ ξ_j ในกรณีที่ common factorไม่สัมพันธ์กัน สมาชิกนอกเส้นทะแยงมุมของ Φ จะเป็นศูนย์ และถ้า common factor ถูกกระทำให้เป็นมาตรฐานเมื่อมี variance เป็น 1 Φ จะเป็น correlation matrix ที่มีค่าตามเส้นทะแยงมุมเป็น 1 และสหสัมพันธ์ระหว่าง common factor นอกเส้นทะแยงมุมเป็นศูนย์

ความแปรปรวนระหว่าง residual factor อยู่ในเมตริคซ์ Θ ซึ่งเป็น symmetric matrix ขนาด q X q สมาชิกตัวที่ i,j ของ Θ คือ Θ_{ij} เป็นความแปรปรวนร่วมระหว่างตัวประกอบคลาดเคลื่อน δ_i และ δ_j และในกรณีที่ตัวประกอบคลาดเคลื่อนไม่สัมพันธ์กันสมาชิกนอกเส้นทะแยงมุมจของเมตริคซ์ Θ จะเป็นศูนย์

ต่อไปนี้จะยกตัวอย่าง structure และข้อตกลงเบื้องต้นของ confirmatory factor ในตาราง multitrait-multimethod ในโมเดลของตัวแปรที่สังเกตได้แต่ละตัว ที่ประกอบด้วยองค์ประกอบ 3 องค์ประกอบ คือ ตัวประกอบลักษณะ (trait factor) ตัวประกอบวิธี (method factor) และตัวประกอบคลาดเคลื่อน (random error factor) ในโมเดลรูปทั่วไปตัวประกอบลักษณะและตัวประกอบวิธีอาจจะสัมพันธ์กันในขณะที่ ตัวประกอบคลาดเคลื่อนที่เกี่ยวข้องกับตัวแปรแต่ละตัวจะไม่สัมพันธ์กับ ตัวประกอบลักษณะ หรือตัวประกอบวิธี แสดงโครงสร้างของโมเดลดังภาพที่ 1

ภาพที่ 1 แสดงค่าตัวประกอบในเชิงทฤษฎีที่ส่งผลต่อตัวแปรที่วัดได้ 9 ตัว ตัวประกอบเชิงทฤษฎีนี้ประกอบด้วย ตัวประกอบที่เกี่ยวกับ trait , method, และ random error ซึ่งมีความสัมพันธ์ภายใน trait และ method และสัมพันธ์กันระหว่าง trait และ method



จากโครงสร้างตัวแปรดังกล่าวสามารถกำหนดค่าที่ต้องประมาณและค่าที่กำหนดให้คงที่ในเมตริคซ์ต่าง ๆ จากสมการ covariance equation ได้ดังตารางที่ 4 ค่าใน 4 A เป็น factor loading ค่าในตาราง 4 B แทน intercorrelation ระหว่าง trait และ method factor (เพราะค่าเหล่านี้เป็นสหสัมพันธ์ของ factor ที่เราสนใจ มันแสดงค่าสหสัมพันธ์ระหว่าง construct ที่ถูกวัดโดยปราศจากความคลาดเคลื่อน) สำหรับค่าต่างๆในตาราง 4 C เป็น unique หรือ random error component ค่า 0.0 และ 1.0 เป็นค่าคงที่ที่ผู้วิจัยกำหนดแทนสมมติฐานที่เกี่ยวกับ structure ของ MTMM Matrix พารามิเตอร์ที่ปล่อยให้เป็นค่าอิสระหรือค่าที่จะถูกประมาณจากพื้นฐานของเมตริคซ์สหสัมพันธ์ที่สังเกตได้ คือจำนวนเลขที่แสดงต่อๆกันในตารางที่ 4 ตั้งแต่ 1-42 ค่า ซึ่งมีสมาชิกที่เป็นค่าเดี่ยวๆไม่ซ้ำกันใน MTMM Matrix จำนวน q(q+1)/2 คือ 9 X (9+1)/2 = 45 ตัว ดังนั้นในโมเดลนี้จึงมีชั้นของความเป็นอิสระเท่ากับ 3 สำหรับใช้ในการทดสอบความเหมาะสมของโมเดลที่กำหนด

จากตัวอย่างการกำหนดค่าพารามิเตอร์ต่างๆในการให้พารามิเตอร์บางตัวเป็นค่าคงที่ และ บางตัวเป็นอิสระ ขึ้นอยู่กับโมเดลเชิงทฤษฎีที่ผู้ศึกษาใช้เป็นหลัก และ โมเดลที่เสนอนี้เป็นโมเดลทั่วไปที่จะสามารถทดสอบได้ บางครั้งผู้วิจัยอาจจะตั้งข้อตกลงบางประการที่สามารถเอามาใช้ทดสอบเมตริคซ์ที่เล็กกว่านี้ เช่น ในกรณีที่ให้ trait และ method factor ไม่สัมพันธ์กัน ตารางที่ 4 ก็จะเกี่ยวกับ การ fix พารามิเตอร์ 22-27 และ 29-31 ให้เป็นศูนย์ หรือหรือการกำหนดข้อตกลงถึงการไม่มีความสัมพันธ์ระหว่าง method factor และระหว่าง trait และ method factor และการที่ method factor มีผลเท่ากันในทุก trait ในกรณีนี้ก็จะเท่ากับการ fix พารามิเตอร์ 19-33 ให้เท่ากับศูนย์

ตารางที่ 4 แสดงค่าพารามิเตอร์ที่จะต้องถูกประมาณใน confirmatory analysis ของ MTMM Matrix ในภาพที่ 1

4 A TRAIT AND METHOD FACTOR LOADING

	T_1	T_2	T_3	M_1	M_2	M_3
x_1	1	0.0	0.0	2	0.0	0.0
x_2	0.0	3	0.0	4	0.0	0.0
x_3	0.0	0.0	5	6	0.0	0.0
x_4	7	0.0	0.0	0.0	8	0.0
x_5	0.0	9	0.0	0.0	10	0.0
x_6	0.0	0.0	11	0.0	12	0.0
x_7	13	0.0	0.0	0.0	0.0	14
x_8	0.0	15	0.0	0.0	0.0	16
x_9	0.0	0.0	17	0.0	0.0	18

4 B INTERCORRELATION OF TRAIT AND METHOD FACTORS

	T_1	T_2	T_3	M_1	M_2	M_3
T_1	1.0					
T_2	19	1.0				
T_3	20	21	1.0			
M_1	22	23	24	1.0		
M_2	25	26	27	28	1.0	
M_3	29	30	31	32	33	1.0

4 C RANDOM ERRORS ASSOCIATED WITH EACH MEASURED VARIABLES

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
34	35	36	37	38	39	40	41	42

2) <u>Identification</u>

เมื่อผู้วิจัยกำหนดค่าพารามิเตอร์ต่างๆใน matrix ด้านขวาของ covariance equation แล้ว ปัญหาที่จะถามต่อไปก็คือว่าเป็นไปได้หรือไม่ว่าพารามิเตอร์ที่จะถูกประมาณจะได้ค่าพารามิเตอร์เพียงค่าเดียว แล้วทำให้สอดคล้องกับ covariance equation เริ่มแรกปัญหาดังกล่าวนี้คือปัญหาของidentification ที่จะต้องดำเนินการต่อไป

identification เป็นกระบวนการที่สนใจว่าพารามิเตอร์ของโมเดลที่ถูกกำหนดขึ้นมาจะสามารถกำหนดได้อย่าง unique หรือไม่ ถ้าโมเดลไม่ถูก identify ก็เป็นไปไม่ได้ที่จะกำหนดค่าพารามิเตอร์ได้อย่าง unique เงื่อนไขที่ทำให้โมเดล identify จะสามารถพิจารณาได้จากเงื่อนไข ต่อไปนี้

ก. เงื่อนไขที่จำเป็น (necessary constriant) เป็นเงื่อนไขที่จำเป็นต้องกำหนดโครงสร้างของโมเดลตั้งแต่แรก เมื่อเงื่อนไขนี้ satisfy แล้วโมเดลจะถูก identify แต่ก็ไม่จำเป็นเสมอไปที่ว่าเงื่อนไขนี้ satisfy แล้ว จะทำให้โมเดล identify ยกตัวอย่างเงื่อนไขเกี่ยวกับจำนวน covariance equation ที่เป็นอิสระกันจะต้องมากกว่าหรือเท่ากับจำนวนพารามิเตอร์ที่กำหนดให้ถูกประมาณจาก covariance equation ตัวแปร q ตัว มี covariance equation q(q+1)/2 ตัว แต่มีพารามิเตอร์ที่เป็นอิสระที่จะต้องถูกประมาณเป็นไปได้ถึง (qs=(s(s+1)/2 +(q(q+1)/2) ตัว ในเมตริคซ์ Λ , Φ , และ Θ ตามลำดับ ดังนั้นโมเดลที่ถูก identify ก็เมื่อมีพารามิเตอร์อย่างน้อยที่สุด (qs+(s(s+1)/2) ตัว จะต้องถูกจำกัดค่า เงื่อนไขนี้เป็นเงื่อนไขที่จำเป็นแต่ไม่ทำให้โมเดล identify เสมอไป

ข. เงื่อนไขที่พอเพียง (sufficient constraint) เงื่อนไขนี้ถ้า satisfy โมเดลจะ identify แต่ในกรณีที่ โมเดลถูก identify บางครั้งพบว่าเงื่อนไขนี้ อาจจะไม่ satisfy ก็ได้

ค. เงื่อนไขที่ทั้งจำเป็นและพอเพียง (necessary and sufficient constraint) ถ้าเงื่อนไขนี้ satisfy แล้ว โมเดลจะถูก identify แต่ถ้าเงื่อนไขนี้ไม่ satisfy โมเดลก็จะไม่ถูก identify

โดยทั่วไปวิธีการที่มีประสิทธิภาพมากที่สุดที่แสดงว่าโมเดลถูก identify คือ การแสดงโดยผ่านการจัดกระทำทางพีชคณิตของ covariance equation ของโมเดล ในแต่ละพารามิเตอร์ที่สามารถแก้ไขได้ในเทอมของ ความแปรปรวนและความแปรปรวนร่วมของตัวแปรที่สังเกตได้ นี้เป็นเงื่อนไขทั้งจำเป็นและพอเพียง สำหรับ identification ด้วยวิธีการทางพีชคณิตนี้ พารามิเตอร์จะถูก identify บนพื้นฐานของแต่ละค่า ถ้าพารามิเตอร์สามารถถูกแก้ได้ในเทอมของ ความแปรปรวน และความแปรปรวนร่วม ของตัวแปรที่สังเกตได้แล้ว พารามิเตอร์นั้นจะถูก identify ในกรณีที่พารามิเตอร์ถูกแก้ด้วยวิธีการมากกว่าหนึ่งวิธีพารามิเตอร์นั้นจะถูก overidentify ในกรณีที่พารามิเตอร์ทุกตัวถูก identify โมเดลก็จะถูก identify ด้วย แต่ถ้าพารามิเตอร์บางตัวถูก overidentify โมเดลก็จะถูก overidentify ด้วย

การแก้ สมการความแปรปรวนต้องการเวลามาก และสำหรับโปรแกรมคอมพิวเตอร์ในปัจจุบันสามารถที่นำมากำหนด identification ได้โดยการคำนวณค่าประมาณ maximum liklihood ของพารามิเตอร์ และ information matrix ของพารามิเตอร์จะสามารถคำนวณได้ information matirx นี้จะสอดคล้องกับเมตริคซ์ของความแปรปรวน และ ความแปรปรวนร่วมสำหรับค่าพารามิเตอร์ Joreskog, and Sorbum (1985) กล่าวว่า information matrix เป็น positive definite ก็แทบจะแน่ใจได้ว่าโมเดลจะถูก identify

สำหรับ identification ใน MTMM Matrix วิธีที่โมเดลจะ identify ถ้าโมเดล satisfy เงื่อนไขที่พอเพียงของ Joreskog ในข้อใดข้อหนึ่งต่อไปนี้

ก. เมตริคซ์ Φ ต้องเป็น symmetric positive definite matrix มีสมาชิกบนเส้นทะแยงมุมเป็นหนึ่ง

ข. เมตริคซ์ Θ เป็น diagonal matrix

ค. Λ อย่างน้อย (s-1) ตัวถูก fix ให้เท่ากับ 0 ในแต่ละคอลัมภ์

ง. Λ^{k} มี rank เท่ากับ (s-1) เมื่อ (k=1,...s) เป็น symmetric ที่ประกอบด้วยแถวของ Λ มีค่าสมาชิก fix=0 ในคอลัมภ์ที่ k

3) <u>Estimation</u>

หลังจากที่โมเดลถูก identify แล้ว กระบวนการประมาณค่าพารามิเตอร์จะถูกกระทำต่อไป โดยการใช้ข้อมูลตัวอย่างที่อยู่ในรูป S (sample covariance matrix) และสารสนเทศที่เกี่ยวกับโครงสร้างของโมเดล ซึ่งคือความรู้เกี่ยวกับข้อจำกัดของพารามิเตอร์ที่จะถูกประมาณ เป็นข้อมูลที่จะนำมาใช้หาค่าประมาณ $\hat{\Lambda}$, $\hat{\Phi}$, และ $\hat{\Theta}$ ของพารามิเตอร์ประชากร

ค่าประมาณเหล่านี้ให้ผลในการทำนายความแปรปรวนและความแปรปรวนร่วมประชากรของตัวแปรที่สังเกตได้ตามสมการ $\hat{\Sigma} = \hat{\Lambda}\hat{\Phi}\hat{\Lambda}' + \hat{\Theta}$ ปัญหาของการประมาณคือการหา $\hat{\Lambda}$, $\hat{\Phi}$, และ $\hat{\Theta}$ ที่ทำให้ เมตริคซ์ความแปรปรวนร่วมที่ถูกทำนาย $(\hat{\Sigma})$ เข้าใกล้เท่าที่จะเป็นไปได้กับความแปรปรวนและความแปรปรวนร่วมที่ถูกสังเกตได้ในเมตริคซ์ของค่าสังเกต (S)

ในการประมาณค่ามีชุดของเมตริคซ์ที่เป็นไปได้ทั้งหมด กับ $\hat{\Lambda}$, $\hat{\Phi}$, และ $\hat{\Theta}$ แต่ต้องแยกเอาเมตริคซ์ที่รวมเอาเงื่อนไขหรือข้อตกลงที่กำหนดขึ้นเข้าไว้ ชุดของพารามิเตอร์นี้เรียกว่า Σ^* จะเป็นไปตามสูตร $\Sigma^* = \Lambda^*\Phi^*\Lambda^{*\prime} + \Theta^*$ ถ้า Σ^* เข้าใกล้ S ปัญหาของการประมาณเป็นการวัดว่า Σ^* ใกล้กับ S อย่างไร และ การหาค่าของ Λ^*, Φ^*, และ Θ^* ที่ให้ Σ^* ที่เข้าใกล้เท่าที่จะเป็นไปได้กับ S ทำอย่างไร

ฟังก์ชั่นที่วัดว่า Σ^* ใกล้เท่าที่จะเป็นไปได้กับ S เรียกว่า fitting function ใช้สัญญลักษณ์ว่า $F(S, \Sigma^*)$ หรือ $F(S, \Lambda^*, \Phi^*, \Theta^*)$ ฟังก์ชั่นนี้สามารถใช้ได้ตลอดเมตริคซ์ที่เป็นไปได้ทั้งหมด Λ^*, Φ^* และ Θ^* ที่ satisfy กับเงื่อนไขใน Λ , Φ,และΘ ถ้าชุดที่หนึ่งของเมตริคซ์ "*" ให้เมตริคซ์ Σ^*_1 และชุดที่สองให้เมตริคซ์ Σ^*_2 และถ้า $F(S,\Sigma^*_1)$ น้อยกว่า $F(S,\Sigma^*_2)$ Σ^*_1 จะถูกพิจารณาว่าใกล้ S มากกว่า Σ^*_2 ค่าเหล่านี้ของ Λ^*_1, Φ^*_1 และ Θ^*_1 ที่ทำให้ fitting function น้อยที่สุดสำหรับ S ที่กำหนดเป็นค่าประมาณตัวอย่างของพารามิเตอร์ประชากรและใช้สัญญลักษณ์ว่า $\hat{\Lambda}$, $\hat{\Phi}$ และ $\hat{\Theta}$

มี fitting function 3 function ที่นำมาใช้ใน CFM คือ

ก. <u>Unweighted least squares (ULS)</u>

$$F_{ULS}\ (S, \Sigma^*) = tr\ [\ (S-\Sigma^*)]^2$$

เมื่อ tr เป็นการจัดกระทำบน trace ของเมตริคซ์ที่ชี้ถึงผลรวมของสมาชิกตามเส้นทะแยงมุมของเมตริคซ์ fitting function นี้เป็นการคำนวณ ผลรวมกำลังสองของสมาชิกที่สอดคล้องกันของ S และ Σ^* ค่าประมาณฟังก์ชั่นนี้เกี่ยวกับการทำให้ ผลรวมกำลังสองที่ลดลง สำหรับตัวอย่างใหญ่ ULS จะให้ค่าประมาณที่ไม่ลำเอียง แต่ ULS เป็นค่าประมาณที่มีคุณสมบัติที่เรียกว่า scale dependency นั่นคือผลลัพธ์ที่ได้มาอาจจะแตกต่างจากหน่วยจากการวัดที่แตกต่างกัน ดังนั้นวิธีนี้ควรทำให้ตัวแปรที่สังเกตได้ เป็นมาตรฐานโดยการวิเคราะห์จากเมตริคซ์สหสัมพันธ์

ข. Generalized least square (GLS)

$$F_{GLS}\ (S,\Sigma^{*}) \quad = \quad tr\ [(S-\ \Sigma^{*})\ S^{-1}]^{2}$$

เมื่อ Σ^{*} เข้าใกล้ S ค่า F_{GLS} จะเล็กลง และถ้า S = Σ^{*} ฟังก์ชั่นนี้จะเท่ากับศูนย์ และฟังก์ชั่น F_{GLS} นี้เป็น scale free ต่างกับฟังก์ชั่นแรกในข้อ ก.

ค. Maximum likelihood (ML)

$$F_{ML}\ (S,\ \Sigma^{*}) \ = \ tr\ (\ S\text{-}\Sigma^{*-1}) + [\log\ |\Sigma^{*}| - \log\ |S|\ ] \ - \ q$$

พิจารณา fitting function นี้ในเทอมแรกถ้า Σ^{*} เข้าใกล้ S inverse ของ Σ จะคล้ายกับ inverse ของ S ทำให้ $S\text{-}\Sigma^{*-1}$ เป็นค่าที่ใกล้กันมากกับ identity matrix ทำให้ trace ของเมตริคซ์นี้เข้าใกล้ q (ผลรวมของค่า 1 ตามเส้นทะแยงมุม q ค่า)

ในเทอมที่สองเป็นความแตกต่างของ log ของ determinant ของ S และ Σ^{*} ถ้า S และ Σ^{*} ใกล้กันมาก determinant ของมันก็จะใกล้กันเทอมนี้จะเข้าใกล้ศูนย์

ในเทอมสุดท้ายเป็นค่าคงที่เท่ากับ q นั่นคือสนองตอบต่อค่าในเทอมแรกเมื่อ Σ^{*} ใกล้กับ S เทอมแรกจะเข้าใกล้ q ทำให้ฟังก์ชั่นนี้จะเข้าใกล้ศูนย์ หรือ เท่ากับศูนย์เมื่อ S และ Σ^{*} เป็นเมตริคซ์ที่.ท่ากัน

ถ้า x มีการแจกแจงแบบ multivariate normal distribution ทั้ง GLS และ ML จะมีลักษณะเป็น asymptotic คือจะเป็นค่าประมาณที่ไม่ลำเอียงเมื่อขนาดตัวอย่างมากๆ ซึ่งเป็นปัญหาในการปฏิบัติว่าตัวอย่างที่ใช้ขนาดจะต้องใหญ่เท่าไร จึงจะได้ประโยชน์ของคุณสมบัติ asymptotic ที่ต้องการ Boomsma (1986, cited in Long, 1983) ได้ผลลัพธ์บางประการในการศึกษาถึง two factor model จากค่าตัวแปรที่สังเกตได้ 6 หรือ 8 ตัว ในโมเดลเช่นนี้เขาสรุปว่า จะให้ค่าที่ลำเอียงเมื่อตัวอย่างขนาดเล็กกว่า 100 จึงคาดว่าเมื่อโมเดลซับซ้อนมากคือมีแฟคเตอร์มากขึ้นโมเดลดังกล่าวต้องการตัวอย่างที่ขนาดใหญ่ขึ้น

การพิจารณาการประมาณค่าในทางปฏิบัติควรสังเกตปัญหาเหล่านี้ คือ

1. การที่จะทำให้ fitting function ค่าเล็กลงเท่าที่จะเป็นไปได้ ต้องการการลงทุนในเรื่องเวลา ในการคำนวณมาก เป็นไปไม่ได้ที่จะไม่ใช้คอมพิวเตอร์ในการคำนวณ

2. การที่ค่าของ fitting function ลดลงมากๆ อาจจะทำให้ค่าพารามิเตอร์

ออกนอกช่วง เช่นค่าสหสัมพันธ์มีค่าเกินหนึ่ง หรือได้ค่าความแปรปรวนติดลบ สาเหตุนี้เป็นผลมาจากโมเดลที่ missspecified หรือขนาดตัวอย่างที่ใหญ่อย่าง insufficient

3. การประมาณด้วยโปรแกรมคอมพิวเตอร์ CFM ต้องการค่าเริ่มต้นสำหรับพารามิเตอร์แต่ละตัวที่ถูกประมาณ ซึ่งการเดาค่าเริ่มต้นยิ่งใกล้กับค่าประมาณสุดท้าย การประมาณจะง่ายเข้า ซึ่งผู้วิจัยอาจจะใช้ค่าจากโมเดลที่คล้ายๆกันกับที่เคยถูกวิเคราะห์มาแล้ว อีกวิธีหนึ่งคือผู้วิจัยต้องเดาค่าเริ่มต้นบนพื้นฐานของความรู้ของกระบวนการในการสร้างโมเดล ในปัจจุบัน LISREL (Joreskog, and Sorbum, 1985) ได้รวม algorithm เข้าไว้สำหรับค่าเริ่มแรกที่ทำให้ประหยัดเวลาในการประมาณโมเดลขึ้น

4) Assessment of fit in the CFM

หลังจากที่โมเดลได้ถูกประมาณแล้วกระบวนการต่อไปก็คือการประเมินถึงความ fit ของโมเดลหลังการประมาณ ซึ่งเทคนิควิธีการหลายวิธีที่จะเป็นตัวบ่งชี้ถึงความเหมาะสมของโมเดล วิธีการสามารถทดสอบนัยสำคัญค่าพารามิเตอร์แต่ละตัวได้ โดยการใช้ Z-test และสามารถทดสอบนัยสำคัญสำหรับความเหมาะสมของโมเดลทั้งหมด หรือ การเปรียบเทียบโมเดลซึ่ง nest กันได้ด้วย X^2 - goodness of fit ดัชนีพื้นฐานของ derivatives ของ fitting function สามารถนำมาใช้เพื่อตัดสินถึง better fitting function model ได้เทคนิคต่างๆ ที่แนะนำมาจาก Joreskog, and Sorbum (1981) และ Bentler, and Bonnet (1980) (cited in Long, 1983)

4.1 การตรวจสอบค่าพารามิเตอร์

โปรแกรมในการประมาณค่าส่วนมากของ CFM ไม่มีข้อจำกัดทำให้อาจจะได้ค่าประมาณทางลบในค่าความแปรปรวน หรือ ค่าสหสัมพันธ์ที่มีค่าเกิน 1 ทั้งที่การทดสอบ goodness of fit ให้สารสนเทศว่าโมเดล sufficient ค่าประมาณที่ไม่สมเหตุผลเหล่านี้ชี้ถึงปัญหาใดปัญหาหนึ่งที่อาจจะเกิดขึ้นได้ดังนี้

ประการแรก โมเดลอาจจะ misspecified ในกรณีนี้สามารถเกิดขึ้นได้แม้แต่กรณีที่ความเหมาะสมของโมเดลทั้งหมด satisfy เงื่อนไขที่ sufficient

ประการที่สอง อาจจะมีการละเมิดในคุณสมบัติเกี่ยวกับ normality ของตัวแปรที่สังเกตได้ ซึ่งทำให้ค่าประมาณ ML และ GLS ไม่สมเหตุสมผลแต่ในกรณีของการละเมิด normality ของตัวแปรไม่เป็นสาเหตุให้ค่าประมาณ ULS ไม่สมเหตุผล

ประการที่สาม การใช้ตัวอย่างที่เล็กเกินไปจึงทำให้สมบัติเกี่ยวกับ asymptotic properties ของวิธีการประมาณเป็นไปได้ การใช้ตัวอย่างเล็กทำให้ผลลัพธ์คือค่าประมาณความแปรปรวนเป็นลบ

ประการที่สี่ การที่โมเดลไม่สามารถที่จะพิสูจน์ได้ว่า identify อาจจะเป็นเพราะว่าวิธีการประมาณค่าพารามิเตอร์บางตัวยากและไม่คงที่ เช่นในกรณีที่ไม่สามารถแยกความแตกต่างระหว่างพารามิเตอร์ 2 ตัว หรือมากกว่าได้ ทั้งนี้เกี่ยวกับการมีความสัมพันธ์กันระหว่างค่าประมาณพารามิเตอร์เหล่านี้สูง จึงทำให้เมื่อมีการเปลี่ยนแปลงในพารามิเตอร์หนึ่งมีผลเกี่ยวกับการเปลี่ยนแปลงในพารามิเตอร์อีกค่าหนึ่ง ซึ่งทำให้เกิดปัญหาของ underidentification

ประการที่ห้า การที่เมตริคซ์ความแปรปรวนร่วมคำนวณโดย pair wise deletion ในกรณีของmissing value การเกิด missing data ทำให้เกิดปัญหาว่า ค่าในเมตริคซ์ความแปรปรวนร่วม หรือเมตริคซ์สหสัมพันธ์ที่เกิดจากแต่ละคู่ของตัวแปรวางอยู่บนพื้นฐานของตัวแปรที่แตกต่างกัน จึงนำไปสู่ เมตริคซ์ความแปรปรวนร่วมที่ไม่เหมาะสมที่จะนำมาใช้ในการประมาณค่า ในกรณีเช่นนี้โปรแกรมสำหรับ ML หรือ GLS estimation จะตรวจสอบความไม่เหมาะสมของเมตริคซ์ก่อนและจะไม่คำนวณให้ แต่ในกรณีที่เกิดความผิดปกติเล็กน้อยเมตริคซ์จะถูกวิเคราะห์ต่อไป แต่จะทำให้ค่าประมาณ ULS คลาดเคลื่อน โดยทั่วไป pairwise covariance matrix จะนำมาใช้ได้ควรมีจำนวนค่าสังเกตที่หายไปเพียงเล็กน้อย และกระจายอยู่ทั่วไปไม่ใช่ ณ ตัวแปรใดตัวแปรหนึ่ง หรือ case ใด case หนึ่ง

4.2 ความแปรปรวนและความแปรปรวนร่วมของค่าประมาณ

มีความเป็นไปได้ที่จะประมาณความแปรปรวนของค่าพารามิเตอร์แต่ละตัวซึ่งสามารถนำมาใช้ทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์แต่ละตัวได้

เมื่อกำหนดให้ w เป็นพารามิเตอร์ใดๆ ที่ถูกประมาณจากโมเดล

$\hat{w}$ เป็นค่าประมาณของ w

$\hat{\sigma}$ เป็นค่าประมาณส่วนเบี่ยงเบนมาตรฐานของการแจกแจงสุ่ม w

ภายใต้ข้อตกลงเกี่ยวกับค่าประมาณ ML และ GLS ในตัวอย่างที่ใหญ่ $\hat{w}$ จะแจกแจงเข้าใกล้ปกติด้วยความคลาดเคลื่อนมาตรฐานที่ถูกประมาณโดย $\hat{\sigma}$ เราสามารถทดสอบสมมติฐานว่า H_o : $w = w^*$ เมื่อ w^* เป็นค่า fix ให้คงที่ทั่วไปเท่ากับศูนย์เพื่อที่จะทดสอบสมมติฐานนี้สถิติทดสอบ $Z = (\hat{w} - w^*) / \hat{\sigma}$ Z เป็นค่าประมาณที่แจกแจงใกล้ปกติมี mean=0 และ variance =1

4.3 chi-square goodness of fit tests

ภายใต้ข้อตกลงในการตัดสิน ML และ GLS การใช้ X^2 - goodness of fit สามารถคำนวณได้ การใช้สถิติสามารถใช้ทดสอบสมมติฐานศูนย์ (H_0) ว่าโมเดลที่กำหนดได้ให้ความเหมาะสมที่ยอมรับได้ของข้อมูลที่สังเกตได้ ความเหมาะสมของโมเดลนี้จะประเมินจาก การเปรียบเทียบเมตริคซ์ความแปรปรวนร่วมที่สังเกตได้ (S) ด้วยเมตริคซ์ความแปรปรวนร่วมที่ถูกประมาณโดยสมการ $\Sigma = \hat{\Lambda}\hat{\Phi}\hat{\Lambda}' + \hat{\Theta}$ $\hat{\Sigma}$ ไม่ได้ให้ S อย่างสมบูรณ์เพราะค่าที่อยู่ใน $\hat{\Sigma}$ เกิดจากค่าที่เกิดขึ้นด้วยการจำกัดเงื่อนไข และ ข้อตกลงเบื้องต้นบนพารามิเตอร์ของโมเดล

การทดสอบ X^2 - goodness of fit เป็นการเปรียบเทียบ imperfect fit ภายใต้สมมติฐาน H_0 กับ perfect fit ภายใต้สมมติฐาน H_1 ซึ่ง Σ เป็นเมตริคซ์ความแปรปรวนร่วมใดๆ ยิ่งความแตกต่างระหว่าง S และ Σ มากขึ้น ค่า X^2 จะใหญ่ขึ้น

ค่า degree of freedom ของสถิติ X^2 ใน CFM จะคำนวณได้โดยใช้จำนวน independent parameter ภายใต้ H_1 ลบด้วยจำนวน independent parameter ภายใต้ H_0 จำนวน independent parameter ภายใต้ H_1 คำนวณได้ง่ายเพราะ H_1 ให้ perfect fit ของข้อมูลนั่นคือพารามิเตอร์อิสระ 1 ตัว จะสำหรับสมาชิกที่เป็นอิสระแต่ละตัวใน Σ ซึ่งกำหนดให้พารามิเตอร์อิสระ $q(q+1)/2$ ตัว เมื่อ q คือจำนวนตัวแปรที่สังเกตได้ ส่วนจำนวนพารามิเตอร์อิสระที่เกี่ยวข้องกับ H_0 จะแตกต่างไปในโมเดลแต่ละอัน

ในการทดสอบสมมติฐานจะดำเนินการโดยหา critical values ที่ระดับนัยสำคัญ α ด้วยการแจกแจง X^2 และชั้นของความเป็นอิสระ df ค่าของ X^2 ที่ใหญ่กว่าค่าของ critical values ทำให้ปฏิเสธสมมติฐาน H_0 และให้ข้อสรุปว่าโมเดลที่เสนอไม่ได้ generate ข้อมูลที่สังเกตได้ แต่ถ้า X^2 เล็กกว่าค่า critical values ให้ผลในการยอมรับสะมติฐาน H_0 และสรุปว่าโมเดลที่เสนอได้ generate ข้อมูลที่สังเกตมาได้

การใช้ X^2 ในการทดสอบความเหมาะสมของโมเดลนี้ต้องตั้งอยู่บนข้อตกลงเบื้องต้นที่ว่า ตัวแปรที่สังเกตได้มีการแจกแจงปกติ (multivariate normal distribution), การวิเคราะห์ควรวางอยู่บนพื้นฐานของเมตริคซ์ความแปรปรวนร่วมของตัวอย่างมากกว่าเมตริคซ์สหสัมพันธ์ของตัวอย่าง และ ขนาดของตัวอย่างมีขนาดใหญ่พอที่จะใช้สมบัติ asymtotic ของสถิติทดสอบ อย่างน้อยที่สุดหนึ่งในข้อตกลงนี้อาจถูกละเมิดได้ ในการนำไปใช้สำหรับโมเดล CFM Joreskog and sorbum (1985)ı แนะนำว่าการทดสอบ X^2 เป็นดัชนีที่ชี้ว่าโมเดลได้ให้ reproduce เมตริคซ์ คืออย่างไรเมื่อเทียบกับความแปรปรวนร่วมที่สังเกตได้ (S) มากกว่าการ

ทดสอบสมมติฐานอย่างเป็นทางการ ค่าที่ใหญ่ของ X^2 จะบอกถึงการ reproduce ที่ไม่ดีของ S และค่า X^2 ที่เล็กจะบอกถึงการ reproduce ที่ดี

4.4 Nested models and difference of Chi-square tests

ในหลายกรณีมีประโยชน์ที่จะเปรียบเทียบสมมติฐานที่ถูกอธิบายโดยโมเดลที่กำหนดกับสมมติฐานที่ถูกอธิบายโดยโมเดลที่เอามาเปรียบเทียบกัน การทดสอบเช่นนี้เป็นไปได้ เมื่อโมเดล 2 โมเดลนี้เกี่ยวพันกัน (nest) ยกตัวอย่างโมเดลที่ M_1 nest ใน M_2 ถ้า M_1 ได้มาจาก M_2 ที่ถูก fix ในพารามิเตอร์บางตัวให้เท่ากัน ดังนั้น M_1 จึงสามารถเรียกได้ว่าเป็นกรณีพิเศษคือ nest ใน M_2

โมเดลที่ nest กัน สามารถที่จะนำมาเปรียบเทียบกันได้ในเชิงสถิติ ถ้า M_1 nest ใน M_2 ความแตกต่างของการทดสอบ X^2 สามารถนำมาเปรียบเทียบ M_1 กับ M_2 ซึ่งเป็นโมเดลที่อยู่ในรูปทั่วไปมากกว่าจะ fit เช่นเดียวกัน หรือ ดีกว่าโมเดลที่ถูกวางเงื่อนไขมากกว่า เพราะอย่างน้อยพารามิเตอร์ที่ไม่ได้ถูกวางเงื่อนไข ที่เพิ่มเข้าไปหนึ่งตัวเป็นเครื่องช่วยในการสร้าง (reproducing) เมตริคซ์ความแปรปรวนร่วมที่สังเกตได้ เมตริคซ์ความแปรปรวนร่วมที่ถูกประมาณจากโมเดลดังกล่าว จะเข้าใกล้ S มากกว่าเมตริคซ์ความแปรปรวนที่ถูกประมาณจากโมเดลที่ถูกวางเงื่อนไขมากกว่า ดังนั้น X_1^2 ด้วย df_1 ที่ได้มาจากการทดสอบโมเดล M_1 จะใหญ่กว่า X_2^2 ด้วย df_2 ที่ได้มาจากการทดสอบ M_2 การปรับปรุงความเหมาะสมที่ได้มาโดยการเพิ่มพารามิเตอร์เข้าไปใน M_1 จะมีนัยสำคัญทางสถิติหรือไม่ถูกกำหนดได้จากความแตกต่างของการทดสอบ X^2 ถ้าตัวอย่างขนาดใหญ่ $X^2 = X^2 - X^2$ จะมีการแจกแจงแบบ Chi-square ด้วย $df = df_1 - df_2$ ถ้า M_1 เป็น true model ถ้า X^2 มากกว่าค่าวิกฤติที่เลือกขึ้นมาสำหรับการแจกแจง X^2 ด้วยชั้นแห่งความเป็นอิสระ df สมมติฐานที่ว่าเงื่อนไขที่กำหนดว่า M_2 มาจาก M_1 อย่าง valid จะถูกปฏิเสธนั่นคือการลดหย่อนเงื่อนไขลงจะให้ผลการปรับปรุงอย่างมีนัยสำคัญทางสถิติ ในความเหมาะสม (fit) ของ M_2 เหนือกว่า M_1

ยกตัวอย่างโมเดลที่ nest กันอีกลักษณะหนึ่งใน MTMM Model ที่ทุก trait factor และทุก method factor สัมพันธ์กัน เว้นค่า 1 บนเส้นทะแยงมุมของเมตริคซ์ Φ จะไม่ถูกวางเงื่อนไข เรียก MTMM Model นี้ว่า M_A แต่ด้วยเหตุผลของการนำไปใช้บางประการที่มีเหตุผลจะสมมติเอาได้ว่า trait factor เป็นอิสระจาก method factor ซึ่งสามารถเขียนอธิบายได้ดังนี้

$$\Phi = \left[\begin{array}{ccc|ccc} 1 & \Phi_{12} & \Phi_{13} & 0 & 0 & 0 \\ \Phi_{21} & 1 & \Phi_{23} & 0 & 0 & 0 \\ \Phi_{31} & \Phi_{32} & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & \Phi_{45} & \Phi_{46} \\ 0 & 0 & 0 & \Phi_{54} & 1 & \Phi_{56} \\ 0 & 0 & 0 & \Phi_{65} & \Phi_{65} & 1 \end{array}\right] = \left[\begin{array}{c|c} \text{trait/trait} & \text{trait/method} \\ \hline \text{method/trait} & \text{method/method} \end{array}\right]$$

จากเมตริคซ์แสดงความสัมพันธ์ของตัวประกอบที่กำหนดให้ตัวประกอบลักษณะและตัวประกอบวิธีไม่สัมพันธ์กันข้างต้น จะเรียก MTMM model นี้ว่า M_B และเรียก โมเดล MTMM ซึ่งไม่ยอมให้ทุก แฟคเตอร์สัมพันธ์กันว่า M_C นั่นคือ $\Phi = I$ เป็น identity matrix

M_B สามารถสร้างรูปแบบขึ้นมาได้จาก M_A โดยการวางเงื่อนไขสหสัมพันธ์ในกลุ่ม methcd/trait และ trait/method ของ Φ ให้เท่ากับศูนย์ M_B จึง nest ใน M_A ส่วน M_C nest ใน M_A และ M_C ก็ nest ใน M_B ด้วยเพราะมันสามารถสร้างรูปแบบมาจาก M_B โดยการวางเงื่อนไขว่าสมาชิกนอกเส้นทะแยงมุมที่เหลืออยู่ของ Φ เท่ากับศูนย์ ในกรณีนี้เราสามารถที่จะทดสอบความแตกต่างของการทดสอบ Chi-square ใน MTMM MOdel ได้โดยให้ X^2_A , X^2_B และ X^2_C เป็นสถิติ Chi-square ที่เกี่ยวกับโมเดล M_A M_B และ M_C ซึ่งมีชั้นแห่งความเป็นอิสระ df_A df_B และ df_C ตามลำดับสำหรับ M_A เมตริคซ์ Φ ไม่ถูกวางเงื่อนไขยกเว้นแต่ค่าตามเส้นทะแยงมุมที่เป็น 1 และใน M_B กลุ่ม method/trait ของ Φ จะถูกวางเงื่อนไขให้เท่ากับศูนย์ และใน M_C เมตริคซ์ Φ ถูกวางเงื่อนไขให้เท่ากับ I สมมติฐานที่ว่าแฟคเตอร์ทุกแฟคเตอร์ไม่สัมพันธ์กันสามารถทดสอบได้โดยการเปรียบเทียบ M_A กับ M_C สถิติทดสอบคือ $X^2 = X^2_C - X^2_A$ ด้วย $df = df_C - df_A$ ถ้าค่า X^2 มากกว่าค่าวิกฤติสมมติฐานที่ว่า แฟคเตอร์ไม่สัมพันธ์กันได้รับการปฏิเสธ สมมติฐานที่ว่า สหสัมพันธ์ระหว่าง method factor และ trait factor เป็นศูนย์ สามารถทดสอบได้โดยการเปรียบเทียบ M_A และ M_B สถิติทดสอบที่เหมาะสมคือ $X^2 = X^2_B - X^2_A$ ด้วยชั้นแห่งความเป็นอิสระ $df_B - df_A$

4.5 Specification searches

ถ้าโมเดลที่ถูกกำหนดเป็นสมมติฐานไม่ fit บ่อยครั้งที่เป็นความสนใจของผู้วิจัยที่จะหาโมเดลที่เหมาะสมต่อไป นั่นคือ ถ้าโมเดลที่ถูกแนะนำในทางทฤษฎีได้รับการปฏิเสธไปอาจจะมีข้อแนะนำในเชิงทฤษฎีบางประเด็นที่กล่าวถึงวิธีการในการปรับปรุงความ fit ของโมเดล

ให้ดีขึ้น ทำให้ผลลัพธ์ที่ได้มาจากการประมาณโมเดลที่ได้ถูกปฏิเสธไปแล้วจะได้นำเอามาใช้เป็นข้อแนะนำเพิ่มเติมในการทำให้โมเดล fit ดีขึ้น กระบวนการนี้เรียกว่า Specification Search ซึ่งเป็นกระบวนการที่ช่วยให้ผู้วิจัยค้นหาโมเดลที่เหมาะสมกับข้อมูลมากขึ้น

5. การนำ CFM ไปใช้ในการวิเคราะห์

CFM เป็นโมเดลทางสถิติที่มีอำนาจโมเดลหนึ่ง ความเหมาะสมของโมเดลนี้ใช้การทดสอบโครงสร้างที่กำหนด (specific structure) ที่ได้รับการแนะนำโดยทฤษฎีที่สำคัญ (Substantive Theory) ทำให้ CFM เป็นโมเดลที่ได้เปรียบกว่า Exploratory factor model และสำหรับการวิเคราะห์ MTMM Model Schmitt, and Stults (1986) ได้ทบทวนวิธีการที่เกี่ยวกับการวิเคราะห์ MTMM Matrix พบว่า CFM เป็นโมเดลที่นำมาใช้มากที่สุดโมเดลหนึ่งและมีความเหมาะสมในการตอบปัญหาที่สามารถประเมินได้ในแต่ละ trait/method unit, trait-method interaction, method intercorrelation, และ trait & method factor ซึ่งสามารถทำให้ทดสอบคุณลักษณะต่างๆที่เกี่ยวกับ convergent validity และ discriminant validity Widaman (1985) ได้เสนอเกี่ยวกับการประเมิน convergent และ discriminant validity ใน MTMM Matrix โดยเขาเสนอว่า เมื่อได้ identify ถึง model ที่ fit ที่สุดในครั้งแรกแล้ว ขั้นตอนต่อไป ก็คือการทำการเปรียบเทียบโมเดลอื่นๆ ที่ nest ภายใน โมเดลที่ fit ที่สุดนี้ จากการสร้างชุดของ Structure model ที่เป็นระบบที่ nest กัน Widaman ได้พิจารณาถึงความสัมพันธ์ระหว่างตัวประกอบลักษณะ และ ตัวประกอบวิธี ที่สนใจแยกกัน จากตารางที่ 4 พบว่าความสัมพันธ์ระหว่างตัวประกอบลักษณะสามารถวัดด้วยพารามิเตอร์ 28, 32, และ 33 ในกรณีนี้มี structure ที่เป็นไปได้ 3 ทางคือ (1) ไม่มี trait factor (หรือไม่มี method factor) (2) trait factor (หรือ method factor) ถูก fix ไม่เป็นศูนย์ก็เป็นหนึ่ง ค่าศูนย์ชี้ถึงระดับที่สูงของ discriminant validity และ ค่าหนึ่งชี้ถึงการขาดซึ่ง discriminant validity อย่างสมบูรณ์ และ (3) trait factor (หรือ method factor) กำหนดให้ค่าสหสัมพันธ์ถูกประมาณอย่างอิสระกัน ข้อเสนอของ Widaman ได้เสนอถึงวิธีการที่เป็นระบบให้กับ confirmatory analysis ในเมตริคซ์ MTMM ในโมเดลที่ 3 ของเขาเป็นโมเดลที่มีข้อจำกัดน้อยที่สุด ที่ได้รับการ identify และการรวมเอาพารามิเตอร์ทุกตัว ดังภาพที่ให้ไว้ ในตารางที่ 4 ข้างต้น ยกเว้นแต่พารามิเตอร์ 22-24, 25-27, และ 29-31 ซึ่งแสดงถึง

trait method intercorrelation ซึ่งโมเดลต่อมาจะถูก fix ให้เท่ากับศูนย์ เงื่อนไขข้อจำกัดต่อไปนี้จะใช้ทดสอบ convergent validity , discriminant validity และ method bias สำหรับเมตริคซ์ทั้งหมดได้คือ

1) การเปรียบเทียบโมเดลเต็ม (full model) กับโมเดลที่มีเพียงตัวประกอบวิธีที่สัมพันธ์กัน จะให้การทดสอบ convergent validity ความแตกต่างในการทดสอบ X^2 ของความเหมาะสมเกี่ยวกับโมเดลทั้ง 2 จะมีการแจกแจงแบบ X^2 ที่มีชั้นแห่งความเป็นอิสระเท่ากับความแตกต่างของชั้นแห่งความเป็นอิสระที่เกี่ยวกับโมเดลทั้ง 2 นี้ จากโมเดลที่บรรยายในตารางที่ 4 ตัวประกอบลักษณะ 3 ตัว จะถูกตัดทิ้งไป จะไม่มีการประมาณสหสัมพันธ์ระหว่าง trait (ซึ่งมีพารามิเตอร์ 3 ตัวอยู่) ในกรณีนี้การทดสอบ convergent validity จะมีชั้นแห่งความเป็นอิสระเท่ากับ 12

2) การเปรียบเทียบโมเดลเต็มกับโมเดลที่มีสหสัมพันธ์ระหว่าง trait factor อย่างสมบูรณ์ได้ให้การทดสอบ discriminant validity สำหรับเมตริคซ์ทั้งหมด การเปรียบเทียบนี้จะมีชั้นแห่งความเป็นอิสระ 3 ในเมตริคซ์ที่แสดงในตารางที่ 4 ซึ่งพารามิเตอร์ตัวที่ 19-21 จะถูกให้เท่ากับ 1.0

3) การทดสอบ method bias ในภาพรวมทั้งหมดจะได้มาจากการเปรียบเทียบโมเดลเต็มๆ กับโมเดลที่ไม่มีตัวประกอบวิธี เช่นเดียวกับการทดสอบ convergent validity การเปรียบเทียบนี้จะมีชั้นแห่งความเป็นอิสระเท่ากับ 12 ซึ่งตัวประกอบวิธีและน้ำหนักของตัวประกอบวิธีนั้นและสหสัมพันธ์ระหว่างตัวประกอบวิธีจะถูกตัดออก

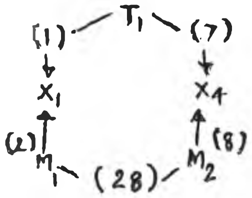
Widaman (1985) ได้กล่าวถึงข้อดีของ CFM ในการแยกส่วนของความแปรปรวนของตัวแปรที่วัดได้แต่ละตัวออก และ การแยกส่วนของสหสัมพันธ์แต่ละตัวใน MTMM Matrix ข้อดีเหล่านี้ได้รับความสนใจในการประเมินตัวแปรที่ถูกวัดได้เฉพาะ สมมติว่าผู้วิจัยสนใจในการประเมินในระดับที่แสดงถึงความแตกต่างแต่ละตัวในตัวแปรที่ถูกวัดได้ตัวแรก (x_1) ในตารางที่ 4 ที่เกี่ยวกับ ความแปรปรวนลักษณะ ความแปรปรวนวิธี และความแปรปรวนคลาดเคลื่อน จะถูกประมาณโดยการยกกำลังสองของพารามิเตอร์ตัวแรกในตารางที่ 4 (ซึ่งเป็น factor loading ของ X_1 บน M_1) method variance จะถูกประมาณด้วยการยกกำลังสองของพารามิเตอร์ตัวที่สอง (factor loading ของ X_1 บน M_1) ส่วน unique variance เกี่ยวกับ X_1

จะถูกประมาณจากพารามิเตอร์ตัวที่ 34 (โดยที่ไม่มีสหสัมพันธ์ระหว่าง trait-method คือ พารามิเตอร์ตัวที่ 22) การแยกส่วนของความแปรปรวนในแต่ละตัวแปรจะทำให้ผู้วิจัยได้ข้อสรุปเกี่ยวกับการวัดแต่ละวิธีใน MTMM Matrix

สหสัมพันธ์ระหว่างตัวแปรที่วัดได้ สามารถถูกแยกส่วนออกเพื่อให้การประเมินถึงระดับที่ซึ่งสหสัมพันธ์เป็นผลมาจาก ตัวประกอบร่วม หรือ ตัวประกอบวิธี Schmitt (1978) ได้ให้ตัวอย่างการแยกส่วนของ MTMM correlation, HTHM correlation, และ HTHM correlation และตัวอย่างของการแยกส่วนสำหรับค่าสหสัมพันธ์ 3 รูปแบบใน MTMM Matrix สหสัมพันธ์ระหว่างตัวแปรที่วัดได้ เป็นผลรวมของผลคูณของพารามิเตอร์ ซึ่งเกี่ยวกับเส้นทางทางทฤษฎี (hypothetical paths) ที่เชื่อมระหว่างตัวแปร 2 ตัว ดังนั้นสำหรับ HTHM correlation ค่าสหสัมพันธ์ระหว่าง trait ที่เกี่ยวกับ common trait variance คือผลคูณของ factor loading ในตัวแปรที่วัดได้แต่ละตัวบน method factor ที่สนใจกับสหสัมพันธ์ระหว่าง method factor 2 ตัว ผลรวมของ common trait และ method correlationคือค่าของ HTHM correlation ที่ถูกแนะนำโดยโมเดลที่แสดงด้วย HTMM Matrix

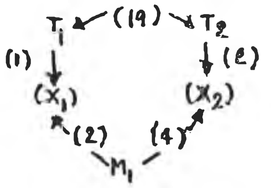
ภาพที่ 2 ตัวอย่างการแยกส่วนของ monotrait-heteromethod correlation, hetero trait-monomethod correlation, และ heterotrait-heteromethod correlation

(A) Monotrait-Heteromethod Decomposition



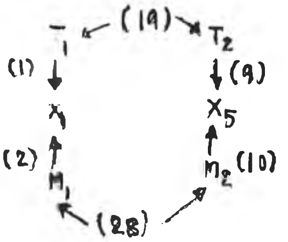
$r_x r_x$ = (1 X 7) common trait variance + (2 X 8 X 28) common method variance

(B) Heterotrait-Monomethod Decomposition



$r_x r_x$ = (1 X 19 X 3) common trait variance + (2 x 4) common method variance

(C) Hetertrait-Heteromethod Decomposition

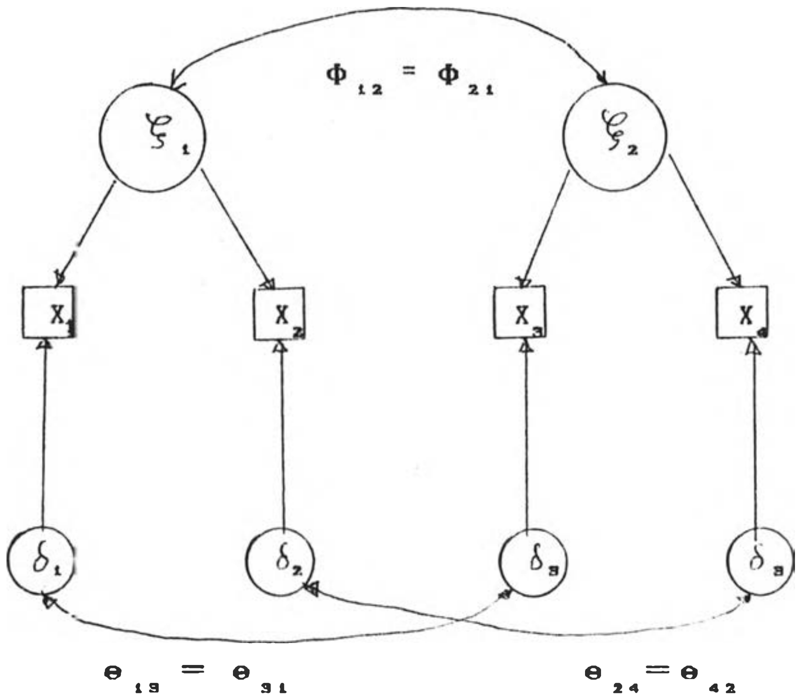


$r_x r_x$ = (1 x 19 x 9) common trait variance + (2 x 28 x 10) common method variance

ในการแยกส่วนตัวแปรที่ประมาณได้จากโมเดลนอกจากสามารถแยกส่วนเพื่อที่จะแสดงความสัมพันธ์ระหว่างตัวแปรดังกล่าวแล้ว การประมาณถึงความเที่ยงของตัวแปรที่สังเกตได้แต่ละตัวสามารถประมาณได้ ความเที่ยงได้รับการนิยามว่าเป็นสหสัมพันธ์กำลังสองระหว่างตัวแปรแฝงและ indicator ที่ถูกสังเกตได้ของมัน ในกรณีเช่นนี้ความเที่ยงได้ชี้ถึงเปอร์เซนต์ของความแปรผันในตัวแปรที่ถูกสังเกตได้ที่ถูกอธิบายโดย common factor ที่มันตั้งใจจะวัด ตัวอย่างที่แสดงถึงการคำนวณในค่าความเที่ยงดังภาพที่ 3 ในโมเดลการวัดความเสื่อมทางจิต เมื่อพิจารณาตัวแปรที่สังเกตได้ X_2 ซึ่งเป็นตัวแปรที่กำหนดโดยตัวแปรแฝง ξ_1 จากสมการตัวประกอบสำหรับ X_2 จะได้ $X_2 = \lambda_{21}\xi_1 + \delta_2$ เพราะตัวแปรที่วัดได้เป็นตัวแปรที่เบี่ยงเบนจากค่า mean ของมัน จึงทำให้ $COV(\xi_1, X_2) = E[\xi_1, X_2] = E[\xi_1(\lambda_{21}\xi_1 + \delta_2)]$ และจากภาพที่ 3 ξ_1 และ δ_2 ได้สมมติให้ไม่สัมพันธ์กันจึงทำให้ $COV(\xi_1, X_2) = \lambda_{21}\phi_{11}$ เนื่องจากสหสัมพันธ์ระหว่าง X_2 และ ξ_1 เป็นความแปรปรวนที่หารด้วยส่วนเบี่ยงเบนมาตรฐานของ X_2 และ ξ_1 ดังนั้นสหสัมพันธ์กำลังสองหรือค่าความเที่ยงจะนิยามได้ดังนี้

$$REL(\xi_1, X_2) = (\lambda_{21}, \Phi_{11})^2 / (\sigma_{22}\Phi_{11})$$
$$= \lambda^2_{21}\Phi_{11} / \sigma_{22}$$

ภาพที่ 3 โมเดลสำหรับการวัดความเสื่อมทางจิต (psychological disorders)



Schriesheim, Solomon, and Kopelman (1989) ได้นำวิธีการ CFA ไปใช้วิเคราะห์เมตริคซ์ลักษณะพหุ-วิธีพหุโดยเขาต้องการตรวจสอบว่ารูปแบบของการจัดข้อคำถามในแบบสอบถาม 2 รูปแบบ คือรูปแบบการจัดกลุ่มข้อคำถามเป็นกลุ่ม โดยชื่อกำหนดลักษณะที่วัด กับรูปแบบการจัดข้อคำถามแบบสุ่ม ว่า 2 รูปแบบนี้จะมีผลต่อความตรงแบบลู่เข้า และความตรงแบบแยกออก และความเที่ยงของการวัดที่เอามาใช้หรือไม่ เขาใช้แบบสอบถาม 2 ชุด ชุดแรกวัดความพึงพอใจ (satisfaction scales) และแบบสอบถามชุดที่ 2 วัดลักษณะของงาน (job Characteristics scale) สำหรับมาตรวัดความพึงพอใจมีคุณลักษณะที่วัด 4 ชนิดคือ ความพอใจในชีวิต (life satisfaction) วัดด้วยข้อคำถาม 2 ข้อ ความพอใจในตนเอง (self satisfaction) วัดด้วยข้อคำถาม 2 ข้อ ความพอใจในงาน (job satisfaction) วัดด้วยข้อคำถาม 2 ข้อ ความพอใจในครอบครัว (family satisfaction) วัดด้วยข้อคำถาม 3 ข้อ คำตอบเป็นแบบลิคเคอร์ต 5 ช่วง ตั้งแต่เห็นด้วยมากที่สุดจนถึงไม่เห็นด้วยมากที่สุด สำหรับมาตรวัดคุณลักษณะของงาน ประกอบด้วยมิติของงาน 5 แบบ คือ Autonomy, Variety, Significance, Identity, และ Feedback อย่างละ 3 ข้อ ในมาตรวัดทั้ง 2 นักศึกษาจะได้รับการสุ่มให้ทำแบบสอบถามเพื่อโอกาสที่แต่ละชุดของแบบสอบถามได้มีอันดับเท่ากันในการกำหนดให้นักศึกษาได้ตอบ

การวิเคราะห์ข้อมูลได้คำนวณหาค่าความเที่ยงแบบความสอดคล้องภายในของมาตรวัดทุกชุด พบว่าความเที่ยงของมาตรที่จัดกลุ่มข้อคำถามและจัดข้อคำถามแบบสุ่มไม่แตกต่างกัน ทั้งมาตรวัดความพอใจและมาตรวัดลักษณะของงาน วิเคราะห์ความตรงแบบลู่เข้า และ แบบแยกออกจากตารางเมตริคซ์ลักษณะพหุ - วิธีพหุ โดยใช้ LISREL maximum likelihood confirmatory factor analysis พบว่า สำหรับมาตรวัดความพอใจในงานชี้ถึงความเหมาะสมของโมเดลเต็มที่แสดงถึงความตรงแบบลู่เข้า และ ความตรงแบบแยกออก ที่มีอยู่ของการวัด และผลของ method bias โดยองค์ประกอบของลักษณะที่วัด (trait factor variance) อธิบาย 75 % ของความแปรปรวนทั้งหมด ความแปรปรวนจากตัวประกอบวิธี (method factor variance) อธิบายได้เพียง 12 % และ ความแปรปรวนคลาดเคลื่อน (error factor variance) อธิบายเพียง 14 % ของความแปรปวน ทั้งหมด ในมาตรแบบจัดกลุ่ม มีเปอร์เซนต์เฉลี่ยของความแปรปรวนลักษณะมากกว่ามาตรแบบสุ่มข้อคำถาม (78 % และ 71 % ตามลำดับ) และ ความแปรปรวนของวิธีน้อยกว่า (17 % และ 18 % ตามลำดับ) แต่มีความแปรปรวนคลาดเคลื่อนมากกว่า (16 % และ 12 % ตามลำดับ) โดยทั่วไปในการวิเคราะห์มาตรที่จัดข้อคำถามแบบกลุ่มในการวัดความพอใจจะมีผลจากความแปรปรวนลักษณะสูงกว่า และผลจากความแปรปรวนวิธีน้อยกว่ามาตรที่จัดข้อคำถามแบบสุ่มส่วนมาตรวัดลักษณะของงาน ผลการ

วิเคราะห์ชี้ว่าในโมเดลเต็มเหมาะสมดีกับข้อมูลซึ่งชี้ถึงความมีอยู่ของความตรงแบบลู่เข้า และความตรงแบบแยกออก และ ความลำเอียงของวิธี ความแปรปรวนของลักษณะ อธิบายความแปรปรวนทั้งหมด 43 % ความแปรปรวนวิธีอธิบาย 35 % และ ความแปรปรวนคลาดเคลื่อนอธิบาย 23 % เมื่อเปรียบเทียบการจัดข้อคำถามแบบกลุ่มกับแบบสุ่มในด้านความแปรปรวนของลักษณะในการจัดข้อแบบกลุ่มจะมากกว่า (51 % & 35 %) และมีความแปรปรวนวิธีพอ ๆ กัน (36 % & 34 %) ในขณะที่มาตรการจัดกลุ่มมีเพียง 16 % ของความแปรปรวนคลาดเคลื่อนและมาตรจัดข้อคำถามแบบสุ่มมี 30 % ซึ่งชี้ว่ามาตรวัดคุณลักษณะของงานที่จัดกลุ่มจะดีกว่า

ผลการวิจัยนี้แสดงว่าการจัดกลุ่มข้อคำถามไม่ส่งเสริมความเที่ยงแบบความสอดคล้องภายในที่เพิ่มขึ้นเมื่อเทียบกับข้อที่จัดข้อคำถามแบบสุ่ม ส่วนการวิเคราะห์ด้วย confirmatory factor analysis แนะนำว่าการจัดกลุ่มจะให้ผลในด้านความตรงแบบลู่เข้า แต่ไม่ช่วยให้เกิดผลในความตรงแบบแยกออก และให้ผลในความลำเอียงของวิธี และมีหลักฐานที่ปรากฏว่ามาตรการจัดกลุ่มจะช่วยปรับปรุงคุณสมบัติทางจิตที่วัดด้วยวิธีที่ยังอ่อน (weak measurement) แต่อย่างไรก็ตามควรได้มีการศึกษาต่อไปโดยเฉพาะการจัดรูปแบบของข้อคำถาม ในปริเขตของการวัดด้านอื่นๆต่อไป งานวิจัยนี้สนับสนุน Schmitt, and Stults (1986); Widaman (1985) และคนอื่นๆว่า confirmatory factor analysis เป็นเทคนิคที่ไว (sensitive) ในการวิเคราะห์สำหรับรูปแบบการวิจัยเช่นนี้ และแนะนำการวิจัยต่อไปที่จะใช้การวิเคราะห์ในลักษณะเช่นนี้

Marsh, and Hocever (1983) ได้ศึกษาถึงการวิเคราะห์เมตริคซ์ลักษณะพหุ-วิธีพหุ โดยการวิเคราะห์ด้วย ANOVA MODEL เปรียบเทียบกับ confirmatory factor analysis model เขาได้สรุปถึงความเหนือกว่าของการใช้ CFA ในการวิเคราะห์เมตริคซ์ลักษณะพหุ-วิธีพหุ ดังนี้

1) ให้การทดสอบที่อ้างอิงถึงตัวแปรแฝงที่สนใจมากกว่าตัวแปรที่สังเกตได้

2) แยกความแตกต่างของความแปรปรวนที่เกี่ยวกับลักษณะ และวิธี ออกจากกัน

3) ยอมให้มีการเปรียบเทียบโมเดลอื่นๆกับโมเดลพื้นฐานที่ nest กัน และการทดสอบความเหมาะสมของโมเดลทั้งหมดกับโมเดลที่เสนอแต่ละคู่

4) ให้การทดสอบทางสถิติแยกกันในค่าพารามิเตอร์ที่ประมาณได้แต่ละตัวกับสมมติฐานศูนย์ที่มีค่าสัมประสิทธิ์เป็นศูนย์

5) ให้ค่าสถิติสรุปของปริมาณของความแปรปรวนที่เกี่ยวกับลักษณะและวิธีในการวัดแต่ละอันที่แยกกัน ในแต่ละชุดของการวัดและทั้งหมดของข้อมูล

6) ยอมให้มีการแยกส่วนของค่าสหสัมพันธ์ที่หามาได้จากค่าประมาณพารามิเตอร์แต่ละตัวที่แสดงถึงองค์ประกอบที่ให้แสดงลักษณะผลของลักษณะและวิธี

7) ให้ค่าประมาณของ uniqueness ในการวัดแต่ละอันที่ได้มาจากการวิเคราะห์

8) ทำให้มีการทดสอบเชิงประจักษ์ถึงความมีอยู่ในสหสัมพันธ์ระหว่างลักษณะ และระหว่างวิธี

9) ทำให้มีการทดสอบเชิงประจักษ์ของจำนวนของตัวประกอบลักษณะและตัวประกอบวิธีที่ให้ความเหมาะสมที่สุดกับข้อมูล

ด้วยประโยชน์และข้อดีของ CFA ในการวิเคราะห์ MTMM Matrix ซึ่งเป็นวิธีทดสอบความตรงเชิงทฤษฎีในการตีความหมายของคะแนน (Marsh, 1990) จึงทำให้ผู้วิจัยได้เลือกวิธี CFA ในการตรวจสอบความตรงเชิงทฤษฎีของคะแนนที่วัด ซึ่งเป็นประเด็นคำถามที่มุ่งแสวงหาคำตอบสำหรับการตรวจสอบคุณภาพของคะแนนที่ได้มาจากสูตรที่พัฒนาขึ้นจากการวิจัยนี้ ซึ่งจะได้กล่าวถึงวิธีการพัฒนาสูตรการให้คะแนน และ การวิเคราะห์ผลคุณภาพของคะแนนจากสูตรที่ให้กับผลการตอบข้อสอบด้วยข้อมูลจริงในบทต่อไป.