บทที่ 2

เอกสารและงายวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียงแตกต่างกัน ดังนั้นเอกสารและงานวิจัยที่เกี่ยวข้องกับการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาในเนื้อหาต่าง ๆ ดังต่อไปนี้

1. ความหมายของความลำเอียงของข้อสอบ
2. แหล่ง/สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ
3. ความเป็นมาของการศึกษาความลำเอียงของข้อสอบ
4. วิธีการตรวจสอบความลำเอียงของข้อสอบทั้ง 3 วิธี คือ
   4.1 การวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีแปลงค่าความยากของข้อสอบ
   4.2 การวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีของ แมนเทล-แฮนส์เซล
   4.3 การวิเคราะห์ความลำเอียงของข้อสอบโดยยึดทฤษฎีการตอบข้อคำถาม
5. งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ
   5.1 งานวิจัยภายในประเทศ
   5.2 งานวิจัยต่างประเทศ

ความหมายของความลำเอียงของข้อสอบ

ความลำเอียงของข้อสอบได้มีผู้ให้ความหมายไว้ต่าง ๆ ดังนี้ คือ

เคลียรีและฮิลตัน (Cleary and Hilton, 1986 : 61) ให้ความหมายของข้อสอบที่มีความลำเอียงว่า เป็นข้อสอบที่เมื่อนำไปทดสอบกับบุคคลกลุ่มใดกลุ่มหนึ่งที่มีความสามารถเท่ากันแล้วปรากฏว่ามีคะแนนเฉลี่ยของข้อสอบข้อเดียวกันของแต่ละกลุ่มไม่เท่ากัน

เอลเลน และ เยน (Allen and Yen, 1979 : 129) ให้ความหมายของข้อสอบที่มีความลำเอียงว่าเป็นข้อสอบที่วัดความสามารถ หรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มได้ไม่ตรงกัน เช่น เมื่อนำข้อสอบวัดความสามารถด้านคณิตศาสตร์ที่ใช้ภาษาสื่อความหมายของโจทย์ปัญหาไปทดสอบกับผู้สอบ 2 กลุ่มที่มีความสามารถทางการอ่านไม่เท่ากัน ข้อสอบนั้นมีความลำเอียงต่อผู้สอบที่มีความสามารถทางการอ่านสูง และข้อสอบที่มีความลำเอียงนี้จะมีโค้งลักษณะของข้อสอบของผู้สอบแต่ละกลุ่มแตกต่างกัน

ลอร์ด (Lord, 1980 : 212) สับโคเวียค และคนอื่น ๆ (Subkoviak and Others, 1984 : 49) แมคลคอลเลอร์ และเมนโดซา (McCanley and Mendoza, 1985 : 389) แฮมเบิลตัน และคณะ (Hambleton et. Al, 1991 : 110) ฮอล์แลนด์ และไวเนอร์ (Holland and Wainer, 1993 : 4, quoting Angoff, 1993) ให้ความหมายของข้อสอบที่มีความลำเอียงว่า เป็นข้อสอบที่ผู้สอบที่มีความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน มีโอกาสขอบข้อสอบข้อนั้นได้ถูกไม่เท่ากัน

ปอบแฮม (Popham, 1981 : 181) ให้ความหมายของความลำเอียงของข้อสอบว่า หมายถึง ความชอบหรือความโน้มเอียงที่ทำให้เกิดการพิจารณาตัดสินเป็นไปอย่างไม่ยุติธรรม

ออสเตอร์ลายด์ (Osterlind, 1983 : 55) ให้ความหมายว่า ความลำเอียงเป็นความคลาดเคลื่อนอย่างมีระบบทางการวัด

คามิลลิ และเชพเพิร์ด (Camilli and Shepard, 1994 : 8) ได้ให้ความหมายของข้อสอบที่มีความลำเอียงว่า เป็นความคลาดเคลื่อนอย่างเป็นระบบหรือความไม่เที่ยงตรงของการวัด ทำให้สมาชิกบางกลุ่มได้ประโยชน์หรือเสียประโยชน์ เช่น แบบทดสอบหรือข้อสอบที่เข้าข้างกลุ่มนักเรียนชายมากกว่านักเรียนหญิง หรือคนผิวขาวกับคนผิวดำหรือระหว่างคนที่อยู่ในภูมิภาคต่างกัน

นอกจากนี้ยังมีผู้ให้ความหมายของความลำเอียงของข้อสอบ เฉพาะด้าน ดังนี้

เมย์เรนส์ และเลย์แมน (Mehrens and Lehmann, 1978 : 435-544) ให้ความหมายของข้อสอบที่มีความลำเอียงด้านวัฒนธรรม (Cultural Bias) ว่าหมายถึง ข้อสอบที่ทำให้กลุ่มผู้สอบสองกลุ่มที่มีความสามารถเท่ากัน แต่ที่มีความแตกต่างทางวัฒนธรรมมีคะแนนเฉลี่ยแตกต่างกัน หรือเป็นข้อสอบที่วัดคุณลักษณะทางจิตวิทยา หรือผลสัมฤทธิ์ในเนื้อหาเดียวกันไม่ตรงกัน และให้ความหมายข้อสอบที่มีความลำเอียงระหว่างเพศ (Sex Bias) ว่าหมายถึง ข้อสอบที่มีเนื้อหาลำเอียงต่อเพศใดเพศหนึ่งโดยเฉพาะ และมีผลต่อการทำข้อสอบแต่ละข้อ

เอลส์ และคนอื่น ๆ (Jensen, 1980 : 369 , quoting Eells and Others, 1951) ให้ความหมายของข้อสอบที่มีความลำเอียงด้านวัฒนธรรมว่า หมายถึง ข้อสอบที่ทำให้ผู้สอบ ซึ่งคุ้นเคยกันวิชาเฉพาะหรือกระบวนการเฉพาะมีโอกาสตอบถูกมากกว่าผู้สอบคนอื่น ที่มีลักษณะของกลุ่มแตกต่างกัน เป็นต้น

จากความหมายดังกล่าวข้างต้นสามารถสรุปได้ว่า ข้อสอบที่มีความลำเอียง หมายถึง ข้อสอบที่เมื่อนำไปทดสอบกับกลุ่มผู้สอบที่มีความสามารถเท่ากันภายใต้ประชากรเดียวกัน แต่มีลักษณะของกลุ่มย่อยแตกต่างกันทางด้านใดด้านหนึ่ง เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม สภาพภูมิศาสตร์ หรือเศรษฐกิจ แล้วปรากฏว่าโอกาสในการที่จะตอบข้อสอบข้อนั้นถูกไม่เท่ากัน หรือทำให้ในแต่ละกลุ่มย่อยมีการได้เปรียบเสียเปรียบจากการสอบข้อสอบเดียวกันไม่เท่ากัน ซึ่งอาจกล่าวได้ว่าเป็นข้อสอบที่ขาดความเชื่อมั่นในคุณลักษณะที่ต้องการวัด นั่นคือ เป็นข้อสอบที่วัด

ความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มได้ไม่ตรงกัน เมื่อทำการวิเคราะห์คุณสมบัติของข้อสอบตามกลุ่มย่อยของผู้สอบแล้ว จะได้คุณสมบัตินั้นไม่เท่ากัน

<u>แหล่ง/สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ</u>

จากการนำข้อสอบที่มีความลำเอียงมาศึกษาภายหลังจากการวิเคราะห์ความลำเอียงด้วยค่าสถิติแบบต่าง ๆ ในแบบทดสอบหลาย ๆ ฉบับ เพื่อศึกษาถึงสาเหตุหรือแหล่งที่ก่อให้เกิดความลำเอียงในตัวคำถามเหล่านั้น พบว่า สาเหตุที่น่าจะก่อให้เกิดความลำเอียงมีหลายสาเหตุด้วยกัน ซึ่งสามารถสรุปสาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบมากที่สุด ได้ดังต่อไปนี้ (Scheunerman, 1982 : 180 – 195)

1. การเดา (Questing) อาจจะเกิดจากข้อสอบนั้นยากเกินไป หรือเวลาไม่เพียงพอจะก่อให้เกิดความไม่เท่าเทียมกันในโอกาสการตอบข้อสอบถูกของผู้สอบแต่ละคน

2. ความเร็ว (Speed) หรือเวลาในการตอบ จะทำให้เกิดการเดาหรือในกรณีเวลาน้อยอาจจะทำข้อสอบไม่ทัน ซึ่งจะมีผลกับข้อสอบข้อหลัง ๆ โดยเฉพาะในการศึกษาความลำเอียงของข้อสอบวัดความถนัด

3. ความกำกวมหรือความไม่ชัดเจน (Unclear) ของข้อคำถาม นั่นคือ ข้อคำถามขาดความเป็นปรนัย การใช้ภาษาถิ่นหรือใช้คำที่ไม่เป็นภาษากลางในการสื่อความหมาย ซึ่งจะก่อให้เกิดความลำเอียงกับกลุ่มภาษาใดภาษาหนึ่งขึ้นได้

4. ลำดับขั้นของคำถาม (Series) อาจจะเป็นสิ่งที่ก่อให้เกิดความสับสน หรือชี้แนะคำตอบข้อสอบบางข้อได้

5. สถานภาพทางสังคมหรือความเกี่ยวข้องทางสังคม (Social Implication) ก็เป็นสิ่งที่ก่อให้เกิดความแตกต่างระหว่างกลุ่มผู้สอบแต่ละกลุ่มได้

6. ประสบการณ์หรือการฝึกฝนของแต่ละกลุ่มย่อย เป็นสิ่งที่ก่อให้เกิดการได้เปรียบเสียเปรียบของแต่ละกลุ่มค่อนข้างจะชัดเจน

7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ ขนบธรรมประเพณี เชื้อชาติ ศาสนา ก็จะเอื้อให้กับบางกลุ่มย่อย จึงทำให้เกิดการได้เปรียบเสียเปรียบในบางเนื้อหาวิชาได้

นอกจากนี้ความลำเอียงของข้อสอบอาจจะเกิดจากสาเหตุหรือแหล่งที่สำคัญ 2 แหล่ง คือ

1. การเลือกเนื้อหา (Bias in Selection) คือ ผู้สร้างข้อสอบเลือกเนื้อหาเฉพาะส่วนใดส่วนหนึ่ง มาสร้างข้อสอบ ทำให้ได้ข้อสอบที่มีเนื้อหาไม่ครอบคลุมและไม่ได้สัดส่วนที่สมดุลกัน

2. การสร้างข้อสอบ (Bias in Contruction) คือ การใช้ภาษาหรือข้อความบางอย่างที่เอื้อให้เกิดประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง

ความเป็นมาของการศึกษาความลำเอียงของข้อสอบ

การศึกษาความลำเอียงของข้อสอบนั้น เริ่มมาจากการสังเกตผลการสอบคัดเลือก ซึ่งพบว่าไม่เป็นไปตามสัดส่วนสติปัญญาหรือโครงสร้างของประชากร กรณีที่ทำให้มีการตื่นตัวมากในเรื่องความลำเอียงของข้อสอบ คือ ในปี ค.ศ. 1971 มาร์โคเดอฟูนิส (Marco Kefunis) และคนอื่น ๆ ซึ่งถูกปฏิเสธจากโรงเรียนกฎหมายของมหาวิทยาลัยวอชิงตัน ได้ฟ้องร้องว่าเขาได้คะแนนการสอบสูงกว่าผู้ที่ได้รับการคัดเลือกบางคน และได้ขึ้นฎีกาฟ้องร้อง ชาร์ลส์ โอเดการ์ด (Charles Odegaurs) เพื่อให้พิจารณาทบทวนการคัดเลือกนักศึกษาใหม่ และจากนั้นเป็นต้นมาการพิจารณาตรวจสอบความลำเอียงของข้อสอบระหว่างผู้สอบกลุ่มย่อยก็ได้รับการปฏิบัติกันจนถึงปัจจุบัน (Breland and Ironson, 1976 : 89) ซึ่งการตรวจสอบความลำเอียงจะทำการตรวจสอบก่อนและหลังจากการทดลองใช้ การตรวจสอบความลำเอียงของข้อสอบก่อนนำไปทดสอบใช้จะใช้การตรวจสอบโดยผู้เชี่ยวชาญ โดยพิจารณาถึงรูปแบบของข้อสอบ เนื้อหา คำที่ใช้และอื่น ๆ เพื่อไม่ให้เกิดความลำเอียง ส่วนการตรวจสอบความลำเอียงของข้อสอบหลังจากการทดลองใช้ผู้เสนอหลายวิธีดังนี้

1. การพิจารณาความแตกต่างกันในค่าเฉลี่ยระหว่างกลุ่ม

2. การวิเคราะห์องค์ประกอบ เป็นการดูความสัมพันธ์ระหว่างคะแนนย่อย ๆ ในแบบทดสอบสำหรับกลุ่มย่อย

3. การเปรียบเทียบค่าความยากเป็นการพิจารณาลำดับที่สัมพัทธ์ของค่าความยากที่ได้จากการวิเคราะห์แยกกลุ่ม พิจารณาปฏิสัมพันธ์ระหว่างข้อสอบและกลุ่มผู้สอบ ถ้าข้อสอบข้อใดมีลักษณะของความสัมพันธ์สำหรับบางกลุ่มแตกต่างไปจากข้ออื่น ๆ ในแบบทดสอบเดียวกัน จะถือว่าข้อสอบข้อนั้นมีความลำเอียง ซึ่งมีวิธีการทำได้ทั้งการวิเคราะห์ความแปรปรวนและการใช้แผนภูมิเส้น

4. การใช้ทฤษฎีการตอบข้อคำถาม เป็นการเปรียบเทียบความน่าจะเป็นในการตอบข้อสอบของผู้สอบกลุ่มย่อย โดยพิจารณาถึงค่าพารามิเตอร์ของข้อสอบและผู้ตอบตามแบบจำลองที่เลือกใช้ทั้ง 1, 2 และ 3 พารามิเตอร์

5. การวิเคราะห์ด้วยค่าไคสแควร์ เป็นวิธีการที่แปลงค่าความสามารถของสมาชิกในกลุ่มเป็นช่วง ๆ แล้วใส่ข้อมูลในตาราง 2 ทาง (จำนวนกลุ่ม x การตอบถูก – ผิด) จากนั้นคำนวณค่าไคสแควร์แต่ละตาราง และเมื่อรวมค่าทุกตารางจะได้ค่าไคสแควร์ที่มีชั้นแห่งความเป็นอิสระเท่ากับจำนวนกลุ่ม ถ้ากลุ่มผู้สอบ 2 กลุ่ม มีคะแนนเฉลี่ยต่างกันแล้วจะเกิดผลที่ดูความลำเอียงได้

6. เทคนิคลอกลีเนียร์ (Loglinear) หรือแบบจำลองโลจีสที่เบี่ยงเบนจากไคสแควร์ และได้มีการนำมาขยายโดยการทำซ้ำ หลังจากการขจัดข้อสอบที่ลำเอียงออกไปแล้วในแต่ละรอบของการวิเคราะห์

7. วิธีของแมนเทล-แฮนส์เซล (Mantel-Haenszel) เป็นการศึกษาตัวแปรประเภทแบ่งสองในกลุ่มที่แตกต่างกันด้วยตัวแปรบล็อค และนำมาใช้เป็นตัวบ่งชี้ความลำเอียงของข้อสอบ

8. วิธีการตอบข้อสอบแบบเทียบ ซึ่งเสนอโดย ลินน์ และฮาร์นิสช์ (Linn and Harnisch, 1981) โดยวิเคราะห์ตามทฤษฎีการตอบข้อสอบ 3 พารามิเตอร์ และรวมกลุ่มย่อยเข้าด้วยกันแล้วแบ่งความสามารถเป็นช่วง พิจารณาความแตกต่างระหว่างความน่าจะเป็นของการตอบตามที่คาดหวัง ซึ่งประมาณได้ด้วยทฤษฎีการตอบข้อคำถามกับความน่าจะเป็นของการตอบถูกจากข้อมูลเชิงประจักษ์พิจารณาเป็นช่วง ๆ ความสามารถ แล้วนำมารวมกันเป็นดัชนีรวมของความลำเอียงของแต่ละข้อ ใช้ได้ในกรณีกลุ่มตัวอย่างน้อย ทำให้ใช้ทฤษฎีการตอบข้อคำถามแบบเต็มรูปไม่ได้

9. วิธี SIBTEST เป็นวิธีดัดแปลงจากวิธีทฤษฎีการตอบข้อคำถามสามารถใช้ได้ทั้งเพื่อการตรวจสอบความลำเอียงของข้อสอบรายข้อ การตรวจสอบความลำเอียงของกลุ่มข้อสอบหรือแบบทดสอบทั้งฉบับ

การวิจัยครั้งนี้ ผู้วิจัยใช้วิธีการตรวจสอบความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีของแมนเทล-แฮนส์เซล

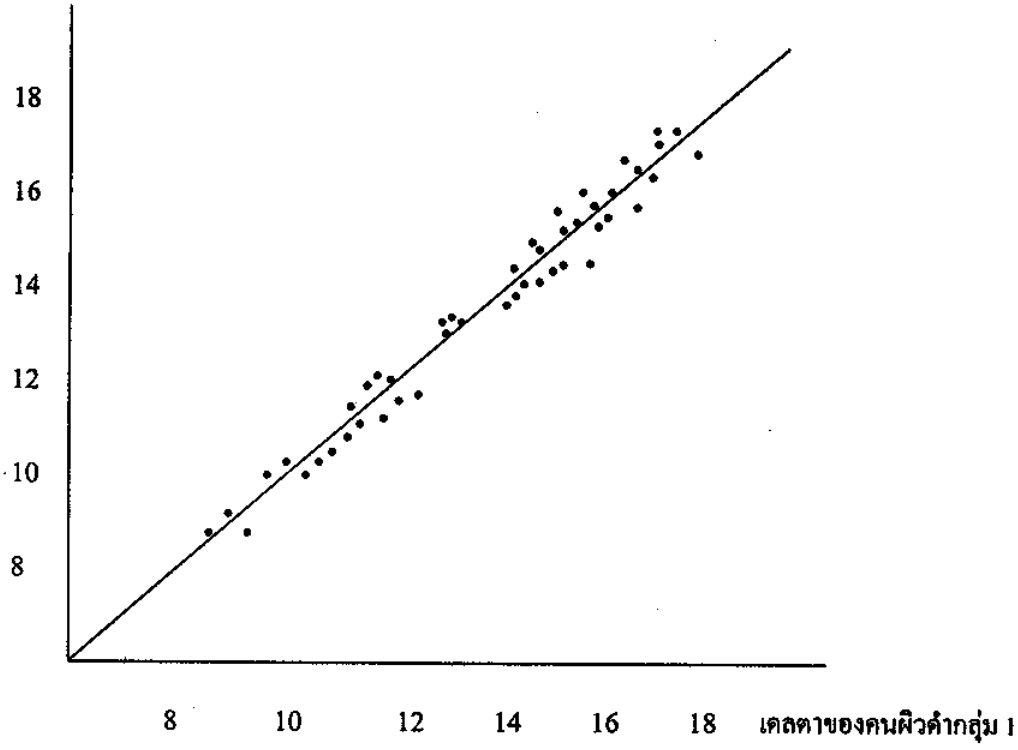
<u>วิธีการตรวจสอบความลำเอียงของข้อสอบ</u>

1. การวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีการแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty)

ผู้ที่เสนอการวิเคราะห์ความลำเอียงของข้อสอบวิธีนี้ คือ แองกอฟฟ์ (Berk, 1982 : 97, quoting Angoff, 1972 ; also Angoff & Ford, 1973) การวิเคราะห์ความลำเอียงของข้อสอบตามวิธีการของ แองกอฟฟ์ ก็คือ จะต้องคำนวณหาค่าความยาก (P-Values) ของข้อสอบแต่ละข้อจากผู้สอบแต่ละกลุ่ม (คือ กลุ่ม 1 กลุ่ม 2 ดังนั้นค่าความยากของข้อสอบแต่ละข้อจะมี 2 ค่า ตามจำนวนกลุ่มที่แบ่ง) ในการเปรียบเทียบค่าความยาก (P-Values) นี้จะใช้วิธีการแปลงค่าความยาก (P-Values) เป็นคะแนนมาตรฐาน Z จากคะแนนมาตรฐาน Z แปลงให้เป็นความยากมาตรฐาน

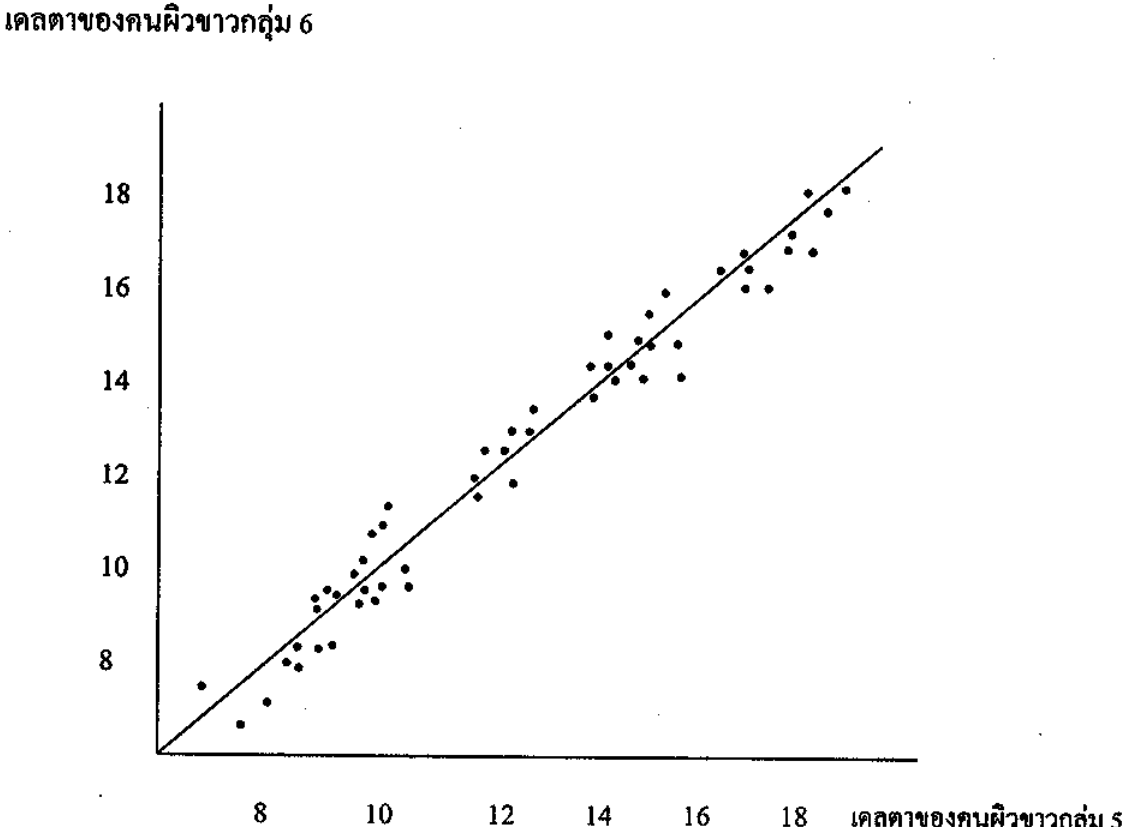
เดลตา (Δ) (Delta) ด้วยสมการ $\Delta = 4Z + 13$ แล้วนำค่าเดลตา (Delta) แต่ละคู่มาลงจุดคู่อันดับบนกราฟ จุดต่าง ๆ บนกราฟ ซึ่งแทนค่าเดลตาจะอยู่ในรูปวงรี แล้วคำนวณหาระยะทางที่จุดต่าง ๆ เหล่านี้ห่างจากเส้นแกนหลัก (Major Axis) ข้อสอบข้อใดที่ห่างจากเส้นแกนหลักน้อยกว่า -3Sd หรือมากกว่า 3Sd จะสรุปว่าข้อสอบข้อนั้นเป็นข้อสอบที่มีความลำเอียง ดังตัวอย่างในภาพประกอบ 1, 2 และ 3

เดลตาของคนผิวดำกลุ่ม 2



ภาพประกอบ 1 ค่าเดลตา ของคนผิวดำกลุ่ม 1 และคนผิวดำกลุ่ม 2

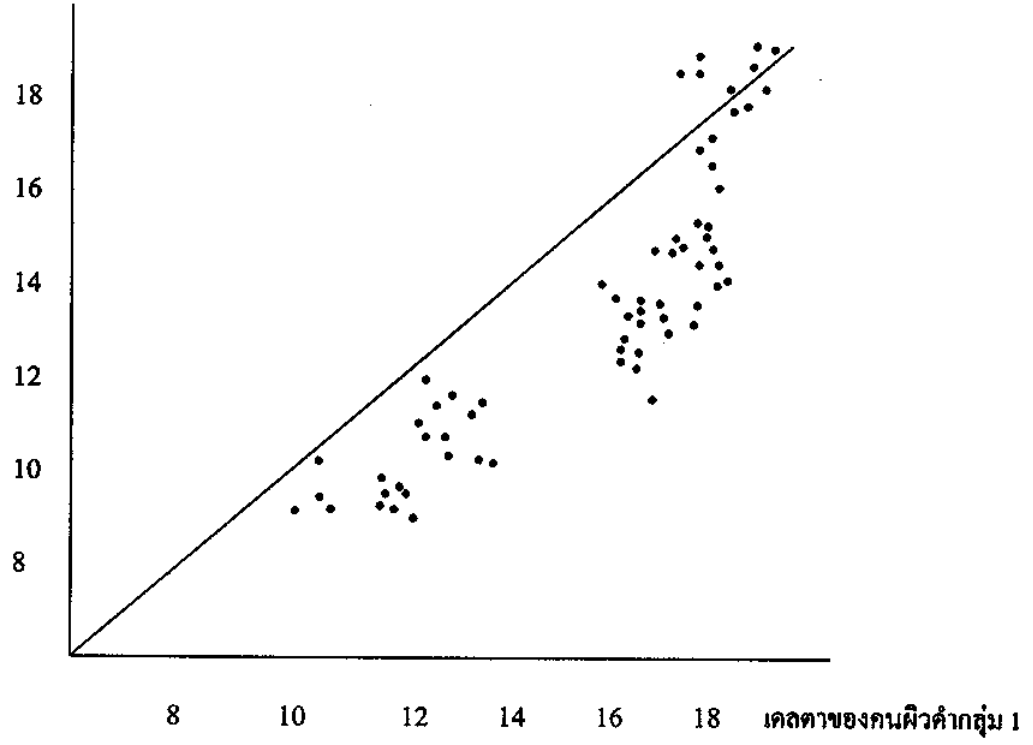
ที่มา : Berk, 1982 : 99 , qouting Angoff & Ford, 1973



ภาพประกอบ 2 ค่าเดลตา ของคนผิวขาวกลุ่ม 5 และคนผิวขาวกลุ่ม 6

ที่มา : Berk, 1982 : 99 , quoting Angoff & Ford, 1973

เดลตาของคนผิวขาวกลุ่ม 5



ภาพประกอบ 3 ค่าเดลตา ของคนผิวขาวกลุ่ม 5 และคนผิวดำกลุ่ม 1

ที่มา : Berk, 1982 : 99 , quoting Angoff & Ford, 1973

เมื่อลงจุดอันดับค่าเดลตาของข้อสอบแต่ละข้อของผู้เข้าสอบแต่ละกลุ่มแล้ว จะเห็นว่าข้อสอบที่ไม่มีความลำเอียงจะได้กราฟ ดังภาพประกอบ 1 และภาพประกอบ 2 ซึ่งภาพประกอบ 1 เป็นกราฟของค่าเดลตาจากกลุ่มตัวอย่างคนผิวดำ 2 กลุ่ม โดยใช้แบบทดสอบวัดความถนัดทางการเรียนด้านภาษา (Berk, 1982 : 99 , quoting Angoff & Ford. 1973) ภาพประกอบ 2 เป็นกราฟของค่าเดลตาที่ได้จากการทดสอบกลุ่มตัวอย่างคนผิวขาว 2 กลุ่ม โดยใช้แบบทดสอบชุดเดียวกัน จะเห็นว่าจุดต่าง ๆ ของค่าเดลตาของกลุ่มผู้เข้าสอบ 2 กลุ่ม จะกระจายอยู่ในแนวเส้นแกนหลัก คือ ไม่มีการเบี่ยงเบนไปจากเส้นแกนหลัก

ส่วนภาพประกอบ 3 เป็นกราฟของค่าเดลตาที่ได้จากการทดสอบกลุ่มตัวอย่างคนผิวดำกับคนผิวขาว โดยใช้แบบทดสอบชุดเดียวกันจะเห็นว่าจุดต่าง ๆ ของค่าเดลตาของผู้เข้าสอบ 2 กลุ่มนี้จะกระจายห่างจากเส้นแกนหลัก คือ มีการเบี่ยงเบนไปจากเส้นแกนหลัก ซึ่งข้อสอบเหล่านี้ยากต่อกลุ่มผู้สอบคนผิวดำ ดังนั้น แบบทดสอบชุดนี้มีปฏิสัมพันธ์ระหว่างข้อสอบกับเชื้อชาติ (Item Race Interaction) และเรียกได้ว่าเป็นข้อสอบที่มีความลำเอียงต่อเชื้อชาติ ซึ่งคำนวณได้ดังนี้

จากสมการเส้นแกนหลัก $y = ax + b$

เมื่อ a แทน ความชันของเส้นแกนหลัก

b แทน ค่าคงที่ของการตัดแกน Y

โดยที่ค่า a และ b คำนวณได้จากสูตร

$$a = \frac{(S_y^2 - S_x^2) + \sqrt{(S_y^2 - S_x^2)^2 + 4r_{xy}^2 S_x^2 S_y^2}}{2r_{xy} S_x S_y}$$

$$b = M_y - aM_x$$

เมื่อ x แทน ค่าเดลตาของกลุ่มที่ 1

y แทน ค่าเดลตาของกลุ่มที่ 2

M_x แทน ค่าเฉลี่ยของค่าเดลตา ของกลุ่มที่ 1

M_y แทน ค่าเฉลี่ยของค่าเดลตา ของกลุ่มที่ 2

S_x^2 แทน ค่าแปรปรวนของค่าเดลตากลุ่มที่ 1

S_y^2 แทน ค่าแปรปรวนของค่าเดลตากลุ่มที่ 2

S_x แทน ค่าความเบี่ยงเบนมาตรฐานของค่าเดลตากลุ่มที่ 1

S_y แทน ค่าความเบี่ยงเบนมาตรฐานของค่าเดลตากลุ่มที่ 2

r_{xy} แทน สหสัมพันธ์ของค่าเดลตา กลุ่มที่ 1 และกลุ่มที่ 2

แองกอฟฟ์ (Angoff) เสนอการคำนวณระยะทางตั้งฉากของจุดคู่อันดับค่าเดลตาของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก จากสูตร

$$d_i = \frac{ax_i - y_i + b}{\sqrt{a^2 + 1}}$$

และ $$S_d = \sqrt{1 - r_{xy}}$$

เมื่อ d_i แทน ระยะทางตั้งฉากของจุดคู่อันดับค่าเดลตาของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก

S_d แทน ค่าเบี่ยงเบนมาตรฐานของระยะ d

สำหรับการวิจัยครั้งนี้ ผู้วิจัยได้เลือกศึกษา วิธีแปลงค่าความยากของข้อสอบ ตามวิธีของแองกอฟฟ์ โดยพิจารณาค่าระยะห่างตั้งฉากจากคู่อันดับเดลตาของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก ข้อสอบที่มีความลำเอียงจะต้องมีค่าระยะห่างตั้งฉากมากกว่า 3Sd หรือน้อยกว่า –3Sd

2. การวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีของแมนเทล-แฮนส์เซล (Mentel-Haenszel)

วิธีแมนเทล-แฮนส์เซล (MH) เป็นวิธีที่มีความคล้ายคลึงกับวิธีไคสแควร์ที่เสนอโดย ซูเนแมน (Scheuneman, 1979) มาราซูโลและสลอเตอร์ (Marasculio and Slayghter, 1981) และเมลเลนเบอร์ก (Mellenberg, 1982) โดยเป็นการเปรียบเทียบผลการตอบข้อสอบของกลุ่มผู้สอบ 2 กลุ่ม หรือที่ฮอลแลนด์ เรียกว่า กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (Reference and Focal Group) โดยที่กลุ่มแรกจะใช้อ้างถึงกลุ่มที่คาดว่าจะได้ประโยชน์จากข้อสอบ และกลุ่มหลังเป็นกลุ่มเสียประโยชน์จากข้อสอบในกรณีที่ข้อสอบมีความลำเอียง โดยมีการตรวจสอบทุก ๆ ระดับคะแนนรวมจากการสอบ ข้อสอบข้อใดที่ผู้สอบทั้ง 2 กลุ่ม ทำได้เท่า ๆ กัน จะเป็นข้อสอบที่ถือได้ว่าไม่มีความลำเอียงต่อกลุ่มใดกลุ่มหนึ่ง วิธีการในการวิเคราะห์มีดังนี้

ถ้าให้ N_{Rj} และ N_{Fj} แทนจำนวนผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีคะแนนรวมในช่วงคะแนน j และให้ T_j ซึ่งเท่ากับ $N_{Rj} + N_{Fj}$ เป็นจำนวนผู้สอบรวมทั้ง 2 กลุ่มที่ได้คะแนนรวมจากการสอบอยู่ในช่วงคะแนน j แล้วนำมาเขียนในรูปตาราง 2 ทาง แสดงผลการตอบถูก (1) และ ผิด (0) ดังตารางที่ 1

ตาราง 1 ข้อมูลการสอบของกลุ่มตัวอย่างกลุ่ม R และ F ที่มีคะแนนรวมอยู่ในช่วงคะแนน j

กลุ่มผู้สอบ	คะแนนที่ได้จากข้อสอบที่ต้องการวิเคราะห์ความฉลาดเรียน			
		1	0	รวม
	R	A_j	B_j	N_{Rj}
	F	C_j	D_j	N_{Fj}
	รวม	M_{1j}	M_{0j}	T_j

เมื่อ A_j , B_j , C_j และ D_j เป็นความถี่ที่สังเกตได้ของการตอบถูก (1) และตอบผิด (0) ในช่วงคะแนน j และแสดงในรูปสัดส่วนได้ ดังตารางที่ 2

ตาราง 2 สัดส่วนการตอบข้อสอบของกลุ่มประชากร 2 กลุ่ม

ผลการตอบ / กลุ่ม	1	0	รวม
R	P_{Rj}	Q_{Rj}	1
F	P_{Fj}	Q_{Fj}	1

โดยที่ P_{Rj} แทน สัดส่วนของกลุ่มอ้างอิงที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก

P_{Fj} แทน สัดส่วนของกลุ่มเปรียบเทียบที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก

Q_{Rj} แทน $1 - P_{Rj}$

Q_{Fj} แทน $1 - P_{Fj}$

จากตาราง 1 จะได้ว่า ในช่วงระดับคะแนน j A_j และ B_j คือ จำนวนผู้สอนกลุ่มอ้างอิงที่ตอบข้อสอบข้อที่ถูกต้อง และไม่ถูกต้องตามลำดับ ในขณะที่ C_j และ D_j คือ จำนวนผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบข้อที่ถูกต้องและไม่ถูกต้องตามลำดับ เช่นกัน

ตามเทคนิควิธีแมนเทล-แฮนส์เซล (MH) ที่เสนอโดยฮอลแลนด์และเธเยอร์นั้น ข้อมูลการตอบข้อสอบแต่ละข้อจะถูกนำมาแสดงในรูปตาราง 2 ทาง เป็นจำนวนตารางเท่ากับคะแนนรวมที่มีผู้สอบทำได้ บวกด้วย 1 จากนั้นจะคำนวณค่าความเป็นไปได้ในรูปของค่าสัดส่วนการตอบข้อสอบถูก-ผิด ระหว่างกลุ่มโดยกำหนดให้เป็นค่า α ดังนี้

$$\alpha_{MH} = \frac{\sum_{j=1}^{S} A_j D_j / T_j}{\sum_{j=1}^{S} B_j C_j / T_j}$$

ค่า α_{MH} จะมีค่าระหว่าง 0 และ α โดยที่ $\alpha_{MH} = 1$ จะแสดงถึงสมมติฐานทางสถิติ ที่ว่าไม่มีความแตกต่างหรือไม่มีความลำเอียง ซึ่งแสดงว่ากลุ่มตัวอย่างทั้ง 2 กลุ่ม ทำข้อสอบ ได้ผลเช่นเดียวกันในแต่ละระดับคะแนนรวม ค่า A_jD_j/T_j และ B_jC_j/T_j มีค่าเท่ากันทั้ง 2 กลุ่ม ทำให้ α_{MH} มีค่าเท่ากับ 1 ถ้าค่า α_{MH} มีค่ามากกว่า 1.00 นั่นคือ สมาชิกของกลุ่มที่ 1 จะตอบข้อสอบข้อนั้นได้ถูกต้องมากกว่าสมาชิกในกลุ่มที่ 2 แต่ถ้าหากว่าค่า α_{MH} น้อยกว่า 1.00 แสดงว่าสมาชิกกลุ่มที่ 2 ตอบข้อสอบข้อนั้นได้ถูกต้องมากกว่าสมาชิกกลุ่มที่ 1 ซึ่งแสดงถึงความแตกต่าง หรือความลำเอียงที่เกิดขึ้นระหว่างกลุ่มตัวอย่าง 2 กลุ่ม

อย่างไรก็ตามในทางปฏิบัติแล้ว แม้ว่าจะมีการควบคุมให้ผู้สอบ 2 กลุ่ม มีความสามารถในสิ่งที่ต้องการวัดเท่ากัน แต่เป็นไปไม่ได้ที่ผลการตอบข้อสอบของผู้สอบทั้ง 2 กลุ่ม จะเป็นเช่นเดียวกัน ดังกล่าว แมนเทล-แฮนส์เซล (Mantel-Haensel, 1959) จึงเสนอค่าสถิติไคสแควร์เพื่อทดสอบค่าที่ได้ว่าจะมีความแตกต่างจาก 1.00 อย่างมีนัยสำคัญทางสถิติหรือไม่ ที่ระดับชั้นแห่งความเป็นอิสระเท่ากับ 1 ดังนี้

$$MH-CHISQ = \frac{\left[\left|\sum_{j=1}^{S}\left[A_j - E(A_j)\right]\right| - 0.5\right]^2}{\sum_{j=1}^{S} Var(A_j)}$$

โดยที่ค่าผลรวมเป็นค่าที่ได้จากการรวมทุกชั้นของคะแนนรวม และค่า $E(A_j)$ มีค่าดังนี้

$$E(A_j) = (N_{Rj})(M_{1j})/T_j$$

$$Var(A_j) = \frac{N_{Rj} N_{Fj} M_{1j} M_{0j}}{T_j^2 (T_j - 1)}$$

และสมมติฐานที่แสดงว่าข้อสอบไม่ลำเอียง ได้แก่ สัดส่วน จำนวนคนตอบถูกในกลุ่มอ้างอิง เท่ากับสัดส่วนจำนวนคนตอบถูกในกลุ่มเปรียบเทียบ สำหรับทุกชั้นของคะแนน j

$$H_\alpha \quad : \quad P_{Rj} = P_{Fj} \qquad \text{สำหรับทุกชั้นของคะแนน j}$$

สมมติฐานทางสถิตินี้เป็นสมมติฐานของความเป็นอิสระอย่างมีเงื่อนไขของสมาชิกกลุ่มและคะแนนที่ได้จากการตอบข้อสอบที่ต้องการตรวจสอบความลำเอียง และภายใต้สมมติฐานดังกล่าว จะได้ค่าคาดหวังของการตอบข้อสอบ ดังนี้

$$\begin{aligned} E(A_j) &= N_{Rj}M_{ij}/T_j \\ E(B_j) &= N_{Rj}M_{oj}/T_j \\ E(C_j) &= N_{Fj}M_{ij}/T_j \\ E(D_j) &= N_{Fj}M_{oj}/T_j \end{aligned}$$

สมมติฐานอื่นของ แมนเทล-แฮนส์เซล ได้แก่

$$H_1 \quad : \quad \frac{P_{Rj}}{Q_{Rj}} \neq \alpha \frac{P_{Fj}}{Q_{Fj}}; j = 1,2...,k \qquad \text{เมื่อ } \alpha \neq 1$$

เมื่อ $\alpha = 1$ ซึ่งสอดคล้องกับสมมติฐานทางสถิติ จะได้ว่า

$$H_1 \quad : \quad \frac{P_{Rj}}{Q_{Rj}} = \frac{P_{Fj}}{Q_{Fj}}$$

และค่าประมาณของ $\alpha_{MH} = \frac{P_{Rj}}{Q_{Rj}} \div \frac{P_{Fj}}{Q_{Fj}} = \frac{P_{Rj}Q_{Fj}}{P_{Fj}Q_{Rj}}$ สำหรับทุก ๆ j = 1, 2, ..., k

นอกจากนั้น ฮอลแลนด์และเธเยอร์ ได้เสนอแนะให้แปลงค่า α_{MH} ให้เป็นเดลตา (Δ) เพื่อให้เป็นค่าที่มีค่าเฉลี่ยเป็น 0 ดังนี้

$$\Delta_{MH} = \frac{-4}{1.7}\ln(\alpha_{MH}) = -2.35\ln(\alpha_{MH})$$

ค่า Δ_{MH} มีค่าอยู่ระหว่าง -2.6 ถึง 2.6 และใช้เป็นดัชนีแสดงความลำเอียงของข้อสอบ ถ้า α_{MH} มีค่าเท่ากับ 1 หรือ Δ_{MH} มีค่าเท่ากับ 0 แล้ว แสดงว่า ข้อสอบข้อนั้นไม่ลำเอียง

จากที่ผ่านมามีผู้นำวิธีการวิเคราะห์ความลำเอียงของแมนเทล-แฮนส์เซล (MH) มาศึกษาหลายท่าน เช่น คลอสเซอร์และคนอื่น ๆ (Clauser and Others, 1991) ใช้วิธีแมนเทล-แฮนส์เซล วิเคราะห์ความลำเอียงอย่างสม่ำเสมอ พบว่า วิธีแมนเทล-แฮนส์เซล จะใช้ได้ดีในกรณีที่ข้อสอบมีค่าอำนาจจำแนกสูง แต่จะไม่สามารถวิเคราะห์ข้อสอบที่มีความยากสูงมากได้ ซึ่งสอดคล้องกับข้อค้นพบของ เมเซอร์ และคนอื่น ๆ (Mazor and Other, 1991) ซุดวิคส์ และทอลแมน (Sudweeks and Tolman, 1990) พบว่า ข้อสอบที่ลำเอียงมักเป็นข้อสอบที่ยาก บากิและเฟอรารา (Baghi and Ferrara, 1990) เมื่อใช้ขนาดกลุ่มตัวอย่าง 750 คนขึ้นไป วิธีแมนเทล-แฮนส์เซล ใช้แทนวิธีใช้ทฤษฎีการตอบข้อคำถาม 3 พารามิเตอร์ได้ สวามินาธาน และโรเจอร์ส (Swaminathan and Rogers, 1990) ซึ่งศึกษาจากข้อมูลจำลอง พบว่า วิธีแมนเทล-แฮนส์เซล วิเคราะห์ได้ดีกว่าการถดถอยแบบโลจีสเล็กน้อย โดยตรวจค้นได้ถูกต้อง ร้อยละ 75 ในกรณีใช้กลุ่มตัวอย่าง 250 คน และตรวจค้นได้ถูกต้อง ร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน กรณีที่ลำเอียงอย่างสม่ำเสมอ และกรณีไม่สม่ำเสมอที่ตัดกันที่ปลายข้างใดข้างหนึ่ง แต่วิธีแมนเทล-แฮนส์เซล มีค่าใช้จ่ายน้อยกว่าโลจีสประมาณ 3 – 4 เท่า แฮมเบิลตันและคนอื่น ๆ (Hambleton and Others, 1986) พบว่า วิธีแมนเทล-แฮนส์เซล ให้ค่าใกล้เคียงกับวิธีของทฤษฎีการตอบข้อคำถาม ทั้งที่ใช้ค่าความแตกต่างของค่าเฉลี่ยกำลังสอง และการตรวจสอบความแตกต่างของพื้นที่รวมใต้โค้ง แต่วิธีแมนเทล-แฮนส์เซล มีค่าใช้จ่ายต่ำกว่า และใช้เวลาน้อยกว่า ลินเดอร์ (Linacre, 1988) พบว่า วิธีแมนเทล-แฮนส์เซล ใช้ได้ดีในสถานการณ์จำลองทุกสถานการณ์พอ ๆ กับโปรแกรม PROX ของ ราส์ช (Rasch) แต่ควรใช้เกณฑ์ในการคำนวณความลำเอียงและความคลาดเคลื่อนหลาย ๆ เกณฑ์ และใช้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานมากกว่า 1 ตัว

สำหรับการวิจัยครั้งนี้ ผู้วิจัยได้เลือกศึกษา วิธีของแมนเทล-แฮนส์เซล โดยพิจารณาเปรียบเทียบผลการตอบข้อสอบของกลุ่มผู้สอบ 2 กลุ่ม ซึ่งคำนวณโดยใช้โปรแกรม MHDIF ข้อสอบที่มีความลำเอียง จะต้องมีค่า α_{MH} มากกว่า 1 หรือ χ^2_{MH} ที่คำนวณได้มีนัยสำคัญทางสถิติที่ระดับ .05

3. ทฤษฎีการตอบข้อคำถาม (Item Response Theory)

ทฤษฎีการตอบข้อคำถาม (Item Response Theory) บางครั้งเรียกว่า ทฤษฎีคุณลักษณะแฝง (Latent Trait Theory) หรือทฤษฎีโค้งลักษณะข้อสอบ (Item Characteristic Curve Theory) หรือทฤษฎีการทดสอบแบบใหม่ (Modern Test Theory) เป็นต้น (ผจงจิต อินทสุวรรณ, 2525 : 54) เป็นทฤษฎีที่ศึกษาเกี่ยวกับความสัมพันธ์ของการตอบข้อสอบของผู้สอบและคุณลักษณะภายใน (Latent Trait) ของผู้สอบ กล่าวคือ การตอบข้อสอบของผู้สอบสามารถได้รับการทำนายได้ด้วยคุณลักษณะเฉพาะของผู้สอบ หรือกล่าวอีกนัยหนึ่งก็คือ เราสามารถนำระดับความสามารถ หรือคุณลักษณะที่ประมาณได้ไปทำนาย หรืออธิบายการตอบข้อสอบถูกของผู้สอบได้ (Hambleton and Cook. 1997 : 75 ; Citing Lord and Novick. 1968 ; Hambleton and Swamianthan. 1985 : 9) เป้าหมายหลักของทฤษฎีการตอบข้อคำถาม คือ ต้องการหาคุณภาพรายข้อของแบบทดสอบ และการประมาณค่าความสามารถของผู้สอบ ซึ่งต้องอาศัยความสอดคล้องกันระหว่างโมเดลและข้อมูลทฤษฎีนี้ได้เข้ามามีบทบาทในวงการวัดผลการศึกษาและจิตวิทยา ตั้งแต่ปี 1950 โดยในระยะแรกการพัฒนาเป็นไปอย่างเชื่องช้ามาก เนื่องจากความยุ่งยากในการคำนวณ และขาดแคลนในโปรแกรมคอมพิวเตอร์ และเครื่องคอมพิวเตอร์ที่ใช้ในการวิเคราะห์ข้อมูลตามทฤษฎีจึงทำให้ชะงักการพัฒนาอยู่ระยะหนึ่ง แต่ในปัจจุบันได้มีการพัฒนาโปรแกรมคอมพิวเตอร์ขึ้นมาเรื่อย ๆ จนทำให้เป็นที่นิยมใช้กันอย่างแพร่หลายและเป็นที่สนใจของนักวัดผลการศึกษามากยิ่งขึ้น

3.1 หลักการของทฤษฎีการตอบข้อคำถาม

ทฤษฎีการตอบข้อคำถามเป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมตอบสนองต่อข้อสอบของบุคคลนั้น ๆ โดยทฤษฎีนี้มีความเชื่อว่าพฤติกรรมตอบสนองต่อข้อสอบของบุคคลจะถูกกำหนด โดยคุณลักษณะภายใน หรือความสามารถที่มีอยู่ในตัวบุคคล (Lord and Novick, 1968 : 358) ซึ่งไม่สามารถวัดหรือสังเกตได้โดยตรง แต่สามารถที่จะอธิบายหรือพยากรณ์ได้ จากพฤติกรรมที่บุคคลนั้นตอบสนองต่อข้อสอบ หรือจากคะแนนการตอบข้อสอบ ซึ่งเป็นสิ่งที่สามารถวัดหรือสังเกตได้ในรูปแบบจำลองทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันการตอบสนองข้อสอบ หรือฟังก์ชันลักษณะข้อสอบ หรือโค้งลักษณะข้อสอบ (Hambleton and Swaminathan, 1985 : 13) ฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างระดับความสามารถกับพฤติกรรมการตอบสนองต่อข้อสอบ สามารถเขียนให้อยู่ในรูปแบบสมการทางคณิตศาสตร์ได้ดังนี้ (Lord and Novick, 1968 : 360)

$$P_i(\theta) = \text{Prob}\ (U_i = 1|\theta) \qquad \text{เมื่อ} \quad U_i = 0,1$$

จากสมการข้างต้นนี้ หมายถึง โอกาสที่ผู้สอบซึ่งมีระดับความสามารถ θ จะตอบคำถามข้อ i ได้ถูกต้อง

3.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อคำถาม

ทฤษฎีการตอบข้อคำถาม มีข้อตกลงเบื้องต้นดังนี้

1. ความเป็นมิติเดียว (Unidimensionality) เป็นแบบทดสอบที่ประกอบด้วยข้อคำถามที่เป็นเอกพันธ์ นั่นคือ แบบทดสอบนั้นจะต้องมุ่งวัดความสามารถอย่างใดอย่างหนึ่งที่มีลักษณะเด่น ๆ เพียงความสามารถเดียว หากไม่กำหนดข้อตกลงเบื้องต้นเช่นนี้จะทำให้โมเดลที่ใช้มีความสลับซับซ้อนมาก (Hambleton and Swaminathan, 1985 : 38) สำหรับการตรวจสอบว่าแบบทดสอบนั้นวัดในมิติเดียวหรือไม่ สามารถทำได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกต ค่าไอเกน (Eigen Value) ขององค์ประกอบสูงสุดว่าแตกต่างจากค่าไอเกนขององค์ประกอบที่เหลืออย่างชัดเจนหรือไม่ในขณะที่ค่าไอเกนขององค์ประกอบที่เหลือทั้งหมดมีค่าใกล้เคียงกัน (Warm, 1978 : 104 ; Lord, 1980 : 21)

2. ความเป็นอิสระในการตอบข้อสอบ (Local Independence) เป็นการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระในการตอบข้อสอบ กล่าวคือ การตอบข้อสอบข้อใดข้อหนึ่งถูกจะไม่มีผลต่อการตอบข้อสอบข้ออื่น ๆ หรือจะกล่าวในเชิงคณิตศาสตร์ได้ว่า ความเป็นอิสระในการตอบข้อสอบ หมายถึง โอกาสที่จะตอบข้อสอบถูกทั้งหมดจะเท่ากับผลคูณของโอกาสการตอบข้อสอบถูกเป็นรายข้อ (Hambleton and Swaminathan, 1985 : 23) ในการตรวจสอบว่าข้อสอบแต่ละข้อเป็นไปตามข้อตกลงเกี่ยวกับความเป็นอิสระในการตอบข้อสอบหรือไม่ ให้สังเกตความเป็นมิติเดียวกันของแบบทดสอบถ้าแบบทดสอบมีความเป็นมิติเดียวแล้ว แบบทดสอบนั้นจะมีความเป็นอิสระในการตอบข้อสอบด้วย (Lord. 1980 : 19 ; Hambleton and Swaminathan, 1985 : 24)

3. โค้งลักษณะข้อสอบ (Item Characteristic Curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบข้อนั้นได้ถูกต้องตรงกับระดับความสามารถที่วัดได้ โดยใช้ชุดของข้อสอบหรือแบบทดสอบนั้น และความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบได้ถูกต้องขึ้นอยู่กับโค้งลักษณะข้อสอบของแต่ละโมเดลที่ใช้แต่จะไม่ขึ้นอยู่กับการแจกแจงความสามารถของกลุ่มตัวอย่าง (Hambleton and Swaminathan, 1985 : 25) กล่าวคือ คุณสมบัติของโค้งลักษณะของข้อสอบจะไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ฉะนั้น จึงทำให้ความน่าจะเป็นในการตอบข้อสอบถูกแต่ละข้อ ไม่แปรเปลี่ยนด้วย โดยผู้สอบที่มีความสามารถในระดับสูง จะมีความน่าจะเป็นในการตอบข้อสอบ ได้ถูกต้องมากกว่าผู้สอบที่มีความสามารถในระดับต่ำ

เนื่องจากมีความเชื่อเกี่ยวกับความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องกับระดับความสามารถที่วัดได้หลายความเชื่อ จึงทำให้เกิด โมเดลของโค้งลักษณะข้อสอบขึ้นมาหลายโมเดลด้วยกัน ดังจะกล่าวต่อไปนี้

3.3 โมเดลของทฤษฎีการตอบข้อคำถาม

ทฤษฎีการตอบข้อคำถาม ได้มีการพัฒนาในรูปแบบต่าง ๆ ทำให้เกิดโมเดลเฉพาะขึ้น ซึ่งแต่ละโมเดลจะแตกต่างกันไปตามฟังก์ชันทางคณิตศาสตร์ และจำนวนพารามิเตอร์ที่ใช้อธิบายโค้งลักษณะข้อสอบของแต่ละโมเดล ซึ่งถ้าจะแบ่งโมเดลตามลักษณะของข้อมูล หรือการตรวจให้คะแนน สามารถแบ่งได้ 3 ประเภท ดังนี้ (Hambleton and Swaminathan, 1985 : 35)

1. โมเดลที่มีการตรวจให้คะแนนแบบสองค่า (Dichotomous) โดยตอบถูกได้ 1 คะแนน และถ้าตอบผิดได้ 0 คะแนน แบ่งออกเป็นโมเดลต่าง ๆ ดังนี้

1.1 กัตต์แมน เฟอร์เฟ็คท์ โมเดล (Guttman Perfect Model)

1.2 ลาเท้นท์ ดิชเท็นซ์ โมเดล (Latent Distance Model)

1.3 ลาเท้นท์ ลิเนียร์ โมเดล (Latent Linear Model)

1.4 นอร์มัล ออร์ไจว์ โมเดล ที่มีพารามิเตอร์ 1, 2, 3 ตัว (One-, Two-, Three – Parameter Normal Ogive Model)

1.5 โลจีสติก โมเดล ที่มีพารามิเตอร์ 1, 2, 3, 4 ตัว (One-, Two-, Three-, Four-Parameter Logistic Model)

1.6 ราสช์ โมเดล (Rasch Model)

2. โมเดลที่มีการตรวจให้คะแนนแบบหลายค่า (Multi – Dichotomous) โดยใช้กับข้อสอบที่ให้คะแนนแบบลำดับขั้น เช่น แบบทดสอบวัดเจตคติ แบบทดสอบวัดภาคปฏิบัติ และแบบทดสอบวัดบุคลิกภาพ เป็นต้น ซึ่งแบ่งออกเป็นโมเดลต่าง ๆ ดังนี้

2.1 นอมินอล เรสพอน โมเดล (Nominal Response Model)

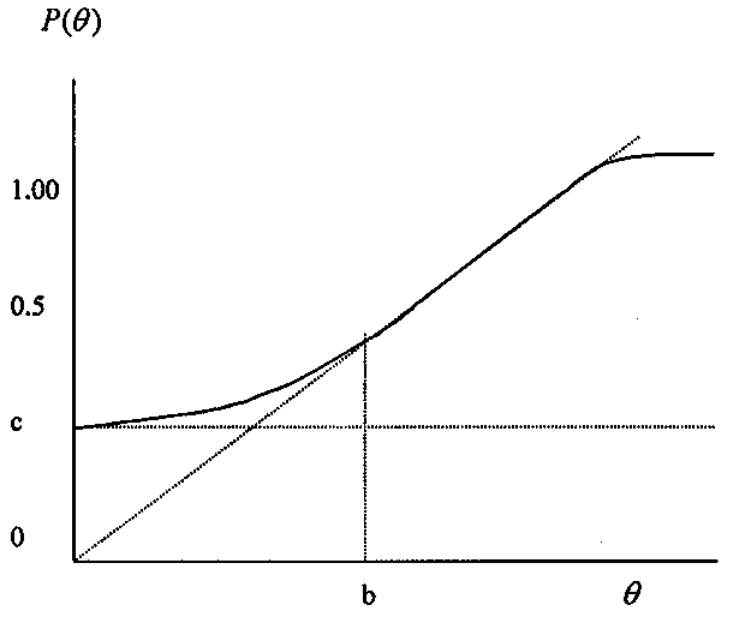
2.2 เกรด เรสพอน โมเดล (Graded Response Model)

2.3 พาเชียล เครดิต โมเดล (Partial Credit Model)

3. โมเดลที่มีการตรวจให้คะแนนแบบต่อเนื่อง (Continuous) ได้แก่ คอนทินิวอัส เรสพอน โมเดล (Continuous Response Model) โมเดลนี้พัฒนาขึ้นโดย ซามีไจมา (Samejima) (Hambleton and Swaminathan, 1985 : 52) สามารถนำไปใช้เป็นเครื่องมือของการทดสอบทางจิตวิทยา หรือการศึกษาทางด้านเจตคติโดยผู้สอนต้องตอบคำถามลงในมาตราวัดแบบต่อเนื่อง (Continuous Scale)

สำหรับในการศึกษาครั้งนี้ จะนำเสนอเฉพาะโมเดลที่มีการตรวจให้คะแนนแบบสองค่า ของ นอร์มัล ออร์ไจว์ โมเดล และโลจีสติก โมเดล ซึ่งเป็นโมเดลที่กล่าวถึงกันมาก และใช้กันแพร่หลาย ดังนี้

1. นอร์มัล ออร์ไจว์ โมเดล (Normal Ogive Model) เส้นโค้งลักษณะข้อสอบถาม โมเดลนี้จะมีลักษณะเป็นรูปตัว S และแต่ละข้อจะมีโค้งลักษณะข้อสอบแตกต่างกันไปทั้งนี้จะขึ้นอยู่กับค่าพารามิเตอร์ของข้อสอบ ดังภาพประกอบ 4



ภาพประกอบ 4 โค้งลักษณะข้อสอบของ นอร์มัล ออร์ไจว์ โมเดล (Normal Ogive Model)

จากภาพประกอบ 4 แสดงให้เห็นว่า โค้งลักษณะข้อสอบเป็นกราฟที่แสดงความสัมพันธ์ระหว่างโอกาสที่จะตอบข้อสอบข้อใดข้อหนึ่งถูกกับระดับความสามารถของผู้สอบ โดยผู้สอบที่มีระดับความสามารถสูงมีโอกาสที่จะตอบข้อสอบถูกมากกว่าผู้สอบที่มีระดับความสามารถต่ำ ค่าความสามารถและค่าความยากของข้อสอบจะอยู่ในสเกลเดียวกัน โดยมีค่าอยู่ระหว่าง $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัติแล้ว ระดับความสามารถจะมีค่าอยู่ระหว่าง -3 ถึง +3 ค่า -3 หมายถึง ผู้สอบมีระดับความสามารถต่ำมาก ค่า +3 หมายถึง ผู้สอบมีระดับความสามารถสูงมาก ค่าความยากจะมีค่าอยู่ระหว่าง -3 ถึง +3 เช่นเดียวกันกับค่าอำนาจจำแนกของข้อสอบ (Hambleton and Swaminathan, 1985 : 50) ส่วนค่าการเดานั้น หมายถึง ค่าปลายต่ำสุด ของโค้งลักษณะข้อสอบซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1

ความสัมพันธ์ดังกล่าวนี้จะอยู่ในรูปฟังก์ชันการแจกแจงแบบปกติ หรือฟังก์ชันปกติออร์ไจว์ (Cumulative Normal Ogive) ดังนี้

$$P_i(\theta) = c_i + (1 - c_i)\int_{-\alpha}^{a_i(\theta - b_i)} (1/\sqrt{2\pi})_e^{-z^2/2}\, dz \qquad (1)$$

เมื่อ $P_i(\theta)$ แทน โอกาสที่ผู้สอบซึ่งมีความสามารถ θ จะตอบคำถามข้อที่ i ได้ถูกต้อง

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i

b_i แทน ค่าความยากของข้อสอบข้อที่ i

c_i แทน ค่าการเดา

θ แทน ระดับความสามารถของผู้สอบ

Z แทน ความเบี่ยงเบนปกติ (Normal Deviation) จากการแจกแจงซึ่งมีค่าเฉลี่ยเท่ากับ b_i และค่าความเบี่ยงเบนมาตรฐานเท่ากับ $1/a_i$

e แทน ค่าคงที่ มีค่าประมาณ 2.71828...

ค่าพารามิเตอร์ของข้อสอบ (a_j , b_i และ c_i) จะเป็นตัวที่ทำให้รูปร่างของโค้งลักษณะของข้อสอบแต่ละข้อแตกต่างกันไป สมการ (1) เป็นสมการที่มีพารามิเตอร์ 3 ตัว คือ พารามิเตอร์ค่าความยาก (b_i) พารามิเตอร์ค่าอำนาจจำแนก (a_i) และพารามิเตอร์ค่าการเดา (c_i) จึงอาจเรียกว่า นอร์มัล ออร์ไจว์ โมเดล แบบ 3 พารามิเตอร์ (Three – Parameter Normal Ogive Model) แต่หากในสถานการณ์การสอบที่มีผู้สอบมีโอกาสตอบถูกโดยการเดามีน้อย สมการ (1) ก็จะเหลือค่าพารามิเตอร์เพียง 2 ตัว คือ พารามิเตอร์ค่าความยาก (b_i) และพารามิเตอร์ค่าอำนาจจำแนก (a_i)

$$P_i(\theta) = \int_{-\alpha}^{a_i(\theta - b_i)} (1/\sqrt{2\pi})_e^{-z^2/2}\, dz$$

เมื่อ $P_i(\theta)$ แทน โอกาสที่ผู้สอบซึ่งมีความสามารถ θ จะตอบคำถามข้อที่ i ได้ถูกต้อง

$(1/\sqrt{2\pi})^{-z^2/2}$ แทน Normal Density Function

b_i แทน ค่าความยากของข้อสอบข้อที่ i

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i

θ แทน ระดับความสามารถของผู้สอบ

ในกรณีที่กำหนดให้ข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันหมด และค่าการเดาเป็นศูนย์ โมเดล ก็จะเหลือ พารามิเตอร์ค่าความยาก (b_i) เพียงตัวเดียว ซึ่งจะเรียกว่า นอร์มัล ออร์ไจว์ โมเดล ที่มีพารามิเตอร์ 1 ตัว โดยมีลักษณะของโมเดล ดังนี้ (Hambleton and Swaminathan. 1985 : 50)

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta-b_i)} (1/\sqrt{2\pi}) e^{-z^2/2} dz$$

นอร์มัล ออร์ไจว์ นั้นมีปัญหาบางประการทั้งในด้านทฤษฎีและคณิตศาสตร์ จึงได้มีผู้พยายามคิดหาแบบจำลองที่มีรูปร่างคล้ายคลึงกันกับ นอร์มัล ออร์ไจว์ โมเดล แต่มีความสะดวกกว่าในการคณิตศาสตร์ โมเดล ดังกล่าวนี้ ก็คือ โลจีสติก โมเดล (Logistic Model)

2. โลจีสติก โมเดล (Logistic Model) โลจีสติก โมเดล เป็นโมเดลที่มีลักษณะใกล้เคียงกันกับนอร์มัล ออร์ไจว์ โมเดล แต่สะดวกในการประมาณค่ามากกว่า นอกจากนี้ยังเป็นโมเดลที่ได้พัฒนาเพื่อให้เหมาะสมกับสถานการณ์การทดสอบ เช่น ในกรณีที่ตอบข้อสอบได้ด้วยการเดาปลายต่ำสุดของโค้งลักษณะข้อสอบไม่เป็นศูนย์ แสดงว่าการทำข้อสอบย่อมมีการเดาและในกรณีผู้สอบทำข้อสอบด้วยความสะเพร่า โมเดลนี้จะเพิ่มค่าความรอบคอบขึ้นอีกหนึ่งค่า ซึ่งจะเห็นได้จากปลายสูงสุดของโค้งลักษณะข้อสอบมีค่าน้อยกว่า 1 โลจีสติก โมเดล แบ่งเป็น 4 โมเดลย่อย ๆ ตามพารามิเตอร์ ดังนี้ (Hambleton and Swaminathan. 1985 : 34 – 49)

2.1 โลจีสติก โมเดล ที่มีพารามิเตอร์สองตัว (Two – Parameter Logistic Model)

เบิร์นบอมได้เสนอโมเดลโลจีสติกที่มีค่าพารามิเตอร์ 2 ตัว คือ พารามิเตอร์ค่าความยาก (b_i) และพารามิเตอร์ค่าอำนาจจำแนก (a_i) ในปี ค.ศ. 1968 มีสมการดังนี้

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_j(\theta-b_i)}}, \quad i = 1, 2, \ldots, n$$

เมื่อ θ แทน ระดับความสามารถที่แท้จริงที่คำนวณจากคะแนนจริง แต่ปรับหน่วยให้เป็นมาตรฐาน

$P_i(\theta)$ แทน ค่าความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ (θ) จะทำข้อสอบข้อที่ i ได้ถูกต้อง

a_i แทน ค่าอำนาจจำแนกซึ่งมีค่าเป็นสัดส่วนโดยตรงกับค่าความชันของโค้ง ณ จุดเปลี่ยนโค้ง หรือจุดที่ชันที่สุด

b_i แทน ค่าความยากที่แสดงระดับความสามารถที่แท้จริง ที่จุดโค้งชันที่สุด หรือในกรณีที่ไม่มีการเดา ค่า b_i คือ 0 ณ จุดความน่าจะเป็น .50

D แทน Scaling Factor ซึ่งมีค่า 1.7 เป็นค่าที่มากที่สุดในการปรับ Logistic Model และ Ogive Model ให้สอดคล้องกัน

e แทน ค่าคงที่ มีค่าประมาณ 2.71828...

2.2 โลจีสติก โมเดล ที่มีพารามิเตอร์สามตัว (Three-Parameter Model)

โมเดลที่มีพารามิเตอร์สามตัว ได้ดัดแปลงมาจากโมเดลที่ใช้พารามิเตอร์สองตัว ค่าพารามิเตอร์ที่เบิร์นบอมได้เพิ่มขึ้นอีก 1 ตัว คือ ค่าสัมประสิทธิ์การเดา (c_i) ซึ่งแสดงได้ด้วยสมการทางคณิตศาสตร์ ดังนี้

$$P_i(\theta) = c_i + (1-c_i)\frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad ; \quad i=1,2,...,n$$

เมื่อ D แทน ค่าคงที่ มีค่าประมาณ 1.7...

c_i แทน ค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำมาก มีโอกาสจะทำข้อสอบข้อนี้ได้ถูกต้อง

2.3 โลจีสติก โมเดล ที่มีพารามิเตอร์ตัวเดียว (One-Parameter Logistic Model)

โมเดลนี้จะกำหนดให้ค่าพารามิเตอร์การเดาของข้อสอบมีค่าเท่ากับศูนย์ $(c_i = 0)$ และถือว่าข้อสอบมีค่าอำนาจจำแนกเท่ากัน $(a_i = \bar{a})$ โมเดลก็จะเหลือค่าพารามิเตอร์ค่าความยาก (b_i) เพียงตัวเดียว ซึ่งตรงกับ โมเดลของทฤษฎีการตอบข้อคำถามที่พัฒนาขึ้นโดย ราส์ช โดยที่ฟังก์ชันของโมเดลนั้นสามารถอธิบายได้ด้วยพารามิเตอร์ของข้อสอบเพียงตัวเดียว คือ พารามิเตอร์ค่าความยาก (b_i) หรือที่นิยม เรียกว่า ราส์ช โมเดล ซึ่งลักษณะของ โมเดลดังนี้ (Hambleton and Swaminathan. 1985 : 47)

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad , \quad i = 1, 2, ..., n$$

เมื่อ $\bar{a}$ แทน ค่าอำนาจจำแนกที่เป็นกลาง

หรือ $$P_i(\theta) = \frac{e^{\theta-b_i}}{1+e^{\theta-b_i}} \quad , \quad i = 1, 2, ..., n$$

2.4 โลจิสติก โมเดล ที่มีพารามิเตอร์ 4 ตัว (Four-Parameter Logistic Model)

โมเดลนี้เสนอโดยแม็คโดแนลด์ (McDonald) ในปี ค.ศ. 1967 พัฒนาโดย ลอร์ด (Lord) และบาร์ตัน (Barton, 1981) โมเดลนี้เพิ่มค่าความรอบคอบ (Careless) หรือพารามิเตอร์อีก 1 ตัว ในโมเดลพารามิเตอร์ 3 ตัว เพื่อแก้ปัญหาที่ผู้สอบที่มีความสามารถในระดับสูงมักตอบข้อสอบผิด อาจจะเป็นเพราะความไม่รอบคอบหรือผู้สอบมีความสามารถอยู่เหนือกว่าแบบทดสอบจึงทำให้ไม่สามารถเลือกคำตอบที่ถูกต้องได้ มีลักษณะของโมเดล ดังนี้ (Hambleton and Sawminathan, 1985 : 34 – 49)

$$P_i(\theta) = c_i + (\gamma_i + c_i)\ \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_j(\theta - b_i)}}\ ,\ i = 1, 2, \ldots, n$$

เมื่อ γ_i แทน ค่าที่ผู้สอบซึ่งมีความสามารถในระดับสูงแต่มีความสะเพร่าตอบข้อสอบไม่ถูกต้อง โดยมีค่าน้อยกว่า 1 เล็กน้อย

3.4 การประยุกต์ทฤษฎีการตอบข้อคำถามเพื่อศึกษาความลำเอียงของข้อสอบ

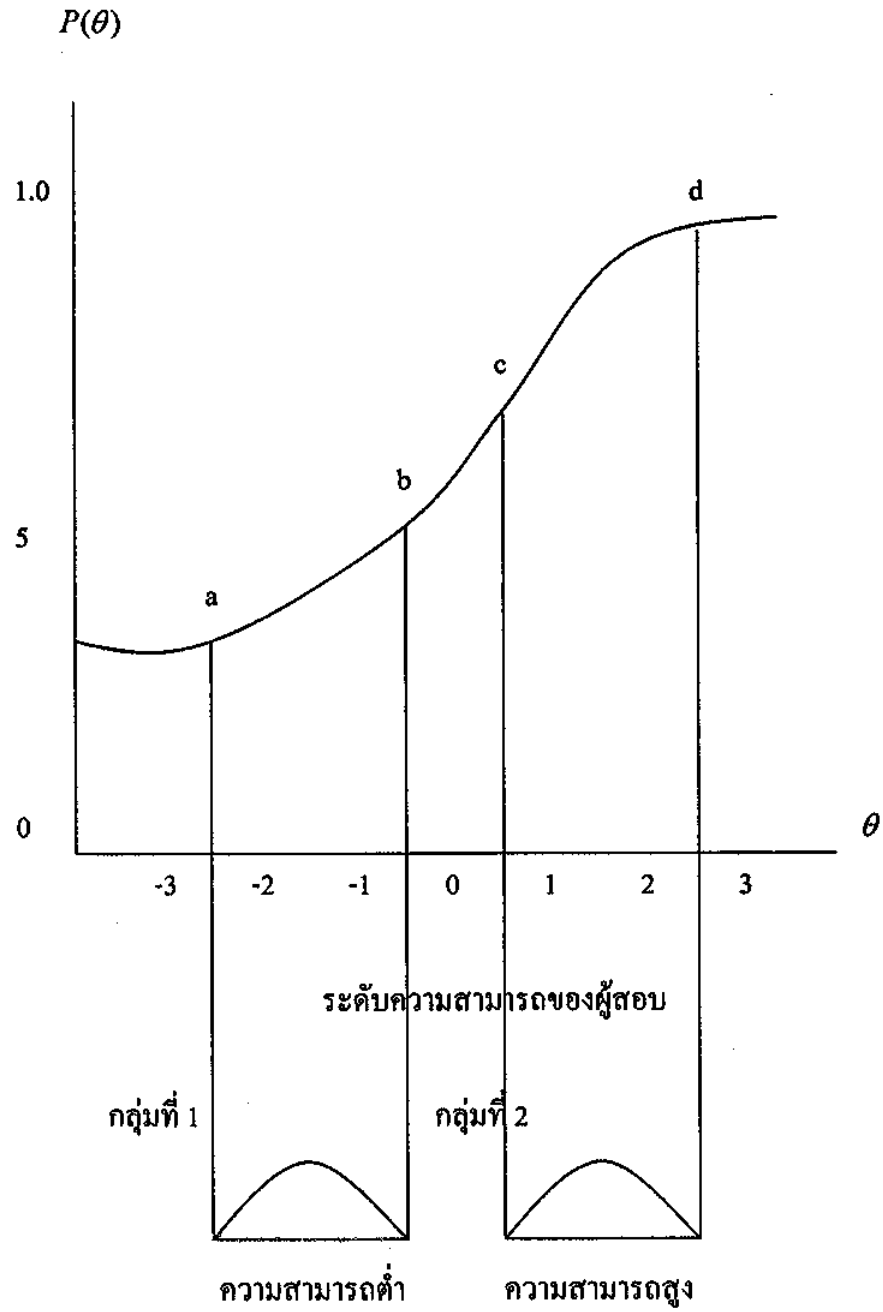
จากข้อตกลงเบื้องต้น 3 ประการของทฤษฎีการตอบข้อคำถาม คือ ความเป็นมิติเดียวของข้อสอบ ความเป็นอิสระในการตอบข้อสอบ และโค้งลักษณะข้อสอบ สามารถนำมาประยุกต์ใช้เป็นแนวทางในการวิเคราะห์ความลำเอียงของข้อสอบ ได้ดังนี้ คือ ถ้ามีแบบทดสอบฉบับหนึ่งวัดความสามารถของผู้สอบมากกว่า 1 กลุ่ม แล้วปรากฏว่า ข้อสอบข้อเดียวกันวัดความสามารถของผู้สอบได้ไม่ตรงกัน เช่น กลุ่มหนึ่ง อาจจะเคยชินกับเนื้อหาของข้อสอบข้อนั้นมาก่อน หรือเมื่อตอบข้อสอบข้ออื่นถูกแล้ว เป็นเหตุทำให้ความน่าจะเป็นของการตอบข้อสอบข้อนั้นถูกมีค่ามากกว่าอีกกลุ่ม ถึงแม้ว่าภายในแต่ละฉบับจะมีความเป็นเอกพันธ์กันทุกข้อ แต่เมื่อพิจารณาเปรียบเทียบข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบทุกกลุ่ม โดยนำผลการสอบของแต่ละกลุ่มมาสร้างโค้งลักษณะข้อสอบบนกราฟเดียวกัน เพื่ออธิบายลักษณะของข้อสอบจะพบว่าโค้งลักษณะข้อสอบจากผู้สอบแต่ละกลุ่มแตกต่างกัน ความแตกต่างของโค้งลักษณะข้อสอบของกลุ่มต่าง ๆ จะเป็นตัวบงชี้ว่าผู้สอบที่มีระดับความสามารถเท่ากันแต่อยู่ต่างกลุ่มกันมีโอกาสจะตอบข้อสอบได้ถูกต้องไม่เท่ากัน เพราะระดับความสามารถที่เท่ากันของผู้สอบนั้นอาจมาจากความสามารถต่างชนิดกัน ข้อสอบที่ลักษณะเช่นนี้ถือว่าเป็นข้อสอบที่มีความลำเอียงต่อกลุ่มที่มีความน่าจะเป็นการตอบข้อสอบถูกมากกว่ากลุ่มที่มีระดับความสามารถต่ำ ซึ่งความลำเอียงของข้อสอบนี้แสดงให้เห็นได้ด้วย โค้งลักษณะข้อสอบ กล่าวคือ ถ้าผู้สอบต่างกลุ่มกันมีโค้งลักษณะข้อสอบข้อเดียวกันแตกต่างกัน ข้อสอบข้อนั้นถือได้ว่าเป็นข้อสอบที่มีความลำเอียงแต่ในทำนองกลับกันหากโค้งลักษณะข้อสอบของ

ผู้สอบต่างกลุ่มกันมีลักษณะโค้งเหมือนกันก็ย่อมแสดงว่าข้อสอบข้อนั้นไม่ลำเอียง สำหรับวิธีการวิเคราะห์ความลำเอียงด้วยโค้งลักษณะข้อสอบนั้นจะใช้คอมพิวเตอร์โปรแกรม LOGISTIC หรือโปรแกรม BILOG 3 ในการประมาณค่าโค้งลักษณะข้อสอบที่มี 2 พารามิเตอร์ หรือ 3 พารามิเตอร์ และใช้โปรแกรม BICAL ในการประมาณค่าโค้งลักษณะข้อสอบที่มี 1 พารามิเตอร์

จากการศึกษารายงานการวิจัยเกี่ยวกับการวิเคราะห์ความลำเอียงของข้อสอบด้วยโค้งลักษณะข้อสอบ พบว่า วิธีวิเคราะห์ที่ใช้กันส่วนมากมี 2 แบบ คือ แบบ 3 พารามิเตอร์ เป็นแบบที่คำนวณพารามิเตอร์ของข้อสอบ 3 ตัว คือ ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i) และค่าสัมประสิทธิ์การเดา (c_i) โดยใช้แบบจำลองโลจีสติก (Three-Parameter Logistic Model) และแบบ 1 พารามิเตอร์ ที่คำนวณค่าพารามิเตอร์ของข้อสอบ 1 ตัว คือ ค่าความยาก (b_i) เพราะกำหนดให้ค่าอำนาจจำแนก (a_i) ของข้อสอบทุกข้อเท่ากัน หรือให้ค่าอำนาจจำแนกมีค่าเท่ากับ 1 และค่าสัมประสิทธิ์การเดา (c_i) เท่ากับ 0 โดยใช้แบบจำลองโลจีสติก หรือแบบจำลองของราส์ช (One-Parameter Logistic Model or Rasch Model) ซึ่งแต่ละแบบต่างก็มีข้อดีข้อเสียแตกต่างกัน กล่าวคือ การวิเคราะห์ด้วยโค้งลักษณะข้อสอบแบบคำนวณค่าพารามิเตอร์ของข้อสอบ 1 ตัว มีความสะดวกง่าย และรวดเร็ว ในการวิเคราะห์ไม่จำเป็นต้องประมาณค่าพารามิเตอร์ของข้อสอบทุกตัว และสามารถหาข้อสอบที่เหมาะสม (Fit) กับวิธีการวิเคราะห์ได้ ถึงแม้ว่าจะมีข้อจำกัดในเรื่องเกี่ยวกับการเดาและอำนาจจำแนกที่ไม่เป็นจริงในทางปฏิบัติ สำหรับการวิเคราะห์ด้วยโค้งลักษณะข้อสอบแบบคำนวณค่าพารามิเตอร์ของข้อสอบ 3 ตัวนั้น แม้ว่าความถูกต้องของการประมาณค่าอำนาจจำแนก (a_i) และสัมประสิทธิ์ในการตอบถูกโดยการเดาค่อนข้างต่ำและต้องวิเคราะห์ด้วยคอมพิวเตอร์ที่ค่อนข้างจะซับซ้อนและเสียค่าใช้จ่ายแพง (Berk, 1982 : 154) แต่ก็ให้ผลการวิเคราะห์ที่มีความถูกต้องและสมบูรณ์กว่า เพราะสามารถวิเคราะห์และอธิบายถึงโอกาสในการตอบข้อสอบถูกของผู้สอบในกลุ่มที่มีความสามารถต่ำมาก ๆ ได้

3.5 คุณสมบัติของโค้งลักษณะ

ในการศึกษาความไม่เปลี่ยนแปลงเกี่ยวกับโค้งลักษณะของข้อสอบนี้ ลอร์ด และโนวิค (Lord and Novick, 1968) ได้กำหนดศึกษาเกี่ยวกับค่าพารามิเตอร์ของข้อสอบข้อเดียวกัน ระหว่างกลุ่มผู้สอบที่มีความสามารถต่างกัน ซึ่งนำมาเขียนสรุปได้ ดังภาพประกอบ 5

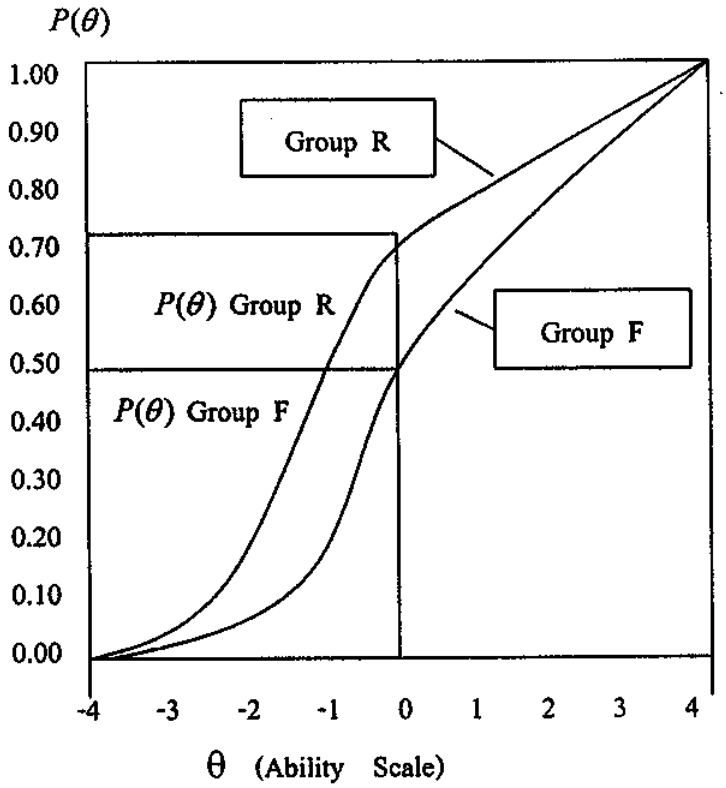


ภาพประกอบ 5 โค้งลักษณะข้อสอบ ซึ่งมีลักษณะไม่เปลี่ยนแปลงตามจำนวน และระดับความสามารถของผู้สอบ

จากภาพประกอบ 5 แสดงว่า ในกรณีที่กลุ่มผู้สอบมีระดับความสามารถต่ำค่าความน่าจะเป็นของการตอบข้อสอบถูกของผู้สอบสามารถจะอธิบายได้ด้วยโค้งลักษณะของข้อสอบในช่วงระยะ $a-b$ และในทำนองเดียวกันเมื่อผู้สอบมีระดับความสามารถสูงความน่าจะเป็นของการตอบข้อสอบถูกก็ย่อมสามารถอธิบายได้ด้วยโค้งลักษณะของข้อสอบในช่วง $c-d$ นั่นคือ ถ้าค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบแล้ว ข้อสอบที่นำไปทดสอบกับผู้ที่มีความสามารถแตกต่างกัน เมื่อนำข้อมูลที่ได้ในแต่ละกลุ่มมาวิเคราะห์จะมีค่าพารามิเตอร์ของข้อสอบเท่ากัน ถ้าข้อสอบข้อนี้วัดความสามารถอย่างเดียวกันของกลุ่มตัวอย่างทั้งสองกลุ่มแล้ว เส้นโค้งลักษณะข้อสอบก็ย่อมมีเส้นเดียว เส้นโค้งในช่วง $a-b$ และ $c-d$ ย่อมเป็นส่วนหนึ่งของโค้งลักษณะข้อสอบข้อนี้ เมื่อเส้นโค้งลักษณะข้อสอบมีได้เส้นเดียว ตัวพารามิเตอร์ซึ่งเป็นตัวกำหนดรูปร่างลักษณะของเส้นโค้งย่อมมีได้ค่าเดียวเช่นกัน หรืออาจกล่าวได้ว่า ถ้าค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบแล้ว ข้อสอบที่นำไปสอบกับผู้สอบที่มีความสามารถแตกต่างกัน เมื่อนำข้อมูลที่ได้จากแต่ละกลุ่มมาวิเคราะห์จะมีค่าพารามิเตอร์ของข้อสอบเป็นค่าเดียวกัน

การวิเคราะห์ความลำเอียงของข้อสอบด้วยโค้งลักษณะข้อสอบจะพิจารณาจากผลการตอบข้อสอบกับโอกาสในการตอบข้อสอบข้อนั้น ๆ ถูก ระหว่างกลุ่มผู้สอบตั้งแต่ 2 กลุ่ม หากโค้งลักษณะข้อสอบมีความแตกต่างกัน แสดงว่าข้อสอบข้อนั้นจะทำหน้าที่เบี่ยงเบนไปจากปกติ หรือทำให้เกิดการได้เปรียบเสียเปรียบระหว่างกลุ่มผู้สอบที่มีลักษณะแตกต่างกัน และค่าพารามิเตอร์ที่เป็นตัวกำหนดโค้งลักษณะข้อสอบตามทฤษฎีการตอบข้อคำถามนั้นมี 3 ค่า คือ ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i) และค่าสัมประสิทธิ์การเดา (c_i) ในการทดสอบว่าข้อสอบข้อใด มีความลำเอียงหรือไม่และจะใช้วิธีใดในการวิเคราะห์ความลำเอียงของข้อสอบให้มีความถูกต้องเหมาะสมเพียงใดนั้น ขึ้นอยู่กับลักษณะความแตกต่างของโค้งลักษณะข้อสอบที่มีความลำเอียง ซึ่งสามารถแบ่งออกเป็น 2 แบบ คือ

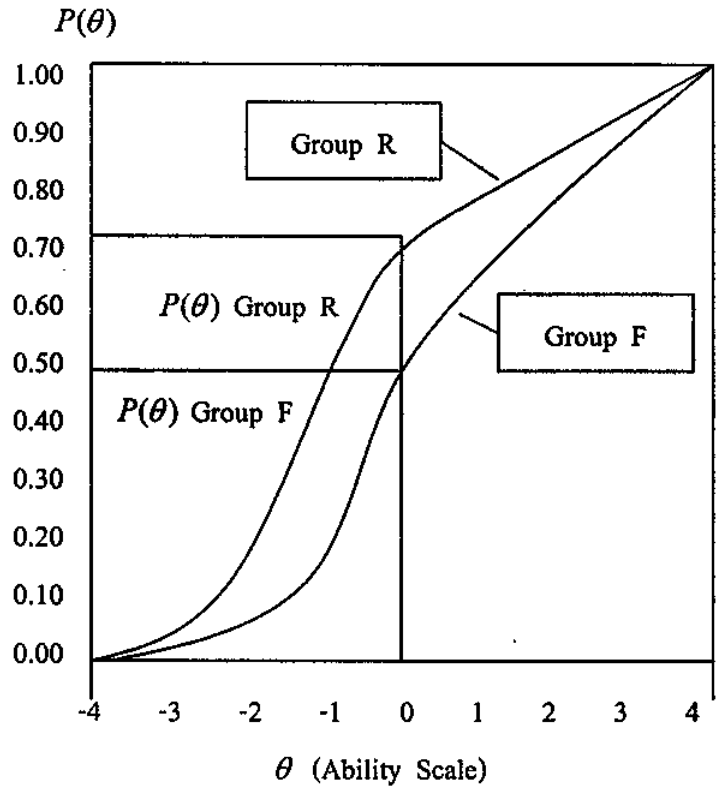
1. แบบลำเอียงอย่างสม่ำเสมอหรือแบบลำเอียงคงที่ คือ โค้งลักษณะข้อสอบของกลุ่มต่าง ๆ จะมีค่าพารามิเตอร์ของข้อสอบเท่ากัน 2 ค่า คือ ค่าอำนาจจำแนกของข้อสอบ (a_i : ความชัน) ของแต่ละกลุ่มเท่ากัน และมีค่าสัมประสิทธิ์การเดา (c_i) ที่ใกล้เคียงกันหรือเท่ากัน วิธีนี้จะมีเฉพาะค่าความยากของข้อสอบ (b_i) ตัวเดียวเท่านั้นที่แตกต่างกันและโค้งลักษณะข้อสอบของกลุ่มต่าง ๆ จะไม่ตัดกันตลอดต่อเนื่องของทุกระดับความสามารถ ดังภาพประกอบ 6



ภาพประกอบ 6 โค้งลักษณะข้อสอบระหว่างกลุ่ม R และกลุ่ม F

จากภาพประกอบ 6 แสดงให้เห็นว่าในกลุ่มผู้สอบที่มีระดับความสามารถ (θ) เดียวกัน ผู้สอบที่อยู่ในกลุ่ม R มีโอกาสในการตอบข้อสอบข้อนี้ได้ถูกต้องมากกว่าผู้สอบที่อยู่ในกลุ่ม F ซึ่งในการวิเคราะห์ความลำเอียงของข้อสอบ จะใช้ค่าพารามิเตอร์เพียงตัวเดียว คือ ค่าความยากของข้อสอบ (b_i) มาคำนวณค่าดัชนีความลำเอียงของข้อสอบ นั่นคือ เป็นการวิเคราะห์ความลำเอียงของข้อสอบตามแบบจำลอง โลจิสติกที่มี 1 พารามิเตอร์ หรือแบบจำลองของ ราส์ช (Rasch Model) นั่นเอง การคำนวณพื้นที่ระหว่างโค้งลักษณะข้อสอบวิธีนี้สามารถคำนวณได้ง่ายเพราะพื้นที่ที่คำนวณได้มีเฉพาะเครื่องหมายบวกเพียงอย่างเดียวเท่านั้น

2. แบบลำเอียงอย่างไม่สม่ำเสมอ หรือแบบลำเอียงไม่คงที่ คือ โค้งลักษณะข้อสอบของกลุ่มต่าง ๆ จะมีค่าพารามิเตอร์ของข้อสอบแตกต่างกัน 3 ค่า คือ ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i) และค่าสัมประสิทธิ์การเดา (c_i) ทำให้โค้งลักษณะข้อสอบของกลุ่มต่าง ๆ มีลักษณะตัดกัน ดังภาพประกอบ 7



ภาพประกอบ 7 โค้งลักษณะข้อสอบระหว่างกลุ่ม R และกลุ่ม F

จากภาพประกอบ 7 แสดงให้เห็นว่าที่ระดับความสามารถสูงผู้สอบในกลุ่ม R มีโอกาสที่จะตอบข้อสอบนี้ได้ถูกต้องมากกว่าผู้สอบในกลุ่ม F ที่มีระดับความสามารถเดียวกัน แต่สำหรับในกลุ่มผู้สอบที่มีระดับความสามารถต่ำแล้ว ผู้สอบในกลุ่ม F มีโอกาสที่จะตอบข้อสอบถูกมากกว่าผู้สอบในกลุ่ม R ในการวิเคราะห์ค่าความลำเอียงของข้อสอบจะนำค่าพารามิเตอร์ทั้ง 3 ค่ามาคำนวณค่าดัชนีความลำเอียงของข้อสอบ นั่นคือ เป็นการวิเคราะห์ความลำเอียงของข้อสอบตามแบบจำลอง 3 พารามิเตอร์ การคำนวณค่าดัชนีความลำเอียงตามวิธีนี้จะมีความยุ่งยากมาก และซับซ้อน เพราะว่าค่าพารามิเตอร์ของโค้งลักษณะข้อสอบแต่ละค่าจะแตกต่างกันทุกค่า ส่วนในกรณีที่ใช้พื้นที่ระหว่างโค้งลักษณะข้อสอบเป็นตัวดัชนีความลำเอียงของข้อสอบนั้นการคำนวณพื้นที่จะมีความยุ่งยากมาก เพราะพื้นที่ ที่คำนวณได้จะต้องมีทั้งทางบวกและทางลบ เนื่องจากโค้งลักษณะข้อสอบของผู้สอบทั้งสองกลุ่มมีลักษณะตัดกัน

3.6 วิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ทฤษฎีการตอบข้อคำถาม

นิยามความลำเอียงของข้อสอบ ในแง่ของทฤษฎีการตอบข้อคำถามนั้นข้อสอบจะลำเอียง ถ้าฟังก์ชันการตอบข้อสอบข้อนั้น ๆ ของประชากรกลุ่มย่อย ๆ 2 กลุ่ม ไม่เป็นเช่นเดียวกัน ซึ่งจากนิยามดังกล่าวข้างต้น การวิเคราะห์ความลำเอียงของข้อสอบตามแนวทฤษฎีการตอบข้อคำถามสามารถทำได้โดยการเปรียบเทียบฟังก์ชันการตอบข้อสอบของผู้สอบกลุ่มย่อยที่ต้องการศึกษา และการเปรียบเทียบ ในทฤษฎีการตอบข้อคำถามมีวิธีการวิเคราะห์หลายวิธี ซึ่งพอสรุปได้ดังนี้ คือ

1. แบบจำลองโลจิสติกชนิด 1 พารามิเตอร์ ซึ่งกำหนดให้ค่าการเดา (c_i) เท่ากับ 0 และค่าอำนาจจำแนก (a_i) ของข้อสอบทุกข้อ เท่ากับ 1.00 มีค่าความยาก (b_i) เพียงตัวเดียวที่แปรเปลี่ยนไปตามกลุ่มผู้สอน วิธีนี้เป็นการวิเคราะห์ความแตกต่างของค่าพารามิเตอร์ความยาก (b_i) ของข้อสอบจากโค้งลักษณะข้อสอบของแต่ละกลุ่ม

เริ่มด้วยการทดสอบความเหมาะสมของข้อมูลทั้ง 2 กลุ่ม กับแบบจำลอง โดยทดสอบรายข้อด้วย $H_i(\theta)$ ซึ่งมีการกระจายแบบไคสแควร์ (χ^2) มี $df = N-1$ เพื่อตัดข้อที่ไม่เหมาะสมออก แล้วจึงนำไปวิเคราะห์ความลำเอียงรายข้อด้วย Z_i

$$H_i(\theta) = \sum_{j=1}^{N} \frac{\left[U_{ij} - \hat{P}_i\right]^2}{\left[\hat{P}_i \cdot Q_i\right]}$$

เมื่อ U_{ij} แทน คะแนนจากการตอบข้อสอบแบบ 0 – 1 ของคนที่ j ข้อที่ i

$$P_i(\theta) = \frac{1}{1 + e\left[-D(\theta - b_i)\right]}$$

$$Z_i = \frac{b_{iA} - b_{iB}}{SE^2_{biA} + SE^2_{biB}}$$

$$SE^2_{bi} = \sqrt{I_{ai}/(I_{ai} I_{bi} I^2_{abi})}$$

ถ้าค่า Z_i ที่คำนวณได้ค่าสูง แสดงว่า ข้อสอบนั้นยากสำหรับกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่ง หรืออาจจะวิเคราะห์ความลำเอียงเป็นรายข้อด้วย t ซึ่งคำนวณได้ จากสูตร

$$t = \frac{b_{iA} - b_{iB}}{SE^2_{biA} + SE^2_{biB}}$$

เมื่อ $b_{iA} + b_{iB}$ แทน ค่าความยากของข้อสอบที่ได้จากการวิเคราะห์ในกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ

$SE^2_{biA} + SE^2_{biB}$ แทน ค่าความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์จากกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ

ค่า t ที่คำนวณได้ คือ ดัชนีความลำเอียงของข้อสอบ ซึ่ง ดราบา (Berk, 1982 : 138 ; quoting Draba, 1977) เสนอให้ใช้ค่า $t \geq 2.4$ เป็นเกณฑ์ในการตัดสินว่าข้อสอบมีความลำเอียง สำหรับวิธีการนี้เรียกอีกอย่างหนึ่งว่า เป็นการวิเคราะห์ความลำเอียงตามแบบจำลองของ ราส์ช (Rasch Model)

2. แบบจำลองโลจีสติกชนิด 2 พารามิเตอร์ กำหนดให้ค่าการเดา (c_i) เท่ากับ 0 ส่วนค่าความยาก (b_i) และค่าอำนาจจำแนก (a_i) แปรเปลี่ยนไปตามกลุ่มผู้สอบ วิธีนี้เริ่มด้วยการประมาณค่าพารามิเตอร์ของข้อสอบแยกตามกลุ่มผู้สอบ กลุ่มย่อย A และ B ด้วยวิธี Maximum Likelihood แล้วเปรียบเทียบโค้งลักษณะการตอบข้อสอบด้วยค่า F ซึ่งคำนวณได้จากสูตร

$$F = \frac{SSE(Pooled) - [SSE(A) - SSE(B)]}{SSE(A) + SSE(B)} \cdot \frac{J_A + J_B - 4}{2}$$

เมื่อ $SSE(A)$ แทน ความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์กลุ่มที่ 1

$SSE(B)$ แทน ความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์กลุ่มที่ 2

$SSE(Pooled)$ แทน ความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์กลุ่มที่ 1 และกลุ่มที่ 2

J_A, J_B แทน จำนวนช่วงความสามารถ

และมีค่าชั้นแห่งความเป็นอิสระ คือ $2,(J_A + J_B - 4)$ ซึ่งการทดสอบด้วยค่า F จะบอกว่าเส้นถดถอยรวมสามารถอธิบายความแปรปรวนของสัดส่วนการตอบของเส้นถดถอยจากแต่ละกลุ่มได้เท่ากันหรือไม่ ถ้าเท่ากันหรือไม่มีนัยสำคัญทางสถิติก็ไม่จำเป็นต้องใช้เส้นถดถอยแยกกันเพราะความสัมพันธ์ระหว่างความสามารถและการตอบข้อสอบของผู้สอบทั้งสองกลุ่มเป็นเช่นเดียวกัน และข้อสอบก็ทำหน้าที่ในการวัดความสามารถของคน 2 กลุ่มนี้ได้เหมือนกัน

3. แบบจำลองโลจีสติกชนิด 3 พารามิเตอร์ เป็นแบบที่ให้ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i) และค่าสัมประสิทธิ์การเดา (c_i) ของข้อสอบ แปรเปลี่ยนไปตามกลุ่มย่อยของผู้สอบ

การคำนวณหาดัชนีความลำเอียงของข้อสอบวิธีนี้ เริ่มจากการประมาณค่าความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบเป็นรายข้อรวมของทุกคนทุกกลุ่ม แล้วแปลงค่าพารามิเตอร์ที่ได้จากผู้สอบ 2 กลุ่ม ให้อยู่บนมาตรฐานเดียวกันบนค่าประมาณความยาก จากนั้นแยกประมาณค่าความยาก และค่าอำนาจจำแนกตามกลุ่มย่อย โดยใช้ค่าการเดาที่ประมาณได้จากการรวมทุกกลุ่ม การทำค่าความยากให้เป็นมาตรฐานนั้นจะตัดข้อสอบที่มีค่าความยากสูงมากหรือต่ำมากออก

การวิเคราะห์ความลำเอียงของข้อสอบตามแบบจำลอง โลจีสติก 3 พารามิเตอร์ มีหลายวิธี เช่น ลินน์ ลีไวน์ ฮาสติ้งส และวอร์ดรอพ (Hulin, Drasgows and Parson, 1983 : 776 ; quoting Linn, Levine, Hastings and Wardrop, 1981) ได้พัฒนาการทดสอบความลำเอียงของข้อสอบจากความแตกต่างระหว่างโค้งลักษณะของข้อสอบของผู้สอบ 2 กลุ่มย่อย เป็นเกณฑ์การตัดสินความลำเอียงของข้อสอบ ผู้ที่ศึกษาความลำเอียงโดยการเปรียบเทียบความแตกต่างของโค้งลักษณะข้อสอบ เช่น รัดเนอร์ (Rudner, 1977 : quoting in Berk, 1982 : 118) รัดเนอร์ เกทสัน และไนท์ (Rudner, Getson and Knight, 1980) ไรออนสัน และสับโคเวียต (Ironson and Subkoviat, 1979) ซึ่งเกณฑ์ที่ใช้พิจารณาความลำเอียง คือ มีพื้นที่ใต้โค้งต่างกันตั้งแต่ .20 ขึ้นไป ในขณะที่ ทัศนีย์ พรมนตรี ใช้ค่ามากกว่า .40 การใช้เกณฑ์ความลำเอียง .20 หรือ .40 แทนที่จะใช้ 0.00 นั้น เนื่องจากยอมให้มีความคลาดเคลื่อนที่อาจจะเกิดจากการสุ่ม ผู้ที่นำดัชนีนี้ไปใช้และใช้เกณฑ์ที่ต่างกัน ได้แก่ คิม และโคเฮน (Kim and Cohen, 1991) โดยมีข้อตกลงเบื้องต้นว่าการกระจายของความแตกต่างของพื้นที่ใต้โค้งที่วัดได้มีลักษณะเป็นโค้งปกติ เกณฑ์การตัดสินความลำเอียง จะพิจารณาค่าที่มากกว่าค่าวิกฤตชนิดทางเดียวที่ระดับ .05 และ .01

นอกจากนี้ ลินน์ และคนอื่น (Linn and Others, 1981) ได้อธิบายเกี่ยวกับการหาดัชนีความลำเอียงของข้อสอบ โดยใช้รากที่สองของค่าเฉลี่ยความแตกต่างกำลังสองของโค้งลักษณะข้อสอบของกลุ่มผู้สอบ 2 กลุ่ม เป็นดัชนีความลำเอียงของข้อสอบในช่วงความสามารถ -3 ถึง 3 ที่ได้จากการคำนวณ จากสูตร ดังนี้

$$RMSD = \frac{1}{600}\left[\sum_{J=1}^{J}\left[P_{iA}(\theta_J) - P_{iB}(\theta_J)\right]^2\right]^{1/2}$$

โดยได้จากการแบ่งช่วงความสามารถจาก -3 ถึง 3 เป็น 600 ช่วง ค่าสัมบูรณ์ของความแตกต่างของโค้งทั้ง 2 ที่จุดกึ่งกลางของ 600 ช่วง จะเป็น $P_{iA}(\theta_j) - P_{iB}(\theta_j)$ คูณด้วยความยาวของช่วงแต่ละช่วง คือ .01 การบวกค่าพื้นที่ทั้ง 600 ช่วง จะได้ค่าประมาณของพื้นที่ ทั้งหมดที่อยู่ระหว่างโค้งทั้งสอง