

การเดากับการพัฒนาโมเดลทฤษฎีการตอบสนองข้อสอบทางเลือกที่คำนึงถึงการเดา Guessing and Development of Alternative IRT Models with Guessing

ธีรุตม์ สุขสกุลวัฒน์^{1*} ณัฐภรณ์ หลาวทอง² และ สิวะโชติ ศรีสุทธียากร³

Teerut Suksakulwat^{1*} Nuttaporn Lawthong² and Siwachoat Srisuttiyakorn³

(Received: July 4, 2020 ; Revised: August 11, 2020 ; Accepted: August 21, 2020)

บทคัดย่อ

การเดาเป็นปัญหาหนึ่งที่หลีกเลี่ยงไม่ได้สำหรับการวัดและประเมินผลด้วยแบบสอบปรนัย แม้ในเบื้องต้นจะมีการป้องกันการเดาด้วยวิธีการต่าง ๆ แต่ก็ไม่สามารถกระทำได้อย่างสมบูรณ์ เนื่องจากปัจจัยที่ส่งผลให้เกิดการเดาอาจมาได้จากหลายสาเหตุ จากข้อจำกัดดังกล่าวจึงเป็นเหตุผลของการพัฒนาโมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) ในรูปแบบต่าง ๆ เพื่อวัดค่าการเดาที่เกิดขึ้น โดยประเด็นที่น่าสนใจในบทความมีดังนี้ 1) การเดาคำตอบในข้อสอบปรนัย แบ่งเป็น 2 รูปแบบคือ การเดาแบบสุ่ม และการเดาแบบอิงความสามารถ 2) ปัจจัยด้านคุณลักษณะของข้อสอบและผู้สอบที่ส่งผลต่อการเดา 3) โมเดล IRT ได้แก่ โมเดล Rasch 2PL และ 3PL 4) โมเดล IRT ทางเลือกที่คำนึงถึงการเดา ได้แก่ โมเดล AG FG3PL 2PLG 3P-RH และ 2PLE และ 5) การเปรียบเทียบความแตกต่างระหว่างโมเดล IRT ทางเลือกที่คำนึงถึงการเดาแต่ละรูปแบบ

คำสำคัญ โมเดลทฤษฎีการตอบสนองข้อสอบทางเลือก การเดา ความรู้บางส่วน

Abstract

Guessing is an unavoidable problem for measurement and evaluation with multiple-choice tests. Although preliminary prevention is performed using various methods, it may not be perfect because the factors that cause the guessing may come from many reasons. These

¹ นิสิตปริญญาเอก สาขาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

² รองศาสตราจารย์ ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

³ อาจารย์ ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

¹ Doctoral degree student in Educational Measurement and Evaluation Program, Faculty of Education, Chulalongkorn University

² Associate Professor, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

³ Lecturer, Department of Educational Research and Psychology, Faculty of Education, Chulalongkorn University

*Corresponding Author E-mail: teerut1234@gmail.com

limitations became the rationale for developing different types of item response theory (IRT) model to measure the guessing. The issues presented in the article are as follows: 1) Guessing in multiple-choice tests has two types: random guessing and ability-based guessing; 2) Characteristics of test and test takers factors that affect guessing; 3) IRT models are Rasch, 2PL and 3PL models; 4) The alternative IRT models with guessing are AG, FG3PL, 2PLG, 3P-RH and 2PLE models, and 5) Comparison of differences between alternative IRT models with guessing.

Keywords: alternative item response theory model, guessing, partial knowledge

บทนำ

แบบสอบปรนัยแบบหลายตัวเลือก (multiple-choice test) เป็นเครื่องมือวัดและประเมินผลที่นิยมใช้อย่างแพร่หลายในการทดสอบวัดความถนัดและผลสัมฤทธิ์ทางการศึกษา เนื่องจากเหตุผลเชิงประสิทธิภาพ เช่น ใช้เวลาบริหารจัดการน้อย มีค่าใช้จ่ายต่ำ ตรวจให้คะแนนได้อย่างเป็นปรนัย และใช้เครื่องจักรตรวจได้อย่างไรก็ตามแบบสอบดังกล่าวยังมีข้อจำกัดสำคัญ 2 ประการ คือ 1) มักเกิดการเดาคำตอบ (guessing) เมื่อผู้ตอบมีความรู้ไม่เพียงพอในการตอบข้อคำถาม ความแปรปรวนของคะแนนความคลาดเคลื่อน (error variance) ที่วัดได้จากแบบสอบจึงมีแนวโน้มสูง ส่งผลให้ความเที่ยงและความตรงต่ำ (Budesu & Bo, 2015) และ 2) ไม่สามารถวัดผู้ตอบที่มีความรู้บางส่วน (partial knowledge) เนื่องจากการตรวจให้คะแนนโดยทั่วไปมีเพียงสองค่าคือ 0 (ผิด) กับ 1 (ถูก) จึงไม่เพียงพอต่อการจำแนกผู้ตอบกลุ่มนี้ออกจากกลุ่มที่มีความรู้สมบูรณ์ (complete knowledge) และกลุ่มที่ไม่มีความรู้เลย (no knowledge) (Bo, Lewis, & Budesu, 2015)

จากที่กล่าวมาข้างต้น การเดาจึงเป็นสิ่งที่ส่งผลกระทบต่อ การวัดและประเมินผลด้วยแบบสอบปรนัยที่มีการตรวจให้คะแนนได้สองค่ามากที่สุด เนื่องจากการเปิดช่องให้ผู้ตอบที่มีความรู้ไม่เพียงพอมีโอกาสได้คะแนนจากการตอบ คะแนนที่วัดได้จึงไม่สะท้อนถึงความรู้จริงของผู้ตอบ และอาจเกิดข้อผิดพลาดในการนำผลคะแนนไปใช้เพื่อประเมินหรือตัดสิน โดยเฉพาะเรื่องที่มีความสำคัญและไม่มีโอกาสติดตามตรวจสอบได้อีก เช่น การสอบเพื่อสำเร็จการศึกษา การสอบคัดเลือกเพื่อเข้าศึกษาต่อ (ศิริชัย กาญจนวาสี, 2556)

ในอดีตแม้จะมีการป้องกันการเดาด้วยวิธีต่าง ๆ เช่น การหักคะแนนเมื่อเดาผิด การเปลี่ยนรูปแบบการตอบจากการเลือกตัวถูกเป็นการเลือกตัวลง การปรับข้อสอบให้มีตัวถูกหลายตัว (Bo et al., 2015) อย่างไรก็ตามวิธีดังกล่าวพบว่ามีข้อจำกัดในเชิงปฏิบัติหลายประการ เช่น การสอบต้องใช้เวลาานกว่าปกติ การสร้างเครื่องมือมีค่าใช้จ่ายสูง การคำนวณคะแนนมีความซับซ้อน รวมถึงผลการวัดยังอาจเปลี่ยนไปตามกลุ่มผู้สอบ เนื่องจากในสถานการณ์จริงอาจมีผู้สอบหลายประเภท เช่น กลุ่มไม่ชอบเสี่ยง กลุ่มไม่ชอบสูญเสีย กลุ่มอนุรักษ์นิยมการทำข้อสอบแบบดั้งเดิม เป็นต้น ผู้สอบแต่ละกลุ่มจึงอาจถูกกลบโห้อย่างไม่เป็นสัดส่วน ส่งผลทำให้การประมาณค่าคะแนนเกิดความลำเอียง (Budesu & Bo, 2015) นอกจากนี้ยังมีปัจจัยอื่น ๆ ที่อาจส่งผลต่อการเดา เช่น ความยาก อำนาจจำแนก จำนวนตัวเลือก ความสามารถ แรงจูงใจในการสอบ จึงเป็นเรื่องยากในเชิงปฏิบัติที่จะป้องกันการเดาด้วยวิธีดังกล่าว (Han, 2012; San Martin, Del Pino, & De Boeck, 2006)

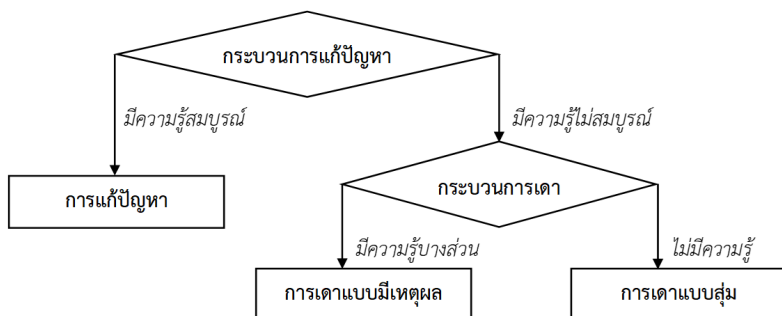
ทฤษฎีการทดสอบแนวใหม่ (modern test theory) จึงเข้ามามีบทบาทในการศึกษาการเดาแทน ทฤษฎีการทดสอบแบบดั้งเดิม ด้วยการอาศัยโมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) เข้ามาอธิบายความสัมพันธ์ระหว่างความสามารถของผู้สอบกับโอกาสในการตอบข้อสอบถูก โดยประเภทของโมเดล IRT ที่นิยมใช้กับแบบสอบมาตรฐานทั่วไปก็คือ โมเดล IRT ที่มีการตรวจให้คะแนนได้สองค่า เช่น โมเดล Rasch โมเดลโลจิสต์ 2 พารามิเตอร์ (2PL: two-parameter logistic) และโมเดลโลจิสต์ 3 พารามิเตอร์ (3PL)

อย่างไรก็ตามโมเดล IRT ที่กล่าวมา พบว่าอาจมีข้อจำกัดเกี่ยวกับการศึกษาการเดาบางประการ ตัวอย่างเช่น การวิเคราะห์ด้วยโมเดล Rasch หรือ 2PL ซึ่งไม่มีการคำนึงถึงการเดา อาจส่งผลทำให้สารสนเทศเกี่ยวกับความรู้บางส่วนของผู้สอบที่ใช้ในการเดาถูกละเลยไป (Andrich & Marais, 2014) ส่วนโมเดล 3PL แม้จะมีการคำนึงถึงการเดา แต่ก็อาจมีข้อจำกัดเกี่ยวกับการประมาณค่าและการแปลผลการเดา (Lee & Bolt, 2018) เนื่องจากผลการศึกษาเชิงจำลองพบว่าค่าประมาณพารามิเตอร์การเดามักมีแนวโน้มคลาดเคลื่อนไปจากค่าจริงในหลายกรณีศึกษา (Thissen & Wainer, 1982) รวมถึงการนิยามให้การเดาเป็นพารามิเตอร์ของข้อสอบ ซึ่งโอกาสในการเดาถูกไม่มีความผันแปรกับความสามารถของผู้สอบ การแปลผลจึงไม่สอดคล้องกับสภาพธรรมชาติเท่าที่ควร (Hambleton, Swaminathan, & Rogers, 1991; San Martín et al., 2006)

จากเหตุผลดังกล่าวจึงเป็นที่มาของการพัฒนาโมเดล IRT ทางเลือกที่คำนึงถึงการเดา (Han, 2012; San Martín et al., 2006) ซึ่งเกิดจากการบูรณาการระหว่างปรัชญาความเชื่อการเดาต่าง ๆ กับโมเดล IRT แบบดั้งเดิม ทำให้โมเดลมีคุณสมบัติสามารถหลีกเลี่ยงปัญหาจากการประมาณค่าพารามิเตอร์การเดา และยังทำให้การแปลผลสอดคล้องกับสภาพธรรมชาติมากขึ้น ทั้งนี้ในบทความจะขอยกตัวอย่างเพียง 5 โมเดล ได้แก่ โมเดล AG FG3PL 2PLG 3P-RH และ 2PLE เพื่อนำเสนอถึงแนวคิดในการบูรณาการระหว่างการเดากับโมเดล IRT รวมถึงจะเปรียบเทียบความแตกต่างระหว่างโมเดลในประเด็นต่าง ๆ ต่อไป

การเดาคำตอบในข้อสอบปรนัย

การเดาแบ่งเป็น 2 รูปแบบ คือ 1) การเดาแบบสุ่ม (blind-, wild-, lucky- or random guessing) เป็นการสุ่มเลือกคำตอบจากบรรดาตัวเลือกทั้งหมด เกิดขึ้นเมื่อผู้ตอบไม่มีความรู้ใดเลยเกี่ยวกับคำถาม และ 2) การเดาแบบอิงความสามารถ (informed-, educated-, logical- or ability-based guessing) เป็นการเลือกคำตอบโดยอาศัยการตัดสินใจเลือกที่คาดว่าจะผิดออกไปก่อน จากนั้นจึงเดาแบบสุ่มจากตัวเลือกที่เหลืออยู่ เกิดขึ้นเมื่อผู้ตอบมีความรู้บางส่วนเกี่ยวกับคำถาม (Hutchinson, 1991; San Martín et al., 2006) ดังภาพ 1



ภาพ 1 โมเดลอธิบายการเดาตามแนวคิดของ Hutchinson (1991)

รูปแบบการเดาทั้งสองนี้มีความแตกต่างกันในการให้สารสนเทศเกี่ยวกับความรู้ของผู้ตอบ กล่าวคือ การเดาแบบสุ่มจะถือว่าไม่มีการให้สารสนเทศใดทั้งสิ้น เนื่องจากการเดาไม่ได้มาจากการใช้ความรู้ของผู้ตอบ แตกต่างจากการเดาแบบอิงความสามารถที่ถือว่ามีการให้สารสนเทศบางส่วน เนื่องจากการเดามีการนำความรู้บางส่วนของผู้ตอบเข้ามาช่วยในการเดาด้วย (Smith, 1993)

ปัจจัยที่ส่งผลต่อการเดา

จากการสังเคราะห์งานวิจัยเกี่ยวกับการเดา พบว่าปัจจัยที่ส่งผลต่อการเดาอาจมาจาก 2 ด้าน คือ 1) คุณลักษณะของข้อสอบ และ 2) คุณลักษณะของผู้สอบ รายละเอียดดังนี้

1) คุณลักษณะของข้อสอบ

1.1) *ความยาก* เมื่อข้อสอบมีระดับความยากสูงกว่าความรู้ของผู้ตอบ ผู้ตอบจะเริ่มมีแนวโน้มคาดเดาคำตอบมากขึ้น โดยข้อสอบระดับง่ายมักเปิดช่องให้มีโอกาสในการเดาถูกสูงกว่าปกติ ตรงกันข้ามกับข้อสอบระดับยาก ซึ่งพบว่าโอกาสในการเดาถูกมักมีแนวโน้มลดต่ำลง (Tsai & Lin, 2015) นอกจากนี้ตำแหน่งของข้อสอบที่มีระดับความยากแตกต่างกันก็อาจส่งผลต่อการเดาเช่นกัน ตัวอย่างเช่น หากข้อสอบในช่วงเริ่มต้นยากเกินไป ผู้ตอบอาจหมดกำลังใจในการสอบและเริ่มเดาคำตอบตั้งแต่ช่วงแรก หรือถ้าข้อสอบในช่วงท้ายยากมากขึ้นและผู้ตอบรู้สึกเหนื่อยล้า ผู้ตอบก็อาจเริ่มเดาคำตอบเช่นกัน (Cao & Stokes, 2008)

1.2) *อำนาจจำแนก* ข้อสอบที่มีอำนาจจำแนกต่ำหรือไม่สามารถจำแนกผู้ตอบที่มีความสามารถสูงและต่ำออกจากกันได้ มักเปิดช่องให้ผู้ตอบทั้งสองกลุ่มมีโอกาสดูถูกใกล้เคียงกัน และการที่ผู้ตอบกลุ่มต่ำมีโอกาสในการตอบถูกใกล้เคียงกับผู้ตอบกลุ่มสูง ย่อมหมายความว่าข้อสอบข้อนั้นมีโอกาสในการเดาถูกสูงหรือเดาถูกได้ง่าย (Hambleton et al., 1991)

1.3) *จำนวนตัวเลือก* หากจำนวนตัวเลือกในข้อสอบมีมากขึ้น โอกาสในการเดาถูกย่อมต่ำลง สังเกตได้จากโอกาสในการเดาถูกแบบสุ่ม ซึ่งมีค่า $1/m$ เมื่อ m คือจำนวนตัวเลือกของข้อสอบ (Han, 2012)

1.4) *คุณภาพตัวลวง* ตัวลวงที่มีคุณภาพจะมีผลต่อการเดาเป็นอย่างมาก โดยเฉพาะในกลุ่มผู้ตอบความรู้ต่ำ ซึ่งพบว่ามีแนวโน้มสามารถล่อลวงผู้ตอบความรู้ต่ำหรือมีความรู้ไม่สมบูรณ์ได้ดี การเดาของผู้ตอบกลุ่มนี้จึงมักไม่ใช้การเดาแบบสุ่ม (Lord, 1974)

1.5) *วิชาที่สอบ/ลักษณะตัวเลือก* โดยทั่วไปในวิชาคณิตศาสตร์ ตัวเลือกในข้อสอบมักมีลักษณะเป็นตัวเลข ซึ่งมีจุดเด่นคือผู้ตอบที่มีความรู้สมบูรณ์มักตอบข้อสอบได้ทั้งหมด แต่หากไม่มีความรู้หรือมีความรู้เพียงบางส่วนจะตอบไม่ได้เลย (all-or-none) การเดาในวิชาคณิตศาสตร์จึงมักเป็นการเดาแบบสุ่ม ส่วนในวิชาทางภาษา ตัวเลือกในข้อสอบมักมีลักษณะเป็นข้อความ ผู้ตอบที่มีความรู้ไม่เพียงพอในการตอบสามารถอ่านข้อความในแต่ละตัวเลือก และใช้ความรู้ที่มีตัดตัวเลือกที่คาดว่าจะไม่ถูกต้องออกไปก่อน แล้วจึงเดาแบบสุ่มจากตัวเลือกที่เหลือ การเดาในวิชาทางภาษาจึงมักเป็นการเดาแบบอิงความสามารถ (San Martín et al., 2006)

1.6) *วิธีการตอบ/การตรวจให้คะแนน* ข้อสอบปรนัยที่มีตัวถูกหลายตัว (multiple correct options) และมีการตรวจให้คะแนนบางส่วน (partial credit) พบว่าสามารถช่วยลดโอกาสในการเดาถูกของผู้ตอบลงไปได้มาก เมื่อเทียบกับข้อสอบปรนัยแบบมีตัวถูกเพียงตัวเดียว (Ben-Simon et al., 1997)

2) คุณลักษณะของผู้สอบ

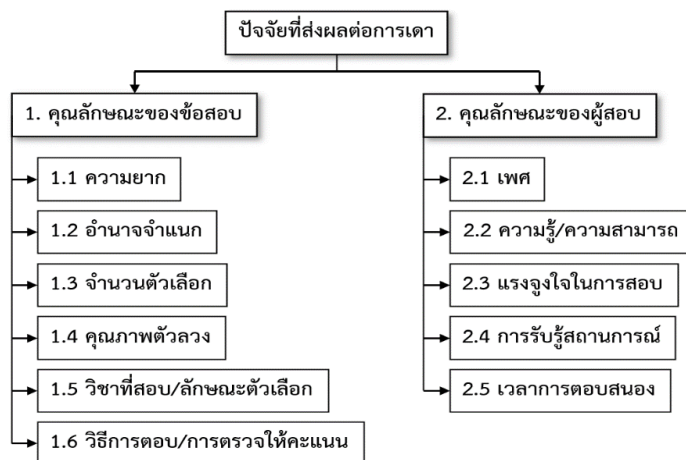
2.1) *เพศ* การศึกษาความแตกต่างของพฤติกรรมการสอบระหว่างเพศชายกับเพศหญิงในบริบทการสอบที่มีความสำคัญสูง (high-stakes testing) เช่น การสอบเลื่อนชั้น การสอบคัดเลือกเพื่อเข้าศึกษาต่อ มักพบว่าเพศชายมีพฤติกรรมชอบเสี่ยงเดาคำตอบมากกว่าเว้นคำตอบไว้ ตรงกันข้ามกับเพศหญิงที่ไม่ชอบเสี่ยงเดาคำตอบแม้จะไม่มีภาระกดดันเมื่อเดาคิดก็ตาม (Baldiga, 2014) ทั้งนี้ผลดังกล่าวอาจมาจากความแตกต่างระหว่างเพศเกี่ยวกับแนวคิดความเป็นเลิศในการแข่งขันในสถานการณ์ที่มีความเสี่ยง โดยเพศชายอาจคิดว่าการได้ผลงานดี คือการเพิ่มคะแนนสอบให้ได้สูงสุดหรือการไม่ยอมรับว่าไม่ทราบคำตอบ จึงตัดสินใจเสี่ยงเดาคำตอบมากกว่าเว้นคำตอบ ส่วนเพศหญิงอาจคิดว่าการได้ผลงานดี คือการไม่ต้องรับการลงโทษใด จึงตัดสินใจเว้นคำตอบมากกว่าเสี่ยงเดาคำตอบ (Croson & Gneezy, 2009)

2.2) *ความรู้/ความสามารถ* ผู้ตอบความสามารถต่ำมักถูกล่อลวงด้วยตัวลวงที่มีคุณภาพ ทำให้โอกาสในการเดาถูกมีแนวโน้มต่ำกว่าโอกาสในการเดาถูกแบบสุ่ม (Lord, 1974) ส่วนผู้ตอบความสามารถสูงมักใช้ความรู้ที่มิดัดตัวเลือกที่คาดว่าจะไม่ถูกต้องออกไปก่อน แล้วจึงเดาคำตอบจากตัวเลือกที่เหลือ โอกาสในการเดาถูกจึงมีแนวโน้มสูงกว่าโอกาสในการเดาถูกแบบสุ่ม (San Martín et al., 2006)

2.3) *แรงจูงใจในการสอบ* แรงจูงใจในการสอบมักมีความสัมพันธ์กับความสำคัญของการสอบและอาจส่งผลต่อการเดา โดยเฉพาะการสอบที่มีความสำคัญต่ำ (low-stakes testing) เช่น การสอบเพื่อติดตามพัฒนาการของผู้เรียน การสอบประเมินระดับชาติที่ไม่มีผลได้ผลเสียต่อตัวผู้สอบโดยตรง เป็นต้น การสอบในลักษณะนี้ผู้ตอบมักให้ความสนใจและพยายามทำข้อสอบในช่วงแรกเท่านั้น เนื่องด้วยความอยากรู้อยากเห็น แต่พอถึงจุดหนึ่งผู้ตอบจะละทิ้งความพยายามและเริ่มเดาคำตอบ นอกจากนี้การสอบดังกล่าวมักใช้ข้อสอบที่มีจำนวนมากและมีรูปแบบการตอบที่หลากหลาย ผู้ตอบจึงเกิดความเมื่อยล้าจากการทำข้อสอบได้ง่าย เป็นผลทำให้เกิดการเดามากขึ้นในช่วงท้ายของการทดสอบ (Cao & Stokes, 2008)

2.4) *การรับรู้สถานการณ์* เมื่อมีการลงโทษหรือหักคะแนนจากการเดาคิด ผู้ตอบที่คาดการณ์ได้ว่ามีคะแนนสูงกว่าคะแนนที่ยอมรับได้ต่ำสุด (minimal acceptable point) หรือคะแนนอ้างอิง มักจะเว้นคำตอบในข้อที่ไม่แน่ใจไว้ เพราะเชื่อมั่นว่ายังได้คะแนนอยู่ในตำแหน่งเดิม จึงไม่จำเป็นต้องเสี่ยงเดาคำตอบ ส่วนผู้ตอบที่คาดการณ์ว่ามีคะแนนต่ำกว่าคะแนนอ้างอิง มักมีแนวโน้มเดาคำตอบมากกว่า เนื่องจากต้องเสี่ยงเพื่อเพิ่มโอกาสให้ได้อยู่ในตำแหน่งที่สูงกว่าคะแนนอ้างอิง (Bereby-Meyer, Meyer, & Flascher, 2002)

2.5) *เวลาการตอบสนอง* เมื่อใช้การสอบผ่านทางคอมพิวเตอร์ (computer-based test) เวลาในการตอบสนองข้อสอบอาจเข้ามามีบทบาทเป็นตัวจำแนกความแตกต่างระหว่างกระบวนการแก้ปัญหากับกระบวนการเดา ด้วยการพิจารณาจากขีดแบ่งเวลา (time threshold) ที่ใช้ในการทำข้อสอบแต่ละข้อ โดยการตอบที่ใช้เวลาน้อยกว่าขีดแบ่งเวลา มีแนวโน้มสูงที่จะเป็นการเดาคำตอบ ส่วนการตอบที่ใช้เวลามากกว่าขีดแบ่งเวลา จะมีแนวโน้มเป็นการแก้ปัญหามากกว่าการเดา (Wise & Kong, 2005) รายละเอียดโดยสรุปดังภาพ 2



ภาพ 2 ปัจจัยที่ส่งผลต่อการเดาในข้อสอบปรนัย

โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT)

ทฤษฎีการทดสอบแนวใหม่ได้เข้ามามีบทบาทสำคัญในการอธิบายถึงความสามารถของผู้สอบด้วยแนวทางที่แตกต่างไปจากทฤษฎีการทดสอบแบบดั้งเดิม ด้วยการอาศัยโมเดล IRT เข้ามาช่วยอธิบายความสัมพันธ์ระหว่างคุณลักษณะของข้อสอบและผู้สอบกับความน่าจะเป็นในการตอบข้อสอบถูก ในรูปโค้งลักษณะข้อสอบ (ICCs: item characteristic curves) หรือฟังก์ชันการตอบสนองข้อสอบ ซึ่งโดยทั่วไปนิยมเขียนอยู่ในรูปฟังก์ชันโลจิส (ศิริชัย กาญจนวาสี, 2555)

โมเดล IRT มีข้อตกลงเบื้องต้นสำคัญ 3 ประการ คือ 1) ความเป็นเอกมิติ (unidimensionality) หมายถึงข้อสอบทุกข้อมุ่งวัดเพียงความสามารถเดียว 2) ความเป็นอิสระ (local independence) ระหว่างผู้สอบกับข้อสอบ กล่าวคือผลการตอบข้อสอบข้อเดียวกันของผู้สอบแต่ละคนเป็นอิสระจากกัน (ความเป็นอิสระระหว่างผู้สอบ) และผลการตอบข้อสอบแต่ละข้อของผู้สอบคนเดียวกันเป็นอิสระจากกัน (ความเป็นอิสระระหว่างข้อสอบ) ด้วยคุณสมบัติดังกล่าว พารามิเตอร์ของผู้สอบและข้อสอบจึงไม่แปรเปลี่ยน (invariance) ไปตามกลุ่มผู้สอบหรือชุดข้อสอบที่นำมาใช้ แตกต่างจากทฤษฎีการทดสอบแบบดั้งเดิมที่มักให้ผลการวัดเปลี่ยนไปตามกลุ่มผู้สอบหรือชุดข้อสอบ และ 3) การเป็นฟังก์ชันต่อเนื่อง (continuous function) ที่สามารถอธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะแฝงกับความน่าจะเป็นในการตอบข้อสอบได้ถูก (ศิริชัย กาญจนวาสี, 2555; Reckase, 2009) โดยบทความนี้จะนำเสนอโมเดล IRT ที่มีการตรวจให้คะแนนได้สองค่าเพียง 3 โมเดล ได้แก่ โมเดล Rasch 2PL และ 3PL เนื่องจากโมเดลทั้งสามเป็นโมเดลมาตรฐานที่นิยมใช้อย่างแพร่หลายในการวิเคราะห์แบบสอบมาตรฐานทั่วไป รายละเอียดของแต่ละโมเดลมีดังนี้

1) โมเดลโลจิส 1 พารามิเตอร์ (1PL) หรือโมเดล Rasch

โมเดล Rasch มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (1) ซึ่งแสดงถึงความน่าจะเป็นในการตอบข้อสอบ i ได้ถูกต้องของผู้สอบ j ($y_{ij} = 1$ หมายถึงตอบถูก; $y_{ij} = 0$ หมายถึงตอบผิด)

$$P(y_{ij}=1 | \theta_j) = \frac{1}{1 + \exp[-(\theta_j - b_i)]} \quad (1)$$

เมื่อ θ_j คือพารามิเตอร์ความสามารถของผู้สอบ j และ b_i คือพารามิเตอร์ความยากของข้อสอบ i

2) โมเดลโลจิสต์ 2 พารามิเตอร์ (2PL)

โมเดล 2PL มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (2)

$$P(y_{ij}=1|\theta_j) = \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \quad (2)$$

เมื่อ a_i คือพารามิเตอร์อำนาจจำแนกของข้อสอบ i

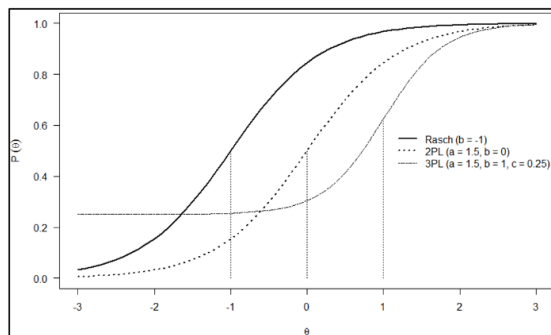
3) โมเดลโลจิสต์ 3 พารามิเตอร์ (3PL)

โมเดล 3PL มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (3)

$$P(y_{ij}=1|\theta_j) = c_i + (1 - c_i) \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \quad (3)$$

เมื่อ c_i คือพารามิเตอร์การเดาของข้อสอบ i หรืออาจเรียกว่าพารามิเตอร์การเดาเทียม (pseudo guessing parameter) เนื่องจากการแปลผลยังไม่สอดคล้องกับสภาพธรรมชาติเท่าที่ควร (Hambleton et al., 1991)

ตัวอย่างของโค้งลักษณะข้อสอบ (ICCs) ของโมเดล Rasch ($b = -1$) โมเดล 2PL ($a = 1.5$, $b = 0$) และโมเดล 3PL ($a = 1.5$, $b = 1$, $c = 0.25$) แสดงดังภาพ 3



ภาพ 3 ตัวอย่าง ICCs ของโมเดล Rasch 2PL และ 3PL

โมเดล IRT ทางเลือกที่คำนึงถึงการเดา (alternative IRT model with guessing)

จากการศึกษาในหลายงานวิจัยพบว่า โมเดล IRT ที่กล่าวมาข้างต้นอาจมีข้อจำกัดเกี่ยวกับการศึกษาการเดา ตัวอย่างเช่น การนำโมเดล Rasch และ 2PL มาวิเคราะห์ข้อสอบปรนัยแบบหลายตัวเลือก ซึ่งข้อสอบรูปแบบนี้เปิดช่องให้ผู้สอบสามารถนำความรู้บางส่วนเข้ามาช่วยในการเดาได้ แต่เนื่องจากโมเดล Rasch และ 2PL ไม่มีการคำนึงถึงการเดา สารสนเทศเกี่ยวกับความรู้บางส่วนของผู้สอบที่ใช้ในการเดาจึงอาจถูกละเลยไป (Andrich & Marais, 2014) ส่วนโมเดล 3PL แม้จะมีการคำนึงถึงการเดา แต่ก็อาจมีข้อจำกัดเกี่ยวกับการประมาณค่าและการแปลผล (Lee & Bolt, 2018) โดยพบว่าค่าประมาณพารามิเตอร์การเดา (c) มักมีแนวโน้มคลาดเคลื่อนไปจากค่าจริงค่อนข้างสูงในหลายกรณีศึกษา ทั้งนี้อาจเป็นเพราะว่า c คือค่าความน่าจะเป็นที่ผู้สอบความสามารถต่ำมากจะตอบข้อสอบได้ถูกต้อง แต่โดยธรรมชาติของข้อมูลทางการศึกษา ผู้สอบที่มีคุณลักษณะดังกล่าวมักมีอยู่น้อยมาก การประมาณค่าโดยอิงจากข้อมูลที่มีอยู่น้อยจึงมักมีความคลาดเคลื่อนสูง (Thissen & Wainer, 1982) นอกจากนี้การนิยามให้ c เป็นพารามิเตอร์ของข้อสอบ ไม่ใช่พารามิเตอร์ของผู้สอบ ยังเป็นการละเลยข้อเท็จจริงที่ว่าโอกาสในการเดาถูกอาจแปรผันตามความสามารถของผู้สอบแต่ละคนได้

ประกอบกับโมเดลยังไม่มีการระบุถึงรูปแบบของการเดาว่าเป็นแบบใด การแปลผลการเดาจึงค่อนข้างคลุมเครือและไม่สอดคล้องกับสภาพธรรมชาติเท่าที่ควร (Hambleton et al., 1991; San Martín et al., 2006)

จากเหตุผลดังกล่าวจึงเป็นที่มาของการพัฒนาโมเดล IRT ทางเลือกที่คำนึงถึงการเดา (Han, 2012; San Martín et al., 2006) ซึ่งเกิดจากการบูรณาการระหว่างปรัชญาความเชื่อการเดาต่าง ๆ กับโมเดล IRT แบบดั้งเดิม ทำให้โมเดลมีคุณสมบัติสามารถจำแนกกระบวนการแก้ปัญหาและกระบวนการเดาออกจากกันได้ รวมถึงบางโมเดล ในกระบวนการเดาก็อาจมีการจำแนกการเดาแบบอิงความสามารถออกจากการเดาแบบสุ่มเช่นกัน ซึ่งคุณสมบัติดังกล่าวอาจมีส่วนช่วยผ่อนคลายข้อจำกัดเกี่ยวกับการประมาณค่าและทำให้การแปลผลการเดาสอดคล้องกับสภาพธรรมชาติมากขึ้น ทั้งนี้ในบทความจะขอยกตัวอย่างโมเดล IRT ทางเลือกที่คำนึงถึงการเดาเพียง 5 โมเดล ได้แก่ โมเดล AG FG3PL 2PLG 3P-RH และ 2PLE รายละเอียดดังนี้

1) โมเดล ability-based guessing (AG)

โมเดล AG (San Martín et al., 2006) ได้รับการขยายแนวคิดมาจากโมเดล Rasch โดยผ่อนคลายให้โอกาสในการเดาถูกมีความสัมพันธ์กับความสามารถของผู้สอบ มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (4)

$$P(y_{ij}=1|\theta_j) = \Lambda[\theta_j - b_i] + (1 - \Lambda[\theta_j - b_i])\Lambda[\alpha\theta_j + g_i] \quad (4)$$

เมื่อ b_i และ g_i คือพารามิเตอร์ความยากและพารามิเตอร์การเดาของข้อสอบ i ตามลำดับ; $g_i \in (-\infty, \infty)$ และ α คือค่าน้ำหนักของความสามารถที่ใช้ในการเดา โดย $\Lambda[x] = 1/(1 + \exp[-x])$

จากสมการ (4) โอกาสในการเดาถูกหรือ $\Lambda[\alpha\theta_j + g_i]$ จะขึ้นอยู่กับพารามิเตอร์การเดา (g) และสัมพันธ์กับความสามารถของผู้สอบ (θ) ด้วยขนาดเท่ากับ α ดังนั้นพารามิเตอร์ α และ g จึงทำหน้าที่เป็นตัวกำหนดรูปแบบของการเดาที่เป็นไปได้ 3 กรณี (ดูภาพ 4) ดังนี้

กรณี 1: $\alpha = 0$ และ g เข้าใกล้ $-\infty \rightarrow$ ไม่มีการเดาเกิดขึ้น (โมเดล Rasch)

กรณี 2: $\alpha = 0$ และ g มีค่าอื่น ๆ $\rightarrow$ การเดาแบบสุ่ม

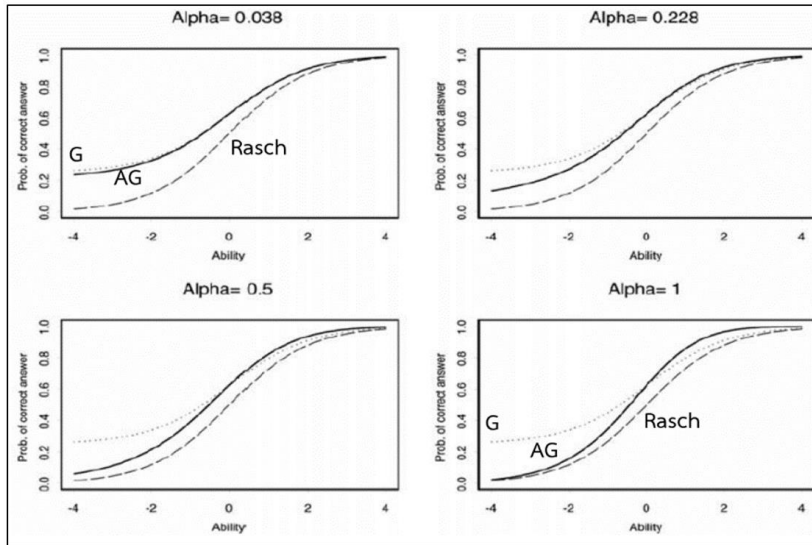
กรณี 3: $\alpha > 0$ และ g มีค่าอื่น ๆ $\rightarrow$ การเดาแบบอิงความสามารถ



ภาพ 4 มโนทัศน์ของกระบวนการแก้ปัญหาและกระบวนการเดาสำหรับโมเดล AG

San Martín และคณะ (2006) ได้ทดลองเปรียบเทียบโมเดล 3 รูปแบบคือ โมเดล AG (เดาแบบอิงความสามารถ) โมเดล G (เดาแบบสุ่ม) และโมเดล Rasch (ไม่มีการเดา) เพื่อวิเคราะห์ข้อสอบปรนัย 4 ตัวเลือกพบว่าเมื่อกำหนด $b = 0$, $g = -1.0986$ และ α เข้าใกล้ 0 (ภาพซ้ายบน) ICCs ของโมเดล AG (เส้นทึบ) จะเข้าใกล้ ICCs ของโมเดล G (เส้นจุด) กล่าวคือการเดาที่เกิดขึ้นจะมีแนวโน้มเป็นการเดาแบบสุ่ม แต่เมื่อ α มีค่าเพิ่มสูงขึ้น ทางซ้ายของโมเดล AG จะค่อย ๆ เข้าใกล้ ICCs ของโมเดล Rasch (เส้นประ) ในขณะที่ทางขวาจะสูงกว่า ICCs ของโมเดล Rasch กล่าวคือการเดาที่เกิดขึ้นจะมีแนวโน้มเป็นการเดาแบบอิงความสามารถ ทั้งนี้ San Martín และคณะให้เหตุผลว่า ผู้สอบความสามารถต่ำมักมีแนวโน้มถูกล่อลวงด้วยตัวลวงที่มีคุณภาพได้ง่าย

โอกาสในการเดาถูกจึงมีค่าต่ำกว่า .25 (โอกาสในการเดาถูกแบบสุ่ม) ส่วนผู้สอบความสามารถสูงมักอาศัยการตัดสินใจเลือกที่คาดว่าจะผิดออกไปก่อน แล้วจึงเดาสุ่มจากตัวเลือกที่เหลือ โอกาสในการเดาถูกจึงสูงขึ้นตามความสามารถที่มีมากขึ้นในรูปฟังก์ชันโลจิส และส่งผลทำให้โอกาสในการตอบถูกสูงขึ้นด้วย ดังภาพ 5



ภาพ 5 ตัวอย่าง ICCs ของโมเดล AG (ที่มา: San Martín et al., 2006)

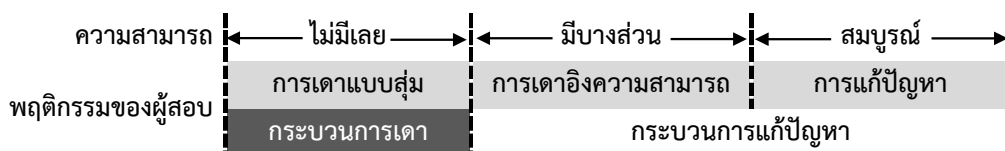
2) โมเดล fixed guessing 3-parameter logistic (FG3PL)

โมเดล FG3PL (Han, 2012) เป็นการปรับนิยามการเดาในโมเดล 3PL ใหม่เป็น “กระบวนการแก้ปัญหาคือแนวทางใดก็ตามที่ช่วยให้ผู้สอบสามารถตอบคำถามได้ ส่วนกระบวนการเดาคือการตอบแบบสุ่มโดยไม่อาศัยความสามารถใดเลย” จากนิยามดังกล่าวโอกาสในการเดาถูกจึงหมายถึงโอกาสในการเดาถูกแบบสุ่มเท่านั้น โดยโมเดลมีฟังก์ชันการตอบสนองข้อสอบดังสมการ (5)

$$P(y_{ij}=1|\theta_j) = \frac{1}{m_i} + \left(1 - \frac{1}{m_i}\right) \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} \quad (5)$$

เมื่อ a_i , b_i และ m_i คือพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก และจำนวนตัวเลือกของข้อสอบ i

กระบวนการแก้ปัญหาสำหรับโมเดล FG3PL จึงหมายรวมถึงการแก้ปัญหาของผู้สอบที่มีความรู้สมบูรณ์และการเดาแบบอิงความสามารถของผู้สอบที่มีความรู้บางส่วน ส่วนกระบวนการเดาจะหมายถึงการเดาแบบสุ่มจากผู้สอบที่ไม่มีความรู้ใดเลย (ดูภาพ 6) ทั้งนี้การที่การเดาแบบอิงความสามารถถูกพิจารณาให้เป็นกระบวนการแก้ปัญหาไม่ใช่กระบวนการเดา เนื่องจากถือว่าผู้สอบได้พยายามใช้ความสามารถในการแก้ปัญหาเพื่อตอบคำถามแล้ว ส่วน ICCs ของโมเดล FG3PL ยังคงมีลักษณะเหมือนกับโมเดล 3PL ทุกประการ



ภาพ 6 มโนทัศน์ของกระบวนการแก้ปัญหาและกระบวนการเดาสำหรับโมเดล FG3PL

3) โมเดล 2-parameter logistic guessing (2PLG)

โมเดล 2PLG (Tsai & Lin, 2015) ได้รับแนวคิดมาจากโมเดล AG โดยปรับปรุงให้โมเดลมีน้อยทั่วไปมากขึ้น ด้วยการนำพารามิเตอร์อำนาจจำแนกเข้ามาพิจารณาในส่วนของการตอบปัญหา กลายเป็นโมเดล 2PL และปรับแก้โอกาสในการเดาถูกให้มีความสัมพันธ์กับความสามารถของผู้สอบ ความยากของข้อสอบ และจำนวนตัวเลือกของข้อสอบ จนกลายเป็นฟังก์ชันการตอบสนองข้อสอบดังสมการ (6)

$$P(y_{ij}=1 | \theta_j) = \Lambda[a_i(\theta_j - b_i)] + (1 - \Lambda[a_i(\theta_j - b_i)]) \left(\frac{\exp[w_i \theta_j - b_i]}{1 + (m_i - 1) \exp[w_i \theta_j - b_i]} \right) \quad (6)$$

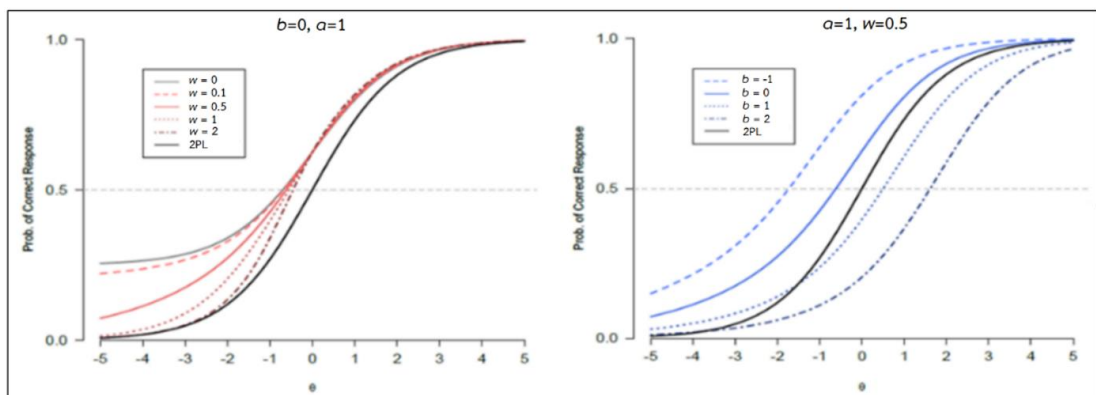
เมื่อ a_i , b_i และ m_i คือพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก และจำนวนตัวเลือกของข้อสอบ i และ w_i คือค่าน้ำหนักของความสามารถของผู้สอบที่ใช้ในการเดาตามความยากของข้อสอบ i

จากสมการ (6) โอกาสในการเดาถูกหรือพจน์สุดท้ายของสมการจะขึ้นอยู่กับค่าพารามิเตอร์ความยาก (b) และสัมพันธ์กับความสามารถของผู้สอบ (θ) ด้วยขนาดเท่ากับ w ดังนั้นพารามิเตอร์ w และ b จึงทำหน้าที่เป็นตัวกำหนดรูปแบบของการเดาที่เป็นไปได้ 2 กรณี ดังนี้

กรณี 1: $w = 0$ และ $b = 0$ → การเดาแบบสุ่ม

กรณี 2: $w > 0$ และ b มีค่าใด ๆ → การเดาแบบอิงความสามารถ

Tsai และ Lin (2015) ได้ศึกษาเปรียบเทียบโมเดล 2PLG และ 2PL เพื่อวิเคราะห์ข้อสอบปรนัย 4 ตัวเลือก โดยแบ่งการศึกษาเป็น 2 กรณีคือ 1) การเปลี่ยนค่าน้ำหนัก (w) และ 2) การเปลี่ยนค่าความยาก (b) ผลการศึกษากกรณี 1 พบว่าเมื่อกำหนด $a = 1$, $b = 0$ และ $w = 0$ หางซ้ายของโมเดล 2PLG จะลู่เข้าสู่ 0.25 การเดาที่เกิดขึ้นจึงมีแนวโน้มเป็นการเดาแบบสุ่ม แต่เมื่อ w มีค่าสูงขึ้น หางซ้ายของโมเดล 2PLG จะค่อย ๆ ลดต่ำเข้าใกล้ ICCs ของโมเดล 2PL ในขณะที่หางขวาจะสูงกว่า ICCs ของโมเดล 2PL ซึ่งลักษณะดังกล่าวมีรูปแบบเดียวกับโมเดล AG ดังได้กล่าวไปข้างต้น การเดาที่เกิดขึ้นจึงมีแนวโน้มเป็นการเดาแบบอิงความสามารถ ดังภาพ 7 (ซ้าย) ส่วนผลการศึกษากรณี 2 พบว่าเมื่อ b มีค่าเปลี่ยนแปลงไป ความยากของข้อสอบจะสัมพันธ์กับโอกาสในการตอบถูกแตกต่างกันตามระดับความสามารถของผู้สอบ กล่าวคือเมื่อข้อสอบยากมากขึ้น ผู้สอบจำเป็นต้องมีความสามารถที่สูงขึ้นด้วยจึงจะมีโอกาสในการตอบถูกเท่าเดิม ดังภาพ 7 (ขวา)



ภาพ 7 ตัวอย่าง ICCs ของโมเดล 2PLG เมื่อ w (ซ้าย) และ b (ขวา) มีค่าต่าง ๆ

(ที่มา: Tsai และ Lin, 2015)

4) โมเดล 3-parameter residual heteroscedasticity (3P-RH)

โมเดล 3P-RH (Lee & Bolt, 2018) ได้รับการขยายแนวคิดมาจากโมเดล 2PNO (two-parameter normal ogive) โดยผ่อนคลายให้ความแปรปรวนของความคลาดเคลื่อนของการประมาณค่าความสามารถของผู้สอบแต่ละระดับจากผลการตอบรายข้อ ($\sigma_{\epsilon_i|\theta}^2$) มีความแตกต่างกันได้ (heteroscedasticity) แตกต่างจากโมเดล 2PNO ที่กำหนดว่าความแปรปรวนดังกล่าวจะคงที่สำหรับทุกระดับความสามารถของผู้สอบ การเดาสำหรับโมเดล 3P-RH จึงสามารถพิจารณาได้จากความโน้มเอียงหรือความเอนของผลการตอบถูกของผู้สอบแต่ละระดับ โดยโมเดลมีฟังก์ชันการตอบสนองข้อสอบในรูปการแจกแจงปกติสะสมดังสมการ (7)

$$P(y_{ij}=1|\theta_j) = \Phi\left(\frac{\alpha_i\theta_j + \beta_i}{\sqrt{2} (1 + \exp[-\delta_i \theta_j])^{-1/2}}\right) \quad (7)$$

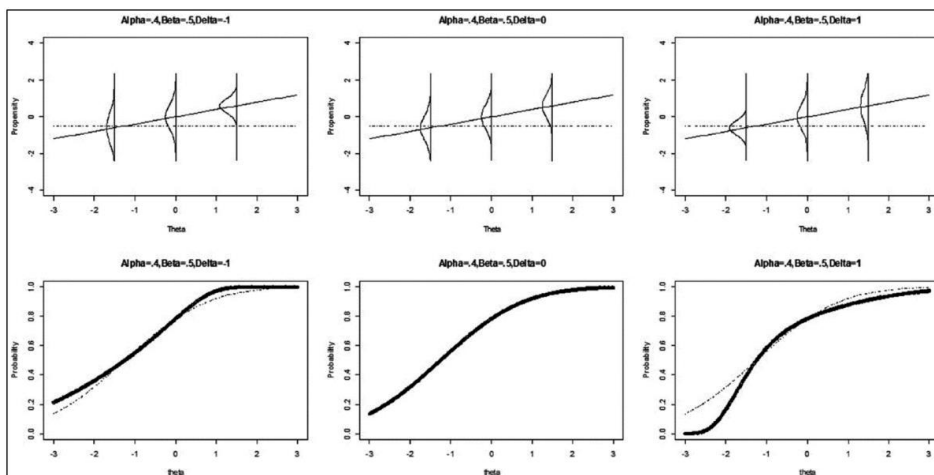
เมื่อ α_i คือพารามิเตอร์อำนาจจำแนกของข้อสอบ i ; $\alpha_i = a_i$ (พารามิเตอร์อำนาจจำแนกบนสเกลโลจิส) และ β_i คือพารามิเตอร์ความง่ายของข้อสอบ i ; $\beta_i = -a_i b_i$ (b_i คือพารามิเตอร์ความยากบนสเกลโลจิส) และ δ_i คือพารามิเตอร์ความเอนของผลการตอบถูกในข้อสอบ i ; $\delta_i \in (-\infty, \infty)$

จากสมการ (7) พารามิเตอร์ δ จึงทำหน้าที่เป็นตัวกำหนดรูปแบบการตอบ 3 กรณี (ดูภาพ 8) ดังนี้

กรณี 1: $\delta < 0$ (เบ้ลบ) $\rightarrow$ มีการเดาเกิดขึ้น (แต่ไม่ทราบรูปแบบ)

กรณี 2: $\delta = 0$ (สมมาตร) $\rightarrow$ ไม่มีการเดาเกิดขึ้น (โมเดล 2PNO)

กรณี 3: $\delta > 0$ (เบ้บวก) $\rightarrow$ ข้อสอบซับซ้อนสูง การตอบน่าจะใช้การแก้ปัญหาหลายขั้นตอน



ภาพ 8 ตัวอย่าง ICCs ของโมเดล 3P-RH (เส้นทึบ) กับ 2PNO (เส้นประ)

(ที่มา: Lee & Bolt, 2018)

5) โมเดล 2-parameter logistic extension (2PLE)

โมเดล 2PLE (Zhu, Wang, & Tao, 2019) ได้รับการปรับปรุงมาจากโมเดล 3PL ให้มีคุณสมบัติในการจำแนกการเดาแบบอิงความสามารถออกจากการเดาแบบสุ่ม ด้วยการกำหนดค่าการเดาหรือโอกาสในการเดาถูก (g) เป็นฟังก์ชันที่มีความสัมพันธ์กับความสามารถของผู้สอบและคุณลักษณะของข้อสอบ ได้แก่ ความยาก อำนาจจำแนก และจำนวนตัวเลือกของข้อสอบ มีฟังก์ชันการตอบสนองข้อสอบดังสมการ (8)

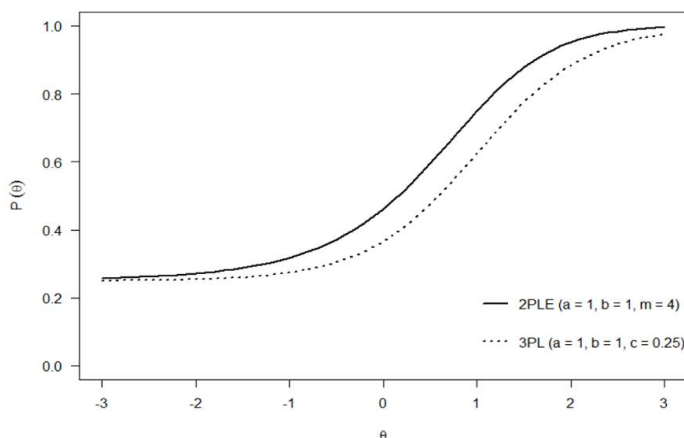
$$P(y_{ij}=1|\theta_j) = g_{ij} + (1 - g_{ij}) \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]} ; g_{ij} = \left(\frac{1}{m_i}\right)^{\frac{1}{1 + \exp[a_i(\theta_j - b_i)]}} \quad (8)$$

เมื่อ a_i , b_i และ m_i คือพารามิเตอร์อำนาจจำแนก พารามิเตอร์ความยาก และจำนวนตัวเลือกของข้อสอบ i

จากสมการ (8) โอกาสในการเดาถูก (g) จะเป็นค่าที่ขึ้นอยู่กับความสามารถของผู้สอบ (θ) อำนาจจำแนก (a) ความยาก (b) และจำนวนตัวเลือกของข้อสอบ (m) ผลจากการนิยามดังกล่าว g จึงหมายถึงโอกาสในการเดาถูกของผู้สอบความสามารถระดับใด ๆ ซึ่งแตกต่างจาก c ในโมเดล 3PL ที่หมายถึงโอกาสในการเดาถูกของผู้สอบความสามารถระดับต่ำมากเท่านั้น การนิยามดังกล่าวจึงเป็นการผ่อนคลายข้อตกลงเกี่ยวกับการเดาที่กำหนดว่าเป็นคุณลักษณะของข้อสอบลง โอกาสในการเดาถูกจึงสามารถผันแปรได้ตามความสามารถของผู้สอบ การแปลผลการเดาจึงมีแนวโน้มสอดคล้องกับสภาพธรรมชาติมากขึ้น และยังเป็นการหลีกเลี่ยงปัญหาที่อาจเกิดจากการประมาณค่าพารามิเตอร์การเดา (c) โดยตรงด้วย ทั้งนี้ θ จะทำหน้าที่เป็นตัวกำหนดรูปแบบของการเดาที่เป็นไปได้ 2 กรณี (ดูภาพ 9) ดังนี้

กรณี 1: θ เข้าใกล้ $-\infty \rightarrow$ การเดาแบบสุ่ม ($g \approx 1/m$ โดยในที่นี้กำหนด $m = 4$)

กรณี 2: $-\infty < \theta < \infty \rightarrow$ การเดาแบบอิงความสามารถ ($1/m < g < 1$)



ภาพ 9 ตัวอย่าง ICCs ของโมเดล 2PLE (เส้นทึบ) และ 3PL (เส้นประ)

การเปรียบเทียบความแตกต่างระหว่างโมเดล IRT ทางเลือกที่คำนึงถึงการเดารูปแบบต่าง ๆ

โมเดล IRT ทางเลือกที่คำนึงถึงการเดาข้างต้น ทุกโมเดลล้วนมีความคล้ายคลึงกันในการพยายามปรับปรุงโมเดล IRT แบบดั้งเดิม (Rasch 2PL และ 3PL) ให้สามารถจำแนกกระบวนการแก้ปัญหาและกระบวนการเดาออกจากกัน เช่น โมเดล FG3PL และ 3P-RH นอกจากนี้บางโมเดล ในกระบวนการเดาก็อาจมีการจำแนกการเดาแบบอิงความสามารถออกจากการเดาแบบสุ่มด้วย เช่น โมเดล AG 2PLG และ 2PLE

จากความหลากหลายของปรัชญาความเชื่อเกี่ยวกับการเดา โมเดล IRT ทางเลือกที่คำนึงถึงการเดาจึงมีรูปแบบในการพัฒนาและแนวคิดที่แตกต่างกัน โดยสรุปเป็นประเด็นที่น่าสนใจได้ 8 ประเด็น ดังนี้ 1) ปัจจัยที่ส่งผลต่อการเดา 2) พารามิเตอร์ที่ต้องประมาณค่าในโมเดล 3) วิธีการประมาณค่าในงานวิจัย 4) โปรแกรมวิเคราะห์ 5) สถิติที่ใช้เป็นเกณฑ์เปรียบเทียบผลการประมาณค่าและการคัดเลือกโมเดล 6) ผลสรุปในงานวิจัย 7) จุดเด่นของโมเดล และ 8) ข้อจำกัดของโมเดล รายละเอียดดังตาราง 1

ตาราง 1 การเปรียบเทียบความแตกต่างระหว่างโมเดล AG FG3PL 2PLG 3P-RH และ 2PLE

ประเด็น	โมเดล AG (2006)	โมเดล FG3PL (2012)	โมเดล 2PLG (2015)	โมเดล 3P-RH (2018)	โมเดล 2PLE (2019)
1) ปัจจัยที่ส่งผลต่อการเดา	- ความสามารถผู้สอบ	- จำนวนตัวเลือก	- ความสามารถผู้สอบ - ความยากข้อสอบ - จำนวนตัวเลือก	- ความเข้มของการตอบถูก	- ความสามารถผู้สอบ - ความยากข้อสอบ - อำนาจจำแนกข้อสอบ - จำนวนตัวเลือก
2) พารามิเตอร์ที่ต้องประมาณค่า	b = ความยาก g = การเดา α = ค่าน้ำหนักของความสามารถที่ใช้ในการเดา	b = ความยาก a = อำนาจจำแนก	b = ความยาก a = อำนาจจำแนก w = ค่าน้ำหนักของความสามารถที่ใช้ในการเดา	β = ความง่าย α = อำนาจจำแนก δ = ความเข้มของการตอบถูก	b = ความยาก a = อำนาจจำแนก
3) วิธีการประมาณค่า	Bayes (MCMC), NLMIXED	ML (EM)	Bayes (MCMC)	Bayes (MCMC)	ML (Hermite-Gauss quadrature)
4) โปรแกรม	SAS (NLMIXED), WINBUGS	PARSCALE, PARAM-3PL, WinGen	R (R2jags)	R (optim), Mx, SAS	ไม่ระบุ
5) สถิติที่ใช้พิจารณา					
- การประมาณค่า	Confidence Band	Bias, RMSE	Credible interval	Correlation	Bias, RMSE
- การคัดเลือกโมเดล	-2LL, AIC, BIC	χ^2	ไม่ระบุ	-2LL, AIC, BIC, DIC	AIC, BIC
6) ผลสรุปในงานวิจัย	- โมเดล AG ประมาณค่าพารามิเตอร์ได้ดี และสอดคล้องกับข้อมูลเชิงประจักษ์	- โมเดล FG3PL ประมาณค่าพารามิเตอร์ได้ดี และใช้ง่ายกว่าโมเดล 3PL	- โมเดล 2PLG ประมาณค่าพารามิเตอร์ได้ดี	- โมเดล 3P-RH ประมาณค่าพารามิเตอร์ได้ดี และมีแนวโน้มสอดคล้องกับข้อมูลมากกว่าโมเดล 3PL	- โมเดล 2PLE ประมาณค่าพารามิเตอร์ได้ดี และมีแนวโน้มสอดคล้องกับข้อมูลมากกว่าโมเดล 3PL
7) จุดเด่น	- เป็นโมเดลขยายของ Rasch - ระบุรูปแบบการเดาได้	- เป็นโมเดลเฉพาะของ 3PL - มีพารามิเตอร์น้อยกว่า 3PL แต่ให้ผลการวิเคราะห์แทบจะไม่แตกต่างจาก 3PL	- เป็นโมเดลขยายของ 2PL - ระบุรูปแบบการเดาได้	- เป็นโมเดลขยายของ 2PNO - แปลผลการเดาได้อย่างอิสระ - บอกความซับซ้อนในการทำข้อสอบแต่ละข้อได้	- เป็นโมเดลขยายของ 2PL - เป็นโมเดลเฉพาะของ 3PL - ระบุรูปแบบการเดาได้
8) ข้อจำกัด	- ไม่มีค่าอำนาจจำแนก - พารามิเตอร์ g อยู่บนสเกลโลจิส จึงนำมาเทียบกับพารามิเตอร์ c ในโมเดล 3PL ไม่ได้โดยตรง	- การเดาเป็นแบบสุ่มเท่านั้น จึงไม่สอดคล้องกับสภาพธรรมชาติเท่าที่ควร	- โมเดลซับซ้อน - การเดาแบบสุ่มเกิดขึ้นเมื่อข้อสอบยากปานกลาง ($b = 0$) จึงไม่สอดคล้องกับสภาพธรรมชาติเท่าที่ควร	- โมเดลซับซ้อน เนื่องจากอยู่ในรูป normal ogive - ระบุรูปแบบการเดาไม่ได้	- ไม่สามารถอธิบายกรณี $g < 1/m$ (โอกาสการเดาถูกน้อยกว่าโอกาสการเดาถูกแบบสุ่ม) ซึ่งอาจเกิดขึ้นเมื่อผู้สอบความสามารถต่ำเจอตัวลวงที่มีคุณภาพ

หมายเหตุ: MCMC = Markov chain Monte Carlo; ML = Maximum likelihood; EM = Expectation-maximization; RMSE = Root mean square error; LL = Log-likelihood; AIC = Akaike information criterion; BIC = Bayesian information criterion; DIC = Deviance information criterion

บทสรุป

การเดาเป็นปัญหาหนึ่งที่หลีกเลี่ยงไม่ได้สำหรับการวัดและประเมินผลด้วยแบบสอบปรนัย ซึ่งเป็นเครื่องมือที่นิยมใช้ทั่วไปในการทดสอบ การเดาแบ่งเป็น 2 รูปแบบคือ 1) การเดาแบบสุ่ม เป็นการสุ่มเลือกคำตอบจากบรรดาตัวเลือกทั้งหมด เกิดขึ้นเมื่อผู้ตอบไม่มีความรู้ใดเลยเกี่ยวกับข้อคำถาม พบได้ทั่วไปในการทดสอบวิชาคณิตศาสตร์ ซึ่งตัวเลือกมีลักษณะเป็นตัวเลข และ 2) การเดาแบบอิงความสามารถ เป็นการเลือกคำตอบโดยอาศัยการตัดตัวเลือกที่คาดว่าผิดออกก่อน แล้วจึงเดาแบบสุ่มจากตัวเลือกที่เหลือ เกิดขึ้นเมื่อผู้ตอบมีความรู้บางส่วนเกี่ยวกับข้อคำถาม พบได้ทั่วไปในการทดสอบวิชาทางภาษา ซึ่งตัวเลือกมีลักษณะเป็นข้อความ

ในเบื้องต้นการเดาอาจป้องกันได้ด้วยวิธีการต่าง ๆ เช่น การหักคะแนน การเปลี่ยนรูปแบบการตอบ การปรับให้มีตัวถูกหลายตัว เป็นต้น อย่างไรก็ตามวิธีดังกล่าวยังมีข้อจำกัดในเชิงปฏิบัติ และยังไม่สามารถ

ป้องกันการเดาได้อย่างสมบูรณ์ เนื่องจากปัจจัยที่ก่อให้เกิดการเดาอาจมาจากหลายสาเหตุ ทั้งคุณลักษณะของข้อสอบ เช่น ความยาก อำนาจจำแนก จำนวนตัวเลือก และคุณลักษณะของผู้สอบ เช่น เพศ ความสามารถ แรงจูงใจในการสอบ เป็นต้น จึงเป็นเรื่องยากในเชิงปฏิบัติที่จะป้องกันการเดาด้วยวิธีดังกล่าว

โมเดลโลจิส 3 พารามิเตอร์ (3PL) จึงเข้ามามีบทบาทในการวัดค่าการเดาที่เกิดขึ้น แทนแนวทางการป้องกันการเดาแบบดั้งเดิม แต่จากการศึกษาในหลายงานวิจัยพบว่าโมเดล 3PL อาจมีข้อจำกัดเกี่ยวกับการประมาณค่าและการแปลผลการเดา ด้วยเหตุผลดังกล่าวจึงเป็นที่มาของการพัฒนาโมเดล IRT ทางเลือกที่คำนึงถึงการเดาเพื่อนำมาใช้แทนโมเดล 3PL อย่างไรก็ตามเนื่องจากการเดาอาจแปรผันได้ตามปัจจัยต่าง ๆ หลายสาเหตุ โมเดล IRT ทางเลือกที่คำนึงถึงการเดาจึงถูกพัฒนาออกมาหลายรูปแบบภายใต้ตามปรัชญาความเชื่อเกี่ยวกับการเดาที่แตกต่างกัน

บทความนี้จึงมีการยกตัวอย่างโมเดล IRT ทางเลือกที่คำนึงถึงการเดาที่มีแนวคิดน่าสนใจและน่าจะนำไปประยุกต์ใช้ในเชิงวิชาการได้ ตัวอย่างเช่น อาจนำแนวคิดไปพัฒนาเป็นโมเดล IRT ทางเลือกรูปแบบใหม่ ซึ่งมีความสอดคล้องกลมกลืน (fit) กับข้อมูลเชิงประจักษ์มากกว่าโมเดล IRT ทั่วไป หรืออาจนำโมเดล IRT ทางเลือกไปบูรณาการเข้ากับโมเดลอื่น ๆ เช่น โมเดลกลุ่มแฝง (latent class) หรือโมเดลพหุระดับ (multilevel) เพื่อผ่อนคลายข้อตกลงหรือข้อจำกัดต่าง ๆ ทำให้โมเดลมีความสอดคล้องกับธรรมชาติของข้อมูลทางการศึกษามากขึ้น เป็นต้น หรือนำไปใช้ในเชิงปฏิบัติ ด้วยการใช้โมเดล IRT ทางเลือกที่คำนึงถึงการเดาแทนโมเดล 3PL ซึ่งอาจช่วยให้การอธิบายข้อมูลที่มีการเดาคำตอบทำได้ดีขึ้น รวมถึงยังเป็นการหลีกเลี่ยงปัญหาการประมาณค่าและการแปลผลการเดาที่เกิดจากโมเดล 3PL ไปพร้อมกัน

เอกสารอ้างอิง

ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่ (Modern test theories)*. พิมพ์ครั้งที่ 4. กรุงเทพฯ:

โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Andrich, D., & Marais, I. (2014). Person proficiency estimates in the dichotomous Rasch model when random guessing is removed from difficulty estimates of multiple choice items. *Applied Psychological Measurement*, 38(6), 432-449.

Baldiga, K. (2014). Gender differences in willingness to guess. *Management Science*, 60(2), 434-448.

Ben-Simon, A., Budescu, D., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement*, 21(1), 65-88.

Bereby-Meyer, Y., Meyer, J., & Flascher, O. (2002). Prospect theory analysis of guessing in multiple choice tests. *Journal of Behavioral Decision Making*, 15(4), 313-327.

Bo, Y., Lewis, C., & Budescu, D. (2015). An option-based partial credit item response model. *Quantitative Psychology Research*, 89, 45-72.

Budescu, D., & Bo, Y. (2015). Analyzing test-taking behavior: decision theory meets psychometric theory. *Psychometrika*, 80(4), 1105-1122.

- Cao, J., & Stokes, S. (2008). Bayesian IRT Guessing Models for Partial Guessing Behaviors. *Psychometrika*, 73(2), 209-230.
- Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, 47(2), 448-474.
- Hambleton, R., Swaminathan, H., & Rogers, H. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.
- Han, K. (2012). Fixing the c Parameter in the Three-Parameter Logistic Model. *Practical Assessment, Research & Evaluation*, 17(1), 1-24.
- Hutchinson, T. P. (1991). *Ability, partial information and guessing: Statistical modelling applied to multiple-choice tests*. Rundle Mall, South Australia: Rumsby Scientific Publishing.
- Lee, S., & Bolt, D. (2018). An Alternative to the 3PL: Using Asymmetric Item Characteristic Curves to Address Guessing Effects. *Journal of Educational Measurement*, 55(1), 90-111.
- Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39(2), 247-264.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.
- San Martín, E., Del Pino, G., & De Boeck, P. (2006). IRT models for ability-based guessing. *Applied Psychological Measurement*, 30(3), 183-203.
- Smith, R. M. (1993). Guessing and the Rasch model. *Rasch Measurement Transactions*, 6(4), 262-263.
- Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika*, 47(4), 397-412.
- Tsai, T., & Lin, C. (2015). Modeling Guessing Properties of Multiple-Choice Items in the Measurement of Political Knowledge. *Paper was presented at the 2015 Asian Political Methodology Conference, Jan 9-10, 2015*. Taipei, Taiwan.
- Wise, S., & Kong, X. (2005). Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests. *Applied Measurement in Education*, 18(2), 163-183.
- Zhu, Z., Wang, C., & Tao, J. (2019). A two-parameter logistic extension model: an efficient variant of the three-parameter logistic model. *Applied Psychological Measurement*, 43(6), 449-463.