

การวิเคราะห์ข้อมูลการประเมินการศึกษานานาชาติใหญ่ระดับนานาชาติโดยใช้ plausible values

รองศาสตราจารย์ ดร. สัมพันธ์ จักรระโท

สถาบันวิจัยและพัฒนา

สาขาวิชาการวัดและประเมินผลการศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช

บทคัดย่อ

โครงการประเมินผลการศึกษาในระดับนานาชาติเป็นโครงการที่ให้ข้อมูลสำคัญสำหรับประเทศต่างๆ จะได้เรียนรู้ความสำเร็จในการจัดการศึกษาของประเทศที่มีประสิทธิภาพการศึกษาดี การเรียนรู้การจัดการศึกษาที่มีประสิทธิภาพ ต้องมีการวิเคราะห์ข้อมูลผลการประเมินที่ถูกต้อง เนื่องจากการประเมินทางการศึกษานานาชาติใหญ่ระดับนานาชาติ มีการออกแบบการเก็บรวบรวมข้อมูลที่เก็บข้อมูลจากนักเรียนบางคน และนักเรียนไม่สามารถทำข้อสอบได้ทั้งหมด และอาจจะไม่ได้ตอบข้อสอบเหมือนกันทุกข้อ เนื่องจากข้อสอบมีจำนวนมาก บทความฉบับนี้จึงต้องการนำเสนอวิธีการวิเคราะห์ข้อมูลที่ถูกต้อง เพื่อให้สามารถใช้ผลการวิเคราะห์ได้ถูกต้อง และเกิดประโยชน์ต่อการจัดการศึกษา

คำสำคัญ: การประเมินการศึกษานานาชาติใหญ่ การวิเคราะห์ข้อมูลทุติยภูมิ ค่าเป็นไปได้

Analysis of International Large-Scale Assessment Data Using Plausible Values

Sungworn Ngudgratoke

STOU Institute for Research and Development

Department of Educational Measurement and Evaluation, STOU

Abstract

International large-scale assessment programs provide essential information that allows countries to learn from the educational successes of high-performing systems. To effectively learn from these systems, it is crucial to analyze assessment data accurately. Because international large-scale educational assessments are designed to collect data from only a subset of students, not all students take the entire test, and students may not respond to the same set of items due to the large number of test questions. Therefore, this article aims to present appropriate methods for analyzing such data to ensure the accuracy and usefulness of analytical results for improving educational management.

Keywords: large-scale educational assessment, secondary data analysis, plausible values

1. โครงการประเมินผลการศึกษาในระดับนานาชาติขนาดใหญ่

โครงการประเมินผลศึกษานานาชาติขนาดใหญ่ เป็นการประเมินความสามารถและทักษะของผู้เรียนในเชิงเปรียบเทียบในกลุ่มประเทศต่างๆ เพื่อนำผลการประเมินมาขึ้นำการจัดการศึกษาที่เป็นเลิศ ซึ่งจะเป็นการดำเนินการจัดการศึกษาเพื่อพัฒนาบุคลากรของประเทศต่างๆ ให้พร้อมรับการเปลี่ยนแปลงที่เกิดขึ้นในอนาคต โครงการประเมินผลศึกษานานาชาติขนาดใหญ่ สามารถแบ่งได้เป็น 2 กลุ่มหลักตามกลุ่มเป้าหมาย คือ กลุ่มที่เน้นการประเมินในโรงเรียน (School-based) และกลุ่มที่เน้นการประเมินในครัวเรือนซึ่งส่วนใหญ่มุ่งเป้าไปที่ประชากรวัยผู้ใหญ่ (Home-based)

โครงการประเมินการศึกษาระดับนานาชาติประเมินสมรรถนะของนักเรียนในระดับชั้นต่างๆ และรวบรวมข้อมูลบริบทที่เกี่ยวข้องกับการเรียนการสอน โครงการหลักๆ ได้แก่ โครงการ TIMSS (Trends in International Mathematics and Science Study) เป็นโครงการศึกษาแนวโน้มการจัดการศึกษาคณิตศาสตร์และวิทยาศาสตร์ ประเมินนักเรียนในระดับชั้นประถมศึกษาปีที่ 4 และมัธยมศึกษาปีที่ 2 (และบางครั้งในระดับชั้นมัธยมศึกษาตอนปลาย) โครงการ PIRLS (Progress in International Reading Literacy Study) เป็นโครงการประเมินความฉลาดรู้ด้านการอ่าน ประเมินนักเรียนในระดับชั้นประถมศึกษาปีที่ 4 โครงการ ICCS (International Civic and Citizenship Education Study) เป็นโครงการศึกษาเรื่องการให้การศึกษาเพื่อสร้างความเป็นพลเมือง ประเมินนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 2 โครงการ ICILS (International Computer and Information Literacy Study) เป็นโครงการประเมินความฉลาดรู้ด้านคอมพิวเตอร์และสารสนเทศ ประเมินนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 2 โครงการ PISA (Programme for International Student Assessment) ซึ่งเป็นโครงการประเมินสมรรถนะนักเรียนมาตรฐานสากล ประเมินนักเรียนอายุ 15 ปี ในด้านการอ่าน คณิตศาสตร์ และวิทยาศาสตร์ รวมถึงสมรรถนะอื่นๆ เช่น การแก้ปัญหา การคิดสร้างสรรค์ และความฉลาดรู้ด้านการเงิน โครงการเหล่านี้มุ่งประเมินทักษะและความสามารถของประชากรวัยผู้ใหญ่ในบริบทชีวิตประจำวันและการทำงาน โครงการ IALS (International Adult Literacy Survey) สำนวความฉลาดรู้ของผู้ใหญ่นานาชาติ (โครงการในอดีต) โครงการ ALLS (Adult Literacy and Life Skills Survey) สำนวความฉลาดรู้และทักษะชีวิตของผู้ใหญ่ (โครงการในอดีต) และโครงการ PIAAC (Programme for the International Assessment of Adult Competencies) เป็นโครงการสมรรถนะผู้ใหญ่มาตรฐานสากล ซึ่งถือเป็นโครงการหลักในปัจจุบันที่ประเมินทักษะที่จำเป็นในศตวรรษที่ 21

โครงการประเมินผลขนาดใหญ่เหล่านี้ดำเนินการโดยองค์กรระหว่างประเทศ 2 แห่งที่มีความเชี่ยวชาญและมีบทบาทสำคัญในการกำหนดทิศทางการศึกษาและนโยบายสาธารณะทั่วโลก ประกอบด้วยสมาคมนานาชาติเพื่อการประเมินผลสัมฤทธิ์ทางการศึกษา (IEA) International Association for the Evaluation of Educational Achievement (IEA) เป็นองค์กรอิสระระดับนานาชาติที่ประกอบด้วยสถาบันวิจัยและหน่วยงานภาครัฐจากประเทศต่างๆ ทั่วโลก IEA มีบทบาทหลักในการวิจัยทางการศึกษาเชิงเปรียบเทียบ โครงการที่ IEA เป็นผู้ดำเนินการโครงการ ได้แก่ TIMSS, PIRLS, ICCS, และ ICILS ส่วนองค์กรเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนา (OECD) หรือ Organisation for Economic Co-operation and Development (OECD) เป็นองค์กรระหว่างประเทศของกลุ่มประเทศที่พัฒนาแล้วซึ่งยอมรับในหลักการประชาธิปไตยและเศรษฐกิจการตลาด OECD มีภารกิจในการส่งเสริมนโยบายที่จะพัฒนาคุณภาพชีวิตทางเศรษฐกิจและสังคมของผู้คนทั่วโลก โครงการที่ OECD เป็นผู้ดำเนินการโครงการ คือ PISA และ PIAAC นอกจากนี้ยังมีส่วนร่วมในโครงการสำหรับผู้ใหญ่ในอดีตอย่าง IALS และ ALLS นอกเหนือจากการดำเนินงานด้านการศึกษาแล้ว OECD ยังรวบรวมข้อมูลในหลากหลายสาขา รวมถึงเศรษฐกิจ สังคม และสิ่งแวดล้อม เพื่อสร้างตัวชี้วัดสำหรับให้ข้อมูลแก่รัฐบาลในการกำหนดนโยบาย

2. ลักษณะสำคัญของโครงการประเมินการศึกษานานาชาติขนาดใหญ่

แม้ว่าโครงการประเมินแต่ละโครงการจะมีเป้าหมายและกลุ่มตัวอย่างที่แตกต่างกัน แต่ทั้งหมดมีลักษณะร่วมทางเทคนิคที่สำคัญ 2 ประการ ซึ่งนักวิจัยจำเป็นต้องทำความเข้าใจอย่างลึกซึ้งเพื่อที่จะวิเคราะห์ข้อมูลได้อย่างถูกต้อง ลักษณะที่สำคัญ มีดังนี้

1. การออกแบบการสุ่มตัวอย่างที่ซับซ้อน (Complex Sample Design)

โครงการเหล่านี้ไม่ได้ทำการสำรวจประชากรทั้งหมด แต่ใช้การสุ่มตัวอย่างเพื่อเลือกกลุ่มตัวแทน การออกแบบการสุ่มตัวอย่างมักเป็นแบบหลายขั้นตอนและมีการแบ่งชั้นภูมิ (Probabilistic, stratified, multistage sample designs) ตัวอย่างเช่น ขั้นตอนที่ 1: สุ่มโรงเรียน (โดยอาจมีการแบ่งกลุ่มโรงเรียนตามภูมิภาค ขนาด หรือประเภทก่อน) ขั้นตอนที่ 2: สุ่มห้องเรียนภายในโรงเรียนที่ได้รับเลือก และขั้นตอนที่ 3: นักเรียนทุกคนในห้องเรียนที่ได้รับเลือกจะเข้าร่วมการประเมิน การออกแบบลักษณะนี้ทำให้ข้อมูลที่ได้อาจมีความเป็นอิสระต่อกัน (นักเรียนในโรงเรียนเดียวกันมักมีความคล้ายคลึงกันมากกว่านักเรียนต่างโรงเรียน) ดังนั้นในการวิเคราะห์ข้อมูล ผู้วิจัยจำเป็นต้องใช้ sampling weights) เพื่อถ่วงน้ำหนักข้อมูลให้สะท้อนภาพของประชากรทั้งหมดได้อย่างถูกต้อง และใช้วิธีการประมาณค่าความคลาดเคลื่อนที่เหมาะสม: เช่น Jackknife Repeated Replication (JRR) หรือ Balanced Repeated Replication (BRR) ซึ่งใช้ตัวแปรที่เรียกว่า Replicate Weights ในการคำนวณค่าความคลาดเคลื่อนมาตรฐาน (Standard Errors) ที่ถูกต้อง การละเลยการออกแบบการสุ่มตัวอย่างที่ซับซ้อนนี้จะนำไปสู่การประมาณค่าที่ผิดพลาดและค่าความคลาดเคลื่อนมาตรฐานที่ต่ำกว่าความเป็นจริง ซึ่งอาจทำให้สรุปผลผิดพลาดได้

2. การออกแบบการประเมินที่ซับซ้อน (Complex Assessment Design)

ดังที่กล่าวไปในส่วนของแต่ละโครงการ ไม่มีนักเรียนคนใดได้ทำข้อสอบทุกข้อในคลังข้อสอบทั้งหมด การออกแบบที่เรียกว่า Multiple Matrix-Sampled Design นี้มีข้อดีคือสามารถประเมินเนื้อหาได้กว้างขวาง แต่ก็สร้างความท้าทายในการวัดระดับความสามารถของนักเรียนแต่ละคน เนื่องจากข้อมูลของนักเรียนแต่ละคนมีอยู่อย่างจำกัด เพื่อแก้ปัญหานี้ โครงการประเมินผลจึงใช้เทคนิคทางสถิติขั้นสูงที่เรียกว่า Plausible Values (PVs) Plausible Values ไม่ใช่คะแนนดิบของนักเรียน แต่เป็นค่าที่เป็นไปได้ของความสามารถที่แท้จริงของนักเรียนแต่ละคน โดยเป็นการสุ่มค่าหลายๆ ค่า (โดยทั่วไปคือ 5 หรือ 10 ค่า) จากการแจกแจงความน่าจะเป็นของความสามารถของนักเรียนคนนั้นๆ ซึ่งคำนวณโดยใช้ข้อมูลการตอบข้อสอบและข้อมูลภูมิหลังของนักเรียน (เช่น เพศ, สถานะทางเศรษฐกิจและสังคม) การใช้ PVs ช่วยให้เราสามารถประมาณค่าสถิติของประชากร (เช่น ค่าเฉลี่ย, ความแปรปรวน) ได้อย่างไม่เอนเอียง (unbiased) ซึ่งการใช้คะแนนที่คำนวณโดยตรงจากข้อสอบที่นักเรียนทำเพียงไม่กี่ข้อจะให้ผลที่เอนเอียง ในการวิเคราะห์ข้อมูลที่มี PVs ผู้วิจัยจะต้องทำการวิเคราะห์ซ้ำสำหรับ PV แต่ละชุด (เช่น วิเคราะห์ 5 ครั้งสำหรับข้อมูลที่มี 5 PVs) จากนั้นจึงนำผลลัพธ์ที่ได้มารวมกันโดยใช้กฎของรูบิน (Rubin's Rules) เพื่อให้ได้ค่าประมาณและค่าความคลาดเคลื่อนมาตรฐานสุดท้ายที่ถูกต้อง ซอฟต์แวร์วิเคราะห์ข้อมูลเฉพาะทาง (เช่น IDB Analyzer) ถูกออกแบบมาเพื่อจัดการกับความซับซ้อนทั้งสองประการนี้โดยอัตโนมัติ

3. ประเภทและโครงสร้างของไฟล์ข้อมูล

โดยทั่วไป ฐานข้อมูลของโครงการประเมินผลเหล่านี้จะประกอบด้วยไฟล์ข้อมูล 2 ประเภทหลัก ซึ่งมักจะแยกจากกันและต้องนำมาเชื่อม (merge) กันเพื่อการวิเคราะห์

3.1 ไฟล์ข้อมูลภูมิหลัง (Background Files) ไฟล์เหล่านี้เก็บรวบรวมข้อมูลจากแบบสอบถามบริบทต่างๆ ซึ่งให้ข้อมูลเชิงลึกเกี่ยวกับปัจจัยที่อาจเกี่ยวข้องกับผลลัพธ์ทางการเรียน ไฟล์ภูมิหลังที่สำคัญ ได้แก่

1) ไฟล์นักเรียน (Student File): เป็นไฟล์หลักที่สำคัญที่สุด ประกอบด้วยข้อมูลส่วนตัวและภูมิหลังทางการศึกษาของนักเรียน ประสบการณ์ในโรงเรียน ทักษะคิดต่อวิชาที่เรียน เป้าหมายในอนาคต ฯลฯ ที่สำคัญไฟล์นี้มักจะรวมคะแนนผลสัมฤทธิ์ (Plausible Values) ไว้ด้วย

2) ไฟล์ครู (Teacher File): มีในโครงการ TIMSS, PIRLS, ICCS, ICILS ประกอบด้วยข้อมูลภูมิหลังของครู การศึกษา ประสบการณ์ แนวปฏิบัติการสอน และความเชื่อทางการสอน

3) ข้อควรระวัง: ใน TIMSS และ PIRLS ไฟล์ครูจะเชื่อมโยงกับนักเรียน (Student-Teacher Linkage) เพื่อวิเคราะห์ว่า "นักเรียนร้อยละเท่าใดที่ถูกสอนโดยครูที่มีคุณลักษณะ..." ในขณะที่ใน ICCS และ ICILS สามารถวิเคราะห์ในระดับครูได้โดยตรงว่า "ครูร้อยละเท่าใดที่มีคุณลักษณะ..."

4) ไฟล์โรงเรียน (School File): มีในทุกโครงการหลัก ประกอบด้วยข้อมูลที่ตอบโดยผู้บริหารโรงเรียน เช่น หลักสูตร บรรยากาศในโรงเรียน ทรัพยากร และจำนวนบุคลากร

5) ไฟล์บ้าน/ผู้ปกครอง (Home/Parent File): มีใน TIMSS (ป.4), PIRLS, และ PISA ประกอบด้วยข้อมูลที่ตอบโดยผู้ปกครอง เช่น ทรัพยากรทางการศึกษาที่บ้าน กิจกรรมการเรียนรู้นอกโรงเรียน การสนับสนุนจากผู้ปกครอง

6) ไฟล์ผู้ประสานงานด้าน ICT (ICT Coordinator File): มีเฉพาะใน ICILS ประกอบด้วยข้อมูลจากผู้รับผิดชอบด้านเทคโนโลยีสารสนเทศของโรงเรียน เกี่ยวกับทรัพยากร นโยบาย และการสนับสนุนด้าน ICT ในโรงเรียน

3.2 ไฟล์ข้อมูลผลสัมฤทธิ์ (Achievement/Cognitive Files) ไฟล์เหล่านี้เก็บข้อมูลการตอบข้อสอบแต่ละข้อของนักเรียน มักเรียกว่า Cognitive Item Files ประกอบด้วย 1) รหัสการตอบข้อสอบแต่ละข้อของนักเรียน (เช่น ตอบถูก, ตอบผิด, ตอบถูกบางส่วน, ไม่ได้ทำ) 2) ไฟล์นี้มีประโยชน์อย่างยิ่งสำหรับการวิเคราะห์ในระดับข้อสอบ (Item Analysis) เพื่อดูว่านักเรียนมีจุดแข็งหรือจุดอ่อนในเนื้อหาหรือทักษะใดโดยเฉพาะ และ 3) เนื่องจากนักเรียนแต่ละคนทำข้อสอบเพียงส่วนหนึ่ง ไฟล์นี้จึงมีข้อมูลที่หายไป (missing data) เป็นจำนวนมากโดยการออกแบบ

3. การออกแบบโครงการประเมินผลการศึกษาในระดับนานาชาติขนาดใหญ่

ความท้าทายของการประเมินการศึกษานานาชาติขนาดใหญ่ คือ ขอบเขตของโดเมนกว้างมาก เวลาในการทดสอบมีจำกัด (ทั้งทางกายภาพและจิตวิทยา) และมี ความท้าทายด้าน 1) จำเป็นต้องดำเนินการรายการข้อสอบด้วยการออกแบบที่เหมาะสม 2) จำเป็นต้องสรุปผลการปฏิบัติงานของรายการข้อสอบ และจำเป็นต้องคำนึงถึงความไม่น่าเชื่อถือของค่าประมาณ การออกแบบโครงการประเมินการศึกษานานาชาติใหญ่ มักดำเนินการโดยใช้การออกแบบตัวอย่างที่ซับซ้อน (Complex Sample Design) ซึ่งออกแบบการสุ่มกลุ่มตัวอย่างแบบอาศัยความน่าจะเป็น แบบแบ่งชั้น และแบบหลายขั้นตอน ซึ่งจำเป็นต้องนำการออกแบบตัวอย่างมาพิจารณาเมื่อคำนวณค่าประมาณ การสุ่มตัวอย่างของโครงการประเมินช่วยแก้ปัญหาที่สำคัญ คือ 1) เป็นไปไม่ได้ที่จะทดสอบทุกคนในทุกรายการ เพราะ มีคนมากเกินไป มีรายการข้อสอบมากเกินไป และมีค่าใช้จ่ายสูงเกินไป 2) ไม่จำเป็นต้องทดสอบทุกคนในทุกรายการ (เช่น ตัวอย่างเลือด, ตัวอย่างน้ำซุบ) และ 3) ผลลัพธ์ควรถูกพิจารณาในบริบทของการออกแบบตัวอย่างของนักเรียนและรายการข้อสอบ

3.1 การออกแบบการสุ่มกลุ่มตัวอย่างนักเรียน

การออกแบบการสุ่มตัวอย่างนักเรียนเป็นการออกแบบตัวอย่างแบบกลุ่มแบ่งชั้นแบบสองขั้นตอน (two-stage stratified cluster sample design) ดังนี้

- ขั้นตอนที่ 1: การเลือกโรงเรียน

- ขั้นตอนที่ 2 (PIRLS & TIMSS): การเลือกชั้นเรียนภายในโรงเรียน
- ขั้นตอนที่ 2 (PISA): การเลือกนักเรียนภายในโรงเรียน
- ขั้นตอนที่ 2 (ICCS): การเลือกชั้นเรียนภายในโรงเรียน และการเลือกครูภายในโรงเรียน

3.2 การกำหนดน้ำหนักการสุ่มตัวอย่าง (Sampling Weights)

น้ำหนักการสุ่มตัวอย่างคือ ส่วนกลับของความน่าจะเป็นของการเลือกสำหรับแต่ละคน (เช่น หากเลือก 2 ใน 20 คน น้ำหนักการสุ่มตัวอย่างคือ 10) น้ำหนักเหล่านี้ต้องถูกปรับให้เข้ากับลักษณะของตัวอย่างและขั้นตอนการเลือก เช่น การแบ่งชั้นหรือการสุ่มตัวอย่างที่ไม่ได้สัดส่วนของกลุ่มย่อย, การปรับค่าสำหรับการไม่ตอบกลับ, และการปรับค่าหลังการแบ่งชั้นไปยังยอดรวมควบคุมภายนอก น้ำหนักการสุ่มตัวอย่างต้องถูกใช้เสมอ (ALWAYS) เพื่อให้ได้ค่าประมาณประชากรที่ถูกต้อง

3.3 การประมาณความแปรปรวนจากการสุ่มตัวอย่าง (Estimating Sampling Variance)

การสุ่มตัวอย่างที่ซับซ้อนทำให้เราไม่มี "การเลือก/ตัวอย่างแบบสุ่มอย่างง่าย" (SRS) จากประชากรและบุคคลภายในกลุ่มมีแนวโน้มที่จะคล้ายคลึงกันมากกว่าบุคคลในกลุ่มอื่น วิธีการประมาณความแปรปรวนของการสุ่มตัวอย่างจากการออกแบบที่ซับซ้อนอย่างถูกต้องคือการใช้ ตัวอย่างจำลอง (Replicate samples) การทำซ้ำ (Replication) จะลบตัวอย่างย่อยที่แตกต่างกันออกจากตัวอย่างเต็มเพื่อสร้างตัวอย่างจำลอง G ชุด Jackknife Repeated Replication (JK2) ถูกใช้ใน TIMSS, PIRLS, และ ICCS และ Balanced Repeated Replication (BRR) ถูกใช้ใน PISA กล่าวโดยสรุป ต้องใช้น้ำหนักการสุ่มตัวอย่างเสมอ และต้องนำการออกแบบมาพิจารณาเสมอเมื่อคำนวณความแปรปรวนของการสุ่มตัวอย่าง (โดยใช้ Replication หรือ Taylor series linearization)

3.4 การประมาณค่าความสามารถของผู้สอบด้วย IRT (IRT scaling of data)

IRT คือ ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ที่ช่วยให้เราสรุปข้อมูลข้ามรายการข้อสอบหลายรายการได้ แม้ว่า/เมื่อคนต่างกันทำรายการข้อสอบต่างกัน โดยความสามารถของแต่ละบุคคลถูกกำหนดโดยอ้างอิงถึงความน่าจะเป็นของการตอบถูกเมื่อพิจารณาจากความยากของงาน IRT ช่วยอำนวยความสะดวกในการจัดการกับการออกแบบตัวอย่างแบบเมทริกซ์หลายชุด การประมาณค่าความสามารถของผู้สอบ จะใช้การคำนวณตามทฤษฎี IRT การศึกษาเหล่านี้ใช้การแทนค่าหลายค่า (multiple imputations) คะแนนความสามารถของผู้สอบหรือ เรียกว่าค่า Pvs (Plausible Values) ซึ่งเป็นค่า เหมือนกับการสุ่มหยิบแบบสุ่มจากการแจกแจงความสามารถที่ประมาณไว้ของนักเรียนที่มีรูปแบบการตอบรายการข้อสอบและลักษณะพื้นฐานที่คล้ายคลึงกัน ข้อควรระวังในการใช้ค่า PVs มีดังนี้ 1) ห้ามถือว่าค่าความสามารถที่เป็นไปได้เป็นคะแนนรายบุคคลเด็ดขาด 2) ห้ามใช้ค่าเฉลี่ยของ PVs ก่อนการวิเคราะห์ 3) ต้องทำซ้ำการวิเคราะห์แยกกันด้วยค่าความสามารถที่เป็นไปได้แต่ละค่าเสมอ และ 4) รายงานค่าเฉลี่ยของสถิติที่คำนวณได้ และรวมความแปรปรวนของการวัดเข้ากับความแปรปรวนจากการสุ่มตัวอย่างเมื่อทำการทดสอบนัยสำคัญ

4. ตัวอย่างการวิเคราะห์ข้อมูลการประเมินโครงการศึกษาขนาดใหญ่

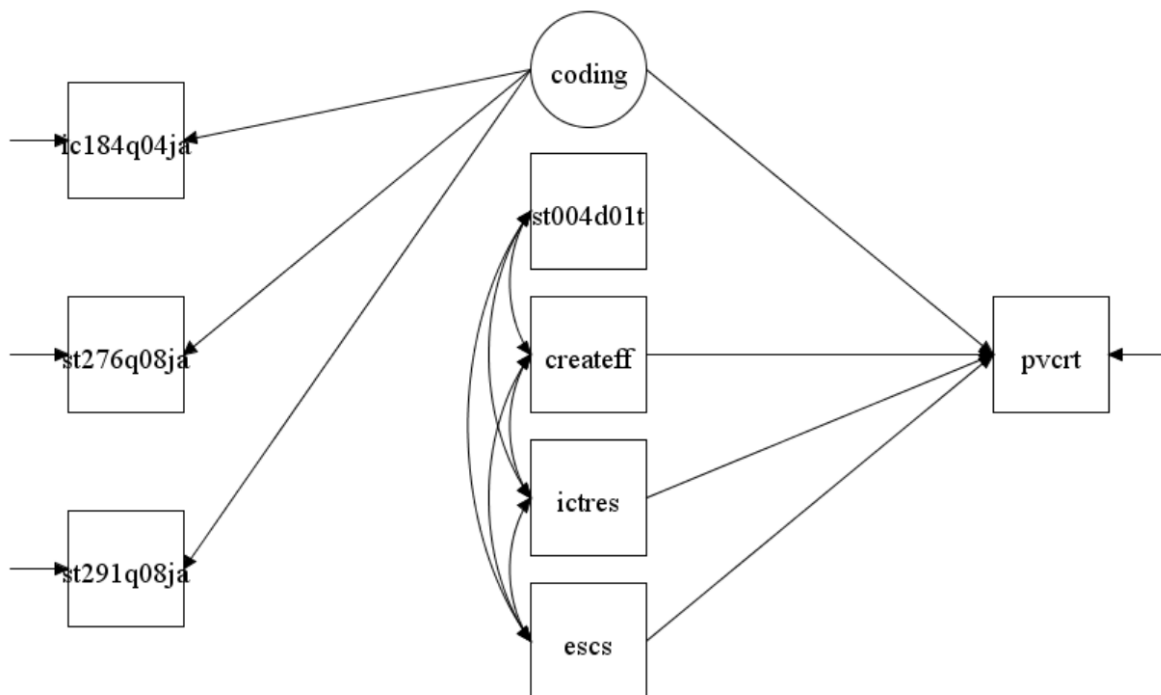
จากที่กล่าวมา การออกแบบการประเมินทางการศึกษาขนาดใหญ่ ใช้แนวคิดของ Multiple Matrix Sampling ทำให้ข้อมูลที่ได้ไม่มีความเป็นอิสระต่อกัน ในการวิเคราะห์ข้อมูล ผู้วิจัยจำเป็นต้องใช้ sampling weights เพื่อถ่วงน้ำหนักข้อมูลให้สะท้อนภาพของประชากรทั้งหมดได้อย่างถูกต้อง และใช้วิธีการ

ประมาณค่าความคลาดเคลื่อนที่เหมาะสม: เช่น Jackknife Repeated Replication (JRR) หรือ Balanced Repeated Replication (BRR) ซึ่งใช้ตัวแปรที่เรียกว่า Replicate Weights ในการคำนวณค่าความคลาดเคลื่อนมาตรฐาน (Standard Errors) ที่ถูกต้อง การละเลยการออกแบบการสุ่มตัวอย่างที่ซับซ้อนนี้จะนำไปสู่การประมาณค่าที่ผิดพลาดและค่าความคลาดเคลื่อนมาตรฐานที่ต่ำกว่าความเป็นจริง ซึ่งอาจทำให้สรุปผลผิดพลาดได้ นอกจากนี้ การออกแบบการประเมินที่ซับซ้อน (Complex Assessment Design) ทำให้พบว่า ไม่มีนักเรียนคนใดได้ทำข้อสอบทุกข้อในคลังข้อสอบทั้งหมด การออกแบบแบบนี้มีข้อดีคือสามารถประเมินเนื้อหาได้กว้างขวาง แต่ก็สร้างความท้าทายในการวัดระดับความสามารถของนักเรียนแต่ละคน เนื่องจากข้อมูลของนักเรียนแต่ละคนมีอยู่อย่างจำกัด เพื่อแก้ปัญหานี้ โครงการประเมินผลจึงใช้เทคนิคทางสถิติขั้นสูงที่เรียกว่า Plausible Values (PVs)

ตัวอย่างนี้ จะนำเสนอการใช้โปรแกรม Mplus ในการวิเคราะห์ข้อมูลประเมินทางการศึกษาขนาดใหญ่ การใช้ Mplus สามารถวิเคราะห์ข้อมูลได้มากกว่าซอฟต์แวร์ของ IEA เช่น IDB Analyzer และสามารถใช้สถิติขั้นสูงวิเคราะห์ข้อมูลได้ ตัวอย่างการวิเคราะห์ เป็นการวิเคราะห์ข้อมูลเรื่อง “Effect of Student Engagement in Coding and Creative Thinking: A Computational Literacy Policy Evaluation Using PISA 2022 Data” ตัวแปรที่ใช้ มีดังนี้

- 1) ตัวแปรต้น คือ Student’s engagement in coding (coding) วัดจากตัวแปรสังเกตได้ 3 ตัวแปร คือ
 - IC184Q04JA: How often: I use [digital resources] for coding or algorithm activities (e.g. using [Scratch], etc.)
 - ST276Q08JA : How often at school: Coding/programming computers
 - ST291Q08JA: How confident in math tasks: Coding/programming computers
- 2) ตัวแปรตาม คือ pvcr1 คือ creativity thinking (10 plausible values in creative thinking)
- 3) ตัวแปรควบคุม ประกอบด้วย Female หรือ st004d01t, Creative self-efficacy (createff), ICT Resources (ictres), และ Index of economic, social and cultural status (escs)

ความสัมพันธ์ของตัวแปรปรากฏดังโมเดลต่อไปนี้



หลักการวิเคราะห์

หลักการสำคัญของการวิเคราะห์ข้อมูล คือ การใช้ sampling weights เพื่อถ่วงน้ำหนักข้อมูลให้สะท้อนภาพของประชากรทั้งหมดได้อย่างถูกต้อง ใช้วิธีการประมาณค่าความคลาดเคลื่อนที่เหมาะสม และใช้ plausible values ค่า plausible values ที่ใช้จำนวน 10 ค่า คือ stu-creative01.dat ถึง stu.creative10.dat ซึ่งระบุในไฟล์ .txt ไฟล์ ดัง ภาพต่อไปนี้ ซึ่งในไฟล์นี้ระบุข้อมูล 10 ข้อมูลของ plausible values จำนวน 10 ข้อมูล และตั้งชื่อไฟล์ว่า datalist.txt

File	Edit	View
stu_creative01.dat		
stu_creative02.dat		
stu_creative03.dat		
stu_creative04.dat		
stu_creative05.dat		
stu_creative06.dat		
stu_creative07.dat		
stu_creative08.dat		
stu_creative09.dat		
stu_creative10.dat		

การวิเคราะห์ข้อมูลด้วย mplus จะต้องเขียนคำสั่งต่อไปนี้ โดยมีคำสั่งหลัก ดังนี้

1. TITLE: เป็นการตั้งชื่อการวิเคราะห์
2. DATA: ระบุชื่อไฟล์ข้อมูล คือ datalist.txt และนำเข้าข้อมูล plausible values ด้วยคำสั่ง Type is Imputation
3. Variable: ระบุตัวแปรในไฟล์ และเลือกตัวแปรที่ใช้ด้วยคำสั่ง USEVARIABLES ARE

