



การสร้างแผนที่ตัวแปรเชิงทฤษฎี:
แนวทางการสร้างโมเดลความคิดในระบบการประเมินที่ถูกต้องและน่าเชื่อถือ
Construct Mapping:
An Approach of Cognition Modeling in Valid Assessment Systems

วิรัช สุกศิริ¹

e-mail: w.suksiri@berkeley.edu

บทคัดย่อ

ระบบการประเมินที่ถูกต้องและน่าเชื่อถือประกอบด้วยองค์ประกอบสำคัญ 3 ส่วนที่เชื่อมโยงต่อเนื่องกันคือ โมเดลความคิด คำสังเกตหรือข้อมูล และการแปลความหมายผลวิเคราะห์ข้อมูลที่เกี่ยวข้องกลับไปยังโมเดลความคิดในตอนต้นของระบบการประเมิน โดยทั่วไป โมเดลความคิดเป็นองค์ประกอบที่ถูกละเลยในระบบการประเมินทางการศึกษาและจิตวิทยา ทำให้คำสังเกตหรือผลวิเคราะห์ข้อมูลไม่สามารถแปลความหมายและสรุปอ้างอิงเกี่ยวกับความสามารถของบุคคลได้อย่างถูกต้องและน่าเชื่อถือ วัตถุประสงค์ของบทความนี้ เพื่อนำเสนอวิธีการสร้างโมเดลความคิดที่เรียกว่าการสร้างแผนที่ตัวแปรเชิงทฤษฎี โดยมีแนวคิดสำคัญคือ การจำแนกความสามารถแฝงหรือคุณลักษณะแฝงของบุคคลตามระดับหรือปริมาณของการมีความสามารถแฝงนั้นหรือคุณลักษณะแฝงนั้นภายใต้กรอบแนวคิดทฤษฎี โดยใช้ผลการตอบข้อคำถามที่สร้างขึ้นตามโมเดลความคิด การสร้างแผนที่ตัวแปรเชิงทฤษฎีเป็นกระบวนการที่ไม่ซับซ้อน มีขั้นตอนที่เป็นระบบชัดเจนและไม่ทำให้ระบบการประเมินที่มีอยู่เดิมยุ่งยาก และยังสามารถประยุกต์ใช้ในระบบการประเมินแทบทุกประเภทได้อีกด้วย ประโยชน์ของการทำแผนที่ตัวแปรเชิงทฤษฎี นอกจากจะช่วยให้การแปลความหมายผลการวัดหรือผลการวิเคราะห์ข้อมูลมีความถูกต้องมากยิ่งขึ้นแล้ว ยังให้ข้อมูลย้อนกลับเพื่อปรับปรุงโมเดลความคิดบนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ ปรับปรุงข้อคำถามให้สอดคล้องกับโมเดลความคิด และปรับปรุงเกณฑ์การให้คะแนนให้มีความเป็นปรนัยมากขึ้น รวมถึงเป็นหลักฐานสำคัญในกระบวนการตรวจสอบความตรงเชิงโครงสร้างของการประเมินอีกด้วย

คำสำคัญ: แผนที่ตัวแปรเชิงทฤษฎี ระบบการประเมินที่ถูกต้องน่าเชื่อถือ การสร้างโมเดลความคิด โมเดลการตอบสนองข้อสอบ

¹ A doctoral student in the Quantitative Methods and Evaluation, University of California-Berkeley, USA



Abstract

A valid assessment system is comprised of three necessary components connected to one another—a model of cognition, observations, and interpretation of analysis results which articulates the model of cognition at the beginning of the assessment system. In general, a model of cognition is overlooked in educational and psychological assessment systems. This yields observations and analysis results that cannot be used for making valid interpretation and inference about the examinee’s ability. The purpose of this article is to introduce an alternative approach to constructing a model of cognition, called “*construct mapping*.” The key concept of construct mapping is to classify attributes, which describe a latent trait, aligned with a postulated continuous scale. This is done by using the examinees’ item responses to items that correspond with a model of cognition to understand the quantity or magnitude of possessing the latent trait under a theoretical framework of interest. Construct mapping is an uncomplicated and systematic process; it is easily adaptable to any kind of existing assessment system. The advantages of construct mapping are not only to support interpreting test results or analysis results more appropriately, but also, construct mapping allows for the following: 1) To give promising feedback in order to improve the model of cognition according to item response theory; 2) To revise items to correspond to the model of cognition; 3) To maximize the objectivity of a scoring guide, and 4) To be used as evidence in investigating the construct validity of an assessment.

Keywords: construct map, valid assessment system, cognition modeling, item response



บทนำ

มุมมองเชิงระเบียบวิธีวิทยา (methodological perspective) ของการประเมินมีความสำคัญไม่น้อยไปกว่ามุมมองที่ใช้นิยามคุณลักษณะต่างๆ แนวทางการสร้างโมเดลความคิดที่นำเสนอในบทความนี้ เป็นตัวอย่างหนึ่งของมุมมองเชิงระเบียบวิธีวิทยาของการประเมินในศตวรรษที่ 21 ซึ่งสามารถนำไปใช้ออกแบบการประเมินตัวแปรแฝงหรือคุณลักษณะแฝงได้อย่างถูกต้องและเหมาะสม เนื้อหาของบทความฉบับนี้เป็นไปตามตรรกะของการพัฒนาการประเมิน (logic of assessment development) (Wilson, et al, 2012) และหลักการพื้นฐานการประเมินแนวใหม่ ซึ่งสามารถศึกษาค้นคว้าเพิ่มเติมได้จาก Borsboom (2005, 2008), Engelhard (2013), Haig & Borsboom (2008), Markus & Borsboom (2013), Michell (2008), Mislavy, Steinberg, & Almond (2003), National Research Council (2001), Pellegrino & Chudowsky (2003), Wilson (2005) เนื้อหาของบทความนี้เริ่มต้นด้วยคำอธิบายและตัวอย่าง เพื่อชี้ให้เห็นความแตกต่างระหว่างการวัด การประเมิน และการประเมินผล ซึ่งมักใช้แทนกันอย่างไม่ถูกต้อง ต่อจากนั้นเป็นการอธิบายให้เห็นถึงความแตกต่างระหว่างขั้นตอนการสร้างเครื่องมือวัดแบบดั้งเดิม ระบบการประเมินที่สมบูรณ์แต่ไม่ถูกต้อง และระบบการประเมินที่ถูกต้องและน่าเชื่อถือ เนื้อหาดังกล่าวชี้ให้เห็นว่าหัวใจสำคัญของบทความนี้ เป็นการบรรยายความสัมพันธ์ระหว่างโมเดลความคิด

แผนที่ตัวแปรเชิงทฤษฎีและการสร้างแผนที่ตัวแปรเชิงทฤษฎี โดยใช้ตัวอย่างสถานการณ์ประกอบการบรรยาย รวมทั้งแสดงให้เห็นถึงตรรกะการคิดที่ช่วยทำให้การสร้างแผนที่ตัวแปรเชิงทฤษฎีเป็นเรื่องปกติ แต่มีความเป็นระบบชัดเจนมากขึ้น เนื้อหาสุดท้ายเป็นตัวอย่างของประโยชน์จากแผนที่ตัวแปรเชิงทฤษฎีที่สามารถนำผลวิเคราะห์ข้อมูลไปปรับปรุงระบบการประเมินได้ทุกขั้นตอน

ระบบการประเมินที่ถูกต้องและน่าเชื่อถือ

การประเมิน (assessment) มีนิยามที่แตกต่างกันตามความเข้าใจของผู้ใช้และบริบทที่เกี่ยวข้อง ตั้งแต่อดีตจนถึงปัจจุบัน นิยามของการประเมินมีความหมายกว้างและเกี่ยวข้องกับทั้งการวัด (measurement) การทดสอบ (testing) และการประเมินผล (evaluation) (Markus & Borsboom, 2013; Popham, 2014; Russell & Airasain, 2012) อย่างไรก็ตาม นิยามของการประเมินที่ใช้กันอย่างแพร่หลายในระบบการศึกษาปัจจุบัน ให้ความสำคัญกับกระบวนการแปลความหมาย (interpretation) ผลการวัดมากขึ้น (Brown & Wilson, 2011; National Research Council, 2001; Pellegrino & Chudowsky, 2003; Wilson et al., 2012; Wyse, 2013) กระบวนการแปลความหมายผลการวัด เป็นส่วนหนึ่งของระบบการประเมินเพื่อบรรยายความสามารถแฝง (latent ability) หรือคุณลักษณะแฝง (latent trait) ของบุคคลจากผลการวัด ตัวอย่างเช่น ในการทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชา



วิทยาศาสตร์ พบว่า นักเรียน A ได้คะแนนสอบ 60 คะแนนจาก 100 คะแนน เมื่อครูนำผลการวัดนี้ไปแปลความหมาย ครูจะสามารถบรรยายความสามารถของนักเรียนได้ว่า นักเรียน A รู้อะไรและไม่รู้อะไร ทำอะไรได้และทำอะไรไม่ได้ และมีทักษะอะไรและไม่มีทักษะอะไร เมื่อนำคะแนนของนักเรียน A ไปเปรียบเทียบกับนักเรียน B และ C ที่ได้คะแนน 90 และ 40 ตามลำดับ ครูก็จะสามารถบรรยายได้ว่านักเรียน B มีความสามารถอะไรที่แตกต่างและดีกว่านักเรียน A และนักเรียน C มีความสามารถอะไรที่เหมือนนักเรียน A และ B เป็นต้น กระบวนการประเมินแบบนี้ จึงไม่ใช่การประเมินผลที่ต้องการ ตัดสินว่า นักเรียนคนใดผ่านเกณฑ์และนักเรียนคนใดไม่ผ่านเกณฑ์ แต่เน้นที่การให้สารสนเทศเกี่ยวกับความสามารถหรือคุณลักษณะของนักเรียนให้มากที่สุด จากนั้นจึงนำสารสนเทศที่ได้นี้ไปใช้ในวัตถุประสงค์ต่างๆ ต่อไป เมื่อครูสามารถบรรยายคุณลักษณะของนักเรียนได้ ครูก็จะสามารถจัดกลุ่มนักเรียนตามความสามารถหรือคุณลักษณะที่คล้ายคลึงกันจากคะแนนสอบออกเป็นระดับต่างๆ และยังสามารถเปรียบเทียบระดับความสามารถที่แตกต่างกันได้ด้วย ทำให้ครูมองเห็นภาพระดับความสามารถของนักเรียนทั้งหมดจากระดับชั้นพื้นฐานไปสู่ระดับสูง ในกรณีที่ครูต้องการประเมินผลความสามารถของนักเรียน ครูจะนำคะแนนของนักเรียน A, B และ C ไปเปรียบเทียบกับเกณฑ์เพื่อตัดสินระดับคุณภาพ นั่นหมายความว่า นักเรียน A และ C ซึ่งอาจจะมีระดับความสามารถที่ต่างกัน

แต่อาจจะอยู่ในระดับคุณภาพเดียวกันก็ได้ หรือนักเรียนทั้งสามคนอาจจะผ่านเกณฑ์ก็ได้ ทั้งนี้ขึ้นอยู่กับเกณฑ์ที่ครูเลือกใช้ จากที่กล่าวมาข้างต้น จึงสรุปได้ว่า นิยามของการประเมินประกอบด้วยอย่างน้อย 2 ส่วนคือ การวัด และการแปลความหมายผลการวัด ส่วนการประเมินผลนั้น อาจจะรวมเป็นส่วนหนึ่งของการประเมินหรือไม่ก็ได้ ขึ้นอยู่กับวัตถุประสงค์ของการประเมินนั้นๆ

นิยามการประเมินข้างต้นมีความสอดคล้องกับสาระสำคัญของระบบการประเมินที่ถูกต้องและน่าเชื่อถือ (valid assessment system) ที่เสนอโดยคณะกรรมการวิจัยแห่งชาติของประเทศไทยสหรัฐอเมริกา (The National Research Council, 2001) ซึ่งประกอบด้วย 3 ส่วน ส่วนแรก เรียกว่า โมเดลความคิด (cognition model) ซึ่งมักจะเกี่ยวข้องกับความรู้ และความเข้าใจในเนื้อหาวิชาต่างๆ การมีทักษะปฏิบัติที่จำเป็น และการแสดงออกถึงคุณลักษณะที่คาดหวังในหลักสูตรการเรียนรู้หรือมาตรฐานการเรียนรู้ ส่วนที่สองเป็นคำสังเกต (observation) ที่ได้มาจากการเก็บรวบรวมข้อมูลของนักเรียนด้วยข้อคำถาม โดยข้อคำถามจะต้องสร้างให้สอดคล้องกับโมเดลความคิด และส่วนสุดท้ายเป็นการแปลความหมายข้อมูล (interpretation) ที่ได้จากส่วนที่สอง โดยเชื่อมโยงกลับไปยังโมเดลความคิดในส่วนที่หนึ่ง อย่างไรก็ตามที่น่าสังเกตว่ากระบวนการประเมินส่วนใหญ่ในปัจจุบันไม่มีส่วนประกอบของโมเดลความคิด ทำให้การแปลความหมายผลการวัดไม่สามารถเชื่อมโยงย้อนกลับไปยังโมเดลความคิดในตอนต้นได้ (Brown & Wilson,

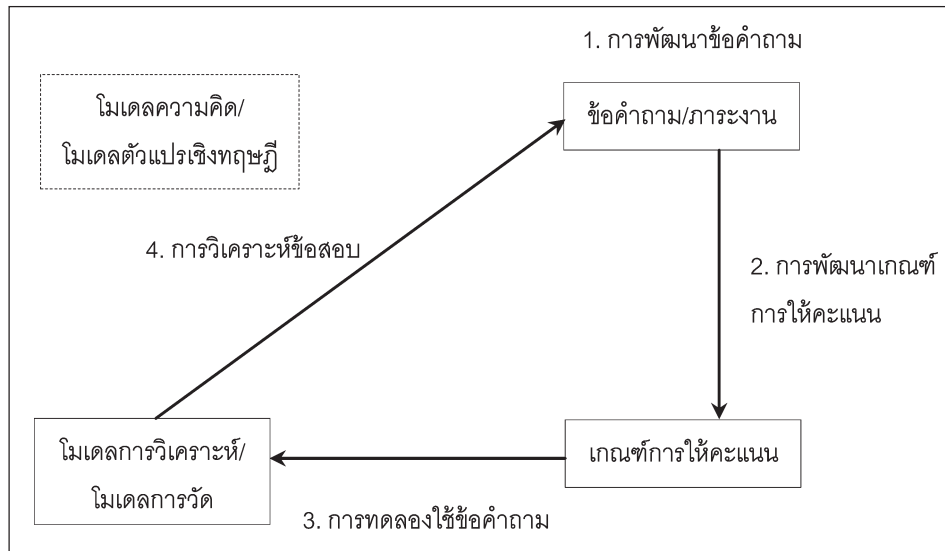


2011; Mislevy, 1996; Wilson et al., 2012) กระบวนการประเมินแบบดั้งเดิม ไม่ว่าจะเป็นการประเมินในชั้นเรียน (classroom assessment) หรือการประเมินขนาดใหญ่ (large-scale assessment) ครูมักจะเริ่มด้วยการระบุว่า ตัวแปรเชิงทฤษฎี (construct) หรือตัวแปรแฝง (latent variable) ที่ต้องการประเมินนั้นคืออะไร เช่น ความสามารถทางการเรียน ความรู้ความเข้าใจ ความสามารถทางภาษา ทักษะทางความคิด ความถนัด พัฒนาการการเรียนรู้ ความคล่องแคล่วในการทำงาน หรือแม้กระทั่งความรับผิดชอบ ความซื่อสัตย์และคุณธรรมจริยธรรม เป็นต้น ซึ่งตัวแปรเหล่านี้มีลักษณะเป็นนิยามเชิงทฤษฎี (theoretical term) เกิดขึ้นภายในสมอง และเกี่ยวข้องกับความคิด สติปัญญา การเรียนรู้ และการรับรู้ของบุคคล จากนั้นจึงส่งผลให้เกิดการตอบสนองต่อสิ่งเร้าในรูปแบบต่างๆ ทำให้ธรรมชาติของตัวแปรเหล่านี้เกี่ยวข้องกับความคิด (cognition) มีลักษณะแฝงอยู่ภายในบุคคล และมีโครงสร้าง หลากหลายตามทฤษฎีที่เกี่ยวข้อง ตั้งแต่โครงสร้างเบื้องต้นไปจนถึงซับซ้อน (Borsboom, 2008; Markus & Borsboom, 2013) จึงอาจเรียกตัวแปรเชิงทฤษฎีว่าเป็น ตัวแปรแฝงด้วยก็ได้ (Embretson & Reise, 2000) ดังนั้นเมื่อก้าวถึงโมเดลความคิดจึงมีความหมายโดยทั่วไปคล้ายคลึงกับโมเดล ตัวแปรแฝงและโมเดลตัวแปรเชิงทฤษฎี โมเดลความคิดเป็นผลที่เกิดขึ้นจากการทำความเข้าใจ ตัวแปรเชิงทฤษฎีที่ครูต้องการประเมิน โดยที่ ตัวแปรเชิงทฤษฎีนั้นมักเป็นสิ่งที่ครูไม่ได้ค้นคว้า เพื่อให้ นิยามเชิงปฏิบัติการ (operational

definition) และไม่ได้บรรยายคุณลักษณะของ นักเรียนที่สอดคล้องกับนิยามเชิงปฏิบัติการ ครู มักใช้ความคุ้นเคยหรือรู้เพียงว่าตัวแปรเชิงทฤษฎี ที่สนใจนั้นคืออะไรและประกอบด้วยอะไรบ้าง แล้วดำเนินการสร้างข้อคำถามเพื่อวัดสิ่งที่สนใจ แทนที่ ทำให้กระบวนการสร้างโมเดลความคิดไม่ ปรากฏชัดเจนและไม่สมบูรณ์ในระบบการ ประเมิน จากนั้นเมื่อครูเริ่มสร้างข้อคำถามตาม กรอบแนวคิดทฤษฎีที่กำหนดขึ้น ครูอาจจะ ขอคำชี้แนะจากผู้เชี่ยวชาญด้านเนื้อหาด้วย เรียกขั้นตอนนี้ว่า การพัฒนาข้อคำถาม (item development) เมื่อข้อคำถามสร้างเสร็จ ครูจึง สร้างเกณฑ์การให้คะแนนกับคำตอบที่เป็นไปได้ ของข้อคำถามแต่ละข้อ ซึ่งอาจจะเป็นแบบสองค่า (dichotomous scoring) เช่น ตอบถูกหรือ ตรงประเด็นได้ 1 คะแนน ตอบผิดหรือไม่ตรง ประเด็นได้ 0 คะแนน หรือเป็นแบบหลายค่า (polytomous scoring) เช่น ตอบถูกต้อง ทั้งหมดได้ 2 คะแนน ตอบถูกบางส่วนหรือตอบ ไม่ครบทุกประเด็นได้ 1 คะแนน และตอบผิด ทั้งหมดได้ 0 คะแนน ซึ่งเรียกขั้นตอนนี้ว่า การ พัฒนาเกณฑ์การให้คะแนน (scoring guide development) หลังจากนั้นจึงนำข้อคำถาม ทุกข้อไปทดลองใช้กับตัวอย่างนักเรียนที่มาจาก ประชากรเป้าหมาย เรียกขั้นตอนนี้ว่า การทดลอง ใช้ข้อคำถาม (pilot testing) และนำคำตอบที่ ได้มาทั้งหมดมาวิเคราะห์ด้วยวิธีการทางสถิติ ตามโมเดลการวัด (measurement model) เพื่อ ตรวจสอบคุณสมบัติของข้อคำถามรายข้อและ แก้ไขปรับปรุงข้อคำถามต่อไป ขั้นตอนนี้เรียกว่า การวิเคราะห์ข้อสอบ (item analysis) ขั้นตอน

ที่กล่าวมาทั้งหมดแสดงดังภาพที่ 1 นอกจากนี้ ในกรณีที่ได้พัฒนาข้อคำถามจนมีคุณภาพเป็นที่พึงพอใจแล้ว จึงนำไปใช้จริงกับนักเรียนเป้าหมาย และให้คะแนนกับคำตอบของนักเรียน เมื่อ

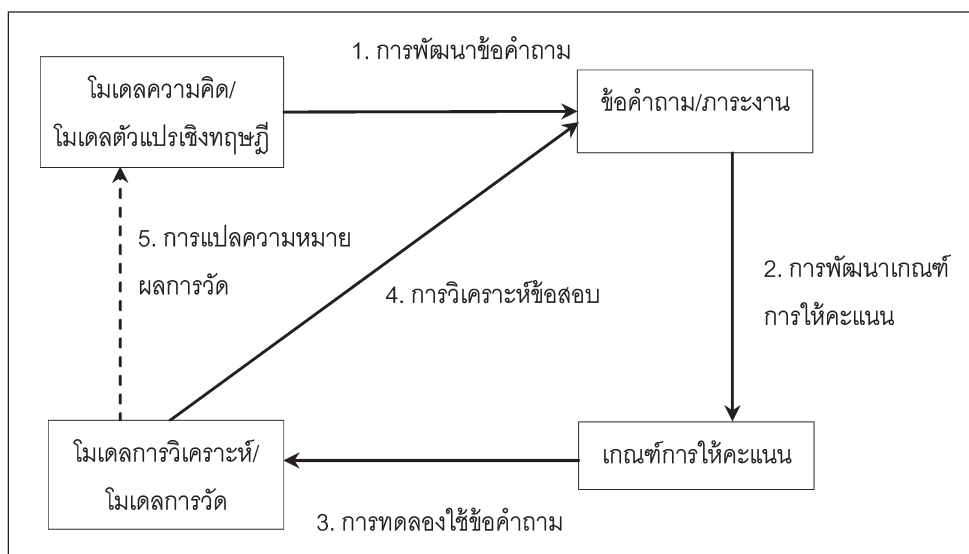
เสร็จสิ้นขั้นตอนเหล่านี้ ครูจะได้ผลการวัดหรือคะแนนที่สามารถนำไปใช้แสดงคุณลักษณะของนักเรียนตามกระบวนการประเมินแบบดั้งเดิม



ภาพที่ 1 กระบวนการพัฒนาเครื่องมือวัดแบบดั้งเดิม (Brown & Wilson, 2011)

ภาพที่ 1 แสดงให้เห็นว่ากระบวนการทั้งหมด “ไม่ใช่” ระบบการวัดประเมินที่ถูกต้องและน่าเชื่อถือตามนิยามของคณะกรรมการวิจัยแห่งชาติ ประเทศสหรัฐอเมริกา แต่เป็นกระบวนการพัฒนาเครื่องมือวัดแบบดั้งเดิม (traditional instrument development) ที่ละเลยการสร้างโมเดลความคิดในตอนแรก และใช้ข้อมูลที่ได้ทั้งหมดจากการทดลองใช้เครื่องมือเพื่อปรับปรุงแก้ไขข้อคำถามหรือสรุปผลเป็นคุณสมบัติของเครื่องมือวัดก่อนนำไปใช้จริง แทนที่จะนำไปใช้ปรับปรุงโมเดลความคิด แม้ว่าในบางกรณีครูจะโต้แย้งว่าได้ทำการค้นคว้า

เกี่ยวกับตัวแปรนั้นจากเอกสารต่างๆ เป็นอย่างดี ทำให้ทราบถึงนิยาม โครงสร้างและองค์ประกอบที่เป็นไปได้ของตัวแปร อย่างไรก็ตาม ภายหลังจากทดลองใช้เครื่องมือแล้วและอาจจะรวมถึงการนำไปใช้จริงด้วย ผลวิเคราะห์ข้อคำถามไม่ได้นำไปแปลผลเพื่อปรับปรุงหรือยืนยันโมเดลความคิดหรือโมเดลตัวแปรเชิงทฤษฎีตามที่ครูได้กำหนดไว้ตั้งแต่ตอนแรก ทำให้กระบวนการทั้งหมดจึงยังไม่เป็นระบบการประเมินที่ถูกต้องและน่าเชื่อถือตามนิยามของคณะกรรมการวิจัยแห่งชาติ ประเทศสหรัฐอเมริกา ดังแสดงในภาพที่ 2



ภาพที่ 2 ระบบการประเมินที่สมบูรณ์แต่ไม่ถูกต้อง

โดยสรุป กระบวนการพัฒนาเครื่องมือวัดแบบดั้งเดิมและระบบการประเมินที่สมบูรณ์แต่ไม่ถูกต้องมีข้อจำกัดที่สำคัญ คือ โมเดลการคิดหรือโมเดลตัวแปรเชิงทฤษฎีอาจจะปรากฏไม่ชัดเจนหรือไม่ปรากฏในระบบการประเมิน และไม่สามารถเชื่อมโยงผลวิเคราะห์ข้อมูลไปยังโมเดลความคิดได้ จึงทำให้กระบวนการแปลความหมายผลการวัดไม่สามารถทำได้อย่างถูกต้อง

โมเดลความคิดและแผนที่ตัวแปรเชิงทฤษฎี

วัตถุประสงค์สำคัญของการสร้างโมเดลความคิด คือเพื่อกำหนดโครงสร้างภายใน (internal structure) ขอบเขต (scope) และคำบรรยาย (descriptor) ของคุณลักษณะแฝงที่สนใจตามทฤษฎีที่เกี่ยวข้อง เพื่อใช้อธิบายการตอบสนองของบุคคลที่เกี่ยวข้องกับคุณลักษณะ

นั้น ตัวอย่างเช่น ครูสนใจวัดความรับผิดชอบของนักเรียน โดยให้ความสนใจที่ความรับผิดชอบทางการเรียน และกำหนดโครงสร้าง ขอบเขต คุณลักษณะความรับผิดชอบจากการทำภาระงานให้เสร็จตามจำนวนที่กำหนดและการส่งงานตรงต่อเวลา จากนั้นครูจึงอธิบายถึงคุณลักษณะที่แสดงถึงการมีความรับผิดชอบตามกรอบแนวคิดทฤษฎีที่เกี่ยวข้อง ซึ่งขั้นตอนเหล่านี้เป็นการสร้างโมเดลความรับผิดชอบ ดังนั้นโมเดลลักษณะนี้ จะแสดงถึงส่วนประกอบของความรับผิดชอบและสารสนเทศเกี่ยวกับระดับหรือปริมาณความรับผิดชอบทางการเรียนที่แตกต่างกันของนักเรียนด้วย เช่น นักเรียน X ทำการบ้านไม่เสร็จครบทุกข้อจึงไม่ส่งการบ้านตามเวลาที่กำหนด ในขณะที่นักเรียน Y ซึ่งทำการบ้านไม่เสร็จครบทุกข้อ แต่ส่งการบ้านตรงตามเวลาที่กำหนด และนักเรียน Z ทำการบ้านเสร็จทุกข้อ แต่กลับส่งการบ้านช้ากว่าเวลาที่



กำหนด เป็นต้น นอกจากนี้ยังใช้เป็นแนวทางสำคัญต่อการพัฒนาโครงสร้างภายในของ คุณลักษณะให้ชัดเจนมากยิ่งขึ้น และสามารถใช้อินโฟร์เมชันที่ได้จากการวัดเป็นหลักฐานยืนยันความตรงเชิงโครงสร้าง (construct validity) และความตรงเชิงแปลความหมาย (interpretive validity) ในกระบวนการประเมินอีกด้วย (Davis-Becker, 2013; McClarty, 2013; Messick, 1995; Mislevy, 1996; Wilson, et al., 2012)

ความตรงของผลที่ได้จากการวัดมีความสำคัญอย่างยิ่ง เนื่องด้วยผลจากการวัดนั้นจะถูกนำไปใช้แปลความหมายและสรุปอ้างอิงเกี่ยวกับคุณลักษณะของนักเรียน มโนทัศน์ของความตรงในปัจจุบัน นอกจากจะเน้นที่ความตรงเชิงโครงสร้างแล้ว ยังให้ความสนใจที่ความตรงเชิงแปลความหมายซึ่งเกี่ยวข้องกับการบรรยายคุณลักษณะของนักเรียนจากผลการวัด ดังนั้นในกระบวนการประเมินจึงจำเป็นต้องอย่างยิ่งที่ครูจะต้องเก็บรวบรวมหลักฐานให้เพียงพอเพื่อให้ผลการวัดมีความถูกต้องและน่าเชื่อถือ เพื่อนำไปใช้สนับสนุนหรือโต้แย้งข้อสรุปอ้างอิงเกี่ยวกับคุณลักษณะของนักเรียน (Messick, 1989, 1995; Mislevy & Riconscente, 2005; Mislevy, Steinberg & Almond, 2003; Wilson, 2005; Wilson, et al., 2012)

ครูต้องมีหลักฐานที่มากพอก่อนที่จะสรุปว่า นักเรียนมีคุณลักษณะอะไรบ้างและไม่มีคุณลักษณะอะไรบ้าง หลักฐานที่กล่าวถึงนี้คือ คำตอบที่มาจากการตอบสนองของนักเรียนต่อข้อคำถาม (item response) ที่สร้างขึ้นให้

สอดคล้องกับโมเดลความคิด (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999) นักเรียนที่มีคุณลักษณะแฝงหรือความสามารถแฝงที่แตกต่างกันจะตอบสนองต่อข้อคำถามแตกต่างกันด้วย เมื่อนักเรียนตอบสนองต่อข้อคำถามแตกต่างกันก็จะนำไปสู่การแปลความหมายผลการวัดที่แตกต่างกันและข้อสรุปอ้างอิงเกี่ยวกับคุณลักษณะของนักเรียนที่แตกต่างกันตามโมเดลความคิดที่ครูได้ออกแบบไว้ หลักฐานที่ครูควรเก็บรวบรวมให้มากพอนั้น ไม่ได้หมายถึงมีข้อคำถามที่วัดเรื่องเดียวกันหลายข้อในเครื่องมือวัดชุดเดียวกันหรือการเก็บข้อมูลซ้ำด้วยเครื่องมือวัดเดียวกับกลุ่มตัวอย่างนักเรียนกลุ่มเดิม หรือการมีข้อมูลเกี่ยวกับนักเรียนจากแหล่งต่างๆ เช่น คะแนนจากแฟ้มสะสมงาน คะแนนจากการทดสอบย่อย และการสังเกตพฤติกรรมนักเรียน แต่หมายถึงผลการตอบของนักเรียนที่มีต่อเครื่องมือวัดใดๆ ที่สามารถนำไปใช้แปลความหมายและโต้แย้งหรือสนับสนุนผลการประเมินนั้นๆ ได้อย่างสมเหตุสมผลและน่าเชื่อถือ ตัวอย่างของระบบการประเมินที่ให้ความสำคัญกับหลักฐานและโมเดลความคิด เช่น การออกแบบการประเมินที่เน้นหลักฐาน (evidence-centered assessment design) (Mislevy & Riconscente, 2005) กรอบการสร้างตัวแปรเชิงทฤษฎี (construct framework) (Messick, 1995) และระบบการประเมินแบบ 4 หน่วยโครงสร้าง (four building blocks



approach) (Wilson, 2005)

การสร้างแผนที่ตัวแปรเชิงทฤษฎี (construct mapping) เป็นส่วนหนึ่งของระบบการประเมินแบบ 4 หน่วยโครงสร้าง (Wilson, 2005) ที่ใช้สำหรับสร้างโมเดลความคิดหรือโมเดลตัวแปรเชิงทฤษฎี จุดเด่นของแผนที่ตัวแปรเชิงทฤษฎีคือ ทำให้ครูมองเห็นภาพโครงสร้างภาวะสันนิษฐานทางทฤษฎีของความสามารถแฝงหรือคุณลักษณะแฝงของนักเรียนอย่างชัดเจน ทำให้ผลการวัดมีความถูกต้องและมีสารสนเทศมากขึ้น ผลลัพธ์ที่ได้จากการทำแผนที่ตัวแปรเชิงทฤษฎีสามารถนำไปใช้พัฒนาระบบการประเมินได้ทั้งระบบ การสร้างแผนที่ตัวแปรเชิงทฤษฎี เป็นวิธีการที่ศาสตราจารย์ Mark Wilson (Wilson, 2005) แห่งมหาวิทยาลัยแห่งรัฐแคลิฟอร์เนีย เมืองเบิร์กลีย์ ประเทศสหรัฐอเมริกา พัฒนาขึ้นเพื่อเชื่อมโมเดลตัวแปรเชิงทฤษฎีเข้ากับผลวิเคราะห์ข้อสอบและคะแนนสอบตามทฤษฎีการตอบสนองข้อสอบแบบหนึ่งพารามิเตอร์ (one-parameter item response model) หรือโมเดลการวัดของ Rasch (Rasch measurement model) แผนที่ตัวแปรเชิงทฤษฎีจะบอกถึงปริมาณ (quantity) หรือความเข้ม (magnitude) ของการมีคุณลักษณะโดยสามารถสังเกตได้จากการตอบสนองต่อข้อคำถาม เมื่อนักเรียนมีความสามารถแตกต่างกัน ทำให้ตอบสนองต่อข้อคำถามที่สร้างตามแผนที่ตัวแปรเชิงทฤษฎีไม่เหมือนกัน และเมื่อนำรูปแบบคำตอบของนักเรียนมาแบ่งกลุ่มตามแผนที่ตัวแปรเชิงทฤษฎี เพื่อแสดงระดับความสามารถแฝงของนักเรียน

ก็จะสามารถทำได้อย่างสมเหตุสมผลและมีความถูกต้อง น่าเชื่อถือ

การสร้างแผนที่ตัวแปรเชิงทฤษฎีสามารถประยุกต์ใช้ในบริบทต่างๆ อย่างกว้างขวาง โดยเฉพาะอย่างยิ่งการประเมินทางการศึกษา เนื่องจากธรรมชาติของศักยภาพ (performance) หรือความสามารถ (ability) ที่นักเรียนได้รับการพัฒนาทั้งความรู้ กระบวนการคิดและทักษะปฏิบัติต่าง ๆ จะเกิดขึ้นตามโครงสร้างของหลักสูตรการเรียนรู้ เนื้อหาความรู้ที่นักเรียนได้รับตามหลักสูตรมักจะเริ่มต้นด้วยหลักการพื้นฐานที่ไม่ซับซ้อนเพื่อให้นักเรียนคุ้นเคยและทำความเข้าใจก่อน ทำให้ความสามารถและทักษะต่างๆ ที่เกี่ยวข้องกับหลักการพื้นฐานจึงเป็นระดับพื้นฐานด้วย เมื่อนักเรียนได้เรียนรู้เนื้อหาที่ซับซ้อนมากขึ้นและเกิดการเรียนรู้เพิ่มขึ้นความสามารถและทักษะต่างๆ ก็จะมีควมสลับซับซ้อนมากขึ้น ทำให้ลักษณะโดยธรรมชาติของคุณลักษณะที่เกี่ยวข้องกับการเรียนรู้ศักยภาพและการรับรู้ เป็นการสะสมอย่างต่อเนื่องจากระดับเบื้องต้นสู่ระดับที่ซับซ้อนและพัฒนาหรือเปลี่ยนแปลงตามแต่ศักยภาพและการรับรู้ของนักเรียนแต่ละคน

แผนที่ตัวแปรเชิงทฤษฎี (construct map) เป็นตัวแทนของโมเดลความคิดที่ช่วยให้ครูรู้วิธีการที่ชัดเจนในการออกแบบการวัดตัวแปรให้สอดคล้องกับโครงสร้างของตัวแปรที่ควรจะเป็นตามทฤษฎีหรือหลักการที่เกี่ยวข้อง (Brown & Wilson, 2011) ผู้เชี่ยวชาญด้านเนื้อหา (content expert) เป็นบุคคลที่ครูมักให้ความสำคัญและขอคำชี้แนะเพื่อตรวจสอบความตรง

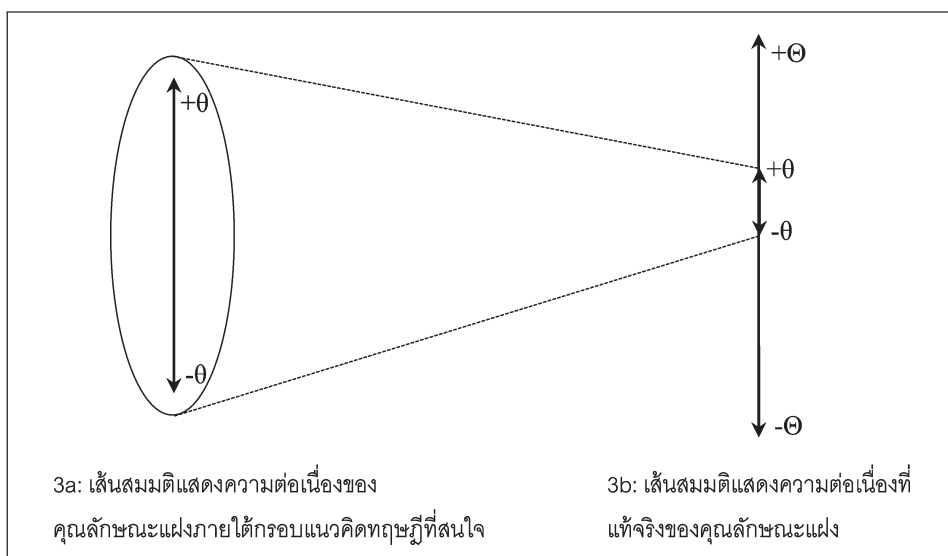


เชิงเนื้อหาของข้อคำถาม (content validation) ในกระบวนการสร้างเครื่องมือ แต่อย่างไรก็ตามผู้เชี่ยวชาญด้านเนื้อหา มักจะไม่ใช่ว่าผู้เชี่ยวชาญด้านโครงสร้างภายในของตัวแปร (construct expert) (Brown & Wilson, 2011) บทบาทของผู้เชี่ยวชาญในกระบวนการประเมินส่วนใหญ่ จึงเป็นการตรวจสอบลักษณะคร่าว ๆ ของขอบเขตและองค์ประกอบโครงสร้างของตัวแปรที่ครูต้องการวัด ซึ่งไม่เพียงพอต่อการยืนยันความถูกต้องของโมเดลตัวแปรเชิงทฤษฎี ดังนั้นในขณะที่ผู้เชี่ยวชาญด้านเนื้อหา กำลังพิจารณาความสอดคล้องของข้อคำถามกับวัตถุประสงค์ของการประเมินหรือสิ่งที่ต้องการประเมิน ผู้เชี่ยวชาญเหล่านี้ อาจจะไม่รู้เลยว่า ครูมีความคิดรวบยอดเกี่ยวกับตัวแปรอย่างไร ครูออกแบบโครงสร้างภายในของตัวแปรอย่างไร และข้อคำถามถูกสร้างตรงตามโครงสร้างภายในของตัวแปรที่เหมาะสมหรือไม่

โดยสรุป แผนที่ตัวแปรเชิงทฤษฎีมีลักษณะเป็นแผนภาพความคิดที่ครูมีต่อตัวแปรเชิงทฤษฎีนั้นๆ ภายใต้กรอบแนวคิดทฤษฎีที่เกี่ยวข้อง ทำให้เข้าใจได้ว่าครูมีมุมมองอย่างไรต่อตัวแปรนั้น แผนที่ตัวแปรเชิงทฤษฎีที่ดีควรให้สารสนเทศที่ชัดเจนเกี่ยวกับสิ่งที่ต้องการประเมินว่า มีโครงสร้างอย่างไร มีองค์ประกอบอะไรบ้าง แต่ละองค์ประกอบมีขอบเขตการประเมินอย่างไร และมีรายละเอียดของคุณลักษณะอย่างไร รวมทั้งคุณลักษณะเหล่านั้น มีความสัมพันธ์กันอย่างไรด้วย

การสร้างแผนที่ตัวแปรเชิงทฤษฎี

คุณลักษณะแฝงหรือความสามารถแฝงที่เกี่ยวข้องกับการเรียนรู้ การรับรู้และความรู้สึกนึกคิด สามารถพิจารณาในรูปของการเปลี่ยนแปลงเชิงสะสม (cumulative change) ที่เกิดขึ้นอย่างต่อเนื่องจากระดับขั้นพื้นฐานสู่ระดับที่ซับซ้อน เส้นตรงที่มีหัวลูกศรทั้งสองด้านดังแสดงในภาพ 3a แสดงถึงความต่อเนื่องของปริมาณการมีคุณลักษณะอย่างใดอย่างหนึ่งจากระดับน้อย (หัวลูกศรด้านล่าง) สู่ระดับมาก (หัวลูกศรด้านบน) หรืออาจจะแสดงถึงการเปลี่ยนแปลงจากคุณลักษณะอย่างหนึ่งเป็นอีกคุณลักษณะหนึ่งที่เกี่ยวข้องกัน ช่วงห่างระหว่างหัวลูกศรทั้งสองอันเป็นสิ่งสมมติที่ครูสามารถกำหนดขึ้นเอง เพื่อแสดงถึงขอบเขตของคุณลักษณะที่สนใจ แม้ว่าเส้นตรงนี้ จะมีความยาวแตกต่างกันได้ แต่ขอบเขตของคุณลักษณะที่ครูต้องการวัดเป็นเพียงส่วนหนึ่งของเส้นความต่อเนื่องที่แท้จริงของคุณลักษณะแฝงนั้น (ดูภาพ 3b) นั้นหมายความว่า ส่วนของเส้นตรงที่อยู่ถัดลงมา จากหัวลูกศรตรงจุด -0 จนถึง -1 เป็นส่วนที่ถูกกำหนดให้เป็นขอบเขตของการไม่มีหรือมีคุณลักษณะน้อยภายใต้กรอบแนวคิดทฤษฎี ในทำนองเดียวกัน ส่วนของเส้นตรงจากจุด -0 จนถึง -1 ถูกกำหนดให้เป็นขอบเขตของการมีคุณลักษณะภายใต้กรอบแนวคิดทฤษฎีเช่นเดียวกัน

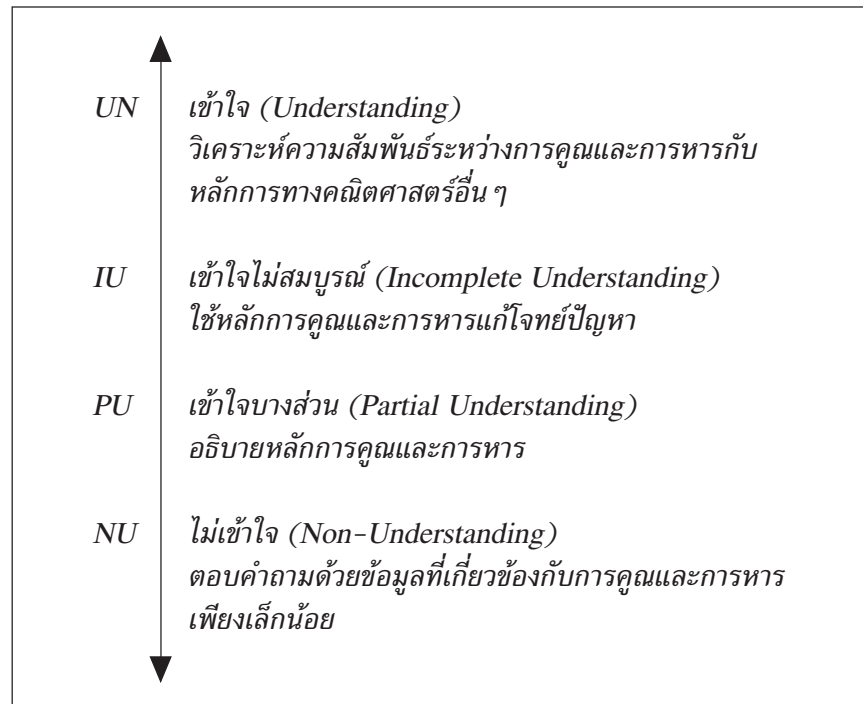


ภาพที่ 3 เส้นสมมติความต่อเนื่องของคุณลักษณะของตัวแปรแฝง

เส้นตรงในภาพ 3a นี้ เป็นเสมือนแกนหลักในการสร้างแผนที่ตัวแปรเชิงทฤษฎี ส่วนล่างของเส้นตรงแสดงถึงการไม่มีคุณลักษณะหรือมีคุณลักษณะในปริมาณน้อย จากนั้นปริมาณของคุณลักษณะที่สนใจจะค่อยๆ เพิ่มขึ้นไปจนถึงส่วนบนของเส้นตรงซึ่งแสดงถึงการมีคุณลักษณะหรือมีปริมาณมาก คุณลักษณะต่างๆ ที่เกิดขึ้นระหว่างจุด $-θ$ ถึง $+θ$ สามารถแบ่งออกเป็นระดับต่างๆ ที่สอดคล้องกับระดับความสามารถแฝงของนักเรียนและรูปแบบการตอบสนองต่อข้อคำถาม ดังนั้นยังมีองค์ประกอบอีก 2 ส่วนที่สัมพันธ์กับแกนเส้นตรง 3a คือ คำบรรยายเกี่ยวกับคำตอบของนักเรียน (description of response) และคำบรรยายเกี่ยวกับความสามารถของนักเรียน (description of respondent) ที่ครูควรสนใจก่อนที่จะสร้างข้อคำถาม

ตัวอย่างเช่น ถ้าครูต้องการวัดความเข้าใจเกี่ยวกับการคูณและการหารของนักเรียนชั้น

ประถมศึกษาปีที่ 4 โดยครูมีแนวคิดว่าคุณความเข้าใจของนักเรียนสามารถวัดได้จากการใช้หลักการคูณและการหารแก้โจทย์ปัญหา สามารถเปรียบเทียบความเหมือนและความแตกต่างระหว่างการคูณกับการหาร และการบวกกับการลบ รวมทั้งเข้าใจถึงความสัมพันธ์ระหว่างการคูณและการหารกับหลักการทางคณิตศาสตร์อื่นๆ นอกจากนี้ความเข้าใจของนักเรียนยังสามารถวัดได้จากทักษะการสื่อสารทางคณิตศาสตร์ โดยนักเรียนสามารถสร้างสมการที่ถูกต้องเพื่อแก้โจทย์ปัญหาที่กำหนด จากนั้นครูจึงเริ่มสร้างแผนที่ตัวแปรเชิงทฤษฎีขึ้นมา โดยออกแบบให้โครงสร้างของตัวแปรความเข้าใจแบ่งออกเป็น 4 ระดับ มีขอบเขตเริ่มต้นจากระดับต่ำที่สุด คือไม่เข้าใจ ไปจนถึงระดับสูงสุดคือเข้าใจ พร้อมกับมีคำอธิบายสั้นๆ เกี่ยวกับความสามารถของนักเรียนในแต่ละระดับตามกรอบแนวคิดที่ครูกำหนดขึ้น ดังแสดงในภาพที่ 4



ภาพที่ 4 แผนที่ตัวแปรเชิงทฤษฎีความเข้าใจเกี่ยวกับการคูณและการหาร

หลักการสำคัญในการสร้างแผนที่ตัวแปรเชิงทฤษฎี มีสองประการคือ ประการแรก ความเหมาะสมของจำนวนระดับความเข้าใจ ที่จะต้องสอดคล้องกับทฤษฎีหรือสอดคล้องกับทฤษฎีสันนิษฐาน และมีโอกาสเกิดขึ้นได้จริง หมายความว่า ครูควรมีหลักการที่สมเหตุสมผลมาสนับสนุนการแบ่งระดับความเข้าใจ และแต่ละระดับก็ควรมีความเป็นไปได้ที่นักเรียนจะตอบสนองออกมา สำหรับเทคนิคที่ช่วยให้ครูสามารถแบ่งระดับความเข้าใจได้ดีขึ้นนั้น อาจจะเริ่มต้นจากการพิจารณาถึงคุณลักษณะที่แสดงว่านักเรียนเข้าใจประกอบด้วยอะไรบ้าง จากนั้นจึงพิจารณาว่าคุณลักษณะที่ตรงกันข้ามกับความเข้าใจประกอบด้วยอะไรบ้าง นั่นคือครูกำลังพิจารณาถึงคุณลักษณะที่แสดงถึงความไม่เข้าใจ

เมื่อครูได้คุณลักษณะสุดโต่งของทั้งสองด้านของแผนที่ตัวแปรเชิงทฤษฎีแล้ว จึงพิจารณาถึงคุณลักษณะที่อยู่ระหว่างระดับเข้าใจและไม่เข้าใจในลำดับต่อไป ประการที่สอง สารสนเทศที่คาดว่าจะได้รับจากระดับความเข้าใจ เนื่องจากครูอาจจะสร้างแผนที่ตัวแปรเชิงทฤษฎีให้มีจำนวนระดับน้อยเกินไป ทำให้ได้สารสนเทศเพียงคร่าวๆ ไม่สามารถวินิจฉัยความสามารถของนักเรียนได้ถูกต้องแม่นยำ ในทางตรงกันข้ามครูอาจจะสร้างแผนที่ตัวแปรเชิงทฤษฎีให้มีจำนวนระดับมากเกินไป แม้ว่าสารสนเทศที่คาดว่าจะได้รับจะมีความละเอียดมากขึ้น แต่ครูก็ไม่สามารถสร้างข้อคำถามที่กระตุ้นให้นักเรียนตอบสนองให้ตรงกับแต่ละระดับได้



ระดับความเข้าใจ

คำบรรยายความสามารถของนักเรียน

คำบรรยายคำตอบของนักเรียน

เข้าใจ (UN): วิเคราะห์ความสัมพันธ์ระหว่างการคุณและการหารกับหลักการทางคณิตศาสตร์อื่นๆ

เข้าใจไม่สมบูรณ์ (IU): ใช้หลักการคุณและการหารแก้โจทย์ปัญหา

เข้าใจบางส่วน (PU): อธิบายหลักการคุณและการหาร

ไม่เข้าใจ (NU): ตอบคำถามด้วยข้อมูลที่เกี่ยวข้องกับการคุณและการหารเพียงเล็กน้อย

นักเรียนสามารถอธิบายคุณสมบัติการคูณและการหาร เข้าใจความสัมพันธ์ระหว่างการคูณและการหารในระบบสมการทางคณิตศาสตร์ สามารถสื่อสารทางคณิตศาสตร์ ใช้หลักการคูณและการหารแก้โจทย์ปัญหาที่ซับซ้อน เปรียบเทียบความแตกต่างของหลักการคูณและการหารกับหลักการทางคณิตศาสตร์อื่นๆ เลือกใช้วิธีที่เหมาะสมในการแก้โจทย์ปัญหา

นักเรียนสามารถอธิบายคุณสมบัติการคูณและการหาร ใช้ทักษะการสื่อสารทางคณิตศาสตร์และหลักการคูณและการหารแก้โจทย์ปัญหา เปรียบเทียบความแตกต่างของหลักการคูณและการหารกับหลักการทางคณิตศาสตร์อื่นๆ สามารถจำแนกวิธีการหรือหลักการที่ไม่เหมาะสมในการแก้โจทย์ปัญหา

นักเรียนสามารถอธิบายหลักการคูณและการหารได้เพียงบางส่วน แก้โจทย์ปัญหาได้อย่างจำกัดโดยอาจจะใช้หลักการบวกและการลบหรือหลักการทางคณิตศาสตร์อื่นๆ ใช้ทักษะการสื่อสารทางคณิตศาสตร์ได้จำกัด แต่สามารถจำแนกวิธีการหรือหลักการบางอย่างที่ไม่เหมาะสมในการแก้โจทย์ปัญหา

นักเรียนไม่สามารถใช้หลักการคูณและการหารแก้โจทย์ปัญหา ไม่สามารถตอบคำถามด้วยข้อมูลที่เกี่ยวข้องกับการคูณและการหาร นักเรียนอาจจะจดจำหลักการคูณและการหารได้ แต่ไม่สามารถนำมาใช้แก้โจทย์ปัญหา

คำตอบที่ตรงกับระดับนี้ ประกอบด้วยข้อสรุปที่ถูกต้องในการเลือกใช้หลักการและวิธีการที่เหมาะสมเพื่อแก้โจทย์ปัญหา มีข้อสนับสนุนหรือข้อโต้แย้งที่สมเหตุสมผลในการแก้โจทย์ปัญหา คำตอบที่เป็นไปได้บ่งชี้ว่า นักเรียนรู้คำตอบที่ถูกต้องและรู้ว่าคำตอบใดไม่ถูกต้อง

คำตอบที่ตรงกับระดับนี้ ประกอบด้วยข้อสรุปที่ถูกต้องในการเลือกใช้หลักการหรือวิธีการที่เหมาะสมเพื่อแก้โจทย์ปัญหา มีข้อโต้แย้งหรือข้อสนับสนุนบางประการในการแก้โจทย์ปัญหา คำตอบที่เป็นไปได้บ่งชี้ว่า นักเรียนรู้คำตอบที่ถูกต้องและรู้เพียงบางคำตอบว่าเป็นคำตอบที่ไม่ถูกต้อง

คำตอบที่ตรงกับระดับนี้ ประกอบด้วยข้อสรุปที่ไม่ถูกต้องจากการเลือกใช้หลักการที่ไม่เหมาะสมเพื่อแก้โจทย์ปัญหา มีข้อสนับสนุนว่านักเรียนไม่สามารถใช้หลักการคูณและการหารแก้โจทย์ปัญหา มีข้อผิดพลาดในการสื่อสารทางคณิตศาสตร์ คำตอบที่เป็นไปได้บ่งชี้ว่านักเรียนไม่รู้คำตอบที่ถูกต้อง แต่รู้ว่าคำตอบใดไม่ถูกต้อง

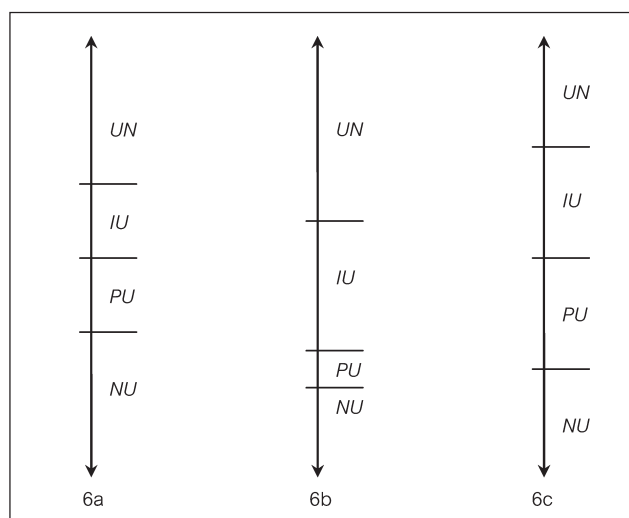
คำตอบที่ตรงกับระดับนี้ ประกอบด้วยข้อมูลที่เกี่ยวข้องกับการคูณและการหารเพียงเล็กน้อยหรือมีข้อมูลที่ไม่เกี่ยวข้องกับการแก้โจทย์ปัญหาจำนวนมาก คำตอบที่เป็นไปได้บ่งชี้ว่านักเรียนไม่รู้คำตอบที่ถูกต้องและไม่รู้ว่าคำตอบใดไม่ถูกต้อง นอกจากนี้ การไม่ตอบคำถาม เว้นว่าง และตอบคำถามไม่สมบูรณ์จนหาข้อสรุปไม่ได้ถือว่าอยู่ในระดับนี้ด้วย

ภาพที่ 5 คำบรรยายสรุปของระดับความเข้าใจเกี่ยวกับการคูณและการหาร



ขั้นตอนต่อไปเป็นการเขียนบรรยายคำตอบของนักเรียนและคำบรรยายความสามารถของนักเรียนที่แสดงถึงความเข้าใจเรื่องการคูณและการหาร ซึ่งครูจะต้องเป็นผู้เขียนขึ้นเอง ขั้นตอนนี้เป็นการให้สารสนเทศในภาพรวมของการประเมินและมีความสำคัญที่สุดต่อการสร้างแผนที่ตัวแปรเชิงทฤษฎีให้สมบูรณ์ รวมทั้งครูจะใช้สารสนเทศนี้เป็นแนวทางในการสร้างข้อคำถาม ใช้แปลความหมายผลการวัดและสรุปอ้างอิงเกี่ยวกับความสามารถของนักเรียนต่อการเขียนบรรยายคำตอบของนักเรียน เป็นกิจกรรมที่ครูจะต้องอธิบายให้ได้ว่า ในคำตอบของนักเรียนจะต้องประกอบด้วยข้อมูลหรือหลักฐานอะไรบ้างที่แสดงว่านักเรียนเข้าใจ เข้าใจไม่สมบูรณ์ เข้าใจบางส่วน และไม่เข้าใจตามแผนที่ตัวแปรเชิงทฤษฎีที่สร้างขึ้น และในทำนองเดียวกัน การสร้างคำบรรยายความสามารถของนักเรียนจึงเป็นความพยายามของครูเพื่อบอกว่า นักเรียนมีคุณลักษณะอย่างไร รู้อะไรบ้างและทำอะไรได้บ้างในแต่ละระดับ ตัวอย่างของคำบรรยายดังกล่าว แสดงดังภาพที่ 5

จากภาพที่ 4 และ 5 จะเห็นว่า มีหลักการอีกประการหนึ่งที่ซ่อนอยู่ภายในแผนภาพทั้งสอง กล่าวคือ เส้นแบ่งความสามารถระหว่างระดับความเข้าใจที่อยู่ติดกันไม่ได้ถูกกำหนดขึ้นเลย เนื่องจากตามแนวคิดในการสร้างแผนที่ตัวแปรเชิงทฤษฎีนั้น ครูไม่สามารถระบุตำแหน่งของเส้นแบ่งแต่ละเส้นได้อย่างถูกต้องและแม่นยำ ครูมีเพียงกรอบแนวคิดที่สามารถระบุในเบื้องต้นว่าคุณลักษณะของนักเรียนนั้น ๆ น่าจะอยู่ระหว่างระดับความเข้าใจใดในแผนที่ตัวแปรเชิงทฤษฎี ดังนั้นจึงเป็นไปได้ว่า แผนที่ตัวแปรเชิงทฤษฎีที่สอดคล้อง (fit) กับคำตอบของนักเรียน อาจจะมีเส้นแบ่งความสามารถที่ทำให้ระดับเข้าใจและระดับไม่เข้าใจกว้างกว่าระดับอื่น (ภาพ 6a) หรืออาจจะมีเส้นแบ่งที่ทำให้ระดับเข้าใจบางส่วนและระดับไม่เข้าใจแคบ (ภาพ 6b) กว่าระดับอื่น หรืออาจจะเป็นไปได้ว่าแผนที่ตัวแปรเชิงทฤษฎีมีเส้นแบ่งระดับความสามารถที่ทำให้ทุกระดับมีความกว้างใกล้เคียงกัน (ภาพ 6c) เป็นต้น



ภาพที่ 6 ตัวอย่างตำแหน่งของเส้นแบ่งระดับความเข้าใจของนักเรียนเกี่ยวกับการคูณและการหาร



ภายหลังจากที่ครูได้สร้างแผนที่ตัวแปรเชิงทฤษฎีพร้อมกับคำบรรยายคำตอบและคำบรรยายความสามารถของนักเรียนแล้ว ขั้นตอนต่อไปเป็นการสร้างข้อคำถามที่จะต้องสามารถกระตุ้นให้นักเรียนตอบสนองได้ใกล้เคียงกับคำบรรยายคำตอบมากที่สุด เพื่อที่ครูจะได้ใช้อ้างอิงความสามารถของนักเรียนตามคำบรรยายความสามารถได้อย่างชัดเจนและสอดคล้องกับโครงสร้างทางทฤษฎีของตัวแปร เมื่อครูสร้างข้อคำถามได้ครบตามจำนวนที่ต้องการ และกำหนดคะแนนให้กับคำตอบแต่ละระดับ และนำข้อคำถามทั้งหมดไปทดลองใช้กับกลุ่มตัวอย่าง และสุดท้ายให้คะแนนกับคำตอบของนักเรียนตามเกณฑ์ที่กำหนดขึ้นเรียบร้อยแล้ว ครูก็จะนำข้อมูลทั้งหมดไปวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ ก็จะได้ค่าประมาณความสามารถของนักเรียนและค่าพารามิเตอร์ข้อสอบที่สัมพันธ์กับแผนที่ตัวแปรเชิงทฤษฎีตรงตามที่ครูตั้งใจไว้ตั้งแต่ต้น จึงสรุปได้ว่า การสร้างแผนที่ตัวแปรเชิงทฤษฎีและกระบวนการที่เกี่ยวข้องในการประเมินที่ครูกำลังดำเนินการอยู่ทั้งหมดนี้เป็นการประเมินที่มีความตรงเชิงโครงสร้างและความตรงเชิงแปลความหมาย

ประโยชน์ของแผนที่ตัวแปรเชิงทฤษฎี

แผนที่ตัวแปรเชิงทฤษฎีสามารถนำไปใช้ได้หลากหลายบริบทและวัตถุประสงค์ เช่น พัฒนามาตรฐานการเรียนรู้ พัฒนาเครื่องมือวัดคุณลักษณะที่คาดหวัง กำหนดเกณฑ์การผ่านขั้นต่ำ (minimum passing score) ในการทดสอบขนาดใหญ่ ปรับปรุงเกณฑ์การประเมิน

คุณภาพสถานศึกษา และสร้างข้อสอบใหม่จากข้อสอบที่เผยแพร่ไปแล้ว เป็นต้น สารสนเทศสำคัญที่ได้จากการวิเคราะห์รูปแบบคำตอบ (response pattern) ด้วยโมเดลการวัดของ Rasch คือ ค่าประมาณความสามารถแฝง (latent ability estimate; θ) และค่าพารามิเตอร์ความยากข้อสอบ (item difficulty parameter; δ) โดยสารสนเทศทั้งสองนี้จะถูกนำไปใช้ในการปรับปรุงกระบวนการต่างๆ ในระบบการประเมินเดิม ทำให้ได้ระบบการประเมินใหม่ที่ต้องการและน่าเชื่อถือมากยิ่งขึ้น การเสนอเนื้อหาต่อไปนี้เป็นตัวอย่างการนำผลวิเคราะห์ข้อมูลที่ได้จากระบบการประเมินที่ต้องการและน่าเชื่อถือไปใช้ปรับปรุงแผนที่ตัวแปรเชิงทฤษฎี ข้อคำถาม และเกณฑ์การให้คะแนน มีรายละเอียดดังนี้

1. ปรับปรุงแผนที่ตัวแปรเชิงทฤษฎี

การวัดคือการกำหนดค่าเชิงปริมาณให้กับสิ่งที่เกิดขึ้นจากกระบวนการสังเกต (Stevens, 1946) เมื่อคำตอบที่ได้จากนักเรียนซึ่งเป็นสิ่งที่เกิดขึ้นจากกระบวนการสังเกต ถูกจัดกลุ่มตามแผนที่ตัวแปรเชิงทฤษฎี แล้วให้ค่าเชิงปริมาณ ก็จะทำให้คุณสมบัติของจำนวนหรือตัวเลขกลายเป็นคุณสมบัติของการวัดและกลายเป็นคุณสมบัติของกลุ่มคำตอบในท้ายที่สุด เมื่อครูนำข้อมูลเหล่านี้มาวิเคราะห์ด้วยโมเดลการวัด ครูจะได้ค่าพารามิเตอร์ความยากของข้อสอบแต่ละข้อ โดยค่าความยากนี้จะทำหน้าที่เป็นเส้นแบ่งความสามารถของนักเรียนออกเป็นกลุ่มย่อยๆ กล่าวคือ นักเรียนที่มีค่าความสามารถแฝงมากกว่าค่าความยากข้อสอบ ($\theta > \delta$) จะมีโอกาสมากกว่า



ร้อยละ 50 ที่จะตอบข้อคำถามได้ดีกว่านักเรียนที่มีค่าความสามารถแฝงน้อยกว่าค่าความยากข้อสอบ ($\theta < \delta$)

จากตัวอย่างการประเมินความเข้าใจของนักเรียนเกี่ยวกับการคูณและการหาร สมมติว่าครูสร้างข้อคำถามจำนวน 10 ข้อ คำตอบที่เป็นไปได้ของข้อคำถามแต่ละข้อมีค่าตั้งแต่ 0 ถึง 3 คะแนนตามแผนที่ตัวแปรเชิงทฤษฎี โดยคำตอบที่ตรงกับระดับไม่เข้าใจ (NU) ได้ 0 คะแนน ระดับเข้าใจบางส่วน (PU) ได้ 1 คะแนน ระดับเข้าใจไม่สมบูรณ์ (IU) ได้ 2 คะแนน และระดับเข้าใจ (UN) ได้ 3 คะแนน ตามลำดับ แล้วนำข้อมูลนี้ไปวิเคราะห์ ก็จะได้ค่าความยากของข้อคำถามจำนวน 3 ค่าต่อหนึ่งข้อ คือ ค่าความยากลำดับที่ 1 (first threshold) เป็นเส้นแบ่งความสามารถระหว่างกลุ่มนักเรียนที่ตอบแบบไม่เข้าใจ ($X_{nu}=0$) และเข้าใจบางส่วน ($X_{pu}=1$) ค่าความยากลำดับที่ 2 (second threshold) เป็นเส้นแบ่งระหว่างกลุ่มนักเรียนที่ตอบแบบเข้าใจบางส่วน ($X_{pu}=1$) และเข้าใจไม่สมบูรณ์ ($X_{iu}=2$) และค่าความยากลำดับที่ 3 (third threshold) เป็นเส้นแบ่งระหว่างกลุ่มนักเรียนที่ตอบแบบเข้าใจไม่สมบูรณ์ ($X_{iu}=2$) และเข้าใจ ($X_{un}=3$) ตามลำดับ เมื่อนำค่าความยากของแต่ละลำดับจากข้อคำถามทุกข้อมาหาค่าเฉลี่ย ก็จะได้เส้นแบ่งความสามารถเฉลี่ยในที่สุด

สมมติว่าตำแหน่งของเส้นแบ่งความสามารถเฉลี่ยเป็นดังภาพ 6b สารสนเทศที่ได้จากแผนภาพดังกล่าว จะชี้แนะครูให้กลับไปทบทวนกรอบแนวคิดทฤษฎีที่ใช้ในการสร้าง

โมเดลความคิดตั้งแต่ต้นว่า ความสามารถของนักเรียนในระดับเข้าใจบางส่วนนั้นเป็นสิ่งที่สามารถสังเกตได้หรือไม่ และมีความแตกต่างจากระดับความเข้าใจที่อยู่ข้างเคียงกันอย่างไร การกำหนดคำบรรยายคำตอบและคำบรรยายความสามารถที่นำไปสู่การให้คะแนนนักเรียน สอดคล้องกับแผนที่ตัวแปรเชิงทฤษฎีหรือไม่ และมีความสมเหตุสมผลหรือไม่ที่ครูจะรวมระดับเข้าใจบางส่วนและระดับไม่เข้าใจเป็นระดับเดียวกัน เพื่อเพิ่มนัยสำคัญในการแปลความหมายความสามารถของนักเรียน เป็นต้น ในกรณีที่ตำแหน่งเส้นแบ่งความสามารถเฉลี่ยเป็นดังภาพ 6a ครูอาจจะตัดสินใจรวมระดับเข้าใจบางส่วนและระดับเข้าใจไม่สมบูรณ์เป็นระดับเดียวหรือไม่ก็ได้ ขึ้นอยู่กับกรอบแนวคิดทฤษฎีที่ครูทบทวนจากเอกสารและงานวิจัยที่เกี่ยวข้อง หรือจากประสบการณ์ที่ครูสังเกตจากคำตอบของนักเรียนโดยตรง ถ้าครูยืนยันว่าแผนที่ตัวแปรเชิงทฤษฎีที่ปรากฏสะท้อนปรากฏการณ์จริงในการประเมินความเข้าใจของนักเรียน เพียงแต่ว่ามีนักเรียนจำนวนไม่มากที่มีความสามารถอยู่ในระดับเข้าใจไม่สมบูรณ์และระดับเข้าใจบางส่วน เป็นต้น การให้เหตุผลในลักษณะนี้เป็นสิ่งที่สามารถทำได้โดยตรงไปตรงมาและสมเหตุสมผลบนพื้นฐานของผลวิเคราะห์ที่ครูได้รับจากการสร้างแผนที่ตัวแปรเชิงทฤษฎีนั้นหมายความว่า การสร้างแผนที่ตัวแปรเชิงทฤษฎีให้ประโยชน์โดยตรงต่อการพัฒนาหรือแก้ไขโมเดลความคิดของการประเมิน และยิ่งให้หลักฐานสำหรับครูนำไปใช้ได้แย่งหรือ

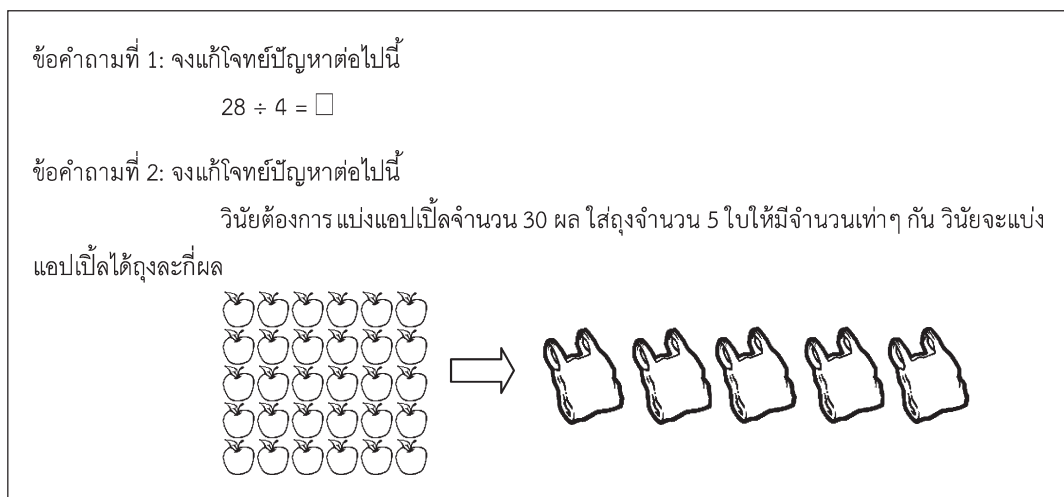


สนับสนุนสิ่งที่ครูจะทำและจะไม่ทำในกระบวนการประเมินความรู้ความสามารถของนักเรียนอีกด้วย

2. ปรับปรุงข้อคำถาม

ลักษณะของข้อสอบที่สร้างขึ้นตามแผนที่ตัวแปรเชิงทฤษฎีนั้นมีคุณลักษณะเด่นที่สำคัญสองประการ ประการแรก คือ การตั้งโจทย์คำถาม (stem) ที่อ้างอิงจากแผนที่ตัวแปรเชิงทฤษฎีจะช่วยให้โจทย์คำถามมีความตรงเชิงโครงสร้างมากขึ้น กล่าวคือ โดยปกติแล้วโจทย์คำถามทั่วไปมุ่งให้ความสนใจที่เนื้อหาและ/หรือทักษะ โดย

ไม่สามารถอธิบายได้ว่า โจทย์คำถามนั้นจะมีระดับความเข้มของเนื้อหาและระดับทักษะการคิดที่เกี่ยวข้องเป็นอย่างไร เมื่อครูใช้แผนที่ตัวแปรเชิงทฤษฎีในการตั้งโจทย์คำถาม ครูจะมีกรอบแนวคิดว่า ต้องการวัดความสามารถของนักเรียนในระดับใดและความสามารถระดับนั้นสัมพันธ์กับเนื้อหาและทักษะอะไรบ้าง ตัวอย่างเช่น ครูสร้างข้อคำถามจำนวน 2 ข้อตามแผนที่ตัวแปรเชิงทฤษฎีความเข้าใจเกี่ยวกับการคูณและการหาร (ภาพที่ 4 และ 5) ดังแสดงในภาพที่ 7



ภาพที่ 7 ตัวอย่างข้อคำถามที่วัดความสามารถของนักเรียนในระดับที่แตกต่างกันตามแผนที่ตัวแปรเชิงทฤษฎี

จากภาพที่ 7 จะเห็นว่า ข้อคำถามทั้งสองข้อวัดเนื้อหาเดียวกัน แต่นักเรียนที่ตอบคำถามข้อที่ 2 ได้ถูกต้องนั้น มีระดับความเข้าใจที่สูงกว่านักเรียนที่ตอบถูกข้อที่ 1 ตามคำบรรยายคำตอบ และคำบรรยายความสามารถในภาพที่ 5 สมมติว่า นักเรียน D และ E ตอบคำถามข้อที่ 1 ได้ถูกต้อง ในขณะที่นักเรียน F ตอบผิด ครูสามารถแปลความหมายคำตอบของนักเรียน D

และ E ได้ว่า มีความสามารถอยู่ในระดับเข้าใจไม่สมบูรณ์ (IU) ส่วนนักเรียน F ก็สามารถระบุได้ว่ามีความสามารถอยู่ในระดับตั้งแต่เข้าใจบางส่วนลงไป (PU) เมื่อพิจารณาข้อคำถามที่ 2 พบว่า นักเรียน D สามารถตอบคำถามได้ถูกต้อง คำตอบของนักเรียน D ประกอบด้วยสมการที่ใช้แก้โจทย์ปัญหา ซึ่งครูแปลความหมายคำตอบของนักเรียน D ได้ว่าอยู่ในระดับเข้าใจไม่สมบูรณ์



(IU) เช่นเดียวกับข้อที่ 1 แต่สำหรับนักเรียน E ซึ่งตอบผิด พบว่า ไม่สามารถใช้ทักษะการสื่อสารทางคณิตศาสตร์ในการแปลโจทย์สถานการณ์ให้เป็นสมการทางคณิตศาสตร์ได้ จึงอยู่ในระดับเข้าใจบางส่วน (PU) ส่วนนักเรียน F นั้น ไม่สามารถตอบคำถามได้เลย คำตอบของนักเรียน F แสดงให้เห็นว่า นักเรียนพยายามเขียนบางอย่างแต่ไม่เกี่ยวข้องกับการแก้โจทย์ปัญหา ซึ่งครูแปลความหมายว่า ความสามารถของนักเรียน F อยู่ในระดับไม่เข้าใจ (NU)

จากตัวอย่างการตอบของนักเรียน D, E และ F จะเห็นได้ว่า ข้อคำถามที่สร้างตามแผนที่ตัวแปรเชิงทฤษฎีโดยคำนึงถึงรูปแบบคำตอบที่เกิดขึ้นด้วย นอกจากจะทำให้ข้อคำถามวัดได้ตรงกับโครงสร้างทางทฤษฎีของตัวแปรแล้ว คำตอบที่เป็นไปได้ ก็สามารถแปลความหมายคำตอบที่ได้อย่างสมเหตุสมผลตามแผนที่ตัวแปรเชิงทฤษฎีที่พัฒนาขึ้น หรือมีความตรงเชิงแปลความหมาย ซึ่งเป็นลักษณะเด่นประการที่สองของข้อคำถามลักษณะนี้

3. ปรับปรุงเกณฑ์การให้คะแนน

ประโยชน์ที่ครูจะได้รับจากการตัดสินใจปรับปรุงแผนที่ตัวแปรเชิงทฤษฎีโดยรวมระดับความเข้าใจบางระดับ หรือแม้กระทั่งแบ่งย่อยระดับความเข้าใจให้ละเอียดมากยิ่งขึ้น คือ เกณฑ์การให้คะแนนคำตอบของนักเรียนที่สอดคล้องกับข้อมูลมากขึ้น เนื่องจากในตอนแรกนั้น เกณฑ์การให้คะแนนเป็นสิ่งที่ครูกำหนดขึ้นเองตามคำอธิบายคำตอบที่สัมพันธ์กับแผนที่ตัวแปรเชิงทฤษฎี เมื่อแผนที่ตัวแปรเชิงทฤษฎีถูกปรับให้สอดคล้องกับข้อมูลมากขึ้น เกณฑ์

การให้คะแนนก็สมควรปรับปรุงเช่นกัน หลักการเบื้องต้นในการปรับปรุงเกณฑ์การให้คะแนนนั้น นิยมให้ระดับต่ำสุดของแผนที่ตัวแปรเชิงทฤษฎีจะมีค่าเท่ากับ 0 คะแนน จากนั้นค่าจะเพิ่มขึ้นทีละ 1 คะแนน เพื่อใช้แทนอันดับที่สูงขึ้น นอกจากนี้ผลการวิเคราะห์ที่ได้จากการทดลองใช้ข้อคำถาม ยังสามารถนำไปใช้ปรับปรุงคำอธิบายคำตอบและคำอธิบายความสามารถของนักเรียนเพื่อให้เห็นความแตกต่างที่ชัดเจนมากขึ้นระหว่างระดับต่างๆ ในแผนที่ตัวแปรเชิงทฤษฎี ตัวอย่างเช่น ตำแหน่งเส้นแบ่งความสามารถเฉลี่ยเป็นดังภาพ 6b เมื่อครูรวมระดับเข้าใจบางส่วนเข้ากับระดับไม่เข้าใจ แผนที่ตัวแปรเชิงทฤษฎีอันใหม่จะมีเพียงสามระดับ คือระดับไม่เข้าใจ ซึ่งมีค่าเท่ากับ 0 คะแนน ระดับเข้าใจไม่สมบูรณ์เท่ากับ 1 คะแนน และระดับเข้าใจ เท่ากับ 2 คะแนน ตามลำดับ จากนั้นครูจะต้องปรับแก้ทั้งคำบรรยาย คำตอบ และคำบรรยายความสามารถนักเรียนใหม่ด้วย เพื่อให้กระบวนการออกแบบการประเมินทั้งหมดสอดคล้องกลมกลืนกัน

บทสรุป

การสร้างแผนที่ตัวแปรเชิงทฤษฎี เป็นวิธีการที่ทำทลายความสามารถของครูในการออกแบบการประเมินตัวแปรต่างๆ แม้ว่าครูอาจจะไม่คุ้นเคยกับขั้นตอนต่างๆ ในการสร้างแผนที่ตัวแปรเชิงทฤษฎี และอาจจะใช้เวลามากเพื่อทำความเข้าใจและฝึกปฏิบัติอย่างต่อเนื่อง แต่ประโยชน์ที่ครูจะได้รับจากแผนที่ตัวแปรเชิงทฤษฎีคุ่มค่า ครูจะมีหลักฐานที่สมเหตุสมผล



สำหรับใช้ในการโต้แย้งหรืออธิบายเพิ่มเติมเมื่อมีบุคคลอื่นไม่เห็นด้วยกับผลประเมินนั้น ๆ รวมทั้งยังใช้สนับสนุนผลการประเมินในหลายกรณีด้วย การสร้างแผนที่ตัวแปรเชิงทฤษฎีสามารถประยุกต์ใช้ได้เป็นอย่างดีกับการประเมินในชั้นเรียนและการประเมินขนาดใหญ่ เนื่องจากมีขั้นตอนที่เป็นระบบชัดเจนและไม่ทำให้ระบบการประเมินที่มีอยู่เดิมยุ่งยาก ประโยชน์ของการทำแผนที่ตัวแปรเชิงทฤษฎี นอกจากจะใช้เป็นหลักฐานสำหรับสรุปอ้างอิงผลการประเมินแล้ว ยังให้ข้อมูลย้อนกลับเพื่อปรับปรุงโมเดลความคิดหรือโครงสร้างทางทฤษฎีของคุณลักษณะที่สนใจ ปรับปรุงข้อคำถามให้สอดคล้องกับโมเดลความคิด และปรับปรุงเกณฑ์การให้คะแนนให้มีความเป็นปรนัยมากขึ้นอีกด้วย

เอกสารอ้างอิง

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press.

- Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research and Perspectives*, 6(1), 25-53.
- Brown, N., & Wilson, M. (2011). A model of cognition: The missing cornerstone of assessment. *Educational Psychological Review*, 23(2), 221-234.
- Davis-Becker, S. (2013). Construct maps: Do they make the unclear clear? *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 174-176.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. New York, NY: Routledge.
- Haig, B. D., & Borsboom, D. (2008). On the conceptual foundations of psychological measurement. *Measurement: Interdisciplinary Research and Perspectives*, 6(1), 1-6.
- Markus, K. A., & Borsboom, D. (2013). *Frontiers of test validity theory*. New York, NY: Edwards Brothers.



- McClarty, K. L. (2013). Construct maps: A tool to organize validity evidence. *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 185-188.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *The American Psychologist*, 50(9), 741-749.
- Michell, J. (2008). Is psychometrics pathological science? *Measurement: Interdisciplinary Research and Perspectives*, 6(1), 7-24.
- Mislevy, R. J. (1996). Test theory reconceived. *Journal of Educational Measurement*, 33(4), 379-416.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-62.
- Mislevy, R. J., & Riconscente, M. M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology*. (PADI Technical Report 9). Menlo Park, CA: SRI international.
- National Research Council (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. J. Pellegrino, N. Chudowsky, & R. Glaser (Eds). Washington, DC: National Academies Press.
- Pellegrino, J. W., & Chudowsky, N. (2003). The foundations of assessment. *Measurement: Interdisciplinary Research and Perspectives*, 1(2), 103-148.
- Popham, W.J. (2014). *Classroom assessment: What teachers need to know*. Upper Saddle River, NJ: Pearson Education.
- Russell, M. K., & Airasian, P. W. (2012). *Classroom assessment: Concepts and applications*. New York, NY: McGraw-Hill.



- Steven, S. S. (1946). On the theory of scales of measurement. *Science*, 103 (2684), 677-680.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Wilson, M., Bejar, I., Scalise, K., Templin, J., William, D., & Irrirbrara, T. (2012). Perspectives on methodological issues in P. Griffi, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 67-141). New York, NY: Springer.
- Wyse, A. E. (2013). Construct maps as a foundation for standard setting. *Measurement: Interdisciplinary Research and Perspectives*, 11(4), 139-170.