

มาตรฐานการทดสอบทางด้านการศึกษาและจิตวิทยา Standard Testing of Education and Psychological

นายเจนรบ โกรธา รหัสนักศึกษา 595050089-7

นักศึกษาระดับปริญญาโท สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษามาตรฐานการทดสอบทางด้านการศึกษาและจิตวิทยา โดยที่ศาสตร์ทางด้านจิตวิทยาและการศึกษามีความสัมพันธ์กันอย่างมาก โดยเฉพาะอย่างยิ่งการวัดทางการศึกษาและจิตวิทยาเป็นการวัดคุณลักษณะภายในของมนุษย์ ซึ่งไม่สามารถสังเกตได้โดยตรง แต่มีความสำคัญและจำเป็นต้องศึกษา เพราะการวัดคุณลักษณะภายในจะทำให้เข้าใจการเกิดพฤติกรรมภายนอกของมนุษย์ที่สามารถสังเกตได้โดยตรง อันจะนำไปสู่การทำนายควบคุม และพัฒนาพฤติกรรมมนุษย์ การวัดคุณลักษณะภายในของมนุษย์จำเป็นต้องอาศัย ทฤษฎีการทดสอบ เพื่อทำความเข้าใจคุณลักษณะของสิ่งที่มุ่งวัด โครงสร้างของการวัด และการพัฒนาเครื่องมือสำหรับทดสอบ จากการศึกษาพบว่า การวัดทางการศึกษาเป็นการวัดสมรรถภาพของมนุษย์ซึ่งอาจเป็นความสามารถทางสมอง พฤติกรรม ความรู้สึกนึกคิดในตัวบุคคล ลักษณะดังกล่าวต้องมีการกำหนดขอบข่าย โครงสร้างหรือคุณลักษณะ เนื่องจากเป็นลักษณะของสิ่งที่ไม่มิตัวตน ดังนั้นการวัดจะแตกต่างจากการวัดที่เป็นรูปธรรมหรือการวัดทางกายภาพเช่นเดียวกันกับการทดสอบทางจิตวิทยา สำหรับหลักการสำคัญมาตรฐานการทดสอบทางจิตวิทยา คือ ความเป็นมาตรฐาน (Standardization) ความเป็นปรนัย (Objectivity) มีค่าระดับมาตรฐานกลาง (Test Norms) มีความน่าเชื่อถือ (Reliability) ความเที่ยงตรง (Validity) ทั้งเที่ยงตรงตามเกณฑ์ที่กำหนด (Criterion – related) และความเที่ยงตรงในการทำนาย (Predictive validity) และหลักการสำคัญมาตรฐานการทดสอบทางการศึกษา สามารถนำไปสู่การปฏิบัติเพื่อให้เกิดผลเป็นรูปธรรมได้โดยมาตรฐานประกอบด้วย 5 ด้าน คือ มาตรฐานการบริหารการทดสอบ มาตรฐานบุคลากรด้านการทดสอบ มาตรฐานการพัฒนาแบบทดสอบ มาตรฐานการพิมพ์ การรับ ส่ง การตรวจ และการประมวลผล และมาตรฐานการรายงานผล และการนำผลไปใช้

คำสำคัญ: มาตรฐานการทดสอบ, การทดสอบทางด้านการศึกษา, การทดสอบทางจิตวิทยา

Corresponding author Tel.: 09-0852-9136

E-mail address: jenroprotha@kkumail.com

บทนำ

พัฒนาการของทฤษฎีการทดสอบนั้นมีความสัมพันธ์กันอย่างมากกับศาสตร์ทางด้านจิตวิทยา โดยเริ่มต้นจากนักจิตวิทยาเองที่ต้องการทดสอบและหาวิธีการแก้ปัญหาหารวัดต่างๆ เพื่อพัฒนาจนเป็นศาสตร์การวัดและตรวจสอบที่มีความน่าเชื่อถืออย่างที่เราพบเห็นและใช้กันในปัจจุบัน

ศาสตร์ทางด้าน การวัดและประเมินผลนั้น มีมายาวนาน แต่ตอนปลายศตวรรษที่ 19 ที่นักวิทยาศาสตร์และนักสถิติได้ให้ความสำคัญ มีการพัฒนาทฤษฎีการทดสอบอย่างมากและมีความรวดเร็วและกว้างขวางรวมทั้งมีความเป็นระบบมาก เนื่องมาจากการรวมตัวกันของนักวัดผลเพื่อแลกเปลี่ยนความคิดเห็นงานวิชาการกันจนทำให้เกิดเป็นแนวคิดใหม่ๆ ทฤษฎีการทดสอบอย่างมากมาจนเป็นที่ยอมรับและสนใจอย่างกว้างขวาง โดยเริ่มต้นนั้นทฤษฎีการทดสอบแบบดั้งเดิม (CTT) เป็นทฤษฎีที่ถือปฏิบัติใช้กันมานานแล้ว ส่วนใหญ่จะรู้จักกันในชื่อ ทฤษฎีคะแนนจริงแบบดั้งเดิม ซึ่งองค์ความรู้ทั่วไปจะเน้นเกี่ยวกับการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนจริง รวมทั้งวิเคราะห์คุณภาพโดยส่วนรวมของข้อสอบและแบบสอบสำหรับผู้สอบเฉพาะกลุ่ม ทฤษฎีคะแนนจริงแบบดั้งเดิมนี้อาศัยอยู่บนพื้นฐานของโมเดลการวัดและข้อตกลงเบื้องต้นที่สำคัญหลายประการ โดยทฤษฎีนี้หลักๆ จะเน้นเพื่อการพัฒนาหลักการสร้างข้อสอบ การวิเคราะห์คุณภาพของแบบสอบ และการวิเคราะห์คุณภาพของข้อสอบ

ทฤษฎีการทดสอบแบบดั้งเดิม Classical Test Theory (CTT) ถูกใช้เป็นแนวคิดพื้นฐานสำหรับการพัฒนาแบบทดสอบทางการศึกษาและจิตวิทยาเป็นส่วนใหญ่ ตลอดช่วงศตวรรษที่ 20 ที่ผ่านมา ในปัจจุบันก็ยังมีการใช้กันอย่างกว้างขวาง แต่ด้วยข้อจำกัดหลายประการของการวัดตามทฤษฎีการทดสอบแบบดั้งเดิม ทำให้มีนักทฤษฎีทางการทดสอบหลายท่านได้ก่อกระแสการวัดคุณลักษณะภายในแนวใหม่เพื่อที่จะแก้ไขจุดอ่อนเกี่ยวกับข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบดั้งเดิม เริ่มต้นด้วยเทอร์สตัน (L.L. Thurston, 1927, 1928) เป็นผู้บุกเบิกการวัดคุณลักษณะภายในบุคคล และพัฒนาเทคนิคการวิเคราะห์ตัวประกอบสำหรับศึกษาคุณลักษณะทางจิตวิทยา อันเป็นการวางรากฐานความคิดที่สำคัญเกี่ยวกับ ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory: MTT) จากผลงานของ ครอนบาคและคณะ (Cronbach et.al., 1963, 1972) ได้เขียนบทความเรื่อง “The Dependability of Behavioral Measurement: Theory of Generalizability of Score and Profiles” ได้เสนอแนวคิดเกี่ยวกับโมเดลความเที่ยงทั่วไปภายใต้เงื่อนไขแบบต่างๆของการทดสอบ รวมทั้งผลงานของ ลอร์ดและโนวีค (Lord and Novice, 1968) เรื่อง “Statistical Theories of Mental Test Score” ซึ่งได้นำเสนอหลักการและแนวคิดสำคัญที่ช่วยผลักดันกระแสการปฏิรูประบบความคิดของการวัดสู่ทฤษฎีการทดสอบแนวใหม่ รวมทั้งการมีส่วนร่วมของนักทฤษฎีการทดสอบในยุคใหม่อีกหลายท่าน จนได้พัฒนาการทดสอบแนวใหม่ซึ่งสามารถจำแนกออกเป็น 2 แนวทางสำคัญๆ ได้แก่ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability Theory) หรือ G-Theory และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) หรือ IRT

ทฤษฎีการทดสอบเป็นองค์ความรู้ที่มีนัยทั่วไปเกี่ยวกับการทดสอบ วิธีการแก้ปัญหา การทดสอบและพัฒนาเครื่องมือการทดสอบ ซึ่งจะช่วยให้นักวัดผลสามารถทำการสร้างและพัฒนาแบบสอบให้มีคุณภาพสามารถแปลความหมายผลการวัดได้อย่างถูกต้องและสามารถนำเสนอสารสนเทศไปใช้สำหรับการตัดสินใจได้อย่างเหมาะสม ประเภททฤษฎีการทดสอบสามารถแบ่งได้เป็น 2 ประเภทหลักๆ ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งจะเน้นการตรวจสอบความสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนที่แท้จริง โดยการวิเคราะห์คุณภาพรวมของข้อสอบและแบบสอบ ส่วนทฤษฎีการทดสอบแนวใหม่ มุ่งเน้นการขยายแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิม เพื่อให้ได้แนวทางการวัดที่น่าเชื่อถือยิ่งขึ้นซึ่งมีพัฒนาการที่

สำคัญ 2 แนวทาง ได้แก่ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory) และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT)

ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) จะมุ่งเน้นการขยายแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิมให้มีขอบเขตที่กว้างขวาง และมีความน่าเชื่อถือมากยิ่งขึ้น โดยปรับข้อตกลงให้มีความสมเหตุสมผล ตามความจริงมากขึ้น เกี่ยวกับความคลาดเคลื่อนของการวัด ยอมให้มีความคลาดเคลื่อนจากการวัดนั้นมีความแตกต่างกันออกไปตามสถานการณ์การวัด เพื่อศึกษาความเที่ยงทั่วไป ภายใต้เงื่อนไขของการวัดตามแนวของ Generalizability Theory: G-Theory รวมทั้งศึกษาความคลาดเคลื่อนจากการวัดที่มีความแตกต่างกันไปตามลักษณะของข้อสอบและแบบสอบ ขึ้นกับระดับความสามารถของผู้สอบ และของข้อสอบและแบบสอบ ทั้งยังพยายามวัดคุณลักษณะภายในหรือความสามารถที่แท้จริงของบุคคลตามแนว Item Response Theory (IRT) ทฤษฎีการทดสอบแนวใหม่จึงน่าจะให้ผลการวัดที่ชัดเจนตรงประเด็นมากขึ้น อันจะเป็นประโยชน์ต่อการพัฒนาข้อสอบและแบบสอบให้สามารถนำไปใช้ในสถานการณ์ต่างๆของการทดสอบได้อย่างมีประสิทธิภาพ สามารถระบุศักยภาพของคะแนนจริง คุณของแบบสอบตามเงื่อนไขของการทดสอบ รวมทั้งการประเมินความสามารถที่แท้จริงของผู้สอบ และการบรรยายพฤติกรรมการตอบสนองข้อสอบของผู้สอบได้เป็นอย่างดี

จากการที่กล่าวมาข้างต้นจะพบว่า การวัดทางการศึกษาและจิตวิทยาเป็นการวัดคุณลักษณะภายในของมนุษย์ ซึ่งไม่สามารถสังเกตได้โดยตรง แต่มีความสำคัญและจำเป็นต้องศึกษา เพราะการวัดคุณลักษณะภายในจะทำให้เข้าใจการเกิดพฤติกรรมภายนอกของมนุษย์ที่สามารถสังเกตได้โดยตรง อันจะนำไปสู่การทำนายควบคุม และพัฒนาพฤติกรรมมนุษย์ การวัดคุณลักษณะภายในของมนุษย์จำเป็นต้องอาศัย ทฤษฎีการทดสอบ เพื่อทำความเข้าใจคุณลักษณะของสิ่งที่มุ่งวัด โครงสร้างของการวัด และการพัฒนาเครื่องมือสำหรับทดสอบ ดังนั้นการทดสอบทางด้านการศึกษาและจิตวิทยาควรจะมีรูปแบบหรือมาตรฐานในการบริหารจัดการในทุกกระบวนการที่เหมือนหรือแตกต่างกันอย่างไร ในการศึกษามาตรฐานการทดสอบทางด้านการศึกษาและจิตวิทยาในครั้งนี้จะขอกล่าวถึง ความเป็นมาของมาตรฐานการทดสอบทางด้านการศึกษาและจิตวิทยา แนวคิดและหลักการของมาตรฐานการทดสอบทางด้านการศึกษาและจิตวิทยา และตัวอย่างของมาตรฐานการทดสอบ

1. ความเป็นมาของการทดสอบทางด้านการศึกษาและจิตวิทยา

1.1 ความเป็นมาของการทดสอบทางด้านการศึกษา

ยุค 210 ปีก่อนคริสตกาลในสมัยราชวงศ์ ฮั่น (HAN) จีนได้มีระบบการสอนเพื่อคัดเลือกบุคคลเข้ารับราชการเป็นครั้งแรก โดยการให้คำตอบคำถามด้วยความเรียงและแต่งร้อยกรอง

ในยุโรปศตวรรษที่ 14 พ่อค้าในเมืองทรีวิโซ (Treviso) ประเทศอิตาลีต้องการความมั่นใจว่าลูกของเขาจะสามารถบริหารธุรกิจของพ่อต่อไปได้ จึงได้กำหนดเงินเดือนครูระดับการสอบ ในหลักสูตรยุคนั้นซึ่ง ใช้การสอบปากเปล่า

ศตวรรษที่ 16 ถึงศตวรรษที่ 20 โดยใช้คำถามสอบถามความรู้ โดยเฉพาะเรื่องเกี่ยวกับศาสนา คำตอบจะได้รับการประเมินถูก หรือผิดซึ่งเป็นวิธีการที่ไม่เป็นทางการมากนัก สมัยต่อมาจึงได้มีการประเมินแบบเป็นทางการมากขึ้น

ปลายศตวรรษที่ 18 ได้เริ่มมีพัฒนาการของการสอบวัดในสาขาจิตวิทยาในอเมริกา ปี ค.ศ. 1845 ฮอแรช แมน ได้ใช้ข้อสอบความเรียงแทนการสอบปากเปล่าในโรงเรียนในเมืองบอสตัน

ต่อมาในปี ค.ศ. 1916 เฮอร์แมน ได้ปรับปรุงและหาค่าปกติพิสัยแบบทดสอบของบินด์ใหม่ เพื่อใช้กับเด็กอเมริกันเรียกว่าแอสตันพอร์ต บินด์ แบบทดสอบเพื่อวัดความสามารถของชาวปัญญาจึง

กลายเป็นการทดสอบในระบบอเมริกัน

องค์ประกอบประการที่สองที่ทำให้การสอบวัดเป็นวิทยาศาสตร์มากขึ้น ได้แก่ การจัดการทางธุรกิจและอุตสาหกรรม เนื่องจากคนเป็นองค์ประกอบสำคัญในการผลิตและให้บริการนักเรียนถูกมองว่าเป็นวัตถุดิบที่จะต้องผ่านกระบวนการในโรงเรียน การทดสอบจึงมีบทบาทมากขึ้น

ในปี ค.ศ. 1908 นับเป็นปีที่เริ่มบุกเบิกและตีพิมพ์แบบทดสอบวัดผลสัมฤทธิ์ในโรงเรียน โดยมี แบบทดสอบเลขคณิต เป็นครั้งแรก

ในปี ค.ศ. 1993 แบบทดสอบ Eight Year Study ได้ถูกนำมาใช้ในโรงเรียนมัธยมเพื่อศึกษาว่าในแต่ละช่วงของ 4 ปี หลักสูตรซึ่งนักเรียนได้เรียนประสบความสำเร็จหรือไม่ ซึ่งต่อมา ไทเลอร์ ได้เก็บรวบรวมข้อมูลจากการทดสอบและได้เสนอแนะการประเมินโครงการขึ้น

ตั้งแต่ปี ค.ศ. 1920 เป็นต้นมา แบบทดสอบชนิดเลือกตอบและแบบทดสอบมาตรฐานได้ถูกนำมาใช้อย่างกว้างขวางกับผู้เรียนในสหรัฐอเมริกา แม้ว่าจะมีการวิพากษ์วิจารณ์จากฝ่ายที่ไม่เห็นด้วย แต่แบบทดสอบชนิดนี้สามารถจัดการบริหารการสอบได้สะดวกและสิ้นเปลืองน้อย

ในปี ค.ศ. 1955 ลินดควิสต์ (Lindquist) ได้ทำให้การใช้แบบทดสอบชนิดเลือกตอบแพร่หลายทั่วไป ใช้ประเมินความก้าวหน้าของผู้เรียนที่มีเชื้อชาติ วัฒนธรรม ภาษาเดิม หรือเพศที่ต่างกันได้

ปี ค.ศ. 1992 วูดรuff (Woodruff) ประธานกรรมการของ Graduate Record Examination ได้เสนอการทดสอบโดยประยุกต์ใช้กับคอมพิวเตอร์ เพื่อให้การสอบมีประสิทธิภาพมากขึ้น และใช้เวลาเฉลี่ยน้อยลง

1.2 ความเป็นมาของการทดสอบทางด้านจิตวิทยา

Sir Francis Galton นักชีววิทยาชาวอังกฤษได้ค้นคว้าเรื่องกรรมพันธุ์ของมนุษย์ โดยได้ตระหนักถึงความจำเป็นในการวัดลักษณะทางชีววิทยาทั้งที่เกี่ยวข้องและเป็นญาติและไม่เกี่ยวข้องกันเป็นญาติ ความแตกต่างและความคล้ายคลึงกันระหว่างพ่อแม่และลูก, พี่กับน้อง ศึกษาส่วนต่าง ๆ ในร่างกายมนุษย์ มีการบันทึกและเก็บข้อมูลการทดสอบเรื่องการมองเห็นและการได้ยิน, ความแข็งแรงของกล้ามเนื้อ และปฏิกิริยาตอบสนอง

James Cattle ได้นำหลักของ Galton ไปใช้เพื่อศึกษาวิจัยเกี่ยวกับเชาวน์ปัญญา โดยวัดความสามารถที่อยู่บนพื้นฐานเกี่ยวกับระบบประสาท

Hermann Ebbinghaus นักจิตวิทยาชาวเยอรมันได้พัฒนานำเอาคณิตศาสตร์และการทำประโยชน์ให้สมบูรณ์มาใช้ทดสอบกับเด็กนักเรียน

Alfred Binet นักจิตวิทยาชาวฝรั่งเศส ศึกษาเกี่ยวกับสาเหตุที่ทำให้เด็กเรียนช้า Binet ร่วมมือกับ Theodore Simon พัฒนาแบบทดสอบทางสติปัญญา (Test of Intelligence) ซึ่งประกอบด้วยคำถาม 30 ข้อ ใช้วัดความสามารถในการตัดสินใจ, ความเข้าใจ และการใช้เหตุผล แบบทดสอบนี้ยังได้รับการพัฒนาอย่างต่อเนื่องโดยนักจิตวิทยาชาวอเมริกัน Lewis Terman ในแนวคิดเกี่ยวกับ IQ (Intelligence Quotient) ซึ่งได้รับการปรับปรุงอีกหลายครั้งจนมีการนำมาใช้กันอย่างแพร่หลาย

นักจิตวิทยาเหล่านี้เป็นผู้บุกเบิกทำให้เกิดแบบทดสอบอีกมากมายในปีต่อ ๆ มา ถึงแม้แบบทดสอบในช่วงแรกจะทดสอบความสามารถทางด้านสติปัญญา แต่ต่อมาก็ได้ขยายมาถึงการทดสอบเรื่อง ทักษะคิด, ความสามารถ, ความสนใจ และบุคลิกภาพด้วย

2. แนวคิดและหลักการของมาตรฐานการทดสอบทางการศึกษาและจิตวิทยา

2.1 แนวคิดและหลักการการทดสอบทางการศึกษา

2.1.1 ลักษณะของการวัดทางการศึกษา

การวัดทางการศึกษาเป็นการวัดสมรรถภาพของมนุษย์ซึ่งอาจเป็นความสามารถทางสมอง พฤติกรรม ความรู้สึกนึกคิดในตัวบุคคล ลักษณะดังกล่าวต้องมีการกำหนดขอบข่าย โครงสร้างหรือคุณลักษณะ เนื่องจากเป็นลักษณะของสิ่งที่ไม่เป็นตัวตน ดังนั้นการวัดจะแตกต่างจากการวัดที่เป็นรูปธรรม หรือการวัดทางกายภาพ เช่น การวัดความยาวของโต๊ะ การวัดความสูงของห้องเรียน วัดถ่วงกล่าวมีโครงสร้างแน่นอนตายตัว แต่การวัดในสิ่งที่เป็นนามธรรมไม่สามารถมองเห็นรูปร่างได้ จึงจำเป็นต้องกำหนดกรอบของสิ่งที่ต้องการวัดก่อน การวัดคุณลักษณะต่างๆ ทางการศึกษา มีลักษณะทั่วไปดังนี้

การวัดทางการศึกษาเป็นการวัดทางอ้อม เนื่องจากการวัดทางการศึกษาจะมีลักษณะเป็นการวัดสิ่งที่ไม่เป็นนามธรรม จับต้องไม่ได้ นอกจากนี้คนอยู่ในสภาพสังคมที่มีการเปลี่ยนแปลงได้ง่ายและมีความซับซ้อน การวัดความสามารถหรือลักษณะดังกล่าวจึงไม่สามารถวัดได้โดยตรง เช่น การวัดผลสัมฤทธิ์ทางการเรียน การวัดทัศนคติต่อวิชาชีพ การวัดความถนัด หรือการวัดความสนใจ เป็นต้น เครื่องมือที่จะใช้ นำไปวัดจะต้องสร้างขึ้นโดยกำหนดหรือสมมุติช่วงให้เท่ากัน และไม่สามารถกำหนดค่าศูนย์ที่แท้จริงให้กับคุณลักษณะนั้นๆ หรือกล่าวได้ว่า ไม่มีศูนย์แท้และไม่สามารถกำหนดตำแหน่งที่ชัดเจนตายตัวให้กับคุณลักษณะนั้นๆ ได้ การกำหนดกรอบหรือโครงสร้างให้กับสิ่งที่วัดซึ่งต้องมีหลักการหรือทฤษฎีรองรับจึงเป็นความพยายามในการสร้างความตรงให้กับการวัด แต่ในสภาพความเป็นจริงไม่มีใครบอกได้ว่าปริมาณหรือขอบข่ายที่แท้จริงของสิ่งนั้นๆ อยู่ที่ใด เช่นการวัดความสามารถในการอ่านเราสามารถบอกได้ว่าประชากรมีความสามารถในการอ่านที่แท้จริงเท่าไร การกำหนดกรอบเนื้อหาหรือความสามารถเป็นการกำหนดเพื่อสอบวัดและใช้เพื่ออ้างในการอธิบายความสามารถของผู้สอบตามกรอบดังกล่าวซึ่งเป็นการวัดทางอ้อม

การวัดเป็นการสุ่มตัวอย่างของคุณลักษณะที่ต้องการวัด ซึ่งอาจทำได้ไม่สมบูรณ์ การวัดทางการศึกษาเป็นการวัดคุณลักษณะที่เป็นนามธรรม จึงมีวิธีการวัดที่แตกต่างจากการวัดทางกายภาพและอาจไม่สมบูรณ์ เนื่องจากข้อจำกัดในการวัดและความไม่ชัดเจนเหมือนกับการวัดทางกายภาพ ตัวอย่างเช่น การวัดผลสัมฤทธิ์ทางการเรียน แม้จะกำหนดขอบข่ายเนื้อหาและจุดประสงค์การเรียนรู้ไว้อย่างชัดเจน แต่การใช้คำถามเพื่อวัดความรู้ทั้งหมดย่อมทำไม่ได้ เพราะอาจมีคำถามมากมายจนไม่สามารถถามได้ในเวลาจำกัด จำเป็นต้องสุ่มตัวอย่างคำถามบางส่วนจากคำถามจำนวนมากนั้น ทำให้การวัดไม่สมบูรณ์ อย่างไรก็ตาม เพื่อให้การวัดมีความหมายอย่างแท้จริง ผู้ออกข้อสอบควรสุ่มตัวอย่างคำถามที่เป็นตัวแทนของความรู้หรือเป็นตัวแทนของสิ่งที่ต้องการวัดเพื่อจะสามารถนำไปอ้างอิงถึงประชากรความรู้หรือคุณลักษณะที่ต้องการวัดได้คะแนนที่ได้จากการวัดมีความคลาดเคลื่อน ประกอบด้วยสิ่งที่เป็นคะแนนหรือความสามารถที่แท้จริงและส่วนที่เป็นความคลาดเคลื่อน ถ้าการวัดในครั้งนั้นมีความคลาดเคลื่อนน้อย คะแนนความสามารถจากการวัดจะใกล้เคียงความสามารถที่แท้จริง

ผลของการวัดจะมีความหมายเมื่ออยู่ในรูปแบบของความสัมพัทธ์หรือเปรียบเทียบ ตัวอย่างเช่น ถ้าสมศักดิ์ สอบวิชาภาษาอังกฤษได้ 50 คะแนน ผลการสอบจะยังไม่มี ความหมาย แต่ถ้านำผลไปเปรียบเทียบกับกลุ่มหรือเกณฑ์ จะทำให้มีความหมาย กล่าวคือ คะแนน 50 ถ้าเปรียบเทียบกับกลุ่มเป็นคะแนนสูงสุดในกลุ่มนั้น หมายความว่า สมศักดิ์ เรียนอังกฤษเก่งที่สุดในห้อง แต่ถ้าเปรียบเทียบกับกลุ่มเป็นอันดับท้าย หมายความว่า สมศักดิ์เรียนอ่อนที่สุดในห้อง หรือนำไปเปรียบเทียบกับเกณฑ์ โดยกำหนดเกณฑ์ไว้ 40 คะแนน จะกล่าวได้ว่า สมศักดิ์ สอบผ่านเกณฑ์

2.1.2 หลักการวัดผลทางการศึกษา

วัดให้ตรงจุดประสงค์ สิ่งสำคัญที่ต้องคำนึงถึงในการวัดผลแต่ละครั้งคือวัตถุประสงค์ การวัดต้องตรงจุดประสงค์ หรือสิ่งที่ต้องการวัด หากการวัดไม่ตรงจุดประสงค์จะไม่สามารถสื่อความหมายเกี่ยวกับปริมาณหรือขนาดของสิ่งที่วัดได้

สาเหตุที่ทำให้การวัดไม่ตรงจุดประสงค์อาจเนื่องจากผู้วัดไม่รู้จักคุณลักษณะของสิ่งที่จะวัด จึงไม่รู้ว่าอะไรจะวัดอะไร เนื่องจากการวัดทางการศึกษามีลักษณะหรือกรอบของสิ่งนั้นและทำให้เข้าใจผิดพลาด เช่น กำหนดขอบข่ายของสิ่งที่วัดผิดหรือเข้าใจคุณลักษณะบางอย่างได้ไม่ถูกต้อง ใช้เครื่องมือวัดไม่ถูกต้อง ลักษณะของสิ่งที่ต้องการวัด ซึ่งเป็นนามธรรม มีความแตกต่างกัน

การใช้เครื่องมือวัดจะต้องเหมาะสม และตรงกับสิ่งที่จะวัด เช่น การใช้คำถามเพื่อวัดความสามารถในการอ่านของนักเรียนระดับชั้น ป.4 และ ม.1 ระดับความยากง่ายของคำถามอาจต้องแตกต่างกันตามระดับชั้น หรือคำถามเพื่อวัดความรู้สึกบางประการในกลุ่มตัวอย่างที่ต่างกันคำถามอาจแตกต่างกัน เพื่อที่จะได้ความรู้สึกที่แท้จริงของกลุ่ม ในบางครั้ง แม้เครื่องมือวัดจะตรงจุดประสงค์ เป็นเครื่องมือที่มีคุณภาพ แต่ถ้าผู้ใช้นำเครื่องมือไปใช้ไม่ถูกวิธี ใช้ผิดกลุ่มผิดระดับ หรือขาดความยุติธรรมมีการแนะนำคำตอบด้วยวิธีใดวิธีหนึ่ง ทำให้เกิดความคลาดเคลื่อนลำเอียงไปจากความเป็นจริง ทำให้การวัดขาดความตรง ผู้ใช้เครื่องมือจำควรระวังในเรื่องดังกล่าววัดได้ไม่ครบถ้วน การวัดไม่ครบถ้วนอาจเกิดจากเครื่องมือวัดไม่ดี การวัดไม่ครอบคลุมหรือไม่เป็นตัวแทนของลักษณะที่ต้องการวัดทำให้วัดได้เพียงบางส่วน ผลจากการวัดจึงสรุปแน่นอนไม่ได้

การใช้เครื่องมือที่มีคุณภาพ เครื่องมือวัดผลไม่ว่าจะเป็นแบบใดก็ตาม ต้องมีคุณภาพ กล่าวคือ วัดได้เที่ยงหรือมีความเที่ยง (reliability) สูง วัดได้ครบถ้วนมีความตรง (validity) กับวัตถุประสงค์หรือสิ่งที่ต้องการวัด เครื่องมือที่ใช้วัดมีความเป็นปรนัย (objectivity) การใช้ภาษาชัดเจนไม่กำกวม ผู้อ่านๆ แล้วเข้าใจตรงกัน มีระบบการให้คะแนนที่เที่ยงธรรม เมื่อเครื่องมือดี มีคุณภาพ ในเวลาที่จะนำไปใช้ต้องนำไปใช้ให้ถูกวิธี เหมาะสมกับกลุ่มและจุดประสงค์ในการใช้เครื่องมืออื่นๆ และเลือกใช้เครื่องมือที่เหมาะสม มีคุณภาพ ควรทำความเข้าใจกับเครื่องมือก่อนนำไปใช้แปลผลได้ถูกต้อง

ผลจากการวัดทางการศึกษาส่วนมากจะเป็นคะแนนหรือระดับคะแนน ซึ่งถ้าใช้เครื่องมือวัดที่ดีและคะแนนที่ได้เป็นตัวแทนประชากรของสิ่งที่ต้องการวัด คะแนนจะมีความตรงและจะมีความหมายเมื่อนำไปใช้เปรียบเทียบกับกลุ่มหรือเกณฑ์ ดังนั้นการแปลผลให้มีความหมายถูกต้องควรต้องพิจารณาเกณฑ์หรือกลุ่มที่นำมาเปรียบเทียบ รวมทั้งต้องคำนึงถึงคุณลักษณะที่เป็นตัวแทนที่จะอ้างถึงด้วย ไม่แปลผลนอกกรอบของความเป็นตัวแทนนั้นๆ ใช้ผลการวัดให้คุ้มค่า ถ้าเครื่องมือวัดมีคุณภาพ ผลจากการวัดอาจเป็นดัชนีบอกความสามารถของผู้สอบ ในขณะที่เดียวกันหากใช้ข้อมูลเป็น จะสามารถเป็นเครื่องมือบอกจุดเด่น-ด้อย ของผู้สอบ และสามารถนำผลนี้เพื่อพัฒนาจุดเด่นของผู้ได้รับการทดสอบหรือแก้ไขข้อบกพร่องในส่วนที่ยังด้อย เพื่อให้ผลสุดท้ายของการจัดการศึกษามีประสิทธิภาพตรงตามจุดประสงค์ที่กำหนดไว้ ผลจากการวัดควรให้ข้อมูลของผู้เรียนเกี่ยวกับระดับความสามารถโดยภาพรวม ความสามารถเด่น-ด้อย ในบางจุด ความเจริญงอกงามหรือความก้าวหน้าของผู้เรียนและสามารถใช้ผลการสอบเพื่อบอกถึงคุณภาพของเครื่องมือ

2.2 แนวคิดและหลักการการทดสอบทางด้านจิตวิทยา

2.2.1 หลักการของการทดสอบทางจิตวิทยา (Principal of Psychological Testing)

Psychological tests แบบทดสอบทางจิตวิทยาได้ถูกพัฒนาและค้นคว้าในหลาย ๆ รูปแบบอย่างระมัดระวัง ซึ่งแตกต่างจากแบบทดสอบที่ถูกตีพิมพ์ในหนังสือพิมพ์ชั้นนำ หรือแบบทดสอบในเว็บไซต์ที่มักถามว่า “คุณเป็นคู่ที่ดีหรือไม่” หรือ “คาร์ระดับเรื่องเพศของคุณอยู่ในระดับใด” แบบทดสอบที่ดีไม่ใช่เพียงกลุ่มของคำถามแต่จะต้องมีเนื้อหาที่ตรงประเด็นและวัดผลได้ แบบทดสอบทางจิตวิทยาที่เหมาะสมจะต้องมีวัตถุประสงค์ มีมาตรฐานกลาง เชื่อถือได้ และเที่ยงตรง ซึ่งมีหลักการสำคัญคือ

1. ความเป็นมาตรฐาน (Standardization) คือ มีรูปแบบและแนวทางการใช้ที่แน่นอนภายใต้สภาวะที่เหมือนกัน ภายในกำหนดเวลาที่เท่ากัน และอยู่ภายใต้สภาพแวดล้อมที่เหมือนกันด้วย

2. ความเป็นปรนัย (Objectivity) คือ ความเป็นตรงในการให้ระดับคะแนนแบบทดสอบปรนัย หรือ Objective tests ของผู้สอบคนหนึ่งคนใด ไม่ว่าใครตรวจยอมได้คะแนนเท่ากัน ไม่มีอคติ ไม่เหมือนกับการทดสอบแบบอัตนัย (Subjective tests) คุณสมบัติของผู้เข้าสอบและท่าทีความรู้สึกหรือเจตคติ (Attitude) ของผู้ตรวจมีผลต่อคะแนนที่ได้ แต่ละคนตรวจก็ให้คะแนนต่างกัน

3. มีค่าระดับมาตรฐานกลาง (Test Norms) ที่สามารถนำไปเปรียบเทียบผลได้ คือ ผลการทดสอบของคนใดคนหนึ่งแม้ว่าจะทำการทดสอบเพียงคนเดียว ก็สามารถทราบได้ว่าระดับคะแนนของเขาถูกแจกแจง (Distribute) อยู่ในระดับใด หากมีการทดสอบกันเป็นกลุ่มใหญ่ ๆ ซึ่งค่าระดับกลางนี้ได้จากการทดสอบกลุ่มตัวอย่างที่ได้มาตรฐาน (Standardization sample) มาแล้ว โดยส่วนใหญ่แบบทดสอบทางจิตวิทยาได้ตั้งค่ามัธยฐานสำหรับผู้ชายและผู้หญิง ความแตกต่างด้านอายุ เชื้อชาติ ชาติพันธุ์ ระดับการศึกษา ค่ามัธยฐานที่เหมาะสมเป็นประโยชน์อย่างมากในการคัดเลือกพนักงาน

4. มีความน่าเชื่อถือ (Reliability) คือ มีความแน่นอนและคงเส้นคงวาในแบบทดสอบ การทดสอบด้วยแบบทดสอบในแบบเดียวกันสำหรับคน ๆ เดียวกัน ในเวลาที่ใกล้เคียงกัน การพิจารณาความน่าเชื่อถือมักจะทำ 3 วิธี คือ

4.1 วิธีสอบซ้ำ (Test - retest method) คือ ใช้แบบทดสอบเดียวกันสอบซ้ำกับคนกลุ่มเดียวกันสองหน หากค่าสหสัมพันธ์ (Correlation) สูง คือ คะแนนแต่ละคนที่ได้เท่าเดิมมากเท่าใด แสดงว่าแบบทดสอบมีความน่าเชื่อถือสูง

4.2 วิธีใช้แบบทดสอบเทียบเท่า (Equivalent forms) เป็นวิธีการสอบซ้ำเหมือนกัน แต่ใช้แบบทดสอบเทียบเท่าที่จัดทำในแนวเดียวกันและมีความยากง่ายเท่ากันสองชุด วิธีการนี้ก็ยุ่งยากและสิ้นเปลืองที่ต้องสร้างแบบทดสอบเทียบเท่าสองชุด

4.3 วิธีแบ่งครึ่งแบบทดสอบ (Split - halves method) เป็นการแบ่งแบบทดสอบเดียวกันออกเป็นสองส่วน (อาจให้ฉบับหนึ่งเป็นข้อเลขคี่ อีกฉบับหนึ่งเป็นข้อเลขคู่) ให้คนกลุ่มเดียวกันทดสอบทั้งสองฉบับในคราวเดียวกัน จะได้คะแนนสองชุด จากนั้นจึงมาหาค่าสหสัมพันธ์ความน่าเชื่อถือ

5. ความเที่ยงตรง (Validity) เป็นสิ่งสำคัญที่สุดของแบบทดสอบ คือ หากต้องการจะทดสอบอะไร แบบทดสอบต้องสามารถใช้ทดสอบสิ่งนั้นได้ นักจิตวิทยาอุตสาหกรรมและองค์การกำหนดความเที่ยงตรงไว้หลายอย่าง คือ

5.1 เที่ยงตรงตามเกณฑ์ที่กำหนด (Criterion - related) แบบทดสอบ แต่เราให้ความสำคัญกับผลคะแนนที่ออก มาสอดคล้องกับผลการปฏิบัติงานจริง ซึ่งการวัดความเที่ยงตรงมักใช้วิธีวัดความเที่ยงตรงในการทำนาย (Predictive validity) คือวิธีการที่ทดสอบผู้สมัครทุกคนในวันแรกและรับทุกคนเข้าทำงานหลังจากเข้าปฏิบัติงานเราจึงทำการทดสอบอีกครั้งเพื่อวัดประสิทธิภาพในการทำงานอนาคต หลายแห่งมีการทดสอบความเที่ยงตรงในความสอดคล้องกับงาน (Concurrent validity) ของพนักงานเมื่อบรรจุเข้าทำงานแล้วแทนการทดสอบในลักษณะการทำนาย วิธีนี้มักจะได้ผลดี เพราะคนที่ไม่ถนัดงานอาจลาออก ถูกไล่ออก หรืออาจถูกปรับเปลี่ยนงานไปแล้ว คนที่ทำงานอยู่จึงค่อนข้างดี

5.2 ความเที่ยงตรงเชิงเหตุผล (Rational validity) เป็นความเที่ยงตรงที่มีส่วนสัมพันธ์กับธรรมชาติ คุณสมบัติและเนื้อหาของแบบทดสอบกับงานที่จะทำมากกว่าการวัดเกณฑ์ที่จะทำนายหรือชี้สมรรถนะในการทำงานโดยตรง ความเที่ยงตรงเชิงเนื้อหา (Content validity) คือแบบทดสอบที่มีเนื้อหาตรงกับงานที่จะทำ ซึ่งสามารถประเมินความรู้และทักษะที่จำเป็นที่จะต้องใช้ในการปฏิบัติงาน โครงสร้างของความเที่ยงตรง (Construct Validity) คือความพยายามที่จะกำหนดบุคลิกภาพทางจิตวิทยาด้วยแบบทดสอบ ทำอย่างไรถึงจะรู้ว่าแบบทดสอบใหม่ที่พัฒนาขึ้นเพื่อวัดความฉลาด แรงจูงใจ

หรือความมั่นคงทางอารมณ์ จะมีความเที่ยงตรง วิธีเดียวที่รู้คือนำค่าสถิติไปเปรียบเทียบกับแบบทดสอบเก่าที่ได้รับการยอมรับว่ามีความเที่ยงตรง หากว่าผลคะแนนมีความสัมพันธ์กันสูง แสดงว่าแบบทดสอบใหม่มีความเที่ยงตรง

5.3 ความเที่ยงตรงตามสภาพงาน (Face validity) ไม่ใช่การประเมินในทางสถิติ แต่เป็นความตรงของข้อคำถามในแบบทดสอบที่สัมพันธ์กับลักษณะงาน การคัดเลือกนักบินอาจจะต้องถูกถามว่ารักพ่อแม่หรือไม่ หรือจะเป็นอะไรไหมหากต้องนอนเปิดไฟ คำถามเหล่านั้นจะเกี่ยวพันกับลักษณะความมั่นคงของอารมณ์ซึ่งนักบินจะต้องมีแบบทดสอบด้านจิตวิทยาที่ดีจะมีข้อมูลและค่าต่าง ๆ เกี่ยวกับความเที่ยงตรงของเครื่องมือแนบอยู่ในชุดเสมอ หากไม่มีค่าความเที่ยงตรงอยู่ด้วย แบบทดสอบหรือเครื่องมือทดสอบนั้นจะไม่ใช่ที่เชื่อมั่นของผู้ใช้

2.2.2 ประเภทของแบบทดสอบทางจิตวิทยา (Types of Psychological Tests)

แบบทดสอบด้านจิตวิทยาที่ใช้คัดเลือกลูกจ้างจะเน้นการวัดคุณลักษณะและพฤติกรรมของคน แบบทดสอบพื้นฐานทั่วไปจะใช้วัดความสามารถด้านความรู้ความเข้าใจ (cognitive abilities), ความสนใจ (interests), ความถนัด (aptitudes), ทักษะด้านการเคลื่อนไหว (motor skills) และบุคลิกภาพ (personality) ซึ่งพอจะอธิบายได้ ดังนี้ แบบทดสอบความสามารถด้านความรู้ความเข้าใจ (Cognitive Ability Tests) บางทีก็เรียกว่าแบบทดสอบเชาวน์ปัญญา (Intelligence Tests) แบบทดสอบความสนใจ (Interest Tests) แบบทดสอบความถนัด (Aptitude Tests) แบบทดสอบทักษะเกี่ยวกับการรับรู้และเคลื่อนไหว (Sensory / Motor Skill Tests) แบบทดสอบบุคลิกภาพ (Personality Tests) แบบทดสอบความซื่อสัตย์ (Integrity Tests) แบบทดสอบความสามารถทางกายภาพ (Physical Ability Testing) และ Multiple - Aptitude Test Batteries

Sternberg (1997) ได้เสนอทฤษฎี Triarchic (หรือ three - part) theory of intelligence ซึ่งแบ่งออกเป็น 3 ส่วน คือ การวัดระดับความรู้, การวัดระดับความสามารถทางด้านการปฏิบัติ และการวัดระดับความสามารถทางด้านการคิดสร้างสรรค์

สติปัญญาเป็นเรื่องที่มีความซับซ้อนมาก นักทฤษฎีบางคนออกมาแสดงความไม่เห็นด้วยกับการนิยาม “สติปัญญา” ออกมาในรูปแบบเดียว เพราะสติปัญญาจริง ๆ แล้วมีหลายรูปแบบ เช่น ความสามารถในการวิเคราะห์, ความสามารถทางภาษา เป็นต้น แบบทดสอบสติปัญญาถูกออกแบบและตีความโดยตั้งอยู่บนพื้นฐานตามหลักวิชาการ คำถามต่าง ๆ ที่อยู่ในแบบทดสอบจะมีคำตอบที่ถูกเพียงข้อเดียว ทั้ง ๆ ที่ในชีวิตจริงปัญหาต่าง ๆ ที่เกิดขึ้นอาจจะไม่ได้ต้องการคำตอบหรือทางแก้ปัญหาเพียงคำตอบเดียว บางทีความฉลาดรอบรู้นี้ซึ่งเป็นรูปแบบหนึ่งของสติปัญญาก็ได้มาจากความสามารถในการแก้ไขหรือจัดการปัญหาที่สั่งสมมาเป็นเวลาหลายปีนั่นเอง เพราะฉะนั้นความสามารถที่เกิดจากการแก้ปัญหาที่ซับซ้อนได้และเป็นที่ยอมรับ ก็เป็นหนึ่งในรูปแบบของสติปัญญาเช่นกัน

2.2.3 การจัดทำโครงการทดสอบ (Establishing a Testing Program)

การทดสอบเพื่อคัดสรรคนเป็นความจำเป็นก้าวแรก เรื่องแรกที่จะต้องทำคือการตรวจสอบงานให้เข้าใจถ่องแท้ว่า ควรจะนำเครื่องมือใดหรือนำวิธีการใดมาใช้ทดสอบคน และเมื่อมีการวิเคราะห์ลักษณะงานและลักษณะคนที่ทำงานแล้ว ความจำเป็นที่ตามมาก็คือการเลือกหรือการสร้างแบบทดสอบที่เหมาะสม เพื่อนำมาใช้วัดพฤติกรรมและความสามารถที่สัมพันธ์กับความสำเร็จของงาน นั่นคือ นักจิตวิทยาอุตสาหกรรมและองค์การอาจใช้แบบทดสอบมาตรฐานสำเร็จรูปที่มีอยู่แล้ว หรือสร้างขึ้นมาใช้เองสำหรับงานหรือหน่วยงานของตนโดยเฉพาะ

แบบทดสอบที่ดีจะมีข้อมูลเกี่ยวกับ ค่าความน่าเชื่อถือ (Reliability), ค่าความเที่ยงตรง (Validity) และค่ามาตรฐานกลาง (Test Norms) แบบทดสอบสำเร็จรูปเหล่านี้มีรายการตีพิมพ์

ใน Mental Measurements Yearbook ซึ่งตอนนี้มีแบบทดสอบมากกว่า 700 แบบแล้ว สำหรับการตัดสินใจว่าจะเลือกซื้อหรือสร้างเองนั้นก็ขึ้นอยู่กับว่าจะทดสอบอะไร เพียงใด การซื้อมาใช้ย่อมถูกกว่าการสร้างเอง ถ้าเป็นการคัดสรรคนไม่กี่คนด้วยแล้ว การสร้างเครื่องมือย่อมไม่คุ้ม ทั้งเงินและเวลา ยังมีเวลาจำกัด การสร้างเครื่องมือที่มีคุณภาพย่อมเป็นไปได้

หากทุกคนที่รับการทดสอบได้รับการบรรจุงาน การทดสอบคนก่อนเข้าทำงานจึงเป็นเพียงการทำนายสมรรถนะในการทำงาน และเป็นตัวช่วยในการมอบหมายงานให้ตรงกับลักษณะคนได้มากขึ้น ในกรณีนี้คะแนนทดสอบมีผลน้อยมากต่อการตัดสินใจจ้างงาน แต่หลังจากที่บุคคลเหล่านี้ได้ทำงานนานพอสมควร เขาก็จะได้รับการทดสอบสมรรถนะในการทำงานเป็นระยะ ๆ ซึ่งเมื่อเปรียบเทียบกับผลการทดสอบก่อนเข้าทำงาน ก็พอจะทราบว่า การทดสอบครั้งแรกแม่นยำเพียงใด และคนได้รับการพัฒนาเพียงใด

เมื่อวิเคราะห์ที่ได้ค่าความเชื่อมั่น ค่าความเที่ยงตรง และจุดตัดค่ามาตรฐานกลางของแบบทดสอบแล้ว หากจำเป็นต้องคัดสรรคน เราก็พอจะกำหนดได้ว่าควรจะได้รับคนในระดับคะแนนที่เท่าใด แต่ทั้งนี้ย่อมขึ้นอยู่กับปริมาณคนที่ต้องการจะจ้างและจำนวนผู้มาสมัครงานอยู่ที่ ถ้าคนมาสมัครมากก็เลือกได้มาก จุดที่จะตัดคะแนนเอาคนออกก็จะสูงขึ้น คุณภาพคนก็จะดีขึ้น ค่าใช้จ่ายในการฝึกงานก็จะลดลง

3. ตัวอย่างมาตรฐานการทดสอบทางการศึกษา

มาตรฐานการทดสอบทางการศึกษาแห่งชาติ (Ordinary National Education Test : O-NET)

3.1 มาตรฐานการทดสอบทางการศึกษาแห่งชาติ

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) กำหนดมาตรฐานการทดสอบแห่งชาติขึ้นครอบคลุมโครงสร้างและการดำเนินงานของ สทศ. และสามารถนำไปสู่การปฏิบัติเพื่อให้เกิดผลเป็นรูปธรรมได้โดยมาตรฐานประกอบด้วย 5 ด้าน ดังนี้

1. มาตรฐานการบริหารการทดสอบ เป็นมาตรฐานเพื่อการประกันคุณภาพว่า ระบบการบริหารการทดสอบมีความชัดเจน สามารถปฏิบัติได้จริงมีคุณภาพเป็นที่ยอมรับและเชื่อถือได้ในระดับชาติและระดับนานาชาติ

2. มาตรฐานบุคลากรด้านการทดสอบ เป็นมาตรฐานเพื่อการประกันคุณภาพว่า บุคลากรที่เกี่ยวข้องกับการทดสอบมีคุณภาพ คุณสมบัตินี้และหน้าที่เทียบตามมาตรฐานสากล

3. มาตรฐานการพัฒนาแบบทดสอบ เป็นมาตรฐานเพื่อการประกันคุณภาพว่าแบบทดสอบที่ใช้ในการวัดและประเมิน มีการพัฒนาอย่างมีระบบ มีขั้นตอนการดำเนินการที่เป็นมาตรฐาน เชื่อถือได้ มีสารสนเทศเชิงประจักษ์และเป็นแบบทดสอบที่มีคุณภาพด้านความตรง ความเที่ยง และมีความยุติธรรม

4. มาตรฐานการพิมพ์ การรับ ส่ง การตรวจ และการประมวลผล เป็นมาตรฐานเพื่อการประกันคุณภาพว่า ระบบการพิมพ์แบบทดสอบและกระดาษคำตอบมีการควบคุมและกำกับกับการดำเนินงานอย่างมีขั้นตอนที่ชัดเจน มีมาตรการด้านความปลอดภัยและมีการเก็บรักษาความลับอย่างรัดกุม มีกระบวนการตรวจให้คะแนนที่ถูกต้อง มีระบบชัดเจนและสามารถตรวจสอบความผิดพลาดได้และมีการรายงานผลด้วยคะแนนที่มีความหมายและมีการแปลคะแนนอย่างเหมาะสมเทียบเคียงกับผลการทดสอบแต่ละปี

5. มาตรฐานการรายงานผลและการนำผลไปใช้เป็นมาตรฐานเพื่อการประกันคุณภาพว่าการรายงานผลและการนำผลไปใช้มีความถูกต้องเหมาะสมและเป็นธรรม

3.2 จุดมุ่งหมายของแต่ละมาตรฐานการทดสอบทางการศึกษาแห่งชาติ

1. มาตรฐานการบริหารการทดสอบ จุดมุ่งหมาย
 - (1) เพื่อให้มีระบบการบริหารการทดสอบที่ชัดเจนสามารถปฏิบัติได้จริง
 - (2) เพื่อให้ระบบการบริหารการทดสอบมีคุณภาพเป็นที่ยอมรับ เชื่อถือได้ในระดับชาติและระดับนานาชาติ
2. มาตรฐานบุคลากรด้านการทดสอบ จุดมุ่งหมาย
 - (1) เพื่อให้ได้บุคลากรที่มีคุณภาพและเป็นที่ยอมรับทั้งในระดับชาติและระดับนานาชาติมาารวมดำเนินการทดสอบ
 - (2) เพื่อให้บุคลากรที่เกี่ยวข้องกับการทดสอบมีคุณสมบัติและหน้าที่ตามมาตรฐานสากล
3. มาตรฐานการพัฒนาแบบทดสอบ จุดมุ่งหมาย
 - (1) เพื่อให้การพัฒนาแบบทดสอบมีระบบ มีขั้นตอนการดำเนินการที่ชัดเจน เชื่อถือได้ มีสารสนเทศเชิงประจักษ์
 - (2) เพื่อให้ได้แบบทดสอบที่มีคุณภาพด้านความตรง ความเที่ยง และมีความยุติธรรม
4. มาตรฐานการพิมพ์การรับ/ส่ง การตรวจ การประมวลผลและการแปลผล จุดมุ่งหมาย
 - (1) เพื่อให้การจัดพิมพ์และการรับ/ส่งแบบทดสอบ มีการควบคุมและกำกับการทำงาน อย่างมีขั้นตอน ที่ชัดเจน มีมาตรการด้านความปลอดภัยและมีการเก็บรักษาความลับอย่างรัดกุม
 - (2) เพื่อให้การตรวจให้คะแนนที่ยึดหลักการความถูกต้อง และความเหมาะสมขององค์ความรู้ของศาสตร์แต่ละสาขา มีกระบวนการตรวจให้คะแนนเป็นระบบชัดเจน และสามารถตรวจสอบความผิดพลาดคลาดเคลื่อนที่อาจเกิดขึ้นได้
 - (3) เพื่อให้มีการรายงานผลการทดสอบด้วยคะแนนมีความหมาย มีวิธีการแปลผลคะแนนอย่างเหมาะสม สามารถแปลผลคะแนนเทียบเคียงกับผลการทดสอบที่แตกต่างกันได้อย่างมีความหมาย
5. มาตรฐานการรายงานผลและการนำผลไปใช้ จุดมุ่งหมาย
 - (1) เพื่อให้การรายงานผลมีความถูกต้องและเหมาะสมต่อการนำไปใช้
 - (2) เพื่อให้การนำผลการทดสอบไปใช้เป็นไปอย่างถูกต้อง และเป็นธรรม

ข้อเสนอแนะ

1. มาตรฐานการทดสอบไม่ว่าจะเป็นการทดสอบทางการศึกษาหรือการทดสอบทางด้านจิตวิทยา หากดำเนินการตามมาตรฐานที่เหมาะสมหรือสอดคล้องกับกระบวนการวัดก็จะส่งผลให้ผลและความหมายจากการทดสอบนั้นมีความน่าเชื่อถือ และเป็นผลที่สอดคล้องกับความสามารถมากที่สุด
2. การมีมาตรฐานการทดสอบที่มีความชัดเจนจะสามารถช่วยให้การดำเนินการทดสอบเป็นไปในแนวทางที่ต้องการและเป็นมาตรฐานเดียวกันและตรงตามวัตถุประสงค์ของการทดสอบ

เอกสารอ้างอิง

- กนกรัตน์ สุชะตุงคะ, สุชีรา ภัทรายุตวรรัตน์, เชิดศักดิ์ โฆวาสินธุ์และคณะ. **การพัฒนาเกณฑ์มาตรฐานแบบทดสอบบุคลิกภาพคนไทย**. วารสารจิตวิทยาคลินิก 35(2): กค.-ชค. 47
- เชิดศักดิ์ โฆวาสินธุ์. (2549) **การวิจัยทางการศึกษา**. กรุงเทพฯ : สำนักพิมพ์โอเดียนสโตร์.
- แอน อนาस्ताซี. (2519). **การตรวจสอบเชิงจิตวิทยา**. กรุงเทพฯ : ไทยวัฒนาพานิช จำกัด.
- สุชีรา ภัทรายุตวรรัตน์. (2545) **คู่มือการวัดทางจิตวิทยา**. กรุงเทพฯ: เมติคัล มีเดีย.
- Kaplan, Robert M. and Saccuzzo, Dennis P. (1993) **Psychological Testing: Principles, Applications and Issues**. Brooks/cole Publishing Company, California.
- Kline, Paul (1993) **The Handbook of Psychological Testing**. Routledge, London
- Murphy, Kevin R. and Davidshofer, Charles o. (1994) **Psychological Testing: Principles and Applications**. Prentice Hall Englewood Cliffs, New Jersey.