



การประมาณค่าความเชื่อมั่นของแบบทดสอบตามโมเดลคะแนนจริงสัมพัทธ์ และโมเดลโลจิสติก 3 พารามิเตอร์

กิตติศักดิ์ นีวรัตน์¹

รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์²

รองศาสตราจารย์ ดร.องอาจ นัยพัฒน์³

ผู้ช่วยศาสตราจารย์ ดร.สุรัชย์ มีชาญ⁴

บทคัดย่อ

ในการวิจัยครั้งนี้ มีความมุ่งหมายหลักเพื่อศึกษาการประมาณค่าความเชื่อมั่นของแบบทดสอบตามโมเดลคะแนนจริงสัมพัทธ์ ซึ่งประกอบด้วยสัมประสิทธิ์ r_B กับสัมประสิทธิ์โครงสร้าง และโมเดลโลจิสติก 3 พารามิเตอร์ ได้แก่ สัมประสิทธิ์ P_{IRT} และใช้สัมประสิทธิ์ KR-20 เป็นเกณฑ์เทียบ

ประชากรเทียมที่ใช้เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 สำนักงานเขตพื้นที่การศึกษาจังหวัดเชียงราย จำนวน 4 เขตพื้นที่ ปีการศึกษา 2550 โดยใช้วิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) จำนวน 6,000 คน และใช้วิธีการสุ่มแบบใส่คืนจากประชากรเทียม แบ่งเป็นกลุ่มตัวอย่างขนาดกลาง 4 ขนาด ได้แก่ 300, 500, 1,000 และ 1,500 คน กลุ่มตัวอย่างขนาดเล็ก 4 ขนาด ได้แก่ 30, 60, 120 และ 240 คน แต่ละขนาดจะมีจำนวน 30 กลุ่ม ใช้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 5 ที่ผู้วิจัยสร้างขึ้น วิเคราะห์ข้อมูลโดยหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน การประมาณค่าความเชื่อมั่น ความคลาดเคลื่อนมาตรฐาน ความลำเอียง และการวิเคราะห์เส้นภาพ ผลการวิจัยสรุปได้ดังนี้

1. การประมาณค่าความเชื่อมั่นของแบบทดสอบ จากประชากรเทียมและจากกลุ่มตัวอย่างขนาดกลาง 4 ขนาด ได้แก่ ขนาด 300 500 1,000 และ 1,500 คน พบว่า ค่าความเชื่อมั่นของแบบทดสอบที่ประมาณค่าด้วยสัมประสิทธิ์ P_{IRT} ได้ค่าความเชื่อมั่นสูงสุด รองลงมา คือ สัมประสิทธิ์ r_B สัมประสิทธิ์ KR-20 และสัมประสิทธิ์โครงสร้างตามลำดับ
2. การประมาณค่าความเชื่อมั่นของแบบทดสอบ จากกลุ่มตัวอย่างขนาดเล็ก 4 ขนาด ได้แก่ ขนาด 30, 60, 120 และ 240 คน พบว่า สัมประสิทธิ์ r_B ประมาณค่าความเชื่อมั่นของแบบทดสอบได้ค่าสูงสุด รองลงมา คือ สัมประสิทธิ์ KR-20 และสัมประสิทธิ์โครงสร้างตามลำดับ
3. ที่กลุ่มตัวอย่างขนาด 300 500 1,000 1,500 คน ค่าความเชื่อมั่นที่ประมาณค่าได้จากสูตรสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B และสัมประสิทธิ์โครงสร้างมีค่าอยู่ในช่วงความเชื่อมั่น 95%
4. ความคลาดเคลื่อนมาตรฐานและความลำเอียงของการประมาณค่าความเชื่อมั่นของแบบทดสอบมีค่าต่ำ เข้าใกล้ศูนย์ แสดงว่าสูตรประมาณค่าความเชื่อมั่น มีความเหมาะสมและมีคุณภาพในการประมาณค่า
5. การเปรียบเทียบความเชื่อมั่นของแบบทดสอบที่ประมาณค่าด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้างและสัมประสิทธิ์ P_{IRT} จากกลุ่มตัวอย่างขนาดกลางมีเส้นภาพความเชื่อมั่นที่ไม่ขนานกัน

คำสำคัญ : คะแนนจริงสัมพัทธ์, ความเชื่อมั่นทางโครงสร้าง, สัมประสิทธิ์ P_{IRT}



The Estimation of The Reliability of A Test Under The Congeneric Model and Three - Parameter Logistic Model

Abstract

The main purpose of this dissertation was to investigate the reliability of the test estimated with the congeneric model (r_B and $r_{\text{construct}}$ co-efficiencies) and three - parameter logistic model (and KR-20 co-efficiencies).

For data gathering, 6,000 Prathomsuksa 5 students out of the 4 Chiangrai Education Services Area Offices divided into 4 population sizes: 300, 500, 1,000, and 1,500 were selected by the stratified random sampling technique. Otherwise, 30 sampling groups out of each the small-sized sample divided into 4 sizes: 30, 60, 120, and 240 were all conducted with an achievement test on mathematical learning achievement designed for Prathomsuksa 5 students. The data were systematically analyzed through mean, standard deviation, the estimation of reliability, the standard error of the estimate, partiality, and Profile analysis.

The findings of the study were as follows:

1. In terms of the estimation of the test's reliability compared the psudo population with the 4 sample groups, including 300, 500, 1,000, and 1,500, it was stated that the reliability of the test with, P_{IRT} , r_B , KR-20, and $r_{\text{construct}}$ co-efficiencies were respectively estimated;

2. In terms of the estimation of the test's reliability of a test compared with the 4 small-sized sample groups of 30, 60, 120, and 240, it was stated that the r_B , KR-20, and $r_{\text{construct}}$ co-efficiencies were respectively estimated;

3. In terms of the reliability of the test with the sample groups of 300, 500, 1,000, and 1,500 estimated with the KR-20 co-efficiency, it was stated that the r_B co-efficiency and $r_{\text{construct}}$ co-efficiency was averaged at its reliability of 95%;

4. In terms of the estimation of the test's reliability found in the standard error and partiality of the estimate with its reliability of $0 \geq 1$, it was stated that the appropriateness and quality of the test were found in the estimation of the reliability;

5. As compared the reliability of the test that was estimated with the KR-20, r_B , $r_{\text{construct}}$, and P_{IRT} co-efficiencies with the medium-sized sample group, it was also stated that the unparallel.

Key Word : congeneric, Construct Reliability, P_{IRT} co-efficient



บทนำ

ในการประมาณค่าความเชื่อมั่นของเครื่องมือวัดนั้นสามารถประมาณค่าความเชื่อมั่นได้หลายวิธี มีวิธีประมาณค่าความเชื่อมั่นที่น่าสนใจ 2 วิธี คือ ทฤษฎีทดสอบคะแนนจริงสัมพันธ์ (Congeneric Test Theory) และทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT)

ทฤษฎีทดสอบคะแนนจริงสัมพันธ์ เป็นทฤษฎีย่อยของทฤษฎีการทดสอบมาตรฐานเดิม (Classical Test Theory : CTT) ซึ่งเป็นทฤษฎีแรกที่ได้พัฒนาขึ้นและมีการพัฒนาอย่างต่อเนื่องจนถึงปัจจุบันทฤษฎีการทดสอบมาตรฐานเดิมได้กล่าวถึงความคู่ขนาน 3 ระดับ คือ แบบมาตรฐานเดิม แบบคะแนนจริงสมมูลและแบบคะแนนจริงสัมพันธ์ ซึ่งมีการผ่อนปรนข้อตกลงเบื้องต้นให้เหลือน้อยลงตามลำดับความคู่ขนาน

สูตรการประมาณค่าความเชื่อมั่นตามโมเดลคะแนนจริงสมมูล (Tau-Equivalent Model) ที่เป็นที่รู้จักและนิยมใช้กัน ได้แก่สูตรของคูเตอร์และริชาร์ดสัน เรียกว่าสัมประสิทธิ์ KR-20

โมเดลคะแนนจริงสัมพันธ์สามารถประมาณค่าความเชื่อมั่นได้จากการคำนวณค่าความแปรปรวน - ความแปรปรวนร่วมและจากโมเดลการวัด(Measurement Model) โดยการประมาณค่าความเชื่อมั่นจากการคำนวณค่าความแปรปรวน-ความแปรปรวนร่วมสามารถคำนวณตามสูตรต่าง ๆ ตามเงื่อนไขของการแบ่งส่วนของแบบทดสอบออกเป็น 2 ส่วน 3 ส่วน และหลายส่วนโดยในปี 2538 บุญเชิด ภิญญอนันตพงษ์ ได้พัฒนาสัมประสิทธิ์ r_{L2} ของเลียว(Liou. 1989) เรียกสัมประสิทธิ์ r_B โดยสัมประสิทธิ์ r_B จะคล้ายกับสัมประสิทธิ์ KR-20 แต่ใช้เงื่อนไขข้อตกลงของโมเดลคะแนนจริงสัมพันธ์

การประมาณค่าความเชื่อมั่นจากโมเดลการวัด (Measurement Model) โจเรสกอก (Jöreskog. 1971 : 107-112) เป็นบุคคลแรกที่ใช้คำว่าทดสอบคะแนนจริงสัมพันธ์ (Congeneric Tests) ในทฤษฎีการทดสอบมาตรฐานเดิม (CTT) โดยได้เสนอโมเดลการทดสอบคะแนนจริงสัมพันธ์ที่วัดองค์ประกอบเดียว (Model for One Set of Congeneric Test Score) ซึ่งใช้วิธีการประมาณค่าด้วยวิธีความน่าจะเป็นสูงสุด (maximum likelihood method) ต่อมาเรย์คอฟ (Raykov. 1997 : 174-178) ได้นำวิธีการของ โจเรสกอก (Jöreskog. 1971) มาใช้ในการประมาณค่าความเชื่อมั่นรวม(composite reliability) สำหรับการวัดคะแนนจริงสัมพันธ์ เรียกว่า สัมประสิทธิ์โครงสร้าง ที่ได้จากการประมาณค่าความเชื่อมั่นทางโครงสร้าง (Construct Reliability)

สำหรับโมเดลการตอบข้อสอบของทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) เมื่อมีการให้คะแนนแบบ 2 ค่า (Binary or Dichotomous IRT) ได้แก่ โมเดล 3 พารามิเตอร์ที่มีพารามิเตอร์ของค่าความยาก (b), ค่าอำนาจจำแนก (a) และค่าการเดา (c) ในปี 2003 ดิมิทروف (Dimitrov. 2003 : 440-458) ได้นำเสนอการแปลงค่าพารามิเตอร์ที่ได้จากทฤษฎีการตอบข้อสอบ (IRT) มาเป็นค่าความเชื่อมั่น เรียกว่า P_{IRT}

เมื่อนำสูตรต่าง ๆ ของการประมาณค่าความเชื่อมั่นมาพิจารณาร่วมกับบริบทของประเทศไทยที่มีการวัดประเมินในระดับชั้นเรียน (Classroom Assessment) เพื่อสนับสนุนส่งเสริมการเรียนรู้ของผู้เรียนกับการวัดประเมินในระดับใหญ่ (Large Scale) ที่เป็นการวัดประเมินมาตรฐานความรู้ในระดับกลุ่มโรงเรียนเขตพื้นที่การศึกษาและระดับประเทศ เป็นการวัดประเมินเพื่อสนับสนุนคุณภาพโดยรวมของโรงเรียนในการวัดประเมินแต่ละระดับมีจำนวนนักเรียนต่างกัน สามารถใช้ทฤษฎีการทดสอบมาตรฐานเดิมโมเดลคะแนนจริงสมมูลด้วยสัมประสิทธิ์ KR-20 โมเดลคะแนนจริงสัมพันธ์ ด้วยสัมประสิทธิ์ r_B โมเดลการวัดคะแนนจริงสัมพันธ์ ด้วยสัมประสิทธิ์โครงสร้างและทฤษฎีการตอบข้อสอบโมเดลโลจิสติก 3 พารามิเตอร์ ด้วยสัมประสิทธิ์ P_{IRT}



มาใช้ในการประมาณค่าความเชื่อมั่นได้เหมาะสมกับบางระดับเท่านั้น ผู้วิจัยจึงมีความสนใจที่จะศึกษาการประมาณค่าความเชื่อมั่นของแบบทดสอบตามโมเดลคะแนนจริงสัมพันธ์ด้วยสัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้าง และโมเดลโลจิสติก 3 พารามิเตอร์ ด้วยสัมประสิทธิ์ P_{IRT} ว่าสามารถประมาณค่าความเชื่อมั่นได้เมื่อกลุ่มตัวอย่างมีขนาดใด โดยใช้โมเดลคะแนนจริงสมมูลด้วยสัมประสิทธิ์ KR-20 เป็นเกณฑ์เทียบแต่ละวิธีให้ค่าความคลาดเคลื่อนมาตรฐานและความลำเอียงมากน้อยเพียงใดและเมื่อกลุ่มตัวอย่างมีขนาดต่างกันจะประมาณค่าความเชื่อมั่นของแบบทดสอบจาก 4 สูตร แตกต่างกันหรือไม่

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีความมุ่งหมายหลักเพื่อศึกษาการประมาณค่าความเชื่อมั่นของแบบทดสอบตามโมเดลคะแนนจริงสัมพันธ์ ซึ่งประกอบด้วยสัมประสิทธิ์ r_B กับสัมประสิทธิ์โครงสร้าง และโมเดลโลจิสติก 3 พารามิเตอร์ ได้แก่ สัมประสิทธิ์ P_{IRT} และใช้สัมประสิทธิ์ KR-20 เป็นเกณฑ์เทียบ โดยมีความมุ่งหมายเฉพาะดังนี้

1. เพื่อประมาณค่าความเชื่อมั่นของแบบทดสอบ ด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้าง และสัมประสิทธิ์ P_{IRT} จากประชากรเทียม
2. เพื่อประมาณค่าความเชื่อมั่นของแบบทดสอบ ด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้าง และสัมประสิทธิ์ P_{IRT} จากกลุ่มตัวอย่างขนาดกลาง 4 ขนาด ได้แก่ ขนาด 300, 500, 1,000 และ 1,500 คน
3. เพื่อประมาณค่าความเชื่อมั่นของแบบทดสอบด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้าง จากกลุ่มตัวอย่างขนาดเล็ก 4 ขนาด ได้แก่ ขนาด 30, 60, 120 และ 240 คน
4. เพื่อศึกษาความคลาดเคลื่อนมาตรฐานและความลำเอียงของการประมาณค่าความเชื่อมั่นของแบบทดสอบด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้างและสัมประสิทธิ์ P_{IRT}
5. เพื่อเปรียบเทียบความเชื่อมั่นของแบบทดสอบที่ประมาณค่าด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้างและสัมประสิทธิ์ P_{IRT} จากกลุ่มตัวอย่างขนาดกลาง

สมมติฐานของการวิจัย

1. สูตรการประมาณค่าความเชื่อมั่น 4 สูตร สามารถประมาณค่าความเชื่อมั่นอยู่ในช่วงความเชื่อมั่น 95% ของความเชื่อมั่นที่แท้จริง
2. สมมติฐานของการวิเคราะห์เส้นภาพ (Profile Analysis)
 - 2.1 กลุ่มตัวอย่างแต่ละขนาดมีเส้นภาพความเชื่อมั่น ขนานกัน
 - 2.2 ค่าความเชื่อมั่นที่คำนวณได้แต่ละสูตรเป็นระนาบกัน
 - 2.3 กลุ่มตัวอย่างแต่ละขนาดมีเส้นภาพความเชื่อมั่นอยู่ในระดับเดียวกัน



ระเบียบวิธีวิจัย

ประชากรที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 สำนักงานเขตพื้นที่การศึกษา จังหวัดเชียงราย ปีการศึกษา 2550 จำนวน 15,395 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างขนาดใหญ่ที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 สำนักงานเขตพื้นที่การศึกษา จังหวัดเชียงราย จำนวน 4 เขตพื้นที่ ปีการศึกษา 2550 โดยใช้วิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) จำนวน 6,000 คน ซึ่งกลุ่มตัวอย่างขนาดใหญ่นี้ ผู้วิจัยกำหนดให้เป็นประชากรเทียม

กลุ่มตัวอย่างขนาดกลางและขนาดเล็กที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 สำนักงานเขตพื้นที่การศึกษา จังหวัดเชียงราย จำนวน 4 เขตพื้นที่ ปีการศึกษา 2550 โดยใช้วิธีการสุ่มแบบใส่คืนจากประชากรเทียม แบ่งเป็นกลุ่มตัวอย่างขนาดกลาง 4 ขนาด ได้แก่ 300, 500, 1,000 และ 1,500 คน กลุ่มตัวอย่างขนาดเล็ก 4 ขนาด ได้แก่ 30, 60, 120 และ 240 คน แต่ละขนาดจะมีจำนวน 30 กลุ่ม

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ คือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 5 ที่ผู้วิจัยสร้างขึ้น มีลักษณะเป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก มีค่าดัชนีความสอดคล้องมากกว่า .50 วิเคราะห์คุณภาพข้อสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม มีค่าความยากง่ายอยู่ในช่วง 0.2 ถึง 0.8 ค่าอำนาจจำแนกมีค่าตั้งแต่ +0.2 ขึ้นไป ตรวจสอบความเป็นมิติเดียว(Unidimensional) ด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis) โดยสกัดองค์ประกอบด้วยวิธีวิเคราะห์องค์ประกอบหลัก (Principle Component) และวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) ที่มีตัวแปรภายนอกแฝงเพียงตัวเดียว เพื่อทดสอบความสอดคล้องระหว่างโมเดลกับข้อมูลเชิงประจักษ์วิเคราะห์ค่าความยากอำนาจจำแนก การเดาของแบบทดสอบตามทฤษฎีการตอบข้อสอบ(IRT) โดยมีค่าความยาก(b)อยู่ระหว่าง -2.5 ถึง +2.5 ค่าอำนาจจำแนก (a) มากกว่า 0.8 และค่าการเดา (c) น้อยกว่า 0.3 ด้วยโมเดลโลจิสติก 3 พารามิเตอร์ จากโปรแกรม BOLOG-MG3 ได้ข้อสอบที่มีคุณภาพจำนวน 50 ข้อ

การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล

1. วิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด คะแนนต่ำสุด มัธยฐาน ฐานนิยม ความเบ้ ความโด่งและความคลาดเคลื่อนมาตรฐาน

2. ประมาณค่าความเชื่อมั่นของแบบทดสอบ ด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B โดยใช้โปรแกรม Formula r_B



3. ประมาณค่าความเชื่อมั่นของแบบทดสอบ ด้วยสัมประสิทธิ์โครงสร้าง จากการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลการทดสอบคะแนนจริงสัมพันธ์ที่วัดองค์ประกอบเดี่ยว แล้วนำค่าน้ำหนักองค์ประกอบมาประมาณค่าความเชื่อมั่นตามวิธีการของเรย์คอฟ (Raykov and Grayson. 2003 : 143-159)

4. ประมาณค่าความเชื่อมั่นของแบบทดสอบ ด้วยสัมประสิทธิ์ P_{IRT} จากโมเดลโลจิสติก 3 พารามิเตอร์ วิเคราะห์ข้อมูลโดยใช้โปรแกรม BILOG-MG3 เพื่อประมาณค่าพารามิเตอร์ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) แล้วนำค่าพารามิเตอร์ทั้ง 3 ตัวมาวิเคราะห์โดยใช้โปรแกรม SPSS ในการหาค่าความแปรปรวนของคะแนนจริงและความแปรปรวนของความคลาดเคลื่อนแล้วนำมาประมาณค่าความเชื่อมั่น โดยทำการประมาณค่าความเชื่อมั่นตามวิธีการของดิมิทรอฟ (Dimitrov. 2003 : 440)

5. ศึกษาความคลาดเคลื่อนมาตรฐานและความลำเอียงของการประมาณค่าความเชื่อมั่นของแบบทดสอบ ด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้างและสัมประสิทธิ์ P_{IRT}

6. เปรียบเทียบความเชื่อมั่น จากการประมาณค่าความเชื่อมั่นของแบบทดสอบด้วยสัมประสิทธิ์ KR-20 ด้วยสัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้างและสัมประสิทธิ์ P_{IRT} ที่กลุ่มตัวอย่างขนาดกลางโดยใช้วิธีการวิเคราะห์เส้นภาพ (Profile Analysis)

ผลการวิจัย

1. การประมาณค่าความเชื่อมั่นของแบบทดสอบด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้าง และสัมประสิทธิ์ P_{IRT} จากประชากรเทียม พบว่า ค่าความเชื่อมั่นของแบบทดสอบที่ประมาณค่าด้วยสัมประสิทธิ์ P_{IRT} ได้ค่าความเชื่อมั่นสูงสุด รองลงมาคือสัมประสิทธิ์ r_B สัมประสิทธิ์ KR-20 และสัมประสิทธิ์โครงสร้างตามลำดับ

2. การประมาณค่าความเชื่อมั่นของแบบทดสอบ ด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้าง และสัมประสิทธิ์ P_{IRT} จากกลุ่มตัวอย่างขนาดกลาง 4 ขนาด พบว่า

2.1 ค่าความเชื่อมั่นของแบบทดสอบจากกลุ่มตัวอย่างขนาด 300 คนที่ประมาณค่าด้วยสัมประสิทธิ์ KR-20 มีค่าอยู่ระหว่าง .895 ถึง .921 สัมประสิทธิ์ r_B มีค่าอยู่ระหว่าง .896 ถึง .922 สัมประสิทธิ์โครงสร้าง มีค่าอยู่ระหว่าง .789 ถึง .921 และสัมประสิทธิ์ P_{IRT} มีค่าอยู่ระหว่าง .959 ถึง .969 โดยสัมประสิทธิ์ P_{IRT} ประมาณค่าความเชื่อมั่นของแบบทดสอบได้ค่าสูงสุดในทุกกลุ่ม รองลงมาคือสัมประสิทธิ์ r_B สัมประสิทธิ์ KR-20 และสัมประสิทธิ์โครงสร้างตามลำดับ

ค่าความเชื่อมั่นที่ประมาณค่าได้จากสูตรสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้าง ส่วนใหญ่ประมาณค่าอยู่ในช่วงความเชื่อมั่น 95% มีเพียง 1 ค่าที่ประมาณค่าได้ต่ำกว่าช่วงความเชื่อมั่น 95% และสัมประสิทธิ์ P_{IRT} ส่วนใหญ่ประมาณค่าไม่ได้อยู่ในช่วงความเชื่อมั่น 95% โดยค่าที่ประมาณได้มีค่าต่ำกว่าช่วงความเชื่อมั่น 95%

2.2 ค่าความเชื่อมั่นของแบบทดสอบจากกลุ่มตัวอย่างขนาด 500 คนที่ประมาณค่าด้วยสัมประสิทธิ์ KR-20 มีค่าอยู่ระหว่าง .900 ถึง .923 สัมประสิทธิ์ r_B มีค่าอยู่ระหว่าง .901 ถึง .923 สัมประสิทธิ์โครงสร้าง มีค่าอยู่ระหว่าง .899 ถึง .922 และสัมประสิทธิ์ P_{IRT} มีค่าอยู่ระหว่าง .964 ถึง .973 โดยสัมประสิทธิ์ P_{IRT} ประมาณค่าความเชื่อมั่นของแบบทดสอบได้ค่าสูงสุดในทุกกลุ่ม รองลงมาคือสัมประสิทธิ์ r_B สัมประสิทธิ์ KR-20 และสัมประสิทธิ์โครงสร้างตามลำดับ



3.3 ค่าความเชื่อมั่นของแบบทดสอบจากกลุ่มตัวอย่างขนาด 120 คนที่ประมาณค่าด้วยสัมประสิทธิ์ KR-20 มีค่าอยู่ระหว่าง .882 ถึง .922 สัมประสิทธิ์ r_B มีค่าอยู่ระหว่าง .882 ถึง .922 สัมประสิทธิ์โครงสร้างมีค่าอยู่ระหว่าง .880 ถึง .916 โดยสัมประสิทธิ์ r_B และสัมประสิทธิ์ KR-20 ประมาณค่าความเชื่อมั่นของแบบทดสอบได้ค่าสูงสุดในทุกกลุ่ม รองลงมาคือสัมประสิทธิ์โครงสร้าง ตามลำดับ

3.4 ค่าความเชื่อมั่นของแบบทดสอบจากกลุ่มตัวอย่างขนาด 240 คนที่ประมาณค่าด้วยสัมประสิทธิ์ KR-20 มีค่าอยู่ระหว่าง .903 ถึง .920 สัมประสิทธิ์ r_B มีค่าอยู่ระหว่าง .903 ถึง .921 สัมประสิทธิ์โครงสร้างมีค่าอยู่ระหว่าง .902 ถึง .921 โดยสัมประสิทธิ์ r_B และสัมประสิทธิ์โครงสร้างประมาณค่าความเชื่อมั่นของแบบทดสอบได้ค่าสูงสุดในทุกกลุ่มรองลงมาคือสัมประสิทธิ์ KR-20 ตามลำดับ

4. ความคลาดเคลื่อนมาตรฐานและความลำเอียงของการประมาณค่าความเชื่อมั่นของแบบทดสอบ พบว่า

4.1 ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความเชื่อมั่นของแบบทดสอบ ด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้างและสัมประสิทธิ์ P_{IRT} ที่กลุ่มตัวอย่างขนาดกลาง 300, 500, 1,000, 1,500 คน มีค่าต่ำเข้าใกล้ศูนย์แสดงว่าสูตรประมาณค่าความเชื่อมั่น มีความเหมาะสมและมีคุณภาพในการประมาณค่า

4.2 ความลำเอียงของการประมาณค่าความเชื่อมั่นของแบบทดสอบด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้างและสัมประสิทธิ์ P_{IRT} ที่กลุ่มตัวอย่างขนาดกลาง 300, 500, 1,000, 1,500 คน มีค่าต่ำเข้าใกล้ศูนย์แสดงว่าสูตรประมาณค่าความเชื่อมั่น มีความเหมาะสมและมีคุณภาพในการประมาณค่า

5. การเปรียบเทียบความเชื่อมั่นของแบบทดสอบที่ประมาณค่าด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้างและสัมประสิทธิ์ P_{IRT} จากกลุ่มตัวอย่างขนาดกลาง

5.1 ความคู่ขนานกันของเส้นภาพ (Parallelism of Profiles) พบว่า ค่า Wilks F = 13.585 มีนัยสำคัญที่ระดับ .01 แสดงว่ากลุ่มตัวอย่างแต่ละขนาดมีเส้นภาพความเชื่อมั่นที่ไม่ขนานกัน

5.2 ความเรียบของเส้นภาพ (Test of Flatness) พบว่า Wilks F = 15226.414 มีนัยสำคัญที่ระดับ .01 แสดงว่าค่าความเชื่อมั่นที่คำนวณได้แต่ละสูตรมีเส้นภาพที่ไม่ระนาบกัน

5.3. ระดับของเส้นภาพ (Level Test) พบว่า กลุ่มตัวอย่างแต่ละขนาดจะให้เส้นภาพ (Profile) ของความเชื่อมั่นรวมในทุกขนาดแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าอยู่ระดับเดียวกันเนื่องจากเส้นภาพความเชื่อมั่นที่ไม่ขนานกันผู้วิจัยจึงทำการทดสอบ Simple effect พบว่า

5.4 เมื่อสูตรการประมาณค่าความเชื่อมั่นเป็นสูตรเดียวกัน ขนาดกลุ่มตัวอย่างต่างกัน มีค่าความเชื่อมั่นของแบบทดสอบแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

5.5 เมื่อขนาดกลุ่มตัวอย่างเป็นขนาดเดียวกัน สูตรการประมาณค่าความเชื่อมั่นต่างกัน มีค่าความเชื่อมั่นของแบบทดสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผู้วิจัยจึงทำการเปรียบเทียบรายคู่ พบว่า สัมประสิทธิ์ KR-20 กับสัมประสิทธิ์ r_B , สัมประสิทธิ์ P_{IRT} กับสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้างมีค่าความเชื่อมั่นของแบบทดสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01



อภิปรายผล

1. การประมาณค่าโดยสัมประสิทธิ์ r_B สัมประสิทธิ์ KR-20 สามารถประมาณค่าได้กับกลุ่มตัวอย่างขนาดเล็กและขนาดกลาง คือ 30, 60, 120, 240, 300, 500, 1,000 และ 1,500 คน โดยสัมประสิทธิ์ r_B สามารถประมาณค่าความเชื่อมั่นได้สูงกว่าสัมประสิทธิ์ KR-20 ทั้ง 2 สูตรสามารถประมาณค่าความเชื่อมั่นอยู่ในช่วงความเชื่อมั่น 95% ของความเชื่อมั่นที่แท้จริงความคลาดเคลื่อนมาตรฐานและความลำเอียงของการประมาณค่าความเชื่อมั่นของแบบทดสอบมีค่าต่ำเข้าใกล้ศูนย์แสดงว่าสัมประสิทธิ์ r_B สัมประสิทธิ์ KR-20 มีความเหมาะสมในการประมาณค่าความเชื่อมั่นและมีคุณภาพในการประมาณค่าผลการวิจัย ส่วนนี้อาจเนื่องมาจากทฤษฎีการทดสอบมาตรฐานเดิม (CTT) ได้แบ่งระดับความคู่ขนานไว้ 3 ระดับ คือ

1.1 ความคู่ขนานแบบมาตรฐานเดิม (Classical Parallel)

1.2 ความคู่ขนานแบบคะแนนจริงสมมูล (Tau-Equivalent)

1.3 ความคู่ขนานแบบคะแนนจริงสัมพันธ์ (Congeneric) แต่ละระดับก็มีข้อตกลงเบื้องต้นและมีการผ่อนปรนข้อตกลงเบื้องต้นบางข้อในระดับที่ 2 และ 3 สัมประสิทธิ์ r_B เป็นสูตรที่อยู่ในระดับความคู่ขนานแบบคะแนนจริงสัมพันธ์ (Congeneric) ที่มีการผ่อนปรนข้อตกลงของความคู่ขนานลงมาตามลำดับเหลือเงื่อนไขเพียง ข้อ 1. แบบทดสอบต้องมีเนื้อหาเป็นเอกพันธ์กันหรือวัดคุณลักษณะเดียวกัน เพียงข้อเดียวและในการผ่อนปรนเงื่อนไขข้อ 2 ให้แต่ละส่วนมีคะแนนจริงต่างกันเท่ากับค่าคงที่ที่มาสัมพันธ์สมบูรณ์แบบหรือมีความสัมพันธ์เป็นเส้นตรง ได้โมเดลคะแนนจริงสัมพันธ์คือ $T_i = a_{ij}T_j + b_{ij}$ ส่วนสัมประสิทธิ์ KR-20 เป็นสูตรที่อยู่ในระดับความคู่ขนานแบบคะแนนจริงสมมูล (Tau-Equivalent) ที่มีการผ่อนปรนข้อตกลงของความคู่ขนานลงมาแต่ก็มากกว่าความคู่ขนานแบบคะแนนจริงสัมพันธ์ คือต้องมีความแปรปรวนร่วมของคะแนนสอบ (X) เท่ากัน ($\sigma_{12} = \sigma_{13} = \sigma_{23} = \dots$) และความแปรปรวนร่วมของคะแนนสอบ (X) กับคะแนนเกณฑ์ภายนอกเท่ากัน ($\sigma_{1Y} = \sigma_{2Y} = \sigma_{3Y} = \dots$) จึงทำให้ค่าความเชื่อมั่นที่ประมาณค่าได้จากสัมประสิทธิ์ r_B มีค่าสูงกว่าสัมประสิทธิ์ KR-20 สอดคล้องกับ บุญเชิด ภิญโญอนันตพงษ์ (2542 : บทคัดย่อ) ได้ทำการศึกษาสัมประสิทธิ์ r_B : การประมาณค่าความเชื่อมั่นสำหรับแบบทดสอบผลสัมฤทธิ์ชนิดเลือกตอบที่ประกอบด้วยความยากรายข้อต่างกันเพื่อพัฒนาและศึกษาผลการใช้สูตร r_B ที่ผู้วิจัยพัฒนาขึ้นมาสำหรับประมาณค่าความเชื่อมั่นของแบบทดสอบผลสัมฤทธิ์ชนิดเลือกตอบที่ประกอบด้วยข้อสอบที่ให้คะแนนระบบ 0, 1 และมีความยากรายข้อต่างกัน ผลการศึกษาพบว่า สูตร r_B คำนวณค่าความเชื่อมั่นของแบบทดสอบจำนวนข้อต่างกันจากกลุ่มตัวอย่างขนาดใหญ่ได้สูงสอดคล้องสัมพันธ์กับสูตร KR-20 โดยสูตร r_B คำนวณค่าความเชื่อมั่นได้สูงกว่าสูตร KR-20 ทุกฉบับแสดงให้เห็นว่าสัมประสิทธิ์ KR-20 และ r_B สามารถนำไปใช้ในการคำนวณค่าความเชื่อมั่นจากกลุ่มตัวอย่างขนาดเล็ก คือ 30, 50 และ 80 คนและกลุ่มตัวอย่างขนาดใหญ่ได้

2. การประมาณค่าโดยสัมประสิทธิ์โครงสร้าง สามารถประมาณค่าได้กับกลุ่มตัวอย่างขนาดเล็กและขนาดกลาง คือ 30, 60, 120, 240 คน และ 300, 500, 1,000 และ 1,500 คน แต่เมื่อพิจารณาความสามารถประมาณค่าความเชื่อมั่นอยู่ในช่วงความเชื่อมั่น 95% ของความเชื่อมั่นที่แท้จริง พบว่าสามารถประมาณค่าอยู่ในช่วงความเชื่อมั่น 95% ของความเชื่อมั่นที่แท้จริง เมื่อกลุ่มตัวอย่างมีขนาด 300, 500, 1,000 และ 1,500 คน ความคลาดเคลื่อนมาตรฐานและความลำเอียงของการประมาณค่าความเชื่อมั่นของแบบทดสอบ มีค่าต่ำเข้าใกล้ศูนย์แสดงว่าสัมประสิทธิ์โครงสร้าง มีความเหมาะสมในการประมาณค่าความเชื่อมั่นและมีคุณภาพในการ



ประมาณค่า ผลการวิจัยส่วนนี้อาจเนื่องมาจากการประมาณค่าโดยสัมประสิทธิ์โครงสร้าง ได้ทำการวิเคราะห์โดยใช้โมเดลลิสเรล ซึ่งมีขนาดของกลุ่มตัวอย่างเข้ามาเกี่ยวข้อง โดยซุมัคเกอร์และโลแมค (นงลักษณ์ วิรัชชัย. 2542 : 311 อ้างอิงจาก Schumacker and Lomax. 1996) ได้สรุปงานวิจัยที่มีการศึกษาเรื่องขนาดของกลุ่มตัวอย่างที่เหมาะสมสำหรับใช้โมเดลลิสเรลว่า จากงานวิจัยของดิง เวลิเซอร์และฮาร์โล (นงลักษณ์ วิรัชชัย. 2542 : 311 อ้างอิงจาก Ding, Velicer and Harlow. 1995) ในปี 1995 พบว่า งานวิจัยที่ใช้โมเดลลิสเรลส่วนใหญ่ใช้ขนาดกลุ่มตัวอย่าง 100-150 คนและให้ผลการวิจัยที่น่าพอใจ บูมสมมา (นงลักษณ์ วิรัชชัย. 2542 : 311 อ้างอิงจาก Boomsma. 1983) เสนอไว้ในบทความเมื่อปี 1983 ว่าขนาดกลุ่มตัวอย่างที่เหมาะสมควรเป็น 400 คน นอกจากนี้ ซุมัคเกอร์และโลแมค(นงลักษณ์ วิรัชชัย. 2542 : 311 อ้างอิงจาก Schumacker and Lomax. 1996) แฮร์และคณะ (Hair, & et. al. 1995) เสนอให้ใช้กฎแห่งความชัดเจน (rule of thumb) ที่นักสถิติตัวแปรพหุใช้กันมากคือ ใช้ขนาดกลุ่มตัวอย่าง 10-20 คนต่อตัวแปรในการวิจัยหนึ่งตัวแปร ซึ่งจากงานวิจัยครั้งนี้ได้ทำการศึกษาจากข้อสอบจำนวน 50 ข้อ เมื่อทำการวิเคราะห์โดยใช้โมเดลลิสเรล จึงทำให้ค่าความเชื่อมั่นที่ประมาณค่าได้จะมีความเหมาะสมเมื่อกลุ่มตัวอย่างมีขนาด 300, 500, 1,000 และ 1,500 คน

3. การประมาณค่าโดยสัมประสิทธิ์ P_{IRT} สามารถประมาณค่าได้กับกลุ่มตัวอย่างขนาดกลาง คือ 300, 500, 1,000 และ 1,500 คน โดยสัมประสิทธิ์ P_{IRT} สามารถประมาณค่าความเชื่อมั่นได้สูงกว่าสัมประสิทธิ์ r_B สัมประสิทธิ์ KR-20 และสัมประสิทธิ์โครงสร้าง สามารถประมาณค่าความเชื่อมั่นอยู่ในช่วงความเชื่อมั่น 95% ของความเชื่อมั่นที่แท้จริงที่กลุ่มตัวอย่างขนาด 1,000 และ 1,500 คน ความคลาดเคลื่อนมาตรฐานและความลำเอียงของการประมาณค่าความเชื่อมั่นของแบบทดสอบมีค่าต่ำเข้าใกล้ศูนย์ แสดงว่าสัมประสิทธิ์ P_{IRT} มีความเหมาะสมในการประมาณค่าความเชื่อมั่น และมีคุณภาพในการประมาณค่า ผลการวิจัยส่วนนี้อาจเนื่องมาจากการประมาณค่าสัมประสิทธิ์ P_{IRT} ได้ทำการวิเคราะห์จากโมเดลโลจิสติก 3 พารามิเตอร์ ซึ่งเป็นโมเดลหนึ่งของทฤษฎีการตอบข้อสอบ (IRT) เมื่อมีการให้คะแนนแบบ 2 ค่า เมื่อได้ค่าพารามิเตอร์จากการประมาณค่าแล้วก็จะนำมาแปลงเป็นค่าความเชื่อมั่น โดยสูตรของ ดิมิทروف (Dimitrov. 2003 : 440) ขนาดกลุ่มตัวอย่างจึงส่งผลต่อค่าพารามิเตอร์ทำให้ส่งผลต่อการประมาณค่าความเชื่อมั่นด้วย สอดคล้องกับ ฮูลิน ดราสโกวและพาร์สัน (Hulin, Drasgow and Parson. 1983 : 99-105) ที่ได้กล่าวถึงปัจจัยที่เกี่ยวข้องกับการประมาณค่าพารามิเตอร์ ได้แก่ ความยาวแบบทดสอบ ขนาดกลุ่มตัวอย่างและจำนวนพารามิเตอร์ของโมเดลโลจิสติก จากผลการวิจัย พบว่า ความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่างเป็นปัจจัยสำคัญที่ส่งผลต่อการประมาณค่าพารามิเตอร์โดยได้เสนอแนะว่าโมเดลโลจิสติก 3 พารามิเตอร์เพื่อประมาณค่าพารามิเตอร์ด้วยวิธีความน่าจะเป็นสูงสุด (ML) ให้ได้ผลถูกต้องนั้นควรใช้แบบทดสอบที่มีจำนวน 60 ข้อกับกลุ่มตัวอย่างจำนวนไม่น้อยกว่า 200 คน หรือถ้าใช้แบบทดสอบที่มีจำนวนข้อน้อยกว่า 30 ข้อ จะต้องใช้กับกลุ่มตัวอย่างไม่น้อยกว่า 1,000 คน ในขณะที่ลอร์ด (Hulin, Drasgow and Parson. 1983 : 100-139 อ้างอิงจาก Lord. 1968) ได้กล่าวว่าการประมาณค่าพารามิเตอร์ด้วยวิธีความน่าจะเป็นสูงสุด(ML)ในโมเดลโลจิสติก 3 พารามิเตอร์ ให้มีความถูกต้องนั้น ควรใช้ข้อสอบประมาณ 50 ข้อกับกลุ่มตัวอย่างอย่างน้อย 1,000 คน ดังนั้นผลการวิจัยครั้งนี้ที่ใช้ข้อสอบจำนวน 50 ข้อ จึงได้กลุ่มตัวอย่างที่เหมาะสมในการประมาณค่าความเชื่อมั่นคือ 1,000 และ 1,500 คน

4. การเปรียบเทียบความเชื่อมั่นของแบบทดสอบที่ประมาณค่าด้วยสัมประสิทธิ์ KR-20 สัมประสิทธิ์ r_B สัมประสิทธิ์โครงสร้าง และสัมประสิทธิ์ P_{IRT} จากกลุ่มตัวอย่างขนาดกลาง มีเส้นภาพความเชื่อมั่นที่ไม่ขนานกัน



ไม่ความสอดคล้องกันและอยู่ในระดับเดียวกัน ผลการวิจัยส่วนนี้อาจเนื่องมาจาก การประมาณค่าความเชื่อมั่นของทั้ง 4 วิธีมีข้อตกลงเบื้องต้นที่ต่างกัน ขนาดกลุ่มตัวอย่างมีผลต่อการวิเคราะห์ข้อมูล เมื่อนำมาเปรียบเทียบกันจึงมีเส้นภาพที่ไม่ขนานกัน เมื่อนำมาพิจารณาพร้อมกับบริบทของการประเมินในระดับชั้นเรียน (Classroom Assessment) เพื่อสนับสนุนส่งเสริมการเรียนรู้ของผู้เรียน การใช้สัมประสิทธิ์ r_B สัมประสิทธิ์ KR-20 จึงมีความเหมาะสมที่จะนำไปใช้ในระดับนี ซึ่งวิธีการทั้ง 2 วิธีนี้ มีขั้นตอนและวิธีคำนวณที่ง่ายต่อการนำไปใช้ ไม่ต้องอาศัยทักษะทางคอมพิวเตอร์มากนัก อีกทั้งครูผู้สอนโดยส่วนใหญ่มีความคุ้นเคยกับการใช้ KR-20 ในการประมาณค่าความเชื่อมั่น ส่วนการวัดประเมินในระดับใหญ่ (Large Scale) ที่เป็นการวัดประเมินมาตรฐานความรู้ในระดับกลุ่มโรงเรียน เขตพื้นที่การศึกษาและระดับประเทศ เป็นการวัดประเมินเพื่อสนับสนุนคุณภาพโดยรวมของโรงเรียนการใช้สัมประสิทธิ์โครงสร้างและสัมประสิทธิ์ P_{IRT} จึงมีความเหมาะสมที่จะนำไปใช้ในระดับนีซึ่งวิธีการทั้ง 2 วิธีสามารถประมาณค่าได้จากนักเรียนจำนวนมาก ได้ค่าพารามิเตอร์ที่บ่งบอกถึงประชากรและความไม่แปรเปลี่ยนตามกลุ่มตัวอย่างซึ่งเป็นข้อตกลงของทฤษฎีการตอบข้อสอบ (IRT) แต่ต้องใช้ความรู้และทักษะทางคอมพิวเตอร์ในการศึกษาทฤษฎีและการวิเคราะห์ข้อมูลจากโมเดลลิสเรลและทฤษฎีการตอบข้อสอบ (IRT)

เอกสารอ้างอิง

- นางลักษณ์ วิรัชชัย. (2542). โมเดลลิสเรล : สถิติวิเคราะห์สำหรับการวิจัย. กรุงเทพฯ : ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- บุญเชิด ภิญโญอนันต์พงษ์. (2542). รายงานการวิจัย สัมประสิทธิ์ r_B : การประมาณค่าความเชื่อมั่นสำหรับแบบทดสอบผลสัมฤทธิ์ชนิดเลือกตอบที่ประกอบด้วยความยากรายข้อต่างกัน. ภาควิชาการวัดผลและการวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Dimitrov, D. M. (2003). "Marginal True-Score Measures and Reliability for Binary Item as a function of Their IRT Parameters". Applied Psychological Measurement. 27, 440-458.
- Hair, J., Anderson, R., Tatham, R. & Black, W. (1995). Multivariate Data Analysis With Readings. Fourth Edition. Prentice-Hall, Inc. New Jersey.
- Hulin, C. L., Drasgow, F., and Parson, C.K. (1983). Item response theory : Application to psychological measurement. Homewood, Ill. : Dow Jones-Irwin.
- Jöreskog, K.G. (1971). "Statistical Analysis of Sets of Congeneric Test". Psychometrika, 36, 10-133.
- Liou, M. (1989). "A note on reliability estimation for a test with components of unknown functional lengths". Psychometrika. 54, 153-163.
- Raykov, T. (1997). "Estimation of Composite Reliability for Congeneric Measures". Applied Psychological Measurement. 21 , 173-184.
- Raykov, T. and Grayson, David. (2003). "A Test for Change of Composite Reliability in Scale Development". Multivariate Behavioral Research. 38, 143-159.