

ชื่อเรื่องวิทยานิพนธ์

การศึกษาความเที่ยงตรงในการจำแนกความรู้และ
เปรียบเทียบผลการทดสอบบางประการจากการใช้แบบ
ทดสอบอิงเกณฑ์และแบบทดสอบเทเลอร์ (Tailored
Test) บางรูปแบบ

ชื่อผู้เขียน

นายประเวศ เวชชะ

ศึกษาศาสตรมหาบัณฑิต

สาขาวิชาการวัดและประเมินผลการศึกษา

คณะกรรมการตรวจสอบวิทยานิพนธ์

รศ.ดร.น	หาญตระกูล	ประธานกรรมการ
รศ.ดร.สุธรรม	จันทร์หอม	กรรมการ
รศ.ดร.บุญส่ง	นิลแก้ว	กรรมการ

บทคัดย่อ

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาความเที่ยงตรงในการจำแนกความรู้ และ
เปรียบเทียบความถูกต้องของการจำแนกความรู้ ค่าประมาณความสามารถ ค่าสาระ
ประโยชน์ของข้อสอบ (Item Information Function) และค่าความคลาดเคลื่อนมาตรฐาน
ของค่าความสามารถจากการใช้แบบทดสอบอิงเกณฑ์ 10, 20 และ 30 ข้อ แบบทดสอบเทเลอร์
แบบแตกแขนง (Branched Test) แบบทดสอบเทเลอร์แบบทดสอบตามสภาพ (Status
Tailoring) แบบทดสอบเทเลอร์แบบทดลองทำ (Performance Tailoring) และแบบ
ทดสอบแบบให้ตัดสินใจเลือกเอง (Self-Tailoring)

กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2529 ใน
โรงเรียนสังกัดสำนักงานการประถมศึกษาจังหวัดเชียงราย จำนวน 115 คน

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลประกอบด้วยแบบทดสอบอิงเกณฑ์ 10, 20 และ
30 ข้อ แบบทดสอบเทเลอร์แบบแตกแขนง แบบทดสอบเทเลอร์แบบทดสอบตามสภาพ แบบทดสอบ
เทเลอร์แบบทดลองทำและแบบทดสอบเทเลอร์แบบให้ตัดสินใจเลือกเอง ซึ่งวัดในจุดประสงค์

การแก้โจทย์ปัญหาทศนิยม ผลจากการทดสอบของนักเรียนในแบบทดสอบแต่ละฉบับนำไปประมาณค่าความสามารถ ค่าความคลาดเคลื่อนมาตรฐานของค่าความสามารถ และค่าสาระประโยชน์ของข้อสอบ (Item Information Function) การจำแนกนักเรียนเป็นผู้รอบรู้หรือไม่รอบรู้กระทำ โดยเปรียบเทียบค่าความสามารถที่ได้จากผลการทดสอบกับเกณฑ์ (ค่าระดับความสามารถที่เป็นเกณฑ์เท่ากับ .54) ถ้าสูงกว่าเกณฑ์ก็จำแนกหรือตัดสินว่าเป็นผู้รอบรู้ และถ้าต่ำกว่าเกณฑ์ก็แสดงว่าเป็นผู้ไม่รอบรู้ การรอบรู้หรือไม่รอบรู้นี้ถือว่าเป็นสถานะการรอบรู้ที่กำหนดจากแบบทดสอบ สำหรับสถานะการรอบรู้ที่แท้จริงของนักเรียนได้จากการทดสอบโดยวิธีเบย์ส (Bayesian Strategies)

สถิติที่ใช้ในการหาความเที่ยงตรงในการจำแนกความรอบรู้ คือ สหสัมพันธ์แบบพาย (Phi Correlation) เปรียบเทียบความถูกต้องของการจำแนกความรอบรู้ โดยสถิติคอกแครน (Cochran Test) และทดสอบรายคู่หลักการวิเคราะห์โดยวิธี Dunn-Bonferroni เปรียบเทียบค่าประมาณความสามารถ ค่าสาระประโยชน์ของข้อสอบ ค่าความคลาดเคลื่อนมาตรฐานของค่าความสามารถ โดยการวิเคราะห์ความแปรปรวนขององค์ประกอบเดียวเมื่อมีการวัดซ้ำ (One Factor Analysis : Repeated Measurement) และทดสอบรายคู่หลังการวิเคราะห์โดยวิธีเชฟเฟ (Scheffe)

ผลการวิเคราะห์ข้อมูลปรากฏผลดังนี้

1. ค่าความเที่ยงตรงในการจำแนกความรอบรู้ของแบบทดสอบอิงเกณฑ์ทั้ง 3 ฉบับและแบบทดสอบเทเลอร์ทั้ง 4 แบบ มีนัยสำคัญที่ระดับ .01 และพบว่าแบบทดสอบเทเลอร์ทั้ง 4 แบบให้ความเที่ยงตรงในการจำแนกความรอบรู้มีแนวโน้มสูง เรียงลำดับจากมากไปน้อยดังนี้คือแบบทดสอบเทเลอร์แบบแตกแขนง แบบทดสอบเทเลอร์แบบทดสอบตามสภาพ แบบทดสอบเทเลอร์แบบทดลองทำ และแบบทดสอบเทเลอร์แบบให้ตัดสินใจเลือกทำเอง (สัดส่วนค่าพายกับพายสูงสุด 1.00, .89, .76 และ .71 ตามลำดับ) สำหรับแบบทดสอบอิงเกณฑ์ 10, 20 และ 30 ข้อ ให้ความเที่ยงตรงในการจำแนกความรอบรู้อยู่ในระดับปานกลาง (สัดส่วนค่าพายกับพายสูงสุด .59, .69 และ .67 ตามลำดับ)

2. แบบทดสอบอิงเกณฑ์ 10 ข้อ ให้ความถูกต้องในการจำแนกความรอบรู้ต่ำกว่าแบบทดสอบเทเลออร์แบบทดสอบตามสภาพอย่างมีนัยสำคัญที่ระดับ .05 ส่วนคู่อื่น ๆ ไม่แตกต่างกันอย่างมีนัยสำคัญ

3. ค่าความสามารถที่ประมาณได้จากแบบทดสอบอิงเกณฑ์ 10, 20 และ 30 ข้อ แบบทดสอบเทเลออร์แบบแตกแขนง แบบทดสอบเทเลออร์แบบทดสอบตามสภาพ แบบทดสอบเทเลออร์แบบทดสอบทำและแบบทดสอบเทเลออร์แบบให้ตัดสินใจทำเองแตกต่างกันอย่างไม่มีนัยสำคัญ

4. แบบทดสอบอิงเกณฑ์ 10, 20 และ 30 ข้อ มีค่าสาระประโยชน์เฉลี่ยของข้อสอบแตกต่างกันอย่างไม่มีนัยสำคัญ และต่ำกว่าแบบทดสอบเทเลออร์ทั้ง 4 แบบอย่างมีนัยสำคัญที่ระดับ .01 โดยค่าสาระประโยชน์เฉลี่ยของข้อสอบของแบบทดสอบเทเลออร์แบบแตกแขนง แบบทดสอบเทเลออร์แบบทดสอบตามสภาพและแบบทดสอบเทเลออร์แบบให้ตัดสินใจเลือกทำเอง แตกต่างกันอย่างไม่มีนัยสำคัญ ค่าสาระประโยชน์เฉลี่ยของข้อสอบของแบบทดสอบเทเลออร์แบบทดสอบทำต่ำกว่าแบบทดสอบเทเลออร์แบบแตกแขนง และแบบทดสอบเทเลออร์แบบทดสอบตามสภาพอย่างมีนัยสำคัญที่ระดับ .01 และ .05 ตามลำดับ

5. ความคลาดเคลื่อนมาตรฐานของค่าความสามารถในแบบทดสอบอิงเกณฑ์ทั้ง 3 ฉบับ และแบบทดสอบเทเลออร์ทั้ง 4 แบบ แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 เมื่อทดสอบค่าเฉลี่ยรายคู่พบว่าเกือบทุกคู่แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 ซึ่งเรียงลำดับจากค่าความคลาดเคลื่อนมาตรฐานของค่าความสามารถจากน้อยไปมากคือ 1) แบบทดสอบเทเลออร์แบบแตกแขนง 2) แบบทดสอบอิงเกณฑ์ 30 ข้อ 3) แบบทดสอบอิงเกณฑ์ 20 ข้อ และแบบทดสอบเทเลออร์แบบให้ตัดสินใจเลือกเอง 4) แบบทดสอบเทเลออร์แบบทดสอบตามสภาพและแบบทดสอบเทเลออร์แบบทดสอบทำ 5) แบบทดสอบอิงเกณฑ์ 10 ข้อ

จากการศึกษาในครั้งนี้ แสดงให้เห็นว่าการใช้แบบทดสอบเพื่อจำแนกความรอบรู้ของนักเรียนในรายจุดประสงค์การเรียนรู้นั้น ควรใช้แบบทดสอบเทเลออร์มากกว่าแบบทดสอบอิงเกณฑ์ ทั้งนี้เพราะมีแนวโน้มให้ความเที่ยงตรงในการจำแนกสูง ให้ค่าสาระประโยชน์ในการตอบข้อสอบแต่ละข้อสูง

Thesis Title Criterion-Referenced and Tailored Tests :
A Study of Classification Validity and
Comparative Test Results.

Author Mr. Praves Vecha

M.Ed Educational Measurement and Evaluation

Examining Committee Assoc.Prof.Daroon Hantrakul Chairman
Assoc.Prof.Dr.Suthan Janhom Member
Assoc.Prof.Dr.Boonsong Nilkaew Member

Abstract

The purposes of this study were as follows : First, to study classification validity of criterion-referenced tests comprised of 10, 20 and 30 items by the methods of branched test, status tailoring, performance tailoring and self-tailoring. Secondly, to compare percision of mastery classification, estimated ability, item information function and standard error of estimated ability for all three criterion-refermced tests and all four methods of tailored test.

The samples were 115 Prathomsuksa 6 students in 1986 which were selected from 4 public schools of Chiangrai Primary Education Bureau.

The instruments used to collect data included, critrion-referenced tests that had length 10, 20 and 30 items, branched test, status tailoring, performance tailoring and self-tailoring which to

establish from the same item pool. The responded in each test of each examinee was estimated ability of examinee, standard error of estimated ability and item information function. Classification / decision to mastery or nonmastery were, if estimated ability was lower than mastery level ($\theta_m = .54$), showed nonmastery and above mastery level, showed mastery. Which mastery or nonmastery was student's estimated mastery states from tests. For student's actual mastery states was classified to use Bayesian Strategies.

The classification validity of mastery used phi correlation, to compare precision of mastery classification by Cochran Test and tested each pair by Dunn-Bonferroni method. To compare estimated ability, item information function and standard error of estimated ability used one factor analysis : repeated measurement and test each pair by Scheffe' method of comparison

The findings are as follow :

1) Mastery classification validity of criterion-referenced test, 10, 20 and 30 item, branched test, status tailoring, performance tailoring and self-tailoring were significantly high at the .01 level and found that all four types of tailored test tended to be the highest (proportion of phi correlation and phi-maximum correlation i.e. 1.00, .89, .76 and .71 as follow). For classification validity of criterion-referenced test (10, 20 and 30 items) tended to be the middle (proportion of phi correlation and phi-maximum correlation i.e. .59, .69, .67 as follow).

2. The precision of mastery classification of criterion-referenced that had 10 items length was lower than status tailoring significantly at the .05 level and the other pairs were not significantly.

3. To estimated ability of all criterion-referenced test and all four types of tailored test were not significantly different.

4. The mean of item information function of criterion-referenced test that had length, i.e. 10, 20 and 30 items were not significantly different but lower than all four types of tailored tests significantly different at the .01 level. The branched test's, status tailoring's and self-tailoring's were not significantly different. And performance tailoring's was lower than branched test's and status tailoring significantly different at the .01 and as level, as follow.

5. The standard error of estimated ability of all three criterion-referenced tests and all four types of tailored test were significantly different at the .01 level. The standard error of estimated ability of the branched test was lowest, criterion-referenced tests composed of 30 items was second criterion-referenced tests that had 20 items length and self-tailoring's were third, status tailoring's and performance tailoring's were fourth, and criterion-referenced tests that had 10 items length was highest.

The findings of this study indicated that all four types of tailored test should be used to classify mastery learning because classification validity and item information function were the highest.