

หัวข้อวิจัย ผลของการตัดข้อสอบที่มีอำนาจจำแนกติดลบออกต่อคะแนนการสอบ
และความเชื่อมั่นของข้อสอบทั้งฉบับ: รายวิชาระบบสุขภาพและการพยาบาล

ผู้ดำเนินการวิจัย ศรีสุตา วงศ์วิเศษกุล

ที่ปรึกษา ดร. เบ็ญจา เตากล้า

หน่วยงาน คณะพยาบาลศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต

ปี พ.ศ. 2558

ค่าความยากง่าย (P) และค่าอำนาจจำแนก (R) เป็นค่าพารามิเตอร์ที่ใช้ในการทดสอบคุณภาพของแบบสอบชนิดหลายตัวเลือกตามทฤษฎีการสอบแบบดั้งเดิม การวิจัยเชิงบรรยายนี้มีวัตถุประสงค์เพื่อประเมินคุณภาพของข้อสอบ ศึกษาการเปลี่ยนแปลงของคะแนนสอบและลำดับคะแนนรวมทั้งค่าความเชื่อมั่นของข้อสอบกลางภาคและปลายภาคเมื่อตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออก ประชากรการวิจัยประกอบด้วยข้อสอบกลางภาคและปลายภาครายวิชาระบบสุขภาพและการพยาบาลปีการศึกษา 2555 สำหรับนักศึกษาพยาบาลที่ลงทะเบียนเรียนจำนวน 92 คน ค่าความยากง่าย (P) ค่าอำนาจจำแนก (R) และค่าความเชื่อมั่นของข้อสอบคำนวณด้วยโปรแกรมสำเร็จรูปคอมพิวเตอร์ การเปลี่ยนแปลงของคะแนนสอบและลำดับคะแนนวิเคราะห์ด้วยสถิติ Paired t-test และ Sign Rank test

ผลการวิจัยพบว่า

1) ในการสอบกลางภาค ข้อสอบร้อยละ 64.29 มีค่าความยากง่ายในเกณฑ์เหมาะสม ($P = .20 - .80$) และร้อยละ 45.71 มีค่าอำนาจจำแนกที่ใช้ได้ ($R \geq .20$) ข้อสอบที่มีค่าอำนาจจำแนกเป็นลบร้อยละ 5.71 สำหรับการสอบปลายภาค ข้อสอบร้อยละ 40 มีค่าความยากง่ายในเกณฑ์เหมาะสม ($P = .20 - .80$) และร้อยละ 47.50 มีค่าอำนาจจำแนกที่ใช้ได้ ($R \geq .20$) ข้อสอบที่มีค่าอำนาจจำแนกเป็นลบร้อยละ 7.50

2) หลังจากตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออก ค่าเฉลี่ยคะแนนสอบกลางภาคและปลายภาคเพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ ($t_{(91)} = -10.225$ ที่ $p \text{ value} < .001$ และ $t_{(91)} = 4.972$ ที่ $p \text{ value} < .001$ ตามลำดับ) เช่นเดียวกับกับลำดับคะแนนที่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญทางสถิติ ($p \text{ value} < .001$)

3) ภายหลังจากตัดข้อสอบที่มีค่าอำนาจจำแนกเป็นลบออก ค่าความเชื่อมั่นของข้อสอบกลางภาคและปลายภาค ทั้งฉบับเพิ่มขึ้น

สรุปได้ว่า การตัดข้อสอบที่มีค่า r ติดลบออกทำให้คะแนนสอบและค่าความเชื่อมั่นของข้อสอบเพิ่มขึ้นรวมทั้งส่งผลต่อการเปลี่ยนแปลงของลำดับคะแนนการสอบเพิ่มขึ้น

คำสำคัญ: คุณภาพข้อสอบ อำนาจจำแนกเป็นลบ ความเชื่อมั่นของแบบสอบชนิดหลายตัวเลือก

Research Title The effects on test scores and the reliability of the test when taking out its items with negative discrimination power: Health System and Nursing subject

Researcher Sresuda Wongwiseskul

Research Consultants Dr. Benja Taoklam

Organization Faculty of Nursing, Suandusit Rajabhat University

Year 2014

Based on classical test theory, difficulty index (P) and discrimination index (R) are the parameters used to evaluate the standard of multiple choices questions (MCQs) in examination. This descriptive research aimed to evaluate the quality of test item, explore changes of test score and students' ranking, and changes of mid-term and final test reliability after removing the test items with negative discrimination power. Populations were midterm and final tests for 92 nursing students enrolled in Health System and Nursing subject, academic year 2012. Difficulty Index, discrimination index, test reliability were calculated using computerized program. Paired t-test and Sign Rank test were applied to test the changes of test score and students' ranking.

The study revealed that:

1) In mid-term test, 64.29% of items had index of difficulty (P) within a proper range ($P = .20 - .80$) and 45.71% had proper index of discrimination ($R \geq .20$). 5.71% of test items had negative discrimination. In the final, 40% of items had index of difficulty (P) within a proper range ($P = .20 - .80$) and 47.50% had proper index of discrimination ($R \geq .20$). 7.50% of test items had negative discrimination.

2) After removing the test items with negative discrimination power, mean score of mid-term and final test were higher and had significantly changes ($t_{(91)} = -10.225$ at p value $< .001$ and $t_{(91)} = 4.972$ at p value $< .001$ respectively). Similarly, the Signed Rank Test showed that students' rankings were significantly changed.

3) The reliability of mid-term and final test was increased after excluding.

In conclusion, excluding test items with negative discrimination power increased test reliability, mean score of the test, and affected score ranking.

Keywords: quality of test items, negative discrimination power, reliability of MEQ examination