ขนาดของผล : ความมีนัยสำคัญทางปฏิบัติในการวิจัย

รองศาสตราจารย์ ดร.สุพัฒน์ สุกมลสันต์

บทคัดย่อ

ขนาดของผล (Effect Size) เป็นแนวคิดที่ค่อนข้างใหม่มากในวงการวิจัย โดยเฉพาะอย่างยิ่งในประเทศไทย แต่เป็นแนวคิดที่สำคัญมากจนทำให้สมาคมนักจิตวิทยาแห่งอเมริกา (American Psychological Association: APA) กำหนดเป็นกรอบให้ผู้ที่ต้องการจะเผยแพร่ผลงานวิจัยบางชนิดจะต้องรายงานเรื่องนี้ในรายงานด้วย และวารสารทางวิชาการที่มีชื่อเสียงจำนวนมากก็ได้กำหนดตามแนวคิดดังกล่าวแล้วด้วย แต่เนื่องจากเรื่องนี้เป็นเรื่องใหม่ จึงทำให้มีปัญหาบางอย่างที่เกี่ยวข้อง เช่น ความหมาย การคำนวณหาขนาดของผล การแปลความ และการเลือกใช้สูตรที่เหมาะสมกับแบบการวิจัยและสถิติที่ใช้ เป็นต้น

Abstract

The idea of effect size (ES) is rather relatively new in research circles, especially in Thailand. It is seen as so important that the APA (American Psychological Association) now requires any authors wanting to have certain types of research articles published in their journal to include a measure of ES or practical significance in their reports. Consequently, many well-known journals are following this practice. However, since it is new, there are some main problems in implementing the concept, for example, its definitions, how to find it, how to interpret its meanings and how to select the appropriate formula that best matches a research design and the statistics used and reported in that research.

บทนำ

เมื่อนักวิจัยทำการวิจัยเชิงเปรียบเทียบแล้วมักจะต้องการรู้ว่าสิ่งที่ตนเองได้ศึกษานั้นดีกว่าของเดิมหรือไม่ ดังนั้น จึงมักนิยมทำการศึกษาเปรียบเทียบของเดิมกับของใหม่ด้วยวิธีการต่างๆ กัน เช่น ศึกษาเปรียบเทียบว่าผลของการสอนรูปแบบใหม่ หรือวิธีการใหม่ หรือตามแนวคิดใหม่ได้ผลดีกว่ารูปแบบเดิม หรือวิธีเดิม หรือแนวคิดเดิมหรือไม่ โดยการทดสอบความมีนัยสำคัญทางสถิติ (Statistical Significance Testing) ที่ระดับความมีนัยสำคัญ (p-value) ที่แตกต่างกัน ผู้ที่ริเริ่มนำวิธีการนี้มาใช้คนแรกคือ Ronald Fisher (ผู้ซึ่งได้ชื่อว่าเป็นบิดาของ ANOVA หรือ Analysis of Variance) ในราวปี ค.ศ. 1920 โดยการทดสอบสมมุติฐานศูนย์ (Null Hypothesis) ต่อมาในราวปี ค.ศ. 1930 Neyman และ Pearson ได้นำวิธีการทดสอบสมมุติฐานทางเลือก (Alternative Hypothesis) มาใช้ ส่วนการกำหนดระดับควรมีนัยสำคัญที่ 0.05 และ 0.01 นั้น ได้รับความนิยมในภายหลัง (Barnette, 2006)

แนวคิดดังกล่าวได้รับความนิยมในหมู่นักวิจัยเรื่อยมา ตั้งแต่ปี ค.ศ. 1940 และได้มีการสำรวจพบว่าผลงานวิจัยที่ตีพิมพ์ในปี ค.ศ. 1960 นั้น ประมาณ 85% ทำการศึกษาเปรียบเทียบด้วยวิธีการดังกล่าวแล้ว และหลังจากนั้นงานวิจัยประมาณ 90% ก็ยังนิยมใช้วิธีดังกล่าว แสดงว่าวิธีการนี้ได้รับความนิยมมากขึ้นเรื่อยๆ แต่ในระยะต่อมาความมีนัยสำคัญทางสถิติได้รับการวิจารณ์มากว่าให้ข้อมูลในการวิจัยจำกัดและมีผู้เสนอ (Thompson, 2000) ให้รายงานวิจัยรายงานขนาดของผล (Effect Size) ซึ่งถือว่าเป็นความมีนัยสำคัญในทางปฏิบัติ (Practical Significance) ด้วย (Thompson, 2000)

ทำไมขนาดผลวิจัยเป็นสิ่งที่นักวิจัยต้องการ

เมื่อประมาณ 20 ปีที่ผ่านมานักวิจัยจำนวนมากมีความเห็นว่าการทดสอบความแตกต่างทางสถิติโดยอาศัยระดับความมีนัยสำคัญเท่านั้นไม่เหมาะสมและไม่เพียงพอด้วยเหตุผลหลายๆ อย่าง (Neil, 2008; Nakagawa and Cuthill, 2007; Barnette, 2006; Coe, 2002) เช่น

1. การทดสอบความมีนัยสำคัญโดยอาศัยสมมุติฐานศูนย์สื่อความหมายผิด เช่น การที่เราตั้งสมมุติฐานเพื่อการทดสอบผลของการสอนวิธีหนึ่ง ภายหลังจากการทดลองสอนเป็นเวลาหนึ่งภาคการศึกษาแล้ว ผู้วิจัยตั้งสมมุติฐานว่า *"สัมฤทธิ์ผลในการเรียนของนักเรียนก่อนการเรียนด้วยวิธีสอนวิธีหนึ่ง และภายหลังการเรียนแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ 0.05"* เป็นสิ่งที่สื่อความหมายผิดจากความเป็นจริง เพราะว่าโดยปกติแล้วการเรียนทุกอย่างผู้เรียนจะมีสัมฤทธิ์ผลสูงขึ้นกว่าเดิมเสมอ เป็นต้น

2. การทดสอบความแตกต่างของผลการทดลองด้วยสถิติแบบเดิมให้ข้อมูลไม่เพียงพอ และขนาดของความแตกต่างระหว่างค่าเฉลี่ยทางคณิตศาสตร์ที่เท่ากัน มีความหมายที่แตกต่างกัน เช่น บางครั้งก็อาจแตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่บางครั้งอาจไม่มีนัยสำคัญ ทั้งนี้เพราะความแตกต่างจะมีนัยสำคัญหรือไม่ขึ้นอยู่กับปัจจัยหลักที่สำคัญ 2 อย่างคือ ขนาดของความแตกต่าง และขนาดของกลุ่มตัวอย่างว่ามีมากน้อยเพียงใด ดังนั้น การทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติ (Statistical Significance Testing) จึงให้ข้อมูลที่จำกัดแก่นักวิจัย

3. การทดสอบความแตกต่างทางสถิติไม่มีหน่วยนับที่สื่อความหมาย เช่น การศึกษาเปรียบเทียบระหว่างความสามารถในการเรียน การฟังภาษาอังกฤษของนักเรียนที่เรียนในตอนช่วงเช้ากับที่เรียนในช่วงบ่ายว่าต่างกันหรือไม่ ถ้านักวิจัยพบว่าโดยเฉลี่ยแล้วนักเรียนกลุ่มที่เรียนในช่วงเช้ามีคะแนนเฉลี่ย 15.20 จาก 20 คะแนน ส่วนนักเรียนที่เรียนช่วงบ่ายมีคะแนนเฉลี่ย 17.90 คะแนน ความแตกต่างเป็น 2.70 คะแนน ความแตกต่างนี้ไม่มีหน่วยของการวัดว่ามีความแตกต่างมากน้อยเพียงใด และมีความหมายว่าอย่างไร

4. ความแตกต่างทางสถิติขึ้นอยู่กับขนาดของกลุ่มตัวอย่างเป็นสำคัญ กล่าวคือ ความแตกต่างของค่าเฉลี่ยจะมีนัยสำคัญหรือไม่ ขนาดของกลุ่มตัวอย่างมีผลต่อการตัดสินดังกล่าว เช่น ความแตกต่างของค่าเฉลี่ยมีมากแต่กลุ่มตัวอย่างมีขนาดเล็กผลที่ได้อาจไม่มีนัยสำคัญทางสถิติ แต่หากว่าความแตกต่างมีน้อยและกลุ่มตัวอย่างมีขนาดใหญ่

(n > 50) ผลที่ได้อาจมีนัยสำคัญ เป็นต้น ดังนั้น อาจทำให้ผลการวิจัยผิดพลาดและสื่อความหมายผิด

5. การทดสอบความแตกต่างทางสถิติไม่ส่งเสริมให้มีการวิจัยซ้ำเรื่องเดิม เนื่องจากงานวิจัยที่ได้ทำแล้วบ่งบอกเพียงว่าผลการทดลองได้ผลแตกต่างกันหรือไม่เท่านั้น แต่ไม่ได้บอกขนาดของความแตกต่างว่ามีมากน้อยเพียงใด หรือผลของการทดลองมีขนาดใด และสมควรที่นักวิจัยจะวิจัยซ้ำในเรื่องนี้อีกหรือไม่

6. ขนาดของระดับมีนัยสำคัญ (p-value) ที่นักวิจัยใช้ไม่ค่อยมีความหมายที่สำคัญมากนักในทางปฏิบัติ เช่น นักวิจัยบางคนอาจกำหนดระดับความมีนัยสำคัญที่ p = 0.001 หรือ p = 0.01 หรือ p = 0.05 เป็นต้น ระดับเหล่านี้บ่งบอกแต่เพียงว่าหากว่ามีการวิจัยซ้ำแล้วจะได้ผลเช่นเดียวกับการวิจัยเดิมอีกกี่ครั้ง แต่ก็ไม่ได้บอกว่าขนาดของความแตกต่างมีมากน้อยเพียงใด

7. การทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติไม่มีประโยชน์โดยตรงในการสร้างองค์ความรู้ (Body of Knowledge) สำหรับเรื่องที่เกี่ยวข้องเพื่อการวิจัยอภิวิเคราะห์ (Meta-Analysis) ในอนาคต

8. การทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติที่มีกลุ่มตัวอย่างขนาดเล็ก (n < 50) อาจทำให้ผลผิดพลาด และสื่อความหมายผิดและทำให้เกิดความผิดพลาดชนิดที่ 2 (Type II Error) คือไม่สามารถปฏิเสธสมมุติฐานศูนย์ (Null Hypothesis) ที่ผิดได้ ในกรณีเช่นนี้นักวิจัยควรใช้ขนาดของผลและค่าระดับความเชื่อมั่น (Confidence Intervals) จะดีกว่า

9. บางครั้งการทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติเป็นการฝ่าฝืนข้อตกลงเบื้องต้น (Basic Assumptions) อย่างหนึ่งของการใช้สถิติพารามิเตอร์ (Parametric Test) เมื่องานวิจัยไม่ได้สุ่มตัวอย่างมาจากประชากรเพราะผู้วิจัยไม่ต้องการอ้างอิงสรุปผลการวิจัย (Generalize) ดังนั้นจึงเป็นการไม่เหมาะสม

ความหมายของค่าว่าขนาดของผล

คำว่าขนาดของผล (Effect size) มีความหมายหลายอย่าง (Nakagawa and Cuthill, 2007) แต่ในการวิจัยทางสังคมศาสตร์และมนุษยศาสตร์ คำว่าขนาดของผลหมายถึงขนาดของผลที่เกิดขึ้นจากตัวแปรต้น (Independente Variable) ต่อตัวแปรตาม (Dependente Variable) ที่ได้จากการศึกษาเชิงเปรียบเทียบหรือเชิงความสัมพันธ์ เช่น ขนาดความแตกต่างของค่าเฉลี่ยทางคณิตศาสตร์ ค่าสัมประสิทธิ์กระถดถอย (Regression Coefficient) ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) และค่า Cohen's d เป็นต้น ค่าขนาดของผลเป็นอิสสระจากขนาดของกลุ่มตัวอย่าง (Becker, 1999b)

ความสำคัญของขนาดของผลในการวิจัยในปัจจุบัน

1. หนังสือคู่มือ APA (American Phonetics Association) บังคับให้รายงานขนาดของผล

เนื่องจากการรายงานผลการวิจัยแต่เพียงเรื่องความมีนัยสำคัญทางสถิติ (Statistical Significance) เท่านั้นให้ข้อมูลไม่เพียงพอและมีปัญหาหลายอย่างดังได้กล่าวมาแล้ว ในปัจจุบันนี้นักวิจัยจำนวนมากกำหนดให้มีการรายงานความมีนัยสำคัญทางปฏิบัติ (Practical Significance) ซึ่งเป็นการรายงานขนาดของผล (Effect Size) ของผลการวิจัยด้วย เช่น ในหนังสือคู่มือ APA (American Phonetics Association Publication Manual) ฉบับพิมพ์ครั้งที่ 4 ในปี ค.ศ.1994 ได้เขียนเสนอแนะเพื่อ "ส่งเสริม" (encourage) ให้นักวิจัยรายงานขนาดของผลในรายงานวิจัยของตน และคู่มือฉบับพิมพ์ครั้งที่ 5 ในปี ค.ศ. 2000 ได้เปลี่ยนข้อความดังกล่าวเป็น "บังคับ" (require) ให้นักวิจัยรายงานขนาดของผล "เสมอ" ทุกครั้งที่มีการรายงานระดับมีนัยสำคัญทางสถิติ (p-value) รวมทั้งได้เสนอแนะสูตรในการเปลี่ยนค่า r_{xy}, d ที่น่าสนใจไว้ด้วย (Thompson, 2000) คือ

1. เมื่อ N รวมมีขนาด > 30

$$r_{xy} = d/[d^2 + 4)\frac{1}{2}]$$

2. เมื่อ N รวมมีขนาด < 30

$$r_{xy} = d/[d^2 + [N^2 + 2N)/(n_1 n_2)]\frac{1}{2}]$$

3. $d = 2_{xy} / [(1 - r_{xy}^{2}) \frac{1}{2}]$

อนึ่ง การที่หนังสือคู่มือ APA บังคับให้นักวิจัยต้องรายงานขนาดของผลก็เพราะว่าการรายงานเฉพาะความมีนัยสำคัญทางสถิติให้ข้อมูลแก่ผู้จะใช้ผลของการวิจัยไม่เพียงพอ และเพื่อประโยชน์ของการวิจัยแบบอภิวิเคราะห์ (Meta-analysis) ต่อไป

2. วารสารชั้นนำในต่างประเทศอย่างน้อย 24 ฉบับ กำหนดเป็นข้อบังคับให้ผู้เขียนบทความวิจัยต้องรายงานขนาดของผลด้วย (Barnette, 2006) เช่น Educational and Psychological Measurement, Early Children and Research Quarterly, Experimental Education, Language Learning, Personality Assessment, Reading and Writing และ Educational Technology Research and Development เป็นต้น และมีจำนวนวารสารที่ใช้ข้อบังคับเดียวกันนี้เพิ่มขึ้นเรื่อยๆ

การคำนวณหาขนาดของผล

ขนาดของผลมีวิธีคำนวณได้หลายอย่างแล้วแต่ชนิดของการวิจัยและสถิติที่ใช้ในการทดสอบ แต่อาจจำแนกออกได้เป็น 3 จำพวก (Barnette, 2006) ตามเกณฑ์ต่อไปนี้

1. ช่วงความเชื่อมั่น (Confidence Interval)

2. หน่วยมาตรฐานของความแตกต่าง (Standardized Units of Difference) และ

3. สถิติการมีส่วนร่วมของความแปรปรวน (Variance-Accounted-for Statistics)

ขนาดของผลแต่ละจำพวกมีรายละเอียดดังต่อไปนี้

ก. ใช้ช่วงความเชื่อมั่น

ตำราทางสถิติโดยมากไม่ได้รวมเรื่องช่วงความเชื่อมั่นเป็นดัชนีการวัดขนาดของผลไว้ด้วย แต่ในปัจจุบันนี้นักวิจัยเริ่มให้ความสนใจเรื่องช่วงความเชื่อมั่นของขนาดของผลมากขึ้น ทั้งนี้ เพราะว่ามีความสนใจเรื่องค่าของความน่าจะเป็น (Probability) ของขนาดของผลในช่วงความเชื่อมั่นระดับต่างๆ มากขึ้น ดังนั้น ช่วงของความเชื่อมั่นของค่าต่างๆ จึงมีความจำเป็น แต่เรื่องนี้ยังเป็นเรื่องใหม่จึงยังไม่มีสูตรที่นิยมใช้แพร่หลายทั่วไปในขณะนี้

ข. ใช้หน่วยมาตรฐานของความแตกต่าง

การคำนวณหาขนาดของผลโดยอาศัยหน่วยมาตรฐานของความแตกต่างของค่าเฉลี่ยเป็นวิธีที่ได้รับความนิยมมากที่สุดโดยมีสูตรพื้นฐานดังนี้

$$\text{ขนาดของผล} = \frac{[\text{ค่าเฉลี่ยของกลุ่มทดลอง}] - [\text{ค่าเฉลี่ยของกลุ่มควบคุม}]}{\text{ค่าเบี่ยงเบนมาตรฐาน}}$$

ปัญหาของการใช้สูตรพื้นฐานข้างต้นก็คือค่าเบี่ยงเบนมาตรฐานเป็นค่าที่ควรได้จากประชากร (Population) ไม่ใช่ค่าสถิติที่ได้จากกลุ่มตัวอย่าง ดังนั้น จึงเป็นค่าที่ผู้วิจัยทั่วไปไม่สามารถหาได้ ด้วยเหตุนี้ จึงมีนักวิจัยและนักสถิติหลายคนได้พัฒนาสูตรการคำนวณหาขนาดของผลที่แตกต่างกัน และสูตรที่ได้รับความนิยมมาก มี 3 สูตร คือ

1. สูตรที่คิดขึ้นโดย Cohen (Cohen's d) ได้แก่ (Furr, 2008)

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_{pooled}}$$

เมื่อ $\bar{x}_1$ = ค่าเฉลี่ยของกลุ่มทดลอง

$\bar{x}_2$ = ค่าเฉลี่ยของกลุ่มควบคุม

$$\sigma^{2}\text{ pooled} = \sqrt{\frac{n_1\sigma_1^{2} + n_2\sigma_2^{2}}{n_1 + n_2}}$$

แต่เนื่องจากค่า σ_1 และ σ_2 เป็นค่าของประชากร ดังนั้นในการปฏิบัตินักวิจัยจึงนิยมใช้ค่า S.D.$_1$ และ S.D.$_2$ มาแทน

3. สูตรที่คิดโดย Glass (Glass's Δ) ได้แก่ (Furr, 2008)

$$\Delta = \frac{\bar{X}_1 - \bar{X}_2}{S.D._{control\ group}}$$

4. สูตรที่คิดโดย Hedges (Hedges's g) ได้แก่ (Furr, 2008)

$$g = \frac{\bar{X}_1 - \bar{X}_2}{S.D._{pooled}}$$

ในเมื่อ $S.D._{pooled} = \sqrt{\frac{(n_1 - 1)S.D_1^{2} + (n_2 - 1)S.D_2^{2}}{(n_1 + n_2 - 2)}}$

อนึ่ง โปรดสังเกตว่าข้อมูลอย่างเดียวกัน เมื่อคำนวณโดยใช้สูตรต่างกันจะได้ค่าขนาดของผลต่างกัน โดยเรียงลำดับจากน้อยไปหามากจากแบบ Hedges, Cohen และ Glass ตามลำดับ (Barnette, 2006)

ค. ใช้สถิติการมีส่วนร่วมของความแปรปรวน

สูตรการหาขนาดของผลตามแนวคิดนี้ ได้แก่ งานวิจัยเชิงความสัมพันธ์ (Correlational Research) และงานวิจัยเชิงทดลองที่วิเคราะห์ข้อมูลโดยใช้วิธีการวิเคราะห์ความแปรปรวน เช่น ANOVA (Analysis of Variance) หรือการวิเคราะห์การกระถดถอยพหุคูณ (Multiple Regression Analysis) เป็นต้น ค่าขนาดของผลตามแนวคิดนี้ นิยมคำนวณหาค่าต่อไปนี้ (Furr, 2008; Becker, 1999b)

1. Eta squared (η^2)

$$\eta^2 = SS_{effect} / SS_{total}$$

2. Partial Eta squared (η_p^2)

$$\eta_p^2 = SS_{effect} /(SS_{effect} + SS_{error})$$

4. Omega squared (ω^2)

$$\omega^2 = ((SS_{effect} - (df_{effect})(MS_{error})) / Ms_{error} + SS_{total}$$

อนึ่ง การคำนวณโดยอาศัยสูตรทั้ง 3 ข้างต้น จะได้ค่าขนาดของผลต่างกันจากน้อยไปหามากตามลำดับ คือ ω^2, η^2 และ η_p^2

การคำนวณหาขนาดของผลโดยวิธีอื่นๆ คือ

นอกจากการคำนวณหาขนาดของผลโดยวิธีการต่างๆ ดังกล่าวแล้วข้างต้น ผู้วิจัยยังอาจคำนวณค่าขนาดของผลโดยใช้สูตรอื่นๆ ได้อีกเช่น

$$R^2, r_{xy}, r_{pb}, r_{bis}$$

ชนิดของขนาดของผล

ขนาดของผลสามารถที่จะแบ่งออกได้หลายชนิด แล้วแต่ชนิดของผลมาตรวัด (scale) รูปแบบในการวิจัย (Research Design) และสถิติที่ใช้ (Statistical Test) เช่น (Furr, 2008; Becker, 1999b; IES, 2009)

1. เมื่อมีการเปรียบเทียบความแตกต่างของค่าเฉลี่ยและมาตรวัดมีความต่อเนื่อง (ได้แก่มาตรวัดแบบ Interval Scale และ Ratio Scale) นิยมใช้สูตร

$$g = \frac{\bar{X}_1 - \bar{X}_2}{S.D._{pooled}}$$

เมื่อ $S.D._{pooled} = \sqrt{\frac{(n_1 - 1)S.D_1^2 + (n_2 - 1)S.D_2^2}{(n_1 + n_2 - 2)}}$

2. เมื่อมีการเปรียบเทียบค่าเฉลี่ยด้วย t-test หรือ ANOVA และ มาตรวัดมีความต่อเนื่องนิยมใช้สูตร

$$g = t\sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad \text{และ}$$

$$g = \sqrt{\frac{F(n_1 + n_2)}{n_1 n_2}}$$

3. เมื่อมีการเปรียบเทียบค่าเฉลี่ยด้วย ANCOVA (Analysis of Covariance) และมาตรวัดต่อเนื่องนิยมใช้สูตร

$$g = \bar{X}_1^{-1} - \bar{X}_2^{-1} / S.D._{pooled}$$

เมื่อ $\bar{X}_1^{-1}$ และ $\bar{X}_2^{-1}$ = ค่าเฉลี่ยของแต่ละกลุ่มที่ปรับแล้ว

ถ้ารูปแบบการวิจัยเป็น Pretest-Posttest Design และรู้ค่า r_{xy} ของคะแนน อาจใช้สูตร

$$g = \sqrt{\frac{F(n_1 + n_2)(1 + r_{xy}^2)}{n_1 n_2}}$$

4. เมื่อผลการวิจัยมีความเป็นไปได้ 2 อย่าง (dichotomous Outcomes) เช่น การสอบได้หรือสอบตกและประสบผลสำเร็จหรือล้มเหลว เป็นต้น นิยมใช้สูตร

$$Odds = \frac{p}{1 - p}$$

เมื่อ p = โอกาสที่เหตุการณ์นั้นจะเกิดขึ้น

$Odds$ = อัตราส่วนของเหตุการณ์ที่อาจแสดงขึ้น (Odds ratio)

5. เมื่องานวิจัยต้องการศึกษาเรื่องความสัมพันธ์ระหว่างตัวแปร นิยมหาขนาดของผลด้วยค่า r_{xy}, R^2 และ Φ เป็นต้น

6. เมื่องานวิจัยต้องการศึกษาการมีส่วนร่วมในการทำนายผลการวิจัย นิยมหาค่าขนาดของผลด้วยค่า η^2, η_p^2, ω^2 และ Intraclass Correlation หรือ ρ_1

การเปลี่ยนค่าขนาดของผลจากสูตรที่ต่างกัน

เนื่องจากการคำนวณหาขนาดของผลทำได้หลายวิธีและใช้สูตรหลายสูตร แต่มีนักวิจัยที่ได้ศึกษาขนาดและความหมายของขนาดของผลไว้แล้วไม่มาก และที่นิยมใช้กันมากได้แก่ ค่าของ Cohen's d และค่า r_{xy} ดังนั้น จึงมีผู้คิดสูตรสำหรับปรับเปลี่ยนค่าที่คำนวณได้จากวิธีต่างๆ ให้เป็นค่า Cohen's d หรือ r_{xy} รวมทั้งการเปลี่ยนค่าต่างๆ ระหว่างกันและกันไว้อีกมาก เช่น

1. เปลี่ยนค่า r_{xy} จากการเปรียบเทียบ 1 กลุ่มกับเกณฑ์ และการเปรียบเทียบ 2 กลุ่มที่สัมพันธ์กัน เช่น การทดสอบซ้ำหรือ Repeated Measures และสองกลุ่มที่ไม่เป็นอิสระจากกัน หรือ Paired/Related Samples/Dependent-Samples โดยใช้สูตร (Furr, 2008)

$$d = \frac{r_{xy}}{\sqrt{1 - r_{xy}}}$$

2. เปลี่ยนค่า Hedge's g เป็นค่า Cohen's d โดยใช้สูตร (Furr, 2008)

$$d = g\sqrt{\frac{N}{df}} = g\sqrt{\frac{N}{N-1}}$$

3. เปลี่ยนค่า t จากการเปรียบเทียบกลุ่มเดียวกับเกณฑ์ และแบบกลุ่มไม่เป็นอิสระ 2 กลุ่ม เป็นค่า d โดยใช้สูตร (Furr, 2008)

$$d = \frac{t}{\sqrt{df}} = \frac{t}{\sqrt{N-1}}$$

นอกจากนี้ยังมีสูตรอีกมากมายที่ใช้เปลี่ยนค่าต่างๆ ที่ใช้วัดขนาดของผล ผู้สนใจสามารถศึกษาเพิ่มเติมได้ที่ R.M. Furr, 2008

การแปลความขนาดของผล

การแปลความหมายของขนาดของผลสามารถทำได้หลายอย่าง (Valentine and Cooper, 2003). ทั้งนี้แล้วแต่ชนิดของค่าที่ได้จากการคำนวณ แต่ค่าที่ได้รับความนิยมมากที่สุดคือค่า Cohen's d ซึ่งสามารถแปลความหมายได้ดังนี้

1. ร้อยละของผลวิจัยในกลุ่มทดลองทำคะแนนได้มากกว่าผลวิจัยในกลุ่มควบคุม

เนื่องจากค่า d มีความหมายเช่นเดียวกับค่า z-score ที่เป็นคะแนนมาตรฐานในการกระจายที่เป็นโค้งปกติ (Normal Curve) ดังนั้น ค่า d จึงมีความหมายเช่นเดียวกับค่า z เช่น d = 0.80 แสดงว่าเนื้อที่ใต้โค้งปกติจาก -3 ถึง +0.80 คลุมเนื้อหาที่ประมาณร้อยละ 79 ดังนั้น หากว่าในการศึกษาเปรียบเทียบผลของการสอน 2 วิธีแก่ผู้เรียน 2 กลุ่มแล้วคำนวณได้ค่าขนาดของผล คือ d = 0.80 แสดงว่าโดยเฉลี่ยแล้วพลวิจัยในกลุ่มทดลองร้อยละ 79 สามารถทำคะแนนได้สูงกว่าพลวิจัยในกลุ่มควบคุม

2. ลำดับที่เมื่อคิดเป็นร้อย (Percentile Rank) แตกต่างกัน

ค่า d สามารถมีความหมายเป็นค่าลำดับที่เมื่อคิดเป็นร้อยละได้ เช่น ถ้าค่า d = 0.0 แสดงว่าโดยเฉลี่ยแล้วพลวิจัยในกลุ่มทดลองและกลุ่มควบคุมต่างมีคะแนนอยู่ที่ลำดับที่ 50 จาก 100 (p_{50}) แต่หากว่าค่า d = 0.80 แสดงว่าโดยเฉลี่ยแล้วพลวิจัยในกลุ่มทดลองมีคะแนนอยู่ที่ลำดับที่ 79 (p_{79}) ในขณะที่พลวิจัยในกลุ่มควบคุมมีคะแนนอยู่ที่ลำดับ 50 (p_{50}) ดังนั้น ลำดับที่ของพลวิจัยในกลุ่มทดลองสูงกว่าลำดับที่ของพลวิจัยในกลุ่มควบคุม = 79-50 = 29 ลำดับที่

3. ร้อยละของคะแนนของกลุ่มแตกต่างกัน

ขนาดของผลยังอาจหมายถึงร้อยละของคะแนนของพลวิจัยในกลุ่มควบคุมกับกลุ่มทดลองแตกต่างกันมากน้อยเพียงใด เช่น ถ้า d = 0.00 แสดงว่าโดยเฉลี่ยแล้วคะแนนของพลวิจัยในกลุ่มควบคุมและกลุ่มทดลองเท่ากัน อยู่ในลำดับที่

เท่ากัน คือ ลำดับที่ 50 (p_{50}) แต่หากว่า d = 0.80 แสดงว่าร้อยละ 47.40 ของพลวิจัยในกลุ่มทดลองมีคะแนนสูงกว่าพลวิจัยในกลุ่มควบคุม เป็นต้น (Becker, 1999b)

4. มีความหมายตามเกณฑ์ที่มีนักวิจัยกำหนดไว้แล้ว

Cohen (1988) และ Hopkins (2002) ได้กำหนดความหมายของขนาดของผลไว้ ดังนี้

d = 0.10 หมายถึง มีผลขนาดน้อยมาก
d = 0.20 หมายถึง มีผลขนาดเล็กน้อย
d = 0.50 หมายถึง มีผลขนาดปานกลาง
d = 0.80 หมายถึง มีผลขนาดมาก
d = 0.90 หมายถึง มีผลขนาดใหญ่มาก

ดังนั้น นักวิจัยอาจเทียบเคียงขนาดของผลที่คำนวณได้กับความหมายตามเกณฑ์ดังกล่าวข้างต้น แต่ว่าการแปลความหมายตามเกณฑ์ดังกล่าวควรกระทำอย่างระมัดระวัง Glass และคณะ (1981) แนะนำว่าการแปลความหมายค่า d ควรพิจารณาความคุ้มค่าและประโยชน์ (Costs and Benefits) มากกว่าขนาดของค่าที่คำนวณได้ เช่น ขนาดของผลมีน้อยมาก คือ d = 0.10 แต่หากว่าการทำให้เกิดผลดังกล่าวมีการลงทุนน้อยก็อาจถือได้ว่ามีประโยชน์ทางการศึกษา โดยเฉพาะอย่างยิ่งหากประโยชน์นั้นเกิดขึ้นกับผู้เรียนทุกๆ คน เป็นต้น

ปัจจัยที่มีผลต่อขนาดของผล

มีปัจจัยที่สำคัญหลายอย่างที่ทำให้ขนาดของผลเปลี่ยนแปลงได้ คือ (Coe, 2002)

1. ค่าเบี่ยงเบนมาตรฐาน (S.D.) เพราะว่าสูตรที่คิดโดย Cohen, Glass และ Hedges ใช้ค่าเบี่ยงเบนมาตรฐานแตกต่างกัน ได้มีนักสถิติเพื่อการวิจัยทำการศึกษาแล้วพบว่าค่า S.D. ที่เหมาะสมที่สุดเพื่อใช้คำนวณหาค่าขนาดของผล คือ

$$S.D._{pooled} = \sqrt{\frac{(N_E - 1)S.D._E^2 + (N_C - 1)S.D._C^2}{N_E + N_C - 2}}$$

2. การกระจายของคะแนนในกลุ่มควบคุมและกลุ่มทดลองไม่เป็นโค้งปกติ การนำคะแนนที่มีการกระจายไม่เป็นโค้งปกติมาเปรียบเทียบกันเป็นการฝ่าฝืนข้อตกลงเบื้องต้นของสถิติพารามิเตอร์ จึงอาจทำให้การทดสอบเบื้องต้นไม่ถูกต้อง โดยเฉพาะหากมีการเลือกใช้สถิติไม่เหมาะสม เช่น แทนที่จะใช้ Welsch's t-test หรือ Welsch's ANOVA แต่ผู้วิจัยใช้ t-test หรือ ANOVA ปกติแทนจะทำให้ผลการทดสอบผิดพลาดตั้งแต่ต้น ซึ่งจะมีผลต่อความถูกต้องของค่าขนาดของผลในภายหลัง

3. ค่าความเที่ยงของเครื่องมือการวัดและประเมินผล กล่าวคือ เครื่องมือวิจัยที่มีค่าความเที่ยงต่ำ ทำให้ค่าขนาดของผลมีน้อยกว่าเครื่องมือที่มีค่าความเที่ยงสูง (Coe, 2002)

ความสัมพันธ์ระหว่างค่าขนาดของผลแบบต่างๆ

ความสัมพันธ์ระหว่างค่า d, r_{xy}, R^2, ω^2, η^2, %-ile Rank, % Non-Overlap และ Odds Ratio มีดังต่อไปนี้ (Barnette, 2006; Coe, 2002; Hopkins, 2002; Becker, 1999b)

Cohen's Standards	d	r_{xy}	R^2	%-ile	% Non-Overlap	ω^2	η^2	Odds Ratio
	2.0	.707	.500	97.7	81.1			
	1.9	.689	.474	97.1	79.4			
	1.8	.669	.448	96.4	77.4			
	1.7	.648	.419	95.5	75.4			
	1.6	.325	.390	94.5	73.1			
	1.5	.600	.360	93.3	70.7			
	1.4	.573	.329	91.9	68.1			
	1.3	.545	.297	90	65.3			
ใหญ่มาก	1.2	.514	.265	88	62.2			-32
	1.1	.482	.232	86	58.9			
	1.0	.447	.200	84	55.4			
	0.9	.410	.168	82	51.9	0.138	01379	9.00
มาก	0.8	.371	.138	79	47.4			
	0.7	.330	.109	76	43.0			
	0.6	.287	.083	73	38.2			
ปานกลาง	0.5	.243	0.059	69	33.0	0.059	.0588	3.50
	0.4	.196	.038	66	27.4			
	0.3	.148	.022	62	21.3			
น้อย	0.2	.100	0.010	58	14.7	0.010	.0099	1.50
	0.1	.050	.002	54	7.7			
น้อยมาก	0.0	.000	.000	50	0			1.00

***หมายเหตุ**

1. เมื่อมีกลุ่มเปรียบเทียบกัน 2 กลุ่ม และกลุ่มตัวอย่างมีขนาดใหญ่ ($(\eta > 100)$ จะเทียบเคียงด้วยค่าอื่น ๆ ได้ถูกต้อง
2. ค่า η^2 และ ω^2 จะมีความสัมพันธ์กันเชิงเส้นตรงไม่ว่ากลุ่มตัวอย่างจะมีขนาดเท่าใด หรือจำนวนกลุ่มเปรียบเทียบจะเป็นเท่าใด

ปัญหาในการใช้ค่าขนาดของผล

แม้ว่าค่าขนาดของผลจะได้รับความนิยมมากในการรายงานผลการวิจัยในปัจจุบันนี้ แต่ก็ยังมีปัญหาในการใช้อยู่บ้าง (Barnette, 2006) อันได้แก่

1. ค่าขนาดของผลมีสูตรในการคำนวณหาระดับความเชื่อมั่น (Confidence Interval) เพียงบางชนิดแต่ยังไม่แพร่หลาย

2. สูตรในการคำนวณค่าขนาดของผลมีหลายสูตร และให้ค่าแตกต่างกัน เช่น จากข้อมูลชุดเดียวกันค่า d, Δ และ g ก็แตกต่างกัน หรือค่า η^2, η^2_p และ ω^2 ก็แตกต่างกัน เป็นต้น

3. ในกรณีที่มีกลุ่มทดลองมากกว่า 2 กลุ่ม ยังไม่มีสูตรในการคำนวณหาขนาดของผลโดยอาศัยค่า d ที่เป็นมาตรฐานในการเปรียบเทียบ Barnette (2006) ได้เสนอแนะวิธีการแก้ปัญหานี้ ดังนี้

ก. ทำการวิเคราะห์ข้อมูลด้วยวิธี ANOVA

ข. ทดสอบค่าเฉลี่ยด้วย Post Hoc Comparison โดยเลือกใช้วิธี Tukey's HSD Method

ค. คำนวณหาพิสัยของความแตกต่างของค่าเฉลี่ยโดยใช้สูตร

$$\text{HSD} = q_{\alpha,k,dfe}\sqrt{\frac{MS_E}{n}}$$

เมื่อ k = จำนวนกลุ่ม
dfe = df ของ Within group
n = ขนาดของกลุ่มตัวอย่าง

ง. ค่าความแตกต่างของค่าเฉลี่ยที่เท่ากับหรือมากกว่าค่า HSD แสดงว่าแตกต่างกันอย่างมีนัยสำคัญ

จ. คำนวณหาค่า d หรือ η^2 เฉพาะคู่ที่มีค่าเฉลี่ยแตกต่างกัน

นอกจากนี้ Morris (2007) ได้ทำการศึกษาเปรียบเทียบวิธีการคำนวณหาขนาดของผลของการวิจัยแบบ Pretest-Posttest-Control Group Design 3 วิธี คือ เช่น (1) ใช้ค่าของ S.D. ของคะแนนจาก Pretest แยกส่วน (Using Separate Pretest SDs) (2) ใช้ค่าของ S.D. ของคะแนนจาก Pretest รวมกัน (Using Pooled Pretest SDs) และ (3) ใช้ค่าของ S.D. ของคะแนนจาก Pretest และ Posttest รวมกัน (Using Pooled Pretest and Posttest SD) ผลปรากฏว่า วิธีที่ 2 ดีที่สุด เพราะว่า ได้ค่าขนาดของผลที่ไม่มีอคติ (Unbiased Estimate) มีความแปรปรวนในการสุ่มน้อย และคำนวณสะดวกกว่าวิธีอื่น จึงแนะนำให้ใช้วิธีดังกล่าวนี้ โดยใช้สูตรดังนี้ (Morris, 2007 : 369) คือ

$$d_{ppc} = c_p\left[\frac{(\bar{X}_{post,T}-\bar{X}_{pre,T})-(\bar{X}_{post,C}-\bar{X}_{pre,C})}{S.D_{pre}}\right]$$

เมื่อ $S.D_{pre} = \sqrt{\frac{(n_T-1)S.D^2_{pre,T}+(n_C-1)S.D^2_{pre,C}}{n_T+n_C-2}}$

$$c_p = \frac{3}{4(n_T+n_C-2)-1}$$

ผู้ที่ต้องการศึกษารายละเอียดเพิ่มเติมอาจดูได้จากบทความของ S.B. Morris (2007)

4. ในกรณีที่การวิจัยเป็นแบบ Random Model ผู้วิจัยควรเลือกใช้ค่า ICC (Intraclass Correlation) เป็นค่าขนาดของผล แทน η^2 หรือ ω^2

5. ปัญหาการเลือกใช้สูตรเมื่อการทดสอบมีเพียงกลุ่มเดียว

ในกรณีที่การทดลองมีเพียงกลุ่มเดียว เช่น Pretest-Posttest Design นักวิจัยยังมีความคิดเห็นต่างกันว่าควรจะใช้สูตรใดในการคำนวณหาขนาดของผล ได้มีผู้เสนอให้ใช้สูตรในกรณีนี้ต่างๆ กัน เช่น

ก. Furr (2008) เสนอให้ใช้สูตรต่อไปนี้สำหรับการวิจัยที่เป็น Repeated Measures เพียง 2 ครั้ง หรือ Paired-Samples คือ

$$d = \frac{\overline{D}}{\sigma_D} \quad \text{และ}$$

$$d = \frac{\overline{D}}{S.D._D}$$

เมื่อ $\bar{D}$ = ความแตกต่างระหว่างคะแนน Pretest กับ Posttest

σ_D = S.D. ของความแตกต่างระหว่างคะแนน Pretest กับ Posttest ของประชากร

$S.D._D$ = S.D. ของความแตกต่างระหว่างคะแนน Pretest กับ Posttest ของกลุ่มตัวอย่าง

ข. Barnette (2006) แนะนำให้ใช้หลายสูตร คือ

1. ทำการวิเคราะห์ข้อมูลด้วยวิธี Repeated Measure ANOVA แล้วคำนวณหาค่า Partial Eta Squared คือ

$$\eta_p^2 = \frac{SS_{time}}{SS_{time} + SS_{residual}}$$

2. วิเคราะห์ข้อมูลด้วยวิธี Repeated Measure ANOVA แล้วคำนวณหาค่า Cohen's d ดังนี้

$$d = \frac{\bar{x}_{post} - \bar{x}_{pre}}{\sqrt{MS_{residual}}}$$

3. คำนวณโดยใช้สูตร Hedges's g คือ

$$g = \frac{\bar{x}_{post} - \bar{x}_{pre}}{S.D._{Pooled}}$$

เมื่อ $S.D._{pooled} = \sqrt{\frac{(n_1 - 1)S.D_1^2 + (n_2 - 1)S.D_2^2}{n_1 + n_2 - 2}}$

อนึ่ง Barnette ได้แนะนำว่าผู้วิจัยไม่ควรหาค่าขนาดของผลโดยสูตรของ Glass ในกรณีนี้เพราะว่าค่าที่ได้จะมีอคติในทางลบ (Negatively Biased Estimate) และในจำนวน 3 สูตรนี้ค่า *d* และ g นับว่าเหมาะสมที่สุด แต่หากว่าต้องการศึกษาเรื่องความสัมพันธ์ของตัวแปรก็ควรใช้ค่า Partial Eta Squared (η_p^2)

1. ปัญหาในการหาค่าขนาดของผลจากสถิติจำพวก Nonparametric Tests

ในระยะแรกๆ นักวิจัยสนใจแต่การคำนวณหาค่าขนาดของผลจากสถิติจำพวก Parametric Tests เช่น t-test และ F-test เป็นต้น เนื่องจากเป็นสถิติที่นิยมใช้แพร่หลายในงานวิจัยเชิงปริมาณ แต่ในระยะต่อมาได้มีนักวิจัยเรียกร้องให้นักสถิติช่วยกันคิดหาวิธีหาค่าขนาดของผลจาก Nonparametric Tests และปรากฏว่านักสถิติมีความเห็นว่า สถิติจำพวก Nonparametric Tests ที่ทดสอบความมีนัยสำคัญด้วยค่าสถิติ Z เช่น Mann-Whitney Test, Two-Sample Kolmogorov-Smirnov Test, Wald-Wolfowitz Test (สำหรับทดสอบความแตกต่างของค่าเฉลี่ยของพลวิจัย 2 กลุ่มที่เป็นอิสระ) Wilcoxon Singed Rank Test และ Sign Test (สำหรับทดสอบความแตกต่างของค่าเฉลี่ยของพลวิจัย 2 กลุ่มที่ไม่เป็นอิสระ) ให้ใช้สูตรต่อไปนี้ (Field, 2009)

$$ES = \frac{Z}{\sqrt{N}}$$

ในเมื่อ

Z = ค่าสถิติ Z ที่ได้จากการคำนวณ (เช่นจากตารางของโปรแกรม SPSS)

N = จำนวนพลวิจัยทั้งหมด (ในกรณีของการทดสอบความแตกต่างของค่าเฉลี่ยของพลวิจัย 2 กลุ่มที่ไม่เป็นอิสระ N = N x 2)

อนึ่ง ค่า ES ที่คำนวณได้จะเป็นค่าโดยประมาณ และสามารถใช้ได้เหมือนค่า d รวมทั้งมีความหมายเช่นเดียวกับค่า d ส่วนการคำนวณหาค่า ES สำหรับสถิติอื่นเพื่อการทดสอบความแตกต่างของค่าเฉลี่ยของพลวิจัยที่เป็นอิสระและไม่เป็นอิสระที่มากกว่า 2 กลุ่มยังคงมีปัญหาอยู่ว่าจะคำนวณหาอย่างไร

งานวิจัยที่ควรต้องรายงานขนาดของผล

มีงานวิจัยเพียง 3 ประเภทเท่านั้นที่สามารถคำนวณหาค่าขนาดของผลได้ (Furr, 2008; Hopkins, 2002; Becker, 1999b) ซึ่งได้แก่

1. งานวิจัยที่ศึกษาเปรียบเทียบค่าเฉลี่ยของกลุ่มตัวอย่างกับเกณฑ์ หรือระหว่างกลุ่มตัวอย่างด้วย t-test, ANOVA หรือ ANCOVA เช่น งานวิจัยเชิงทดลอง (Experimental Research)

2. งานวิจัยที่ศึกษาโอกาสเสี่ยงที่จะเป็นไปได้ของเหตุการณ์อย่างใดอย่างหนึ่ง (Odds Ratio)

3. งานวิจัยที่ศึกษาความสัมพันธ์ระหว่างตัวแปร เช่น งานวิจัยเชิงความสัมพันธ์ (Correlational Research) ที่ใช้การวิเคราะห์ความสัมพันธ์อย่างง่าย (Pearson Correlation) และการวิเคราะห์กระดถดถอยพหุคูณ (Multiple Regression Analysis) เป็นต้น

การใช้เทคโนโลยีในการช่วยการคำนวณหาค่าขนาดของผล

ปัจจุบันการคำนวณหาขนาดของผลทำได้ง่ายและสะดวกมากเนื่องจากมีโปรแกรมคอมพิวเตอร์ช่วยในการคำนวณ ผู้วิจัยอาจใช้วิธีการต่างๆ ดังต่อไปนี้

1. หาสูตรที่เหมาะสมกับงานวิจัยของตนเอง เช่นดูที่ Furr (2008), Becker (1996c) และ IES (2009) เป็นต้น แล้วใช้โปรแกรมคอมพิวเตอร์ทางคณิตศาสตร์ช่วยการคำนวณ เช่น Math Cad 2000, Excel หรือ SPSS แต่หากว่าหาไม่ได้ก็อาจคำนวณเองด้วยมือโดยใช้เครื่องคิดเลขก็ได้

2. ใช้โปรแกรมคอมพิวเตอร์ที่อาจหาได้โดยไม่เสียค่าใช้จ่ายจาก Internet เช่น Coe (2008), G-power Software และ Effect Size Generator (Wikipedia, 2009) เป็นต้น

3. ใช้โปรแกรมคอมพิวเตอร์ทาง Internet (Online Program) เช่น Becker (1999a) เป็นต้น

4. เลือกทางเลือก (option) ในการหาขนาดของผลจากโปรแกรมทางสถิติที่ใช้ทั่วไป เช่น SPSS (Statistical Packages for the Social Sciences) เมื่อวิเคราะห์ด้วย ANOVA, ANCOVA และ Repeated Measures ANOVA เป็นต้น

สรุป

การรายงานผลการวิจัยในปัจจุบันนี้ผู้วิจัยจะรายงานเฉพาะเรื่องความแตกต่างอย่างมีนัยสำคัญทางสถิติ (Statistical Significance) อย่างเดียวซึ่งไม่เพียงพอ ผู้วิจัยควรอย่างยิ่งที่จะต้องรายงานความมีนัยสำคัญในทางปฏิบัติ (Practical Significance) หรือขนาดของผล (Effect Size) ด้วย โดยเฉพาะหากว่าต้องการเผยแพร่ผลงานวิจัยในวารสารต่างประเทศ ทั้งนี้เพราะเป็นข้อบังคับของวารสารชั้นนำจำนวนมากที่ผู้วิจัยพึงต้องปฏิบัติตาม ส่วนวิธีการคำนวณหาขนาดของผลนั้นทำได้ไม่ยากเพราะมีโปรแกรมคอมพิวเตอร์ใช้ได้สะดวก และไม่ต้องเสียค่าใช้จ่ายจำนวนไม่น้อยที่อาจหาได้ทั่วไป

เอกสารอ้างอิง

Barnette, J. (2006). **Effect Size and Measures of Association.** School of Public Health, University of Alabama at Birmingham. Retrieved from http://www.eval.org/SummerInstitute/06SIHandouts/SI06.Barnette.TR2.Online.pdf on November 1, 2009.

Becker, L. (1999a). **Effect Size Calculator.** Retrieved from http://www.uccs.edu/~faculty/lbecker/ on November 1, 2009.

Becker, L. (1999b). **Effect Size.** Retrieved from http://www.uccs.edu/~faculty/lbecker/es.htm on November 1, 2009.

Becker, L. (1999c). **Measures of Effect Size.** Retrieved from http://www.helsinki.fi/~komulain/Tilastokirjat/Becker_Measures%20of%20Effect%20Size.pdf on November 1, 2009.

Coe, R. (2002). **It's Effect Size, Stupid.** School of education, University of Durham. Retrieved from http://www.cemcentre.org/Documents/CEM%20Extra/EBE/ESguide.pdf on November 1, 2009.

Coe, R. (2008). **Effect Size Calculator.** Retrieved from http://www.cemcentre.org/renderpage.asp?linkID=30325017/ on November 1, 2009.

Cohen, J. (1988). **Statistical Power Analysis for the Behavioral Sciences** (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Furr, M. (2008). **Summary of Effect Size and their Links to Inferential Statistics.** Psychology Department, Wake Forest University. Retrieved from http://psych.wfu.edu/furr/EffectSizeFormulas.pdf on November 1, 2009.

Glass, G. V, McGaw, B., and Smith, M. L. (1981). **Meta-analysis in Social Research.** Beverly Hills, CA: SAGE Publications.

Hopkins, W. (2002). **A Scale of Magnitudes for Effect Statistics.** Retrieved from http://www.sportsci.org/resource/stats/effectmag.html on November 1, 2009.

IES. (2009). **Effect Size Computations.** Retrieved from http://ies.ed.gov/ncee/wwc/references/idocviewe/Doc.aspx?docId=19&tocId=8/ on November 1, 2009.

Morris, S.B. (2007). **Estimating Effect Sizes from Pretest-Posttest-Control Groups Designs.** Retrieved from http://orm.sagepub.com/cgi/content/abstract/11/2/364 on November 1, 2009.

Nakagawa, S. and Cuthill, I (2007). **Effect Size, Confidence Interval and Statistical Issues in the Interpretation of Effect Sizes.** Retrieved from http://ies.ed.gov/ncee/wwc/pdf/essig.pdf on November 1, 2009.

Neil, J. (2008). **Why Use Effect Sizes instead of Significance Testing in Program Evaluation?** Retrieved from http://www.wilderdom.com/research/effectsizes.html on November 1, 2009.

Thompson B. L. (2000). **A Suggested Revision to the Forthcoming 5th Edition of the APA Publication Manual.** Retrieved from http://www.coe.tamu.edu/~bthompson/apaeffec.htm on November 1, 2009.

Valentine, J and Cooper, H. (2003). **Effect Size Substantive Interpretation Guidelines: Significance: A Practical Guide for Biologists.** Retrieved from http://www.bio.bris.ac.uk/research/behavior/R_scripts/Nakagawa%20Cuthill%20BR%202007.pdf on November 1, 2009.

Wikipedia (2009). **Effect Size.** Retrieved from http://en.wikipedia.org/wiki/Effect_size on November 1, 2009.

ประวัติผู้เขียน

รองศาสตราจารย์ ดร.สุพัฒน์ สุกมลสันต์

เป็นอาจารย์ประจำสถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย มีความชำนาญด้านการเรียนการสอนภาษาอังกฤษ การทดสอบ ทางภาษา สถิติเพื่อการวิจัย วิธีวิทยาการวิจัยทางภาษาศาสตร์และการศึกษา การเขียนโปรแกรมคอมพิวเตอร์ และการใช้คอมพิวเตอร์เพื่อการวิจัยและเพื่อช่วยการเรียนการสอน มีผลงานทางวิชาการหลายเล่ม และเป็นกรรมการผู้ทรงคุณวุฒิของสถาบันการศึกษาหลายแห่ง