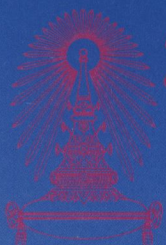


ISSN 0857 - 2933



# วารสาร วิธีวิทยาการวิจัย JOURNAL OF RESEARCH METHODOLOGY

วารสาร

ปีที่ ๗ ฉบับที่ ๒

ภาควิชาวิจัยการศึกษา คณะครุศาสตร์

กรกฎาคม - ธันวาคม ๒๕๓๘

จุฬาลงกรณ์มหาวิทยาลัย

ความรู้ชายแดนด้านการประเมินผลการศึกษา

ความรู้ทางด้านการวัดผลการศึกษา

วิธีวิทยาขั้นสูงด้านการวัดและประเมินผลการศึกษา

วิธีวิทยาขั้นสูงด้านการวิจัย และสถิติการศึกษา



## แนะนำผู้เขียน

- นงลักษณ์ วิรัชชัย, Ph.D.** ผู้ช่วยศาสตราจารย์ ประจำภาควิชาวิจัยการศึกษา คณะครุศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย
- ศิริชัย กาญจนวาสี, Ph.D.** รองศาสตราจารย์ ประจำภาควิชาวิจัยการศึกษา คณะครุศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย
- อุทุมพร จามรมาน, Ph.D.** ศาสตราจารย์ ประจำภาควิชาวิจัยการศึกษา คณะครุศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย
- สุวิมล ว่องวานิช, Ph.D.** รองศาสตราจารย์ ประจำภาควิชาวิจัยการศึกษา คณะครุศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย
- เยาวดี วิบูลย์ศรี, Ph.D.** รองศาสตราจารย์ ประจำภาควิชาวิจัยการศึกษา คณะครุศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย
- ศิริเดช สุชีวะ, ค.ต.** อาจารย์ประจำภาควิชาวิจัยการศึกษา คณะครุศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย
- วีระยุทธ ชาตะกาญจน์** มหائبันฑิต สาขาวิจัยการศึกษา คณะครุศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย

# วารสารวิธีวิทยาการวิจัย

JOURNAL OF RESEARCH METHODOLOGY

ปีที่ 7 ฉบับที่ 2 กรกฎาคม-ธันวาคม 2538

ISSN 0857-2933

วารสารรายหกเดือน ของภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ที่ปรึกษา : คณะกรรมการบริหารเงินทุนภาควิชาวิจัยการศึกษา และศาสตราจารย์ ประคอง กรรณสูต

บรรณาธิการ : ศาสตราจารย์ ดร.สมหวัง พิธิยานุวัฒน์

บรรณาธิการผู้ช่วย : รองศาสตราจารย์ ดร.สุวิมล ว่องวานิช

ผู้จัดการ : อาจารย์อวยพร เรืองตระกูล

ผู้จัดการผู้ช่วย : ดร.ศิริเดช สุชีวะ, นางสาวเกษราพร ภูเขาเทพ, นายอนุชา กอนพวง, นายประสิทธิ์ ไชยกาล

กองบรรณาธิการ :

ศาสตราจารย์ ดร.ธีระ อาชวเมธี

ดร.อำรุง จันทวานิช

รองศาสตราจารย์ อุเทน ปัญญ

รองศาสตราจารย์ ดร.สุวัฒนา อุทัยรัตน์

รองศาสตราจารย์ ดร.สุรศักดิ์ อมรรัตนศักดิ์

รองศาสตราจารย์ ดร.ฉัตรนภา พรหมมา

รองศาสตราจารย์ ดร.สุพัฒน์ สุกมลสันต์

ผู้ช่วยศาสตราจารย์ ดร.ทวีวัฒน์ ปิตยานนท์

ผู้ช่วยศาสตราจารย์ ดร.สุชาติ ตันธนะเดชา

ผู้ช่วยศาสตราจารย์ ดร.นิคม ตั้งคะพิภพ

ผู้ช่วยศาสตราจารย์ ดร. สำราญ มีแจ่ง

ผู้ช่วยศาสตราจารย์ ดร.ภาวิณี ศรีสุขวัฒนานันท์

อาจารย์ ดร.จุมพล พูนภัทรชีวิน

คุณอารีรัตน์ วัฒนสิน

ดร.สุพักตร์ พิบูลย์

ดร.ปัญญา ธีระวิทย์เลิศ

ศาสตราจารย์ บุญธรรม กิจปรีดาบริสุทธิ์

รองศาสตราจารย์ ดร.วรรณมา ปุณฺณโชติ

รองศาสตราจารย์ ศักดา บุญยไวยโรจน์

รองศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี

รองศาสตราจารย์ ดร.ชิตชนก เขิงเขาว์

รองศาสตราจารย์ ดร.พิศิษฐ ตันทวนิช

รองศาสตราจารย์ ดร.สัมพันธ์ พันธุ์พุกษ์

ผู้ช่วยศาสตราจารย์ ประนอม สุรัสวดี

ผู้ช่วยศาสตราจารย์ ดร.ใจทิพย์ เชื้อรัตนพงษ์

ผู้ช่วยศาสตราจารย์ ดร.จักรกฤษณ์ สำราญใจ

ผู้ช่วยศาสตราจารย์ ดร.อรพรรณ ณรงค์ศรีศักดิ์

ผู้ช่วยศาสตราจารย์ ดร.นงลักษณ์ วิรัชชัย

ดร.อรุณศรี อนันตรศิริชัย

ดร.สุวรรณา กาญจนผลิน

อาจารย์ เสรี ชัดรัมย์

คุณदनัย เทียนพุด

วารสารวิธีวิทยาการวิจัย ก้าวเข้าสู่ปีที่ 7 (พ.ศ. 2538) ด้วยความหวังและด้วยความมั่นใจในคุณภาพของผลงาน ขณะนี้กำลังเริ่มเปิดรับสมัครสมาชิกใหม่ โดยมีรายละเอียดดังนี้

สมัครเป็นสมาชิกปี พ.ศ. 2537-2538

ค่าสมาชิก 140 บาท

สมัครเป็นสมาชิกปี พ.ศ. 2537-2540

ค่าสมาชิก 260 บาท

จำหน่ายปลีกเล่มละ 40 บาท กำหนดออกปีละ 2 เล่ม (มกราคม และ กรกฎาคม)

จ่ายเช็ค ธนาคารใด ตัวแลกเงิน สั่งจ่าย ปณ. จุฬาลงกรณ์มหาวิทยาลัย ในนามของอาจารย์อวยพร เรืองตระกูล ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กทม. 10330 โทรศัพท์ 218-2440, 218-2446 โทรสาร 218-2446

ท่านที่ประสงค์จะเผยแพร่บทความทางวิธีวิทยาการวิจัยทั่วไป หรือเฉพาะทาง ตลอดจนผลงานวิจัยด้านวิธีวิทยาการวิจัย ผลการวิจัยทางสังคมศาสตร์ หรือด้านครุศาสตร์/ศึกษาศาสตร์ กรุณาส่งต้นฉบับ พิมพ์ A-4 ประมาณ 10-15 หน้า ไปที่ ศาสตราจารย์ ดร.สมหวัง พิธิยานุวัฒน์ ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ถนนพญาไท เขตปทุมวัน กทม. 10330

# สารบัญ

หน้า

## บทบรรณาธิการแถลง

- วิธีวิทยาขั้นสูงด้านการวิจัยและสถิติ  
    *นงลักษณ์ วิรัชชัย* 1
- วิธีวิทยาขั้นสูงด้านการวัดและประเมินผลการศึกษา  
    *ศิริชัย กาญจนวาสี* 37
- ความรู้ชายแดนด้านการวัดผลการศึกษา  
    *อุทุมพร จามรมาน* 42
- ความรู้ชายแดนด้านการประเมินผลการศึกษา  
    *สุวิมล ว่องวาณิช* 52
- An Application of Rule Assessment Approach for  
    Diagnosing Misconceptions  
    *Siridej Susheva* 68
- ผลของการแจ้งให้ทราบล่วงหน้าต่างวิธี ตำแหน่งหน้าที่ของผู้ตอบ  
    และการติดตามทางไปรษณีย์ที่มีต่ออัตราการตอบกลับและความจริงใจ  
    ในการตอบแบบสอบถามที่ส่งทางไปรษณีย์  
    *วีระยุทธ ชาตะกาญจน์* 81
- การติดตามคุณภาพและประสิทธิภาพของแบบประเมินตนเอง  
    สำหรับครูมัธยมศึกษา  
    *เยาวดี วิบูลย์ศรี* 93

## บรรณาธิการแถลง

ในช่วงเวลานี้ จะเห็นว่าพัฒนาการความรู้ด้านการวิจัย การวัดผล การประเมินผล และ สถิติการศึกษาก้าวหน้าไปมาก วิทยาการใหม่ ๆ เกิดขึ้นมากมาย ไม่ว่าจะเป็นเชิงแนวคิด ทฤษฎี หรือเทคนิควิธีที่ใช้ในการเสาะแสวงหาความรู้ หรือการวิเคราะห์ข้อมูล วารสารวิธีวิทยา การวิจัยฉบับนี้จึงได้รวบรวมบทความที่เกี่ยวข้องกับวิธีวิทยาขั้นสูง ตลอดจนความรู้ชายแดน ในศาสตร์นี้ไว้ด้วยกัน เพื่อให้ผู้อ่านที่สนใจติดตามข่าวสารในเรื่องนี้ได้ศึกษาค้นคว้าเพิ่มเติมได้สะดวก นอกจากนี้ยังได้นำเสนอบทความคัดย่อจากรายงานการวิจัย ซึ่งมีทั้งการพัฒนาวิธีการวิจัย มโนทัศน์ที่คลาดเคลื่อนในวิชาคณิตศาสตร์ การวิจัยที่พยายามหาปัจจัยที่ช่วยเพิ่มอัตราการ ตอบกลับและเพิ่มความจริงจังในการตอบ และการติดตามคุณภาพและประเมินประสิทธิภาพ ของแบบประเมินตนเองสำหรับครุมัธยมศึกษา

ปัจจุบันเทคโนโลยีคอมพิวเตอร์เข้ามามีบทบาทอย่างมากต่อการทำวิจัย เทคนิค วิเคราะห์ข้อมูลแบบใหม่ ๆ ทำให้สามารถตอบคำถามวิจัยซึ่งแต่เดิมอาจมีข้อจำกัดนั้นมีความ ลึกซึ้งขึ้น สามารถทำให้การพัฒนาแบบสอบถามหรือมาตรวัดต่าง ๆ มีประสิทธิภาพและมีความ เหมาะสมกว่าเดิม การก้าวให้ทันวิทยาการเหล่านี้ จำเป็นต้องอาศัยความร่วมมือกันในหมู่นักวิชาการ โดยการแลกเปลี่ยนประสบการณ์และข่าวสารซึ่งกันและกัน ผู้จัดทำวารสารวิธี วิทยาการวิจัย จึงถือเป็นภารกิจสำคัญที่จะทำหน้าที่เป็นสื่อกลางระหว่างเพื่อนสมาชิกให้ สมบูรณ์เท่าที่จะทำได้ จึงขอเชิญชวนสมาชิกช่วยกันส่งบทความทางวิชาการที่ต้องการเผยแพร่ในวารสารฯ กันมามาก ๆ และขอถือโอกาสนี้เชิญชวนให้เพื่อน ๆ ที่ยังไม่ได้เป็นสมาชิกวารสารฯ นี้ หรือผู้ที่หมดสมาชิกภาพริบสมักรเป็นสมาชิกรับนับแต่บัดนี้

สุวิมล ว่องวานิช

บรรณาธิการผู้ช่วย



# วิธีวิทยาขั้นสูงด้านการวิจัยและสถิติ

นางลักษณ์ วิรัชชัย

ช่วงสองสามทศวรรษที่ผ่านมาเป็นช่วงเวลาซึ่งศาสตร์ทุกสาขามีความเจริญก้าวหน้าอย่างรวดเร็วทั้งในส่วนที่เป็นองค์ความรู้ด้านเนื้อหาสาระ และด้านวิธีวิทยาของศาสตร์ วิธีวิทยาและเนื้อหาสาระของศาสตร์ต่างก็เป็นปัจจัยสำคัญที่เอื้อให้เกิดการพัฒนาซึ่งกันและกัน ยิ่งมีการพัฒนา การสร้างสั่งสมความรู้เนื้อหาสาระมากเพียงใด ยิ่งมีการพัฒนาวิธีวิทยามากเพียงนั้น และในทางกลับกัน ยิ่งมีการพัฒนาวิธีวิทยาให้มีประสิทธิภาพมากเพียงใด ยิ่งมีผลให้เกิดความเจริญก้าวหน้าขององค์ความรู้ในศาสตร์มากขึ้นเพียงนั้น กล่าวได้ว่าเครื่องมือสำคัญในการพัฒนาศาสตร์ทุกสาขาก็คือวิธีวิทยา โดยเฉพาะวิธีวิทยาการวิจัยอันเป็นกระบวนการแสวงหา การสร้าง การสะสม และการพัฒนาองค์ความรู้ด้วยวิธีการทางวิทยาศาสตร์ ด้วยเหตุนี้การเรียนรู้ศาสตร์ต่าง ๆ ในระดับสูงจึงให้ความสำคัญกับการเรียนรู้วิธีวิทยาการวิจัยของศาสตร์ นักวิชาการที่มีคุณภาพในศาสตร์แต่ละสาขานอกจากจะมีความรอบรู้ในหลักการ ทฤษฎี และเนื้อหาสาระของศาสตร์แล้วยังต้องมีความสามารถเชี่ยวชาญในวิธีวิทยาการวิจัย รวมทั้งมีความสนใจติดตามความก้าวหน้าของวิธีวิทยาการวิจัย และมีศักยภาพในการพัฒนาวิธี

เอกสารประกอบการสัมมนาคุณวุฒิบัณฑิตสาขาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เรื่อง “วิธีวิทยาการขั้นสูงด้านการวิจัย สถิติ การวัดและประเมินผลการศึกษา” วันจันทร์ที่ 21 สิงหาคม 2538 ณ อาคารศศปฐศาลา จุฬาลงกรณ์มหาวิทยาลัย

ผู้เขียนขอกราบขอบพระคุณ ศ.ดร.สมหวัง พิธิยานุวัฒน์ ประธานกรรมการบริหารหลักสูตรคุณวุฒิบัณฑิตสาขาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยการศึกษา ผู้ซึ่งเปิดโอกาสและเสนอแนะให้ผู้เขียนบทความนี้ ขอบพระคุณ รศ.ดร.สุวิมล ว่องวาณิช ที่ช่วยจัดหาเอกสารตำรา และช่วยตรวจสอบความคิด ขอขอบคุณนิสิตบัณฑิตศึกษาทุกคนที่เรียนสถิติและวิจัยกับผู้เขียน โดยเฉพาะผู้ที่รักเรียน รักการอ่านและมีส่วนในการติดตามความก้าวหน้าของวิธีวิทยาการวิจัยและสถิติร่วมกับผู้เขียน

วิทยาการวิจัยในศาสตร์ของตน สามารถนำวิธีวิทยามาใช้ในการวิจัยบริสุทธิ์เพื่อพัฒนาศาสตร์ และใช้ในการวิจัยประยุกต์เพื่อพัฒนาบุคคลและสังคมได้อย่างมีประสิทธิภาพ

ด้วยตระหนักในความสำคัญของวิธีวิทยาการวิจัย ผู้เขียนในฐานะผู้สอนวิชาวิจัยและสถิติผู้หนึ่ง จึงได้ติดตามความก้าวหน้าและพยายามมีส่วนร่วมในการพัฒนาวิธีวิทยาการวิจัย รวมทั้งกระตุ้นให้นิสิตบัณฑิตศึกษาได้สนใจศึกษาวิธีวิทยาการวิจัยขั้นสูง บทความนี้เป็นผลจากการประมวลความรู้เกี่ยวกับวิธีวิทยาการวิจัยใหม่ ๆ ที่ได้จากการเรียนรู้ตามหลักสูตรการศึกษาในระบบ จากประสบการณ์ในการวิจัยและจากการแลกเปลี่ยนความรู้ระหว่างผู้สอนในสาขาเดียวกันและนิสิตบัณฑิตศึกษาที่มีจิตวิญญาณของการใฝ่รู้ร่วมกัน สาธารณะในบทความอาจไม่ครอบคลุมวิธีวิทยาการวิจัยขั้นสูงที่ได้รับการพัฒนาใหม่ ๆ ทั้งหมด แต่ผู้เขียนหวังว่าจะมีสารครอบคลุมวิธีวิทยาการขั้นสูงมากพอที่จะทำให้เกิดการปรับปรุงการเรียนการสอนสถิติและวิจัยในระดับบัณฑิตศึกษา การนำเสนอบทความแบ่งเป็นสามตอน ตอนแรกกล่าวถึงความหมายและที่มาของวิธีวิทยาขั้นสูงในการวิจัยการศึกษา ตอนที่สองให้รายละเอียดเกี่ยวกับเนื้อหาสาระของวิธีวิทยาขั้นสูงในด้านการวิจัยและสถิติ และในตอนสุดท้ายเป็นการสรุปรวมลักษณะภาพอนาคตของวิธีวิทยาการวิจัยและสถิติ

### ความหมายและที่มาของวิธีวิทยาการวิจัยขั้นสูง

คำว่าวิธีวิทยาตรงกับศัพท์ภาษาอังกฤษว่า methodology ซึ่งมาจากภาษาละตินว่า methodus + logia และมาจากภาษากรีกว่า methodos (meta + hodos = way) + logie (Webster's Ninth New Collegiate Dictionary, 1991) ตามรากศัพท์วิธีวิทยามีความหมายถึงวิทยาการหรือการศึกษาที่มีระบบเกี่ยวกับวิธีการหรือเทคนิควิธี คำว่า "วิธีวิทยาการวิจัย" จึงมีความหมายครอบคลุมระเบียบวิธีดำเนินการทุกขั้นตอนในการวิจัย ได้แก่ การกำหนดปัญหาวิจัย การศึกษาและรายงานเอกสารที่เกี่ยวข้องกับการวิจัย การกำหนดสมมุติฐานวิจัย การกำหนดกลุ่มประชากรและการเลือกกลุ่มตัวอย่าง การสร้างเครื่องมือวิจัย การรวบรวม การนำเสนอ การวิเคราะห์และการแปลความหมายผลการวิเคราะห์ข้อมูลซึ่งรวมอยู่ในวิธีวิทยาทางสถิติ ตลอดจนเทคนิควิธีการวัดและการประเมินผล

วิธีวิทยาการวิจัยและสถิติที่ใช้ในการวิจัยการศึกษามีทั้งวิธีวิทยาขั้นพื้นฐานและวิธีวิทยาขั้นสูง ดังนั้นเมื่อกล่าวถึงวิธีวิทยาขั้นสูงจึงหมายถึงวิธีวิทยาเฉพาะส่วนที่ไม่ใช่วิธีวิทยาขั้นพื้นฐาน เมื่อศึกษารายวิชาวิธีวิทยาการวิจัย หลักการวัดและประเมินผล และสถิติขั้นพื้นฐาน ในหลักสูตร

ปริญญาโทบัณฑิตสาขาครุศาสตร์หรือศึกษาศาสตร์ จะเห็นได้ว่าวิธีวิทยาการวิจัยขั้นพื้นฐาน มีเนื้อหาสาระครอบคลุมเฉพาะด้านการวิจัยที่เป็นการวิจัยแบบบรรยาย (descriptive research) และการวิจัยทดลอง/การวิจัยกึ่งทดลอง (experimental/quasi-experimental research) ที่มีแผนแบบ (design) การวิจัยง่าย ๆ ไม่ซับซ้อน ด้านการวัดที่ใช้ทฤษฎีการวัดแบบดั้งเดิม (classical test theory) ด้านการประเมินผลเฉพาะหลักการและโมเดลการประเมินเบื้องต้น และในด้านสถิติมีเฉพาะสถิติบรรยาย สถิติว่าด้วยการสุ่มตัวอย่าง และสถิติวิเคราะห์เฉพาะส่วนที่เป็นสถิตินั้นพาราเมตริกบางตัวและสถิติอนุมานเบื้องต้นในการเปรียบเทียบและการศึกษาความสัมพันธ์ระหว่างตัวแปร 2-3 ตัวแปร ดังนั้นวิธีวิทยาการวิจัยนอกเหนือจากสาระที่กล่าวมาแล้วจึงถือกันว่าเป็นวิธีวิทยาการวิจัยขั้นสูง ในรายวิชาสัมมนาหรือวิธีวิทยาขั้นสูงด้านการวิจัย การวัดและประเมินผลและสถิติ จึงมีขอบข่ายของรายวิชาครอบคลุมแผนแบบการวิจัยที่ซับซ้อนมากขึ้น มีเทคนิควิธีการที่จะช่วยให้ได้ผลการวิจัยที่เที่ยงตรงมากขึ้น ด้านการวัดมีสาระเกี่ยวกับทฤษฎีการตอบสนองรายข้อ (item response theory) และการประยุกต์ใช้ ด้านการประเมินผลมีเทคนิควิธีการและโมเดลการประเมินใหม่ ๆ และในด้านสถิติมีเทคนิคการวิเคราะห์ที่ช่วยให้นักวิจัยตอบคำถามวิจัยเกี่ยวกับความสัมพันธ์เชิงสาเหตุ (causal relationship) และวิเคราะห์ข้อมูลด้วยสถิติวิเคราะห์ตัวแปรพหุนาม (multivariate statistics)

เพื่อมิให้เกิดความซ้ำซ้อนกับวิธีวิทยาขั้นสูงด้านการวิจัยและสถิติที่มีอยู่ในรายวิชาตามหลักสูตรระดับบัณฑิตศึกษาในสาขาการศึกษาและสาขาพฤติกรรมศาสตร์และสังคมศาสตร์แล้ว ผู้เขียนจึงจำกัดขอบเขตการเสนอวิธีวิทยาการขั้นสูงในบทความนี้ เฉพาะส่วนที่เป็นวิธีวิทยาการวิจัยใหม่ ๆ ที่ได้รับการพัฒนาในช่วงสองสามทศวรรษที่ผ่านมาและยังมีได้มีการนำไปเพิ่มเติมหรือที่จะได้รับการเพิ่มเติมลงในหลักสูตรเท่านั้น นอกจากนี้เพื่อให้ผู้อ่านได้เห็นแนวโน้มของการพัฒนาวิธีวิทยาการวิจัยขั้นสูง ผู้เขียนจึงนำเสนอที่มาของวิธีวิทยาการวิจัยโดยเน้นเฉพาะสาขาการศึกษาซึ่งมีความเกี่ยวข้องกับสาขาสังคมศาสตร์และพฤติกรรมศาสตร์ก่อนดังนี้

ครุศาสตร์ หรือวิทยาการสาขาการศึกษา เป็นศาสตร์ที่อาศัยวิทยาการด้านปรัชญา จิตวิทยา สังคมวิทยา ประวัติศาสตร์ และเศรษฐศาสตร์ เป็นพื้นฐาน วิธีวิทยาการวิจัยการศึกษาจึงเป็นผลจากบูรณาการวิธีวิทยาการวิจัยของศาสตร์ที่เป็นพื้นฐานเหล่านั้น บูรณาการของวิธีวิทยาเกิดขึ้นในโลกตะวันตกและมีลำดับการบูรณาการพอแยกได้เป็นสามช่วงดังนี้คือ ช่วงแรกคือ ช่วงต้นคริสต์ศตวรรษที่ 20 วิธีวิทยาการวิจัยการศึกษาในช่วงนี้เป็นผลจากการผสมผสาน

วิถีวิทยาการวิจัยของศาสตร์ที่เป็นพื้นฐานการศึกษาโดยตรง เช่น จิตวิทยาการศึกษา และ ประวัติการศึกษา การวิจัยการศึกษาส่วนใหญ่จึงเป็นการวิจัยเชิงประจักษ์ตามแนวจิตวิทยาการศึกษา ให้ความสำคัญกับการวิจัยกึ่งทดลองที่มีจุดมุ่งหมายเพื่อศึกษาวิธีการสอน เทคนิคการสอน และการพัฒนาผู้เรียนที่ทำให้ได้ผลการเรียนการสอนที่มีประสิทธิภาพ ในระยะแรกวิถีวิทยาการวิจัย สนใจเฉพาะอิทธิพลหลัก ในระยะหลังจึงให้ความสำคัญกับอิทธิพล เนื่องจากปฏิสัมพันธ์ระหว่างตัวแปรจัดกระทำและลักษณะผู้เรียน นอกจากนี้มีการวิจัยตามแนวประวัติการศึกษาซึ่งเป็นการวิจัยเพื่อศึกษาเปรียบเทียบระบบการศึกษา และบทบาทของการศึกษาต่อการพัฒนาสังคม

ช่วงที่สอง คือ ช่วงกลางคริสต์ศตวรรษที่ 20 อันเป็นช่วงของการพัฒนาหลักสูตร สหวิทยาการสาขาพัฒนาศึกษาศาสตร์ (development education) ซึ่งให้ความสำคัญกับการนำวิถีวิทยาการวิจัยในสาขาสังคมศาสตร์และพฤติกรรมศาสตร์มาใช้ในจัดการศึกษาเพื่อเป้าหมายการพัฒนาประเทศ นักการศึกษาในช่วงนี้มีความเชื่อว่าการศึกษาและสังคมเป็นสิ่งที่แยกกันไม่ออก การพัฒนาสังคมจะเกิดขึ้นไม่ได้ถ้าไม่มีการพัฒนาการศึกษา และนักศึกษาคควรจะต้องมีความรอบรู้มิใช่เฉพาะศาสตร์ที่เป็นพื้นฐานการศึกษา แต่จะต้องรอบรู้ศาสตร์ทุกสาขาที่เกี่ยวข้องสัมพันธ์กับการศึกษา ดังนั้นเศรษฐศาสตร์หรือวิทยาการสาขาการศึกษาในช่วงนี้จึงมีแนวโน้มที่จะเป็นสหวิทยาการและพหุวิทยาการ มีบูรณาการวิถีวิทยาการวิจัยทางเศรษฐศาสตร์ และสังคมวิทยาทุกแขนงเข้ากับวิถีวิทยาการวิจัยการศึกษามากขึ้น การวิจัยการศึกษาขยายขอบเขตจากการจัดการเรียนการสอนครอบคลุมถึงการจัดการศึกษาทั้งระบบให้มีประสิทธิภาพ เพื่อพัฒนาสังคม วัตถุประสงค์สำคัญในการวิจัยการศึกษาคือ การวางแผนการศึกษา การจัดสรรทรัพยากรที่มีอยู่ให้เกิดประโยชน์สูงสุด และการศึกษาลักษณะความสัมพันธ์ระหว่างบุคคลในองค์กร ลักษณะการวิจัยเน้นเรื่องการใช้ฟังก์ชันผลผลิตทางการศึกษา (educational production function) การวิเคราะห์ต้นทุน-ผลประโยชน์ (cost-benefit analysis) การใช้โมเดลเศรษฐมิติทางการศึกษา (educational econometric model) การศึกษาสถานศึกษา และองค์กรทางการศึกษาเพื่อให้ทราบถึงอิทธิพลของตัวแปรต่างระดับขององค์กรที่มีต่อการจัดการศึกษา และการศึกษาเกี่ยวกับวัฒนธรรมการเมืองและการมีส่วนร่วมในการปกครองและการเมือง เป็นต้น กล่าวได้ว่าความก้าวหน้าของวิถีวิทยาการวิจัยการศึกษาในช่วงนี้เป็นผลมาจากการบูรณาการสังคมศาสตร์และพฤติกรรมศาสตร์เพื่อประโยชน์ในการจัดการศึกษาทั้งระบบเป็นเครื่องมือในการพัฒนาสังคมและประเทศชาติ

ช่วงปลายคริสต์ศตวรรษที่ 20 นับเป็นช่วงเวลาที่นักการศึกษาเปลี่ยนความเชื่อจาก “การศึกษาเป็นเครื่องมือสำคัญในการพัฒนาประเทศ” เป็นความเชื่อที่ว่า “ความสัมพันธ์ระหว่างการศึกษากับสังคมเป็นความสัมพันธ์แบบสองทาง (two way relationship) สภาพของระบบต่าง ๆ ในสังคมมีส่วนสำคัญในการกำหนดลักษณะและรูปแบบการศึกษาในอนาคต และขณะเดียวกันระบบการศึกษามีส่วนในการกำหนดวิถีทางการพัฒนาระบบต่าง ๆ ในสังคม” ด้วยเหตุนี้วิชาการด้านครุศาสตร์จึงมีความเกี่ยวข้องสัมพันธ์กับวิชาการด้านสังคมศาสตร์ พฤติกรรมศาสตร์ มนุษยศาสตร์ และวิทยาศาสตร์มากยิ่งขึ้นกว่าสมัยก่อนซึ่งอาศัยศาสตร์สาขาต่าง ๆ เป็นเพียงพื้นฐาน แนวโน้มของบูรณาการระหว่างครุศาสตร์และศาสตร์สาขาต่าง ๆ มีมากขึ้นเพื่อให้เกิดความสมบูรณ์ในการพัฒนาการเรียนการสอน การพัฒนาบุคคล การจัดการศึกษาทั้งระบบ และการพัฒนาสังคม ผลจากบูรณาการของศาสตร์ทุกสาขากับครุศาสตร์ ประกอบกับความเจริญก้าวหน้าของศาสตร์ และเทคโนโลยีคอมพิวเตอร์ทำให้วิธีวิทยาการวิจัยและสถิติการศึกษา มีความเจริญก้าวหน้ามากยิ่งขึ้น โดยเฉพาะอย่างยิ่ง วิธีวิทยาการวิจัยคุณภาพ การวิจัยด้านจิตมิติ (psychometrics) และเศรษฐมิติ (econometrics) วิทยาการจัดการ (management sciences) ซึ่งก้าวหน้าไปอย่างรวดเร็วและมีบทบาทสำคัญทำให้เกิดวิธีวิทยาขั้นสูงด้านการวิจัยและสถิติทั้งสิ้น

การวิจัยไม่ว่าจะเป็นการวิจัยประเภทใดมีวัตถุประสงค์หลักในการวิจัยอยู่ 4 ประการเรียงตามความเข้มข้นของวิธีวิทยาการวิจัยจากน้อยไปมากดังนี้คือ วัตถุประสงค์เพื่อการสำรวจและบรรยาย (exploration and description) เพื่อการเปรียบเทียบ (comparison) และ/หรืออธิบาย (explanation) ความสัมพันธ์เชิงสาเหตุ เพื่อการทำนาย (prediction) และประเมิน (evaluation) และเพื่อการควบคุม (control) ปรากฏการณ์ให้เกิดการพัฒนา (development) ในทางที่พึงประสงค์ ลักษณะการวิจัยที่ใช้วิธีวิทยาการวิจัยขั้นสูงในช่วงปลายคริสต์ศตวรรษที่ 20 นี้จึงเป็นการวิจัยที่มีวัตถุประสงค์เพื่ออธิบายความสัมพันธ์เชิงสาเหตุเป็นอย่างดี และมีการใช้วิธีวิทยาขั้นสูงที่เป็นผลจากบูรณาการวิธีวิทยาในสังคมศาสตร์ พฤติกรรมศาสตร์ มนุษยศาสตร์ และวิทยาศาสตร์ รวมทั้งต้องอาศัยเทคโนโลยีคอมพิวเตอร์เป็นเครื่องมือสำคัญในการดำเนินการวิจัย เพราะการวิจัยการศึกษาเป็นเรื่องของการศึกษาความสัมพันธ์เชิงสาเหตุระหว่างตัวแปรและสารสนเทศที่มีความหลากหลาย มีตัวแปรจำนวนมาก มีระดับการวัดตัวแปรแต่ละตัวต่างกัน มีหน่วยการวิเคราะห์หลายระดับ และมีความเกี่ยวข้องกับศาสตร์สาขาต่าง ๆ กล่าวได้ว่าวิธีวิทยาการวิจัยการศึกษาขั้นสูงในช่วงปลายคริสต์ศตวรรษที่ 20 นี้เป็นผลจากการขยาย

พรมแดนการวิจัยการศึกษาให้สามารถนำวิธีวิทยาการวิจัยศาสตร์สาขาต่าง ๆ รวมทั้งเทคโนโลยีคอมพิวเตอร์ มาใช้ประโยชน์เพื่อให้ได้ผลงานวิจัยที่เป็นการสร้างและพัฒนาองค์ความรู้ใหม่ทั้งในด้านเนื้อหาสาระ และวิธีวิทยาการวิจัยนั่นเอง

### สาระเกี่ยวกับวิธีวิทยาขั้นสูงด้านการวิจัยและสถิติ

วัตถุประสงค์สำคัญของการวิจัย คือ การตอบคำถามวิจัยได้อย่างถูกต้องแม่นยำ และการอนุมานผลการวิจัยไปใช้ได้อย่างกว้างขวาง ดังนั้น ไม่ว่าวิธีวิทยาการวิจัยขั้นสูงจะได้รับการพัฒนาให้ดีขึ้นเพียงไร ลักษณะสำคัญที่เป็นหัวใจของวิธีวิทยาการวิจัยยังคงมีลักษณะเดิม แต่จะมีบางส่วนที่ได้รับการพัฒนาเพื่อให้การดำเนินการวิจัยสามารถตอบคำถามการวิจัยได้อย่างถูกต้องและแม่นยำมากยิ่งขึ้น ดังนั้นวิธีวิทยาขั้นสูงด้านการวิจัยและสถิติในปัจจุบันจึงเป็นเทคนิควิธีที่มีรากฐานจากวิธีวิทยาขั้นสูงในอดีต แต่มีวิธีการที่ละเอียดซับซ้อนมากขึ้นเพื่อให้ผลการวิจัยมีความตรง 4 ประเภทตามแนวคิดของ Kirk (1995: 16 -17) ได้แก่ ความตรงในการสรุปทางสถิติ (statistical conclusion validity) ความตรงภายใน (internal validity) ความตรงเชิงโครงสร้างของสาเหตุและผล (construct validity of causes and effects) และความตรงภายนอก (external validity) เพิ่มมากยิ่งขึ้นนั่นเอง

วิธีวิทยาขั้นสูงด้านการวิจัยและสถิติที่ใช้ในปัจจุบันมีลักษณะที่สำคัญ 5 ประการ ประการแรก คือ ลักษณะของวิธีวิทยาที่มีการคำนวณอย่างเข้มข้นและต้องมีการใช้คอมพิวเตอร์ช่วยในการปฏิบัติงาน ประการที่สองคือลักษณะที่เป็นบูรณาการจากวิธีวิทยาการวิจัยหลายสาขา ประการที่สามถึงห้า เป็นลักษณะเฉพาะของวิธีวิทยาด้านสถิติ และการวัด ได้แก่ ประการที่สามคือ ลักษณะวิธีวิทยาการวิเคราะห์ที่มีการผ่อนคลายข้อตกลงเบื้องต้น (relax assumptions) ของวิธีวิทยาการวิเคราะห์แบบเดิม ประการที่สี่ คือ ลักษณะวิธีวิทยาการวิเคราะห์ที่ใช้งานได้กว้างขวางมีความทั่วไป (generality) มากขึ้น และประการสุดท้าย คือ ลักษณะวิธีวิทยาการวิเคราะห์ที่มีความถูกต้อง และมีความไว (accuracy and sensitivity) มีความแกร่ง (robust) มากขึ้นกว่าเดิม

วิธีวิทยาขั้นสูงด้านการวิจัยการศึกษาเฉพาะส่วนที่เป็นวิธีวิทยาการวิจัยใหม่ ๆ ซึ่งนำเสนอต่อไปนี้ครอบคลุมวิธีวิทยา 4 ด้าน คือ ด้านการกำหนดแผนแบบการวิจัย ด้านการสุ่มตัวอย่าง ด้านการวัดและการประเมินผล และด้านสถิติ การนำเสนอแต่ละด้านแยกตามหัวข้อเรื่องของเทคนิควิธี สาระที่นำเสนอในแต่ละหัวข้อเป็นเพียงการให้สังกับโดยย่อของแต่ละเรื่องให้ทราบถึง

ความหมาย หลักการ และประโยชน์จากการใช้เทคนิควิธีนั้น ๆ ผู้ที่สนใจอาจศึกษาเพิ่มเติมได้จากเอกสารอ้างอิงที่ระบุไว้ในบทความนี้ต่อไป

## 1. วิธีวิทยาด้านการกำหนดแผนแบบการวิจัย

การพัฒนาวิธีวิทยาด้านแผนแบบการวิจัยเป็นผลจากความพยายามของนักวิจัยที่จะปรับปรุงวิธีวิทยาการวิจัยที่มีอยู่ให้มีมาตรฐาน และมีคุณภาพดีมากขึ้นเพื่อสนองความต้องการของนักวิจัย ผลจากการพัฒนาทำให้เกิดการวิจัยที่มีแผนแบบการวิจัยเปลี่ยนไปจากเดิม การวิจัยหลายแบบพัฒนามาจากวิธีการวิเคราะห์ข้อมูล โดยที่วิธีการวิเคราะห์นั้นมึลักษณะปัญหาวิจัยและวิธีดำเนินการวิจัยเฉพาะตัวแตกต่างจากการวิจัยทั่ว ๆ ไป รูปแบบการวิจัยที่มีการพัฒนาใหม่ที่น่าสนใจมีดังต่อไปนี้

### 1.1 การวิจัยนโยบาย (Policy Research) และการวิจัยเพื่อวางแผนการศึกษา

วิธีวิทยาการวิจัยนโยบาย และการวิจัยเพื่อวางแผนการศึกษานับเป็นวิธีวิทยาการวิจัยที่มีการพัฒนาอย่างรวดเร็ว สถาบันการศึกษาระดับอุดมศึกษาหลายแห่งได้ตั้งศูนย์วิจัยรองรับและเปิดสอนหลักสูตรบัณฑิตศึกษาศาขากการวางแผนและการวิจัยนโยบายการศึกษา เหตุผลสำคัญที่ทำให้มีการพัฒนาวิธีวิทยาการวิจัยประเภทนี้อย่างรวดเร็วก็คือ ความต้องการของผู้บริหารที่จะนำผลการวิจัยไปใช้ในการตัดสินใจ และความต้องการนำผลการวิจัยไปใช้ในการปฏิบัติงานเกี่ยวกับการบริหารการจัดการของบุคลากรซึ่งมีทักษะในการวิจัยของหน่วยงานทั้งภาครัฐและเอกชน

ในระยะแรกของการพัฒนา การวิจัยนโยบายหมายถึงการวิจัยที่ศึกษานโยบาย และผลที่เกิดจากการดำเนินการตามนโยบาย ในระยะหลัง การวิจัยนโยบายมีลักษณะเป็นการวิจัยประเภทสหวิทยาการมีความหมายรวมถึงการวิเคราะห์นโยบาย (policy analysis) ซึ่งเป็นการศึกษาประเมินผลและทำนายผลที่เกิดจากการกำหนดนโยบายแบบต่าง ๆ เปรียบเทียบกัน วิธีดำเนินการวิจัยนโยบายมีขั้นตอนการดำเนินงานแตกต่างจากการวิจัยโดยทั่วไปเล็กน้อย งานขั้นตอนแรกเริ่มจากการกำหนดปัญหาวิจัยจากสภาพที่เป็นปัญหาความเดือดร้อนในสังคม การศึกษารายงานเอกสารที่เกี่ยวข้องเพื่อกำหนดสมมุติฐานที่แสดงถึงความสัมพันธ์ระหว่างตัวแปรสาเหตุที่ได้จากนโยบาย และตัวแปรผลที่เป็นปัญหาสังคม และกำหนดสมมุติฐานที่แสดงถึงผลที่เกิดขึ้นจากการใช้นโยบายที่เป็นทางเลือกใหม่ จากนั้นเป็นขั้นตอนการรวบรวมข้อมูล การวิเคราะห์ข้อมูลและแปลความหมาย ตัวอย่างวิธีการวิเคราะห์ข้อมูลที่ใช้กันมากได้แก่ การ

วิเคราะห์ผลประโยชน์-ต้นทุน และการวิเคราะห์ประสิทธิผล-ต้นทุน โปรแกรมเชิงเส้น การพยากรณ์และการวิเคราะห์อนุกรมเวลา การวิเคราะห์เส้นโค้งลอเรนซ์ และการวิเคราะห์ระบบ ขั้นตอนสุดท้ายเป็นการสรุป เสนอทางเลือกในการดำเนินงานต่อผู้บริหารหรือผู้บริหารโครงการวิจัย นโยบาย ในการดำเนินการวิจัยดังกล่าวนี้ นักวิจัยนโยบายต้องมีความรอบรู้และมีความชำนาญ ในการใช้เทคนิควิธีทางเศรษฐศาสตร์ การเงินสาธารณะ พฤติกรรมองค์กร การสื่อสารและการนิเทศ การวิเคราะห์ระบบ (system analysis) และสถิติ (Stecklein, J.E., 2532: 8; Weimer, D.L. and Vining, A.R., 1992: 2-13)

### 1.2 การวิจัยเชิงจำลอง (Simulation Research)

Creno, W.D. และ Brewer, M.B. (1973: 114-120) กล่าวว่า การวิจัยเชิงจำลอง แบ่งออกเป็น 3 กลุ่มตามระยะเวลาการพัฒนา กลุ่มแรก เป็นการวิจัยที่มีการจำลองเลียนแบบสภาพการณ์จริงตามแนวคิดวิทยาในรูปแบบของการเล่นเกมบทบาทสมมติ (manned simulated role playing game research) ในระยะต่อมาจึงมีการพัฒนาการวิจัยเชิงจำลองในแนวรัฐศาสตร์เพื่อตรวจสอบทฤษฎีเกี่ยวกับความสัมพันธ์ระหว่างประเทศ และระบบต่าง ๆ ในสังคม เรียกว่า simulation research of international relations ซึ่งเป็นงานวิจัยในกลุ่มที่สอง กลุ่มที่สามเป็นการวิจัยเชิงจำลองที่ใช้คอมพิวเตอร์เพื่อศึกษาระบบงาน หรือระบบข้อมูลที่มีลักษณะแตกต่างกันตามข้อตกลงเบื้องต้นในการวิเคราะห์ข้อมูล งานวิจัยในกลุ่มนี้เป็นงานวิจัยด้านคณิตศาสตร์ สถิติ และการวิจัยปฏิบัติการ ในเรื่องการรอคอย (queuing) เท่าที่ผ่านมา การวิจัยการศึกษาในประเทศไทยมีการใช้วิธีวิทยาการวิจัยเชิงจำลองในการศึกษาเปรียบเทียบเทคนิควิธีการวิเคราะห์แบบต่าง ๆ เพื่อให้ได้วิธีการที่มีประสิทธิภาพสูงสุด เทคนิคที่ใช้คือ มอนติคาร์โล และนักวิจัยต้องเขียนโปรแกรมเพื่อสร้างข้อมูลจำลองเอง แต่ในปัจจุบันมีโปรแกรมคอมพิวเตอร์ที่ช่วยให้นักวิจัยสร้างหรือจำลองข้อมูลได้สะดวก เช่น โปรแกรม PRELIS พัฒนาโดย Joreskog, K.G. และ Sorbom, D. (1988) เป็นต้น McLean, J.M. (1978: 329-352) Duke, R.D. (1978: 353-368) คาดว่าในอนาคตจะมีการวิจัยเชิงจำลองมากขึ้น และเสนอแนะให้ใช้เทคนิคการวิจัยเชิงจำลองในการวิจัยอนาคตด้วย

### 1.3 การวิจัยอนาคต (Futures Research)

วิธีวิทยาการวิจัยอนาคตมีจุดมุ่งหมายเพื่อบรรยายทางเลือกอนาคตที่เป็นไปได้แบบต่าง ๆ เพื่อประเมินสถานการณ์ในปัจจุบันและบ่งชี้ผลกระทบที่เป็นไปได้ในแต่ละทางเลือกอนาคต เพื่อเตรียมรับสถานการณ์ที่จะเกิดขึ้นในอนาคต และเพื่อวางแผนอนาคตไปในแนวทางที่พึง

ประสงค์ การวิจัยอนาคตที่ใช้กันในปัจจุบันมีแผนแบบการวิจัยที่แตกต่างไปจากเดิมมาก เนื่องจากมีการประยุกต์เทคนิควิธีการวิจัยแบบอื่น ๆ มาใช้ร่วมกันกับการวิจัยอนาคต ดังจะเห็นได้จากแผนแบบการวิจัยต่าง ๆ ดังนี้

### 1.3.1 การวิจัยอนาคตแบบ Ethnographic Delphi Futures Research (EDFR)

การวิจัยแบบ EDFR เป็นวิธีวิทยาที่ไม่ต้องลงทุนมากเหมาะสำหรับประเทศกำลังพัฒนา และเป็นเทคนิคที่สามารถนำไปประยุกต์ใช้ได้อย่างกว้างขวางทั้งในการกำหนดนโยบาย การตัดสินใจหาวิธีการแก้ปัญหา การใช้ข้อเท็จจริงในอดีตมาแก้ปัญหาอนาคต เป็นต้น เทคนิคการวิจัยเป็นการสร้างภาพอนาคตจากฉันทามติของคณะผู้เชี่ยวชาญโดยใช้เทคนิคเดลไฟ ซึ่งอาจใช้ วิธีการสัมภาษณ์ หรือการใช้แบบสอบถามก็ได้ การเลือกกลุ่มผู้เชี่ยวชาญจึงเป็นเรื่องที่มีความสำคัญมากสำหรับการวิจัยนี้ (จุมพล พูนภัทรชิวิน, 2530) การวิจัยอนาคตแบบ EDFR เป็นผล จากการประยุกต์ใช้วิธีวิทยาการวิจัยชาติพันธุ์วรรณา กับการวิจัยอนาคต ประกอบกับวิธีการรวบรวมข้อมูลด้วยเทคนิคเดลไฟ

### 1.3.2 การวิจัยโดยใช้เทคนิคการวิเคราะห์ผลกระทบไขว้ (Cross-Impact Analysis)

การวิจัยอนาคตที่น่าสนใจอีกแบบหนึ่งได้แก่ การวิจัยโดยใช้เทคนิคการวิเคราะห์ผลกระทบไขว้ Stover, J.G. และ Gordon, T.J. (1978: 301-328) สรุปว่าการวิจัยนี้เป็นประโยชน์ในการพยากรณ์ และการสร้างภาพอนาคตให้เห็นภาพของปฏิสัมพันธ์ระหว่างเหตุการณ์ เห็นความเชื่อมโยงระหว่างเหตุการณ์อย่างต่อเนื่องเป็นสายโซ่ ผลการวิจัยให้ภาพอนาคตที่เป็นความเคลื่อนไหวของปรากฏการณ์ เทคนิคการวิจัยแบบนี้ใช้การประมาณค่าความน่าจะเป็นของการเกิดเหตุการณ์แต่ละเหตุการณ์ และความน่าจะเป็นของการเกิดเหตุการณ์หนึ่ง ๆ เมื่อเกิดเหตุการณ์หนึ่งแล้ว จากนั้นนำมาวิเคราะห์หาค่า odds ratio สำหรับการเกิดเหตุการณ์แต่ละคู่ในเมทริกซ์ผลกระทบไขว้ ผลจากการวิเคราะห์ให้ภาพอนาคตของผลกระทบจากแต่ละเหตุการณ์พร้อมด้วยค่าความน่าจะเป็น เทคนิคนี้ยังสามารถนำไปใช้ตรวจสอบความไวของการดำเนินงานตามนโยบายได้อีกด้วย

### 1.3.3 การวิจัยโดยใช้เทคนิคการสร้างภาพอนาคต (Scenarios)

Wilson, I.H. (1978: 225-248) สรุปว่าภาพอนาคตมีลักษณะเป็นภาพรวมของสภาพการณ์ที่เป็นไปได้ในอนาคต หรือแนวโน้มของปรากฏการณ์ตามเงื่อนไขทางเลือกอนาคตแบบต่าง ๆ ขั้นตอนการสร้างภาพอนาคตประกอบด้วยพยากรณ์การเกิดเหตุการณ์

แต่ละส่วน การวิเคราะห์ผลกระทบไขว้หรือปฏิสัมพันธ์ระหว่างเหตุการณ์ และการเขียนภาพอนาคต ภาพอนาคตที่ได้อาจแสดงในรูปของการบรรยายภาพ หรือแผนภาพอนาคตก็ได้

### 1.3.4 การวิจัยโดยใช้กระบวนการอนาคตปริทัศน์ (Future Scanning Process)

พรซูลี อาชาวาร์จ (2537) ได้สรุปแนวคิดของ Morrison, J.L. Renfro, W.L. และ Boucher, W.I. ว่า กระบวนการอนาคตปริทัศน์ คือการจัดการกับความเปลี่ยนแปลงของสิ่งแวดล้อม โดยใช้เรดาร์ชนิดหนึ่งกวาดดูโลกอย่างมีระบบระเบียบให้ทราบถึงปฏิสัมพันธ์ระหว่างองค์ประกอบต่าง ๆ และเตือนว่าจะเกิดอะไรใหม่ หรือมีการเปลี่ยนแปลงอย่างไร เพื่อให้องค์กรสามารถจัดการกับแนวโน้มคัดสรรที่คาดว่าจะเกิดขึ้น และดำเนินการให้องค์กรพัฒนาไปสู่เป้าหมายที่พึงประสงค์ ขั้นตอนการใช้กระบวนการอนาคตปริทัศน์ ได้แก่ การกำหนด เป้าหมาย การกวาดตรวจหรือปริทัศน์อย่างเป็นทางการจากเอกสาร และการสัมภาษณ์ การหาความตรงของข้อมูล การประชุมที่นักวิจัยเพื่อร่างแนวโน้มที่บ่งบอกความเปลี่ยนแปลง การคัดสรรผู้ทรงคุณวุฒิพิจารณาแนวโน้มตามที่รวบรวมได้ การจัดประเภทของสาระให้เห็นแนวโน้ม ความเปลี่ยนแปลงแต่ละด้าน และการพัฒนาว่าจากแนวโน้มดังกล่าวทำให้เกิดผลกระทบอย่างไร รวมทั้งการทำนาย กับการกำหนดกิจกรรมที่น่าจะเกิดจากแนวโน้มนั้น

## 1.4 การวิจัยเชิงคุณภาพ (Qualitative Research)

การวิจัยเชิงคุณภาพมีจุดมุ่งหมายในการศึกษาปัญหาในแนวลึกมากกว่าแนวกว้าง เป็นการวิจัยเพื่อตอบคำถามว่าทำไมจึงเป็นเช่นนั้น ปรากฏการณ์นั้น ๆ เกิดขึ้นได้อย่างไร มีกระบวนการเปลี่ยนแปลงอย่างไร ผลการวิจัยนำไปสู่การสรุปเป็นทฤษฎีที่เรียกว่า grounded theory วิธีวิทยาการวิจัยเชิงคุณภาพที่ได้รับการพัฒนาในระยะหลังมีแผนแบบการวิจัยแตกต่างกันตามแนวปรัชญาที่นักวิจัยใช้ Tierney, W.G. และ Lincoln, Y.S. (1994) สรุปว่าแผนแบบการวิจัยเชิงคุณภาพในปัจจุบันแตกต่างกันเป็นสามแบบตามแนวปรัชญาหน้าที่นิยม (functionalism) โครงสร้างนิยม (structuralism) และวิพากษ์นิยม (criticalism) นอกจากนี้ การวิจัยเชิงคุณภาพยังมุ่งผสมผสานเทคนิควิธีการวิจัยเชิงปริมาณมากขึ้นเพื่อแก้จุดอ่อนของการวิจัยเชิงคุณภาพ เทคนิควิธีการวิจัยที่ได้รับการพัฒนาที่น่าสนใจมีดังนี้

### 1.4.1 การศึกษาหลายกรณี และการศึกษาหลายพื้นที่ (Multicase and Multisite Studies)

เนื่องจากการศึกษากรณีเดียว หรือพื้นที่เดียว มีข้อจำกัดในด้านความตรงภายนอก

(external validity) นักวิจัยคุณภาพจึงได้พัฒนาวิธีวิทยาให้ดีขึ้นโดยการเลือกกลุ่มตัวอย่างหลายกรณี หรือหลายแหล่งเพื่อให้มีความเป็นตัวแทนของกลุ่มประชากร และเพื่อควบคุมสิ่งที่อาจเป็นเบี่ยงปรากฎการณ์ที่นักวิจัยต้องการศึกษา หลักการเลือกกลุ่มตัวอย่างที่ใช้เป็นหลักการเลือกกลุ่มตัวอย่างแบบแบ่งชั้น (stratified random sampling) โดยใช้ลักษณะที่ต้องการควบคุมเป็นเกณฑ์ในการแบ่งชั้น การวิจัยที่เป็นการศึกษาหลายกรณีให้ความสำคัญกับการเปรียบเทียบเพื่อการสรุปอ้างอิง ส่วนการวิจัยที่เป็นการศึกษาหลายพื้นที่ให้ความสำคัญกับการพัฒนาทฤษฎีที่เป็นจริงในทุกพื้นที่ และต้องใช้กระบวนการอุปมานเชิงวิเคราะห์ (analytic induction) ในการวิจัย ในบางกรณีนักวิจัยอาจกำหนดแผนแบบการวิจัยในรูปของการวิจัยหลายกรณีและหลายพื้นที่ (multisite case studies) ก็ได้ซึ่งทำให้ได้ผลการวิจัยชัดเจนยิ่งขึ้น (Bogdan, R.C. and Biklen, S.K., 1992)

#### 1.4.2 เทคนิคการวิจัยโดยการจัดกลุ่มสนทนา (Focus Group Research)

การวิจัยนี้พัฒนามาจากเทคนิคการเก็บรวบรวมข้อมูลแบบการสัมภาษณ์เป็นกลุ่ม และแบบมีจุดเน้น (group and focus interview) ประกอบกับการเลือกกลุ่มตัวอย่างที่มีลักษณะเฉพาะร่วมกัน วิธีการวิจัยจัดว่าเป็นการวิจัยที่ใช้พลวัตรของกลุ่ม 8-12 กลุ่มกระตุ้นให้สมาชิกของกลุ่มแสดงความคิดเห็นและความรู้สึกของตนการจัดกลุ่มสนทนาใช้หลักการให้สมาชิกภายในกลุ่มแต่ละกลุ่มมีลักษณะร่วมกัน และกลุ่มสนทนาทุกกลุ่มต้องมีลักษณะต่างกันตามเงื่อนไขที่นักวิจัยต้องการเปรียบเทียบสาระที่ได้จากการสนทนา กิจกรรมการดำเนินการจัดกลุ่มสนทนา มีหลักการและวิธีการเฉพาะที่ต้องมีการฝึกฝนก่อนการลงมือทำการวิจัย การวิเคราะห์ข้อมูลต้องมีการถอดเทปบันทึกการสนทนาและวิเคราะห์เนื้อหา (Morgan, D., 1988, นภาพรณ ทะวานนท์, 2535)

#### 1.4.3 การวิจัยเชิงปฏิบัติการอย่างมีส่วนร่วม (Participatory Action Research)

การวิจัยเชิงปฏิบัติการอย่างมีส่วนร่วมเป็นการวิจัยที่มุ่งศึกษาชุมชนโดยเน้นการวิเคราะห์ปัญหา การแก้ปัญหา และการติดตามผลการแก้ปัญหาของชุมชน โดยที่สมาชิกของชุมชนนั้นต้องเข้ามามีส่วนร่วมในการปฏิบัติการ และการวิจัยด้วย (อลิศรา ชูชาติ, 2538) การวิจัยรูปแบบนี้เป็นการผลิตความรู้วิทยาการวิจัยแบบปฏิบัติการกับการสังเคราะห์แบบมีส่วนร่วม ทีมนักวิจัยประกอบด้วยนักวิจัยซึ่งมิใช่สมาชิกของชุมชน กับสมาชิกในชุมชนซึ่งเป็นผู้มีความรู้ความเข้าใจเกี่ยวกับชุมชนดีที่สุดในที่สุด และเป็นผู้ที่จะได้รับประโยชน์จากผลการวิจัยโดยตรง

#### 1.4.4 การวิเคราะห์เนื้อหา (Content Analysis)

Weber, R.P. (1985) กล่าวว่า การวิเคราะห์เนื้อหาเป็นวิธีวิทยาการวิจัยที่ใช้กระบวนการหลายแบบมาสรุปอ้างอิงผลที่ได้จากข้อความ หลักการสำคัญของการวิเคราะห์เนื้อหาคือการจำแนกคำ กลุ่มคำ ประโยคจากข้อความเข้ากลุ่ม จากนั้นจึงจัดกลุ่มนำเสนอข้อค้นพบพร้อมทั้งแปลความหมาย ปัจจุบันนี้เทคนิคการวิเคราะห์เนื้อหาในประเทศไทยทำได้สะดวกมากขึ้นโดยการใช้โปรแกรมคอมพิวเตอร์ Ethnograph ในต่างประเทศ Weber กล่าวว่าเทคนิคการวิเคราะห์เนื้อหามีการพัฒนาโดยการเปลี่ยนข้อมูลเชิงคุณภาพในแต่ละกลุ่มให้เป็นข้อมูลเชิงปริมาณ และนำมาวิเคราะห์ด้วยเทคนิคเชิงปริมาณ เช่น การสร้างแผนภาพ การวิเคราะห์องค์ประกอบ การวิเคราะห์จำแนก และการวิเคราะห์ความสัมพันธ์โครงสร้างเชิงเส้น (Isrel) ด้วย

วิธีวิทยาการวิจัยเชิงปริมาณ และการวิจัยเชิงคุณภาพนับวันแต่จะมีการผสมผสานกันมากขึ้น ในการวิจัยแต่ละเรื่องอาจใช้วิธีวิทยาการวิจัยทั้งเชิงปริมาณและคุณภาพประกอบกันได้ เทคนิควิธีการวิจัยเชิงคุณภาพอีกแบบหนึ่งที่เป็นที่นิยมใช้ คือการจัดสัมมนาผู้เชี่ยวชาญ (connoisseurship) หลังจากเสร็จสิ้นการวิจัยเชิงปริมาณเพื่อนำผลการวิจัยมาตรวจสอบความถูกต้องหรือแสวงหาแนวทางปฏิบัติต่อไปโดยคณะผู้เชี่ยวชาญที่เลือกสรรแล้ว

#### 1.5 การสังเคราะห์งานวิจัยด้วยการวิเคราะห์อภิมาน (Meta-analysis)

การวิเคราะห์อภิมานเป็นวิธีวิทยาการวิจัยที่พัฒนาใหม่ล่าสุดในการสังเคราะห์งานวิจัยเชิงปริมาณ หัวใจสำคัญของการวิเคราะห์อภิมานคือการประมาณค่าผลงานวิจัยให้อยู่ในรูปดัชนีมาตรฐาน และใช้วิธีการทางสถิติตรวจสอบว่าความแปรปรวนในดัชนีมาตรฐานเหล่านั้นมีค่าเหมาะสมที่จะสรุปได้หรือไม่ หากยังไม่สามารถสรุปได้ต้องแยกศึกษาตามตัวแปรปรับ (moderator variables) หรือวิเคราะห์ว่าลักษณะงานวิจัยสามารถอธิบายความแตกต่างในดัชนีมาตรฐานได้อย่างไร (นงลักษณ์ วิรัชชัย, 2530) เทคนิคการวิเคราะห์อภิมานนี้สามารถนำไปใช้วิเคราะห์เพื่อสรุปนัยทั่วไปของความตรง (validity generalization) ได้ ทั้งนี้เพราะค่าความตรงอยู่ในรูปของสัมประสิทธิ์สหสัมพันธ์อันเป็นดัชนีมาตรฐานแบบหนึ่ง เทคนิคการสรุปนัยทั่วไปของความตรงได้รับการพัฒนาอย่างรวดเร็ว Law, K.S., Schmidt, F.L. และ Hunter, J.E. (1994) รายงานไว้ว่าทำได้ถึง 5 แบบ คือ interactive procedure, noninteractive procedure, multiplicative procedure, Taylor series approximation 1 and 2, Raju, Burke, Normand and Langlois procedure ทุกแบบใช้หลักการวิเคราะห์อภิมานตามแบบของ Hunter เป็นพื้นฐานทั้งสิ้น

วิธีวิทยาการวิเคราะห์อภิมานยังจะได้รับการพัฒนาให้ดีขึ้นอีกมากในอนาคต โดยเฉพาะในเรื่องการประมาณค่าดัชนีมาตรฐานสำหรับผลการวิจัยที่นำมาสังเคราะห์แต่ละเรื่อง และเรื่องวิธีการวิเคราะห์ข้อมูล สำหรับการวิเคราะห์ข้อมูลนั้นเนื่องจากโมเดลในการวิเคราะห์อภิมานเป็นโมเดลสอดแทรกเป็นลำดับลดหลั่น (hierarchical nested model) ซึ่งเป็นโมเดลพหุระดับ ดังนั้น Draper, D. (1995) จึงกล่าวว่าวิธีการวิเคราะห์ข้อมูลที่ตีวิธีหนึ่งสำหรับการสังเคราะห์งานวิจัยด้วยการวิเคราะห์อภิมาน คือ วิธีการวิเคราะห์พหุระดับ (multi-level analysis) ผลงานวิจัยที่เป็นการวิเคราะห์อภิมานของ Draper จึงจัดว่าเป็นการวิเคราะห์อภิมานเรื่องแรกที่วิเคราะห์ข้อมูลด้วยการวิเคราะห์พหุระดับ

วิธีวิทยาการสังเคราะห์งานวิจัยด้วยการวิเคราะห์อภิมานที่ได้รับการพัฒนาใหม่อีกแนวทางหนึ่งคือการวิเคราะห์อภิมานของการวิจัยที่เป็นการวิเคราะห์อภิมาน หรือการวิจัยที่ Levin, H.M. เรียกว่า mega-meta analysis และงานวิจัยที่เขายกย่องว่าเป็น “mother of all educational research synthesis” คืองานวิจัยเรื่อง “Toward a Knowledge Base for School Learning” โดย Wang, M.C., Haertel, G.D. Walberg, H.J. (1993) ซึ่งเป็นที่รู้จักกันในชื่อของ “WHW article” งานวิจัยนี้สังเคราะห์งานวิจัยอภิมานรวม 91 เรื่อง ประกอบกับการวิเคราะห์เนื้อหาโดยมีการประมาณค่าขนาดอิทธิพลของตัวแปรกว่า 10,000 ค่า (นงลักษณ์ วิรัชชัย, 2537)

### 1.6 การศึกษาหน่วยตัวอย่างเดียว (Single Subject Study)

การวิจัยที่เป็นการศึกษาหน่วยตัวอย่างเดียวมีลักษณะแตกต่างจากการศึกษารายกรณี เพราะการศึกษารายกรณีเป็นการวิจัยเชิงคุณภาพ ส่วนการศึกษาหน่วยตัวอย่างเดียวเป็นการวิจัยเชิงปริมาณที่มีการเก็บรวบรวมข้อมูลอย่างต่อเนื่อง หรือมีการวัดซ้ำ (repeated measures) เป้าหมายสำคัญของการวิจัยคือการศึกษาคงความเจริญเติบโต ความเปลี่ยนแปลง หรือพัฒนาการของหน่วยตัวอย่างในเรื่องใดเรื่องหนึ่ง วิธีดำเนินการวิจัยเป็นผลจากการผสมผสานวิธีการศึกษาระยะยาว (longitudinal study) กับวิธีการวิเคราะห์ข้อมูลแบบวัดซ้ำ โดยที่ข้อมูลจากการวัดซ้ำของหน่วยตัวอย่างแต่ละคนมีลักษณะเป็นโมเดลลำดับลดหลั่น การวิเคราะห์ข้อมูลที่เหมาะสมสำหรับการศึกษาหน่วยตัวอย่างเดียวคือการวิเคราะห์พหุระดับกรณีโมเดลเชิงเส้นลำดับลดหลั่น (hierachical linear model) หรือโมเดลสัมประสิทธิ์สุ่ม (random coefficient Models) (Rogosa, D. and Saner, H., 1995)

## 2. วิธีวิทยาด้านการสุ่มตัวอย่าง

ด้านการสุ่มตัวอย่างมีการพัฒนาเทคนิควิธีนี้้อยมากเมื่อเปรียบเทียบกับด้านอื่น ๆ เหตุที่เป็นเช่นนี้อาจเนื่องมาจากวิธีการที่ใช้อยู่มีความสมบูรณ์แล้ว สิ่งที่ได้รับการพัฒนาอย่างต่อเนื่องคือการประมาณค่าของขนาดกลุ่มตัวอย่างที่เหมาะสมสำหรับการวิจัย สูตรที่ได้รับการยอมรับว่ามีความถูกต้อง และใช้กันมาก คือสูตรของ Cohen, J. (1988) ซึ่งมีการกำหนดขนาดของกลุ่มตัวอย่างโดยใช้สารสนเทศจากค่าขนาดอิทธิพล ค่าความคลาดเคลื่อนประเภทที่ I และประเภทที่ II ( $\alpha$  และ  $\beta$ ) และสถิติที่ต้องการใช้ในการทดสอบสมมุติฐาน สำหรับวิธีการเลือกกลุ่มตัวอย่างนั้นมีการพัฒนาเทคนิคการสุ่มตัวอย่างแบบเมทริกซ์ (matrix sampling) ให้ดีขึ้น เรียกว่าเทคนิคการสุ่มตัวอย่างแบบ balanced incomplete block (BIB) spiraling การสุ่มตัวอย่างแบบเมทริกซ์เดิมมีจุดอ่อนตรงที่สามารถประมาณค่าสหสัมพันธ์ระหว่างแบบวัดย่อยในแต่ละชุดได้ แต่ไม่สามารถประมาณค่าสหสัมพันธ์ระหว่างแบบวัดย่อยต่างชุดได้ แต่เทคนิค BIB จัดแบบวัดแต่ละชุดให้ประกอบด้วยแบบวัดย่อยที่มีอย่างน้อยหนึ่งแบบวัดย่อยตรงกับแบบวัดย่อยในชุดอื่น หลักการจัดมีลักษณะคล้ายกับ Latin square design (Kaplan, D., 1995; Messick, S., 1984)

## 3. วิธีวิทยาด้านการวัดและประเมิน

ในช่วงสองทศวรรษที่ผ่านมาวิธีวิทยาด้านการวัดและประเมินมีการพัฒนาอย่างรวดเร็ว เนื่องจากการพัฒนามโนทัศน์เกี่ยวกับการวัดและการประเมิน และการพัฒนาด้านเทคโนโลยีคอมพิวเตอร์ จากความเปลี่ยนแปลงของสภาพบริบทในการวัดและการประเมิน และจากความต้องการ ความคาดหวังของผู้ต้องการใช้ผลการวัดและการประเมิน วิธีวิทยาใหม่ ๆ ที่น่าสนใจซึ่งจะนำเสนอต่อไปนี้แยกกล่าวเป็นสองด้านคือ ด้านการวัด และการประเมิน ในแต่ละด้านจะให้ความสำคัญเฉพาะวิธีวิทยาที่เกี่ยวข้องกับมโนทัศน์ของการวัดและการประเมิน ในส่วนที่เป็นวิธีการวิเคราะห์จะนำเสนอในหัวข้อวิธีวิทยาด้านการวิเคราะห์ในหัวข้อที่ 4 ต่อไป

### 3.1 วิธีวิทยาด้านการวัด

องค์กร American Council on Education (ACE) ได้เริ่มจัดพิมพ์หนังสือชื่อ Educational Measurement เพื่อรวบรวมองค์ความรู้ และวิธีวิทยาด้านการวัด โดยเฉพาะการวัดทางการศึกษา ฉบับแรกพิมพ์เมื่อ ค.ศ. 1951 ฉบับที่สองพิมพ์เมื่อ ค.ศ. 1971 และได้ร่วมมือกับ

National Council on Measurement in Education ในการพิมพ์ฉบับที่สามเมื่อ ค.ศ. 1989 Linn, R.L. (1989) ได้เปรียบเทียบหนังสือฉบับที่สองและฉบับที่สามให้เห็นว่านับจาก ค.ศ. 1971 อันเป็นปีที่พิมพ์หนังสือฉบับที่สองนั้นวิธีวิทยาด้านการวัดมีการพัฒนาไปมาก วิธีวิทยาระดับสูงที่น่าสนใจมีดังนี้

### 3.1.1 ทฤษฎีการตอบสนองรายข้อ (Item-Response Theory = IRT)

แม้ว่า Lawley จะได้เสนอโมเดลทฤษฎีการตอบสนองรายข้อไว้ตั้งแต่ ค.ศ. 1943 แล้วก็ตาม แต่ในระยะนั้นไม่มีการนำโมเดลไปใช้ในทางปฏิบัติ มีเพียงการเสนอแนวคิดและหลักการ คริสต์ทศวรรษที่ 1970 นับเป็นช่วงที่เริ่มมีการประยุกต์ใช้ทฤษฎีการตอบสนองรายข้อ และทำให้วิธีวิทยาด้านการวัดมีการพัฒนามากขึ้นอย่างรวดเร็ว โดยเฉพาะในเรื่องการปรับเทียบข้อสอบ (test equating) การทำหน้าที่ต่างกัน (differential item functioning) หรือความลำเอียงของข้อสอบ (test bias) การบริหารการสอบด้วยคอมพิวเตอร์ (computerized test administration) การสร้างมาตรวัด และการหาปกติวิสัย (scaling and norming) เรื่องต่าง ๆ ที่กล่าวมานี้ไม่ใช่เรื่องใหม่ แต่วิธีวิทยาที่ใช้ในเรื่องเหล่านี้ได้รับการพัฒนาใหม่ให้มีประสิทธิภาพดีขึ้นเรื่อย ๆ (Linn, R. L., 1989)

Hambleton, R.K. (1989) สรุปลักษณะของโมเดลการตอบสนองรายข้อที่ได้รับการพัฒนาใหม่ ๆ ในช่วงสองทศวรรษที่ผ่านมาว่า โมเดลการตอบสนองรายข้อนอกจากจะมีโมเดลโลจิสติกแบบเอกมิติหนึ่ง สอง สาม และสี่พารามิเตอร์ และโมเดลไอโจ้วปกติซึ่งใช้กับผลสัมฤทธิ์ทางการเรียน และความถนัดที่เป็นข้อมูลทวิภาคแล้ว ยังมีโมเดลอีกหลายแบบที่นักวัดผลการศึกษาได้พัฒนาจากโมเดลการตอบสนองรายข้อ ซึ่งผู้เขียนได้สรุปจากบทความของ R.K. Hambleton และบทความทางวิชาการใหม่ ๆ ไว้ดังนี้

ก. Nominal Response Model = NRM ของ Bock เป็นโมเดลใช้กับข้อมูลแบบพหุวิภาค มีจุดมุ่งหมายเพื่อประมาณค่าความสามารถของผู้สอบให้มีความถูกต้องสูงสุดโดยใช้สารสนเทศจากโค้งลักษณะการตอบแต่ละตัวเลือก ภายใต้เงื่อนไขว่าผลรวมความน่าจะเป็นในการตอบทุกตัวเลือกของผู้สอบที่ระดับความสามารถที่กำหนดให้มีค่าเท่ากับหนึ่ง ต่อมา Thissen ได้นำแนวคิดนี้ไปพัฒนาต่อโดยกำหนดน้ำหนักสำหรับตัวเลือกแต่ละตัวทำให้การประมาณค่าความสามารถผู้สอบที่ระดับความสามารถต่ำมีความถูกต้องมากขึ้น

ข. Graded-Response Model = GRM หรือ Difference Model ของ Samejima โมเดลนี้ใช้กับมาตรวัดแบบประมาณค่าเช่น มาตรวัดแบบ Likert หรือมาตรวัดแบบนัยจำแนก

และมีการใช้สารสนเทศจากโค้งลักษณะการตอบแต่ละตัวเลือกด้วย การที่โมเดลได้ชื่อว่า difference models เพราะการคำนวณความน่าจะเป็นของผู้สอบที่ตอบตัวเลือกที่  $k$  ได้ถูกต้องนั้นต้องคำนวณจากผลต่างระหว่างความน่าจะเป็นจากตัวเลือกที่ 1 และตัวเลือกที่  $k-1$  นั้นเอง การพัฒนาล่าสุดของ GRM คือ multiplicative Poisson Model = MPM ซึ่งพัฒนาโดย Andrich, D. (1995) การพัฒนานี้ทำขึ้นเพื่อแก้ปัญหาจากกรณีที่ไม่สามารถใช้ unidimensional Rasch model = URM ได้เนื่องจากมีข้อจำกัดเมื่อมีการรวมตัวเลือกการตอบสนองของแบบสอบ

ค. Binomial Trials Model และ Rating Scale Model = RSM เป็นโมเดลที่พัฒนาโดย Andrich เพื่อใช้ในการวัดตัวแปรพหุวิภาค (polychotomous variable) โมเดลที่กล่าวมาข้างต้นยังมีการพัฒนา และการนำไปประยุกต์ใช้เรื่อยมา Fischer, G.H. และ Parzer, P. (1991) ได้ประยุกต์ RSM ใช้ในการวัดคะแนนความเปลี่ยนแปลง (change) ซึ่งทฤษฎีการวัดแบบดั้งเดิมใช้ไม่ได้ผลดีเท่าที่ควร เพราะมีปัญหาเนื่องจากอิทธิพลเพดานและพื้น (ceiling and floor effects) และลักษณะการแจกแจงเบ้ โมเดลที่พัฒนาใหม่เรียกว่า linear rating scale model (LRSM) เพราะพารามิเตอร์ข้อคำถามจะอยู่ในรูปของฟังก์ชันเชิงเส้นของพารามิเตอร์อื่น ๆ วิธีการสร้างมาตร (scaling methods) ที่สามารถใช้กับมาตรวัดแบบ Likert เช่นเดียวกับ RSM ได้แก่ dual scaling ซึ่งพัฒนาโดย Nishisato และ Cheung, K.C. กับ Mooi, L.C. ได้เปรียบเทียบให้เห็นว่าทั้งสองวิธีได้ผลคล้ายคลึงกัน แต่ dual scaling ผ่อนคลายข้อตกลงเบื้องต้นได้มากกว่า RSM

ง. Partial Credit Model = PCM พัฒนาโดย Masters ให้สามารถใช้กับแบบสอบที่มีหลายตัวเลือก และการให้คะแนนตัวเลือกแต่ละตัวแตกต่างกัน โมเดลนี้ได้รับการพัฒนาให้ดีขึ้นโดย Wright และ Masters (1984) นักวัดผลการศึกษาทั้งสองยังได้ประมวลโมเดลโลจิสติกหนึ่งพารามิเตอร์ หรือ Rasch model แบบต่าง ๆ ที่มีการพัฒนาขึ้นรวมเรียกว่า โมเดลการวัดขั้นพื้นฐาน (fundamental measurement model) ซึ่งมีคุณสมบัติสำคัญคือการแยกพารามิเตอร์ผู้สอบ และพารามิเตอร์ข้อคำถาม และมีการใช้สถิติที่พอเพียงในการประมาณค่าพารามิเตอร์ของโมเดล การพัฒนาล่าสุด คือ Muraki, E. (1993) ได้พัฒนาโมเดลรวมค่าพารามิเตอร์ความชัน (slope) ของโค้งลักษณะข้อคำถามแต่ละข้อไว้ในโมเดล เรียกว่า generalized partial credit model = GPCM

จ. Linear Logistic Latent Trait Model พัฒนาโดย Fischer และ Formann กับ Multicomponent Latent Trait Model พัฒนาโดย Embreston จัดว่าเป็นโมเดลที่รวมองค์

ประกอบด้านปัญญา (cognitive component) เข้าเป็นพารามิเตอร์ในโมเดลด้วย นอกจากนี้มี linear logistic model ซึ่ง Fischer พัฒนาโดยกำหนดให้มีพารามิเตอร์ความยากของข้อสอบ ในรูปของผลรวมเชิงเส้นขององค์ประกอบที่มีอิทธิพลต่อค่าความยากอยู่ในโมเดลด้วย ผลงานของ Embreston และ Fischer นับว่าเป็นจุดเริ่มต้นของการผสมผสานแนวคิดทาง cognitive psychology กับ psychometric เข้าด้วยกัน โมเดลของ Fischer ใช้ในการวัดตัวแปรเอกมิติ แต่โมเดลของ Embreston สามารถใช้กับปัญหาทักษะ (cognitive skill) หลายองค์ประกอบ ได้โดยที่แต่ละองค์ประกอบมีความสำคัญต่อการที่ผู้สอบจะตอบข้อคำถามแต่ละข้อได้ถูกต้อง

จ. Unfolding Models พัฒนาโดย Andrich และ Hyperbolic Cosine Unfolding Model พัฒนาโดย Andrich, D. และ Luo, G (1993) เป็นโมเดลที่มีการพัฒนาควบคู่กับ probabilistic models for the cumulative models in pair comparison design ที่ Thurstone ได้วางแนวคิดไว้ โมเดลในกลุ่มนี้ได้แก่ Squared Simple Logistic Model = SSLM พัฒนาโดย Andrich ซึ่งให้ผลการวิเคราะห์เท่าเทียมกับ Simple Hyperbolic Cosine Model = SHCM ของ Andrich และ Lou นอกจากนี้มี two-parameter hyperbolic cosine model = 2PHCM

ข. โมเดลที่พัฒนาจากทฤษฎีการตอบสนองรายข้อเพื่อใช้ในสภาพการณ์ที่ต่างจากการ สอบปกติ ได้แก่ โมเดลที่ Bock, R.D., Mislevy, R. Woodson, C (1982) พัฒนาขึ้นเพื่อใช้ในการสอบที่มีหน่วยการวิเคราะห์เป็นกลุ่ม เช่น ชั้นเรียน โรงเรียน จังหวัด เขตการศึกษา random coefficients multinomial logit model = RCMLM พัฒนาจาก Rasch model โดย Wilson, M. (1995) เพื่อใช้ในการวัดที่แบบสอบมีข้อสอบรวมเป็นชุดข้อสอบหลายชุด (item bundles) และแต่ละชุดเกี่ยวข้องกับคุณลักษณะผู้สอบ และเชื่อมโยงกันด้วยโครงสร้าง ข้อสอบร่วมกัน ตัวคำถามร่วมกัน หรือเนื้อหาข้อสอบร่วมกัน

การใช้โมเดลตอบสนองรายข้อเป็นประโยชน์ต่อการวัดผลการศึกษามาก เมื่อข้อมูล สอดคล้องกับข้อตกลงเบื้องต้นของโมเดลที่ใช้ นักวัดผลจะสามารถประมาณค่าความสามารถ ของผู้สอบได้โดยที่พารามิเตอร์นี้เป็นอิสระไม่ขึ้นกับการเปลี่ยนแปลงของข้อสอบ ได้ค่าพารา- มิเตอร์ของข้อสอบที่ไม่ขึ้นกับกลุ่มตัวอย่างผู้สอบ ได้ค่าสถิติที่บ่งชี้ความถูกต้องในการประมาณ ค่าความสามารถผู้สอบที่ขึ้นอยู่กับค่าความสามารถผู้สอบ จำนวน และคุณสมบัติทางสถิติของ ข้อคำถาม และได้มาตรฐานร่วม (common scale) ซึ่งใช้บรรยายคุณสมบัติผู้สอบและข้อสอบได้

การวิจัยด้านการวัดผลการศึกษาเกี่ยวกับทฤษฎีการตอบสนองรายชื่อนอกจากจะเป็นการพัฒนาโมเดลและตรวจสอบโมเดลแล้ว ยังมีการวิจัยเกี่ยวกับการสร้างมาตรวัดคะแนนความสามารถ (ability scores) ของผู้สอบ มีการพัฒนาคะแนนความสามารถในรูปฟังก์ชันของพารามิเตอร์ความสามารถของผู้สอบรูปต่าง ๆ โดยมีการกำหนดน้ำหนักคะแนนแบบต่าง ๆ และมีการวิจัยเกี่ยวกับการพัฒนาวิธีประมาณค่าพารามิเตอร์ความสามารถแบบต่าง ๆ ด้วย

### 3.1.2 วิธีวิทยาการวัดด้วยคอมพิวเตอร์ (Computerized Educational Measurement)

Bunderson, V.V., Inouye, D.K. และ Olsen, J.B. (1989) ได้สรุปไว้ว่าวิธีวิทยาการวัดด้วยคอมพิวเตอร์เป็นผลจากบูรณาการมโนทัศน์ด้านการวัดกับเทคโนโลยีคอมพิวเตอร์ช่วยให้กระบวนการวัดทำได้สะดวกรวดเร็ว ถูกต้อง และประหยัด รวมทั้งมีการสื่อสารโต้ตอบ มีระบบการเก็บหลักฐานการสอบ และการรายงานคะแนนสอบ และการบริหารการสอบที่มีประสิทธิภาพ วิธีวิทยาการวัดด้วยคอมพิวเตอร์แยกตามขั้นตอนของพัฒนาการได้เป็น 4 ประเภท ประเภทแรก คือ การสอบโดยใช้คอมพิวเตอร์ (computerized testing) การบริหารการสอบทุกขั้นตอน การตรวจให้คะแนน การรายงานผล การสร้างธนาคารข้อสอบและการจัดทำข้อสอบใช้คอมพิวเตอร์ช่วยในการทำงานทุกขั้นตอน แต่การวิเคราะห์ข้อสอบยังใช้ทฤษฎีแบบดั้งเดิม ประเภทที่สอง คือ การสอบปรับเหมาะด้วยคอมพิวเตอร์ (computer-adaptive testing) เป็นการสอบที่มีการวิเคราะห์ตามทฤษฎีการตอบสนองรายข้อ มีการจัดข้อสอบให้เหมาะสมกับผู้สอบแต่ละคน และการบริหารการสอบทุกขั้นตอนทำโดยใช้คอมพิวเตอร์ ประเภทที่สาม คือ การวัดต่อเนื่อง (continuous measurement) เพื่อวัดความเจริญหรือความเปลี่ยนแปลงและสร้างโปรไฟล์ของผู้สอบแต่ละคน คอมพิวเตอร์มีหน้าที่สร้างคะแนนพหุมิติ และแปลความหมายโปรไฟล์ของผู้สอบแต่ละคนเพิ่มขึ้นจากประเภทที่สอง ประเภทที่สี่ คือการวัดอย่างเฉลียวชาญ (intelligent measurement) การวัดประเภทนี้ใช้ประโยชน์จากคอมพิวเตอร์สูงสุด กล่าวคือองค์ความรู้ที่มีอยู่ในสมองของผู้เฉลียวชาญด้านการวัดถูกถ่ายทอดลงในคอมพิวเตอร์ และคอมพิวเตอร์ช่วยในการบริหารการสอบทุกขั้นตอนเสมือนหนึ่งมีผู้เฉลียวชาญคอยให้คำปรึกษา

### 3.1.3 โปรแกรมคอมพิวเตอร์สำหรับการวัดผลการศึกษา

เนื่องจากวิธีวิทยาขั้นสูงด้านการวัดผลการศึกษาส่วนใหญ่เป็นวิธีวิทยาที่ใช้ทฤษฎีการตอบสนองรายข้อเป็นพื้นฐาน และการประมาณค่าพารามิเตอร์ในโมเดลต้องใช้คอมพิวเตอร์ช่วยในการคำนวณ ดังนั้นจึงมีการวิจัยที่เป็นการพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการวัด

ผลการศึกษาเป็นจำนวนมาก Hambleton, R.K. (1989) Baker, F.B. (1989) Cheung, K.C. and Mooi, L.C. ได้ให้รายชื่อโปรแกรมคอมพิวเตอร์ที่เป็นผลจากการวิจัยดังกล่าวไว้สรุปได้ดังนี้

ก. โปรแกรมที่ใช้ในการวิเคราะห์ข้อสอบการประมาณค่าพารามิเตอร์และการให้คะแนน ได้แก่ โปรแกรม LOGIST พัฒนาโดย Wingersky ใช้วิเคราะห์ข้อสอบ และประมาณค่าพารามิเตอร์สำหรับโมเดลโลจิสติกหนึ่ง สอง และสามพารามิเตอร์ โปรแกรม BILOG พัฒนาโดย Mislevy and Bock และ Scientific Software, Inc. ได้พัฒนาให้สามารถใช้ได้กับคอมพิวเตอร์ส่วนบุคคล โปรแกรม BILOG นี้ทำงานคล้ายโปรแกรม LOGIST แต่สามารถประมาณค่าพารามิเตอร์ด้วยวิธี maximum likelihood และวิธี bayesian ได้ โปรแกรม MULTILOG พัฒนาโดย Thissen ทำงานเหมือน BILOG แต่สามารถใช้กับข้อมูลพหุภูมิภาคได้ โปรแกรม BICAL พัฒนาโดย Wright and Stone ใช้สำหรับ Rasch model โปรแกรมนี้ได้รับการพัฒนาในระยะต่อมาเป็นโปรแกรม MICROSCALE โดย Mediix Interactive Technologies, Inc. โปรแกรม NOHARM พัฒนาโดย Fraser ใช้สำหรับการวัดตัวแปรทั้งแบบเอกมิติ และพหุมิติ โปรแกรม ANCILLES และ OGIVA พัฒนาโดย Urry ซึ่งเป็นโปรแกรมที่ใช้วิธีการประมาณค่าแบบ heuristic โปรแกรม BIGSCALE พัฒนาโดย Wright, Linacre and Schultz สามารถวิเคราะห์ Rasch model และ RSM ได้ โปรแกรม DUAL3 พัฒนาโดย Nishisato and Nishisato ใช้วิเคราะห์ dual scaling

ข. โปรแกรมใช้สำหรับการวิเคราะห์ที่จำเป็นก่อนหน้าที่จะใช้โปรแกรมคอมพิวเตอร์ในข้อ ก. ได้แก่โปรแกรม DATAGEN พัฒนาโดย Hambleton and Rovieneli สำหรับสร้างข้อมูลจำลองสำหรับการวิเคราะห์ตามทฤษฎีการตอบสนองรายข้อ โปรแกรม TESTFAC Wilson, D.T., Wood, R. and Gibbons, R. (1991) ใช้คำนวณค่าสถิติของข้อสอบ วิเคราะห์องค์ประกอบข้อสอบ (item factor analysis) และสร้างข้อมูลจำลอง

ค. โปรแกรมสำหรับการบริหารการสอบ โปรแกรมในกลุ่มนี้ทำงานเพียงบางหน้าที่ในกระบวนการบริหารการสอบ ได้แก่ โปรแกรม PRISM พัฒนาโดย Psychological Corporation เพื่อสร้างแบบสอบจากธนาคารข้อสอบ โปรแกรม Microcomputer Scoring System พัฒนาโดย TESTCOR, Inc. เพื่อตรวจให้คะแนน วิเคราะห์ และรายงานผลโดยใช้เครื่องกราดตรวจ (scanner) ซึ่งเป็นเครื่องมือที่จำเป็นสำหรับการสอบที่มีจำนวนผู้สอบมาก โปรแกรม Test Bank พัฒนาโดย Advanced Technology Application เพื่อสร้างธนาคารข้อสอบและจัดทำข้อสอบโดยใช้การประมวลผลค่า นอกจากนี้มี optical scanners พัฒนาโดย National Computer systems, SCAN-TRON Corporation เพื่อใช้สำหรับการตรวจให้คะแนน

ง. ระบบคอมพิวเตอร์ที่ช่วยบริหารการสอบทั้งระบบ ได้แก่ โปรแกรม Microcomputer Test Administration System = MTAS พัฒนาโดย Science Research Associates เพื่อสร้างและพิมพ์ข้อสอบ ตรวจให้คะแนน วิเคราะห์ และรายงานผล โปรแกรมนี้มีข้อจำกัดในการสร้างข้อสอบได้ไม่เกิน 100 ข้อ และไม่มีการสร้างธนาคารข้อสอบ ระบบคอมพิวเตอร์ที่สมบูรณ์ได้แก่ PETA พัฒนาโดย Institute for Practice and Research in Education, School of Education, University of Pittsburgh เป็นโปรแกรมสำหรับครูใช้ในการสอบหนึ่งวิชานักเรียนไม่เกิน 10 ห้อง และห้องละไม่เกิน 50 คน ตัวระบบประกอบด้วยโปรแกรมน้อย 3 โปรแกรม คือ โปรแกรมจัดการเรื่องธนาคารข้อสอบ การสร้างและพิมพ์ข้อสอบ โปรแกรมการตรวจให้คะแนน การวิเคราะห์ และประเมินข้อสอบ และโปรแกรมจัดเก็บข้อมูลผลการสอบของนักเรียน ระบบคอมพิวเตอร์ MicroCAT พัฒนาโดย Assessment Systems Corporation โปรแกรมนี้จัดว่าเป็นระบบคอมพิวเตอร์ที่สมบูรณ์ที่สุดสำหรับการวัดผลการศึกษา เพราะมีโปรแกรมที่ช่วยในการบริหารการสอบทุกขั้นตอนตั้งแต่ การพัฒนาข้อสอบและสร้างธนาคารข้อสอบ ซึ่งรวมทั้งการสร้างแบบสอบสำหรับการสอบปรับเหมาะ (adaptive testing) ด้วยการดำเนินการสอบทั้งการสอบรายบุคคล และการสอบเป็นกลุ่ม การประเมินคุณลักษณะของแบบสอบและผู้สอบด้วยโปรแกรม ITEMAN และการประมาณค่าพารามิเตอร์ด้วยโปรแกรม ASCAL การจัดการในการสอบโดยการเชื่อมต่อกับระบบ LAN (local area network) และการรายงานผล

### 3.2 วิธีวิทยาด้านการประเมิน

วิทยาการด้านการประเมินผลการศึกษามีการพัฒนาเป็นอย่างมากในช่วงปลายคริสต์ศตวรรษที่ 19 เช่นเดียวกับด้านการวัดผลการศึกษา จากบทความของ Scriven, M (1995) แสดงให้เห็นว่าลักษณะการพัฒนาเป็นการปรับแนวคิดทฤษฎีให้การประเมินมีความเป็นศาสตร์ และเป็นวิชาชีพมากขึ้น วิธีวิทยาขั้นสูงในการประเมินที่ได้รับการพัฒนาได้แก่ meta evaluation ซึ่งเป็นการประเมินผลการประเมิน qualitative evaluation ซึ่งเป็นการประเมินที่ใช้แนวคิดของการวิจัยคุณภาพเข้ามาใช้ประโยชน์ เป็นต้น ซึ่งผู้เขียนจะไม่กล่าวถึงรายละเอียดของวิธีวิทยาด้านการประเมินในที่นี้

เนื้อหาสาระที่นำเสนอในหัวข้อที่ 3 ทั้งหมดนี้เป็นเพียงส่วนหนึ่งของวิธีวิทยาขั้นสูงด้านการวัดและการประเมินผลการศึกษาเท่านั้น ยังมีวิธีวิทยาขั้นสูงอีกมากที่มีได้นำเสนอไว้ในที่นี้ เช่น วิธีวิทยาการสร้างตัวบ่งชี้แบบไมเออร์-บริกกส์ (Myer-Briggs type indicator =

MBTI) ตามทฤษฎีจุง (Jung) ในงานวิจัยของ Pittener, D.J. (1993) วิธีวิทยาการสร้างดัชนีวัดประสิทธิผลของโรงเรียนในงานวิจัยของ Webster, W.J., Mendro, R.L. and Almagur, T.O. (1994) เป็นต้น ผู้ที่สนใจ อาจติดตามอ่านได้จากบทความในวารสารฉบับใหม่ ๆ และจากบทความของ รศ. ดร.ศิริชัย กาญจนวาสี เรื่อง วิธีวิทยาขั้นสูงด้านการวัดและการประเมินผลการศึกษา บทความของ ศ. ดร. อุทุมพร จามรมาน เรื่อง ความรู้ชายแดนด้านการวัดผลการศึกษา และบทความของ รศ. ดร. สุวิมล ว่องวาณิช เรื่อง ความรู้ชายแดนด้านการประเมินผลการศึกษา ในวารสารฉบับนี้ได้

#### 4. วิธีวิทยาด้านสถิติ

ความเจริญก้าวหน้าของเทคโนโลยีคอมพิวเตอร์ ประกอบกับการพัฒนาโมเดลทางสถิติ และวิธีการประมาณค่าพารามิเตอร์แบบต่าง ๆ รวมทั้งการบูรณาการวิธีการวิเคราะห์ข้อมูลทางเศรษฐมิติ และจิตมิติทำให้วิธีวิทยาด้านสถิติพัฒนาไปอย่างรวดเร็วในช่วงปลายคริสต์ศตวรรษที่ 19 วิธีวิทยาขั้นสูงด้านสถิติวิเคราะห์ใหม่ ๆ ที่มีใช้ในปัจจุบันล้วนแต่พัฒนามาจากเทคนิคการวิเคราะห์ที่มีอยู่เดิม แต่มีวิธีการที่ดีขึ้น ผ่อนคลายข้อตกลงเบื้องต้นได้มากขึ้น ให้ผลการวิเคราะห์ที่ถูกต้องมากขึ้น และเป็นวิธีการที่ต้องใช้คอมพิวเตอร์ช่วยในการวิเคราะห์ทั้งสิ้น วิธีวิทยาที่น่าสนใจมีดังนี้

##### 4.1 ลิสเรล (Linear Structural Relationship = LISREL )

ลิสเรล มีความหมายเป็น 3 นัย นงลักษณ์ วิรัชชัย (2537) สรุปว่านัยแรกหมายถึงโมเดลแสดงความสัมพันธ์โครงสร้างเชิงเส้นระหว่างตัวแปรซึ่งเป็นที่ตั้งตัวแปรสังเกตได้และตัวแปรแฝง โมเดลลิสเรลประกอบด้วยโมเดลการวัด (measurement model) ของชุดตัวแปรที่เป็นสาเหตุ และชุดตัวแปรที่เป็นผล และโมเดลสมการโครงสร้าง (structural equation model) โมเดลลิสเรลเป็นชื่อที่ใช้กันอย่างแพร่หลาย นอกจากชื่อโมเดลลิสเรลแล้วยังเป็นที่รู้จักกันในชื่อของโมเดลโครงสร้างความสัมพันธ์ (covariance structure model) หรือโมเดลโครงสร้างแสดงสาเหตุ (causal structural model)

ความหมายนัยที่สองหมายถึงภาษาลิสเรลที่ใช้ในการเขียนคำสั่งสำหรับโปรแกรมคอมพิวเตอร์ เพื่อวิเคราะห์โมเดลลิสเรล ภาษาลิสเรลประกอบคำที่มีอักษรภาษาอังกฤษสองตัวใช้แทนคำสั่ง และชื่อเมทริกซ์พารามิเตอร์ที่ต้องการประมาณค่า

ความหมายนัยที่สามหมายถึงโปรแกรมคอมพิวเตอร์สำหรับการวิเคราะห์โมเดลลิสเรล ซึ่งเป็นบูรณาการของการวิเคราะห์องค์ประกอบ (factor analysis) การวิเคราะห์อิทธิพล (path

analysis) และการประมาณค่าพารามิเตอร์ (parameter estimation) ตามหลักวิชาเศรษฐมิติ (econometrics) นักวิจัยใช้การวิเคราะห์ลิสเรลเพื่อพัฒนา และตรวจสอบทฤษฎี ดังนั้นการวิจัยที่ใช้ลิสเรลนั้นนักวิจัยต้องสร้างโมเดลลิสเรลตามทฤษฎีเป็นกรอบความคิด หรือสมมุติฐานวิจัยก่อน กระบวนการวิเคราะห์ลิสเรลประกอบด้วย การกำหนดข้อมูลจำเพาะของโมเดล (model specification) การระบุความเป็นได้ค่าเดียวของโมเดล (model identification) การประมาณค่าพารามิเตอร์ (parameter estimation) ในโมเดล การทดสอบเทียบความกลมกลืน (goodness of fit test) ระหว่างโมเดลลิสเรลตามทฤษฎีและข้อมูลเชิงประจักษ์ เพื่อตรวจสอบความตรงของโมเดล (model validation) การปรับโมเดล (model adjustment) และการสรุปและแปลความหมาย

โปรแกรมคอมพิวเตอร์ลิสเรล พัฒนาโดย Jorekog, K. and Sorbom, D. และจัดจำหน่ายโดย Scientific Software International, Inc. โปรแกรมลิสเรลมีคุณลักษณะทั่ว ๆ ไป (generality) ใช้วิเคราะห์ข้อมูลได้กว้างขวางครอบคลุมวิธีการวิเคราะห์ข้อมูลทางสถิติทุกรูปแบบในการวิจัยทางสังคมศาสตร์ และพฤติกรรมศาสตร์ กล่าวได้ว่า วิธีการทางสถิติที่ต้องใช้วิเคราะห์ด้วยการวิเคราะห์ความแปรปรวน (anova) การวิเคราะห์ความแปรปรวนร่วม (ancova) การวิเคราะห์การถดถอยพหุคูณ (multiple regression analysis) ทั้งตัวแปรเอกนามและพหุนาม การวิเคราะห์องค์ประกอบ (factor analysis) การวิเคราะห์สหสัมพันธ์คาโนนิกอล (canonical correlation) การวิเคราะห์อิทธิพล (path analysis) การวิเคราะห์โมเดลที่มีตัวบ่งชี้หลายตัว การวิเคราะห์โมเดลซิมเพล็กซ์ (simplex model) การวิเคราะห์โมเดลมิมีค (multiple indicators multiple causes model = MIMIC model) การวิเคราะห์โมเดลสองคลื่น (two wave model) การวิเคราะห์โมเดลการวัดคอนเจนเนอริค (congeneric measurement model) และการวิเคราะห์โมเดลหลากหลายลักษณะหลายวิธี (multitrait-multimethod model) ล้วนแต่เป็นกรณีหนึ่งของการวิเคราะห์ด้วยโปรแกรมลิสเรลทั้งสิ้น โปรแกรมลิสเรลจึงเป็นประโยชน์มากทั้งในการวิจัยและการวัดผลทางการศึกษา

คุณสมบัติเด่นอีกประการหนึ่งของโปรแกรมลิสเรลคือการประมาณค่าพารามิเตอร์ด้วยวิธีต่าง ๆ กัน นับตั้งแต่วิธีกำลังสองน้อยที่สุดไม่ถ่วงน้ำหนัก (unweighted least squares = ULS) วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักทั่วไป (generalized least squares = GLS) วิธีโลกลีสูงสุด (maximum likelihood = ML) วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักทั่วไป (generalized weighted least squares = WLS) วิธีกำลังสองน้อยที่สุดถ่วงน้ำหนักแนวทแยง (diagonally

weighted least squares = DWLS) และยังใช้วิธีกำหนดตัวแปรเป็นเครื่องมือ (instrumental variable = IV) กับวิธีกำลังสองน้อยที่สุดสองชั้น (two-stage least squares = TSLS) ในการกำหนดค่าตั้งต้นสำหรับการประมาณค่าด้วยวิธีไลค์ลียูต์สูงสุด

โปรแกรมลิสเรลในปัจจุบันมีการพัฒนาถึงเวอร์ชันที่ 8 ซึ่งพัฒนาให้ใช้กับโปรแกรมวินโดวส์ได้ และการเขียนคำสั่งง่ายสะดวกขึ้น ภาษาที่ใช้ในเวอร์ชันที่ 8 นี้ อาจใช้ได้ทั้งภาษาลิสเรล และภาษาซิมพลิส (SIMPLIS) ซึ่งเป็นภาษาอังกฤษทั่วไป นอกจากนี้โปรแกรมยังมีคำสั่งให้คอมพิวเตอร์สร้างแผนภาพแสดงเส้นทางอิทธิพล (path diagram) ให้ด้วย

โปรแกรมคอมพิวเตอร์ที่ใช้วิเคราะห์โมเดลลิสเรลนอกเหนือจากโปรแกรมลิสเรลแล้วยังมีหลายโปรแกรม เช่น โปรแกรม EQS พัฒนาโดย Bentler, P.M. และจัดจำหน่ายโดย BMDP Statistical Software, Inc. โปรแกรม EzPATH พัฒนาโดย Steiger, J.H. และจัดจำหน่ายโดย SYSTAT, Inc. โปรแกรม LINC พัฒนาโดย Schoenberg, R. และ Arminger, G. และจัดจำหน่ายโดย Aptech Systems, Inc. โปรแกรม LISCOMP พัฒนาโดย Muthen, B. และจัดจำหน่ายโดย Scientific Software International, Inc. โปรแกรม CALIS จัดจำหน่ายโดย SAS, Inc. โปรแกรม PLS พัฒนาโดย Wold, H. โปรแกรม COSAN พัฒนาโดย McDonald, R.P. และโปรแกรม AMOS จัดจำหน่ายโดย Survey Design and Analysis Services Pty Ltd., Small Waters Corporation จากประสบการณ์ของผู้เขียนพบว่าโปรแกรมเหล่านี้ มีโปรแกรม LISREL เวอร์ชันที่ 8 และ AMOS ใช้ง่ายและสะดวก และจากรายงานของ Waller (1993) สรุปการวิเคราะห์ห้อยประกอบเชิงยืนยันจากการวิเคราะห์ด้วยโปรแกรม 7 โปรแกรมยกเว้น COSAN, AMOS และ PLS พบว่าโปรแกรมที่ดีและใช้กันมากคือ LISREL, LISCOMP และ EQS

#### 4.2 โมเดลระดับลดหลั่นเชิงเส้น (Hierarchical Linear Model = HLM) หรือ โมเดลสัมประสิทธิ์สุ่ม (Random Coefficient Model = RCM)

ข้อมูลสำหรับการวิจัยทางการศึกษาโดยทั่วไปเมื่อมีการศึกษาแบบมหภาคจะมีลักษณะเป็นข้อมูลที่สอดแทรกเป็นระดับลดหลั่น (hierarchical nested data) หรือข้อมูลหลายระดับ (multi-level data) ตัวอย่างเช่นข้อมูลจากนักเรียนที่เป็นหน่วยย่อยของห้องเรียน ห้องเรียนที่เป็นหน่วยย่อยของโรงเรียน ในที่นี้หน่วยการวิเคราะห์มี 3 ระดับลดหลั่นกันจากโรงเรียน ห้องเรียน และนักเรียน ลักษณะข้อมูลระดับลดหลั่นนี้หน่วยตัวอย่างแต่ละระดับต้องเป็นหน่วย

ตัวอย่างที่สุ่มจากประชากรแบบทราบค่าความน่าจะเป็น หน่วยตัวอย่างอาจเป็นหน่วยทางภูมิศาสตร์ เช่น หมู่บ้าน ตำบล อำเภอ จังหวัด เขต ภาค เป็นต้น อาจเป็นหน่วยทางการจัดองค์กรทางการศึกษา เช่น นักเรียน ห้องเรียน โรงเรียน กลุ่มโรงเรียน เป็นต้น หรืออาจเป็นหน่วยที่เกิดจากการวัดซ้ำ (repeated measures) ก็ได้

นงลักษณ์ วิรัชชัย (2535) สรุปว่าจุดมุ่งหมายของการวิเคราะห์ข้อมูลหลายระดับ หรือข้อมูลระดับลดหลั่นนี้แบ่งออกได้เป็น 4 ประการคือ ประการแรก เพื่อศึกษาความเปลี่ยนแปลง (change) หรือพัฒนาการ หรือการเจริญเติบโต (growth) ในช่วงเวลาหนึ่ง ๆ วิธีวิทยาที่ใช้ในการวิเคราะห์ได้แก่ การวิเคราะห์ข้อมูลอนุกรมเวลา ประการที่สอง เพื่อประมาณค่าส่วนประกอบความแปรปรวนของตัวแปรแต่ละตัวว่าความแปรปรวนในแต่ละระดับมีความแตกต่างกันมากน้อยเพียงใด จุดมุ่งหมายข้อนี้เป็นประโยชน์ต่อการวิเคราะห์เพื่อศึกษาความเสมอภาคในการจัดการศึกษา วิธีวิทยาที่ใช้ในการวิเคราะห์คือการวิเคราะห์ส่วนประกอบความแปรปรวน (variance component analysis) ประการที่สาม เพื่อศึกษาอิทธิพลของตัวแปรต้นที่มีผลต่อตัวแปรตามในแต่ละระดับ วิธีวิทยาที่ใช้คือ การวิเคราะห์การถดถอยแยกแต่ละระดับ เป็น between unit model และ within unit model การวิเคราะห์ดังกล่าวเป็นผลงานของ Cronbach, Burstein และ Eckland และประการที่สี่ เพื่อศึกษาอิทธิพลของตัวแปรต้นที่มีต่อตัวแปรตามในระดับหน่วยการวัดที่เล็กที่สุด และศึกษาว่าอิทธิพลจากความแตกต่างแต่ละระดับมีผลต่อความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตามอย่างไร วิธีวิทยาการวิเคราะห์ที่กำหนดค่าสัมประสิทธิ์การถดถอยจากการวิเคราะห์ระดับหน่วยการวัดที่เล็กที่สุดให้เป็นตัวแปรสุ่ม และนำมาใช้เป็นตัวแปรตามเพื่อศึกษาว่าได้รับอิทธิพลจากตัวแปรในระดับหน่วยการวัดที่ใหญ่กว่าอย่างไร การกำหนดให้ค่าสัมประสิทธิ์การถดถอยเป็นตัวแปรสุ่มนี้เองเป็นที่มาของชื่อเรียกโมเดลอีกชื่อหนึ่งว่า โมเดลสัมประสิทธิ์สุ่ม (RCM) วิธีการวิเคราะห์นี้พัฒนาโดย Bryk, A และ Raudenbush, S ซึ่งเป็นผู้พัฒนาโปรแกรมคอมพิวเตอร์ HLM ใช้ในการวิเคราะห์ด้วยโปรแกรมนี้จัดทำจำหน่ายโดย Scientific Software International, Inc.

จากจุดมุ่งหมายการวิเคราะห์ทั้งสี่ประการที่กล่าวมาแล้วจะเห็นได้ว่า วิธีวิทยาการวิเคราะห์สามประการแรกเป็นวิธีที่ใช้กันมาก่อน ส่วนวิธีวิทยาการวิเคราะห์โมเดล HLM สำหรับจุดมุ่งหมายประการที่ 4 นั้นนับว่าเป็นวิธีวิทยาขั้นสูงที่ได้รับการพัฒนาใหม่ โปรแกรม HLM รวมวิธีการประมาณค่าพารามิเตอร์ด้วยวิธีของเบย์เชิงประจักษ์ (empirical Bayes = EM) สำหรับสัมประสิทธิ์ในระดับที่ 1 ใช้วิธีกำลังสองน้อยที่สุดตามนัยทั่วไป (generalized least squares

= GLS) สำหรับสัมประสิทธิ์ในระดับที่ 2 และใช้วิธีโลคัลลิตีสูงสุด (maximum likelihood = ML) สำหรับการประมาณค่าส่วนประกอบความแปรปรวน-ความแปรปรวนร่วม (variance-covariance components) ทั้งสองระดับ (Bryk, A., Raudenbush, S.W. and Congdon, R.T., 1994)

การวิเคราะห์โมเดลระดับลดหลั่นเชิงเส้น หรือโมเดลสัมประสิทธิ์สุ่มนี้เป็นประโยชน์ต่อการวิจัยและการสังเคราะห์งานวิจัยมาก Journal of Educational and Behavioral Statistics (1995) ได้จัดทำวารสารฉบับพิเศษเกี่ยวกับเรื่องนี้ ในวารสารดังกล่าวสรุปได้ว่า การวิเคราะห์โมเดลระดับลดหลั่นเชิงเส้นนี้ใช้ประโยชน์ในการศึกษาประสิทธิผลของโรงเรียน (school effectiveness studies) และการศึกษามูลค่าเพิ่ม (value added) ใช้ในการสังเคราะห์งานวิจัยด้วยวิธีการวิเคราะห์ห่อภิมาณ (meta-analysis) และใช้ในการวิเคราะห์ข้อมูลจากการศึกษาระยะยาว (longitudinal study) ได้ สำหรับโปรแกรมคอมพิวเตอร์นั้นนอกจากโปรแกรม HLM ซึ่งสามารถวิเคราะห์ข้อมูลได้ 3 ระดับแล้ว ยังมีโปรแกรม GENMOD พัฒนาโดย Wong and Mason และประมาณค่าด้วยวิธี restricted maximum likelihood = REML มีโปรแกรม ML2 และ ML3 พัฒนาโดย Goldstein และประมาณค่าด้วยวิธี RLML กับวิธี iterative generalized least square = IGLS และมีโปรแกรม VARCL พัฒนาโดย Longford และประมาณค่าด้วยวิธี full maximum likelihood = FML

### 4.3 วิธีวิทยาด้านสถิติสำหรับการวิจัยนโยบายและวางแผนการศึกษา

ความต้องการงานวิจัยนโยบายและวางแผนการศึกษานับวันจะมีมากขึ้น และผู้บริหารโศกงานวิจัยนโยบายได้แก่ผู้บริหารมีความต้องการให้มีการวิจัยที่ใช้วิธีวิทยาการวิเคราะห์ทางสถิติที่ดี และลึกซึ้งมากยิ่งขึ้น ความต้องการดังกล่าวเป็นผลให้มีการพัฒนาวิธีวิทยาด้านสถิติสำหรับการวิจัยดังกล่าว และเกิดวิธีวิทยาขั้นสูงขึ้นหลายแบบทั้งที่เป็นวิธีวิทยาใหม่ และวิธีวิทยาที่เป็นผลจากบูรณาการวิธีวิทยาของศาสตร์ในสาขาที่เกี่ยวข้อง เช่น เศรษฐศาสตร์ วิธีวิทยาที่น่าสนใจมีดังนี้

#### 4.3.1 PERT และ CPM

PERT (Program Evaluation and Review Technique) เป็นเทคนิคที่นำมาใช้ในการวางแผน และบริหารโครงการที่มีกิจกรรมซับซ้อนเชื่อมโยงกันให้เสร็จทันตามกำหนดเวลา โดยเสียค่าใช้จ่ายน้อยที่สุด ขั้นตอนการวิเคราะห์เริ่มต้นจากการระบุกิจกรรมที่ต้องทำในโครงการ

การระบุความเกี่ยวข้องระหว่างกิจกรรม การสร้างข่ายงาน (network) และการประมาณค่าช่วงเวลาที่ใช้ในการทำกิจกรรม 3 แบบ คือ เวลาที่สั้นที่สุด เวลาที่นานที่สุด และเวลาที่ใช้โดยปกติ รวมทั้งเส้นทางวิกฤติ การวิเคราะห์ข้อมูลในระยะหลังรวมถึงการประมาณค่าความน่าจะเป็นที่โครงการจะเสร็จสิ้นสมบูรณ์ และการประมาณค่าต้นทุนการสูญเสียแต่ละกิจกรรมเมื่อกิจกรรมนั้นไม่เสร็จสมบูรณ์ตามเวลาด้วย ส่วน CPM (Critical Path Model) เป็นเทคนิคที่คล้ายคลึงกับ PERT ต่างกันที่ CPM เน้นการวิเคราะห์กรณีที่มีความจำเป็นเร่งด่วนต้องการให้โครงการเสร็จเร็วก่อนกำหนด CPM จะช่วยให้ผู้บริหารโครงการตัดสินใจได้ว่าจะเร่งกิจกรรมใดจึงจะเสียค่าใช้จ่ายเพิ่มขึ้นน้อยที่สุด (Bruno, J.E., 1976)

#### 4.3.2 การวิเคราะห์ผลประโยชน์-ต้นทุน และ การวิเคราะห์ต้นทุน-ประสิทธิผล (Benefit-Cost and Cost-Effectiveness Analyses)

เทคนิคการวิเคราะห์ผลประโยชน์-ต้นทุน เป็นเทคนิคการวิเคราะห์ทางเศรษฐศาสตร์ ที่ถูกนำมาใช้ในการวิจัยทางการศึกษาเพื่อเปรียบเทียบประสิทธิภาพของการดำเนินงานตามนโยบายต่าง ๆ ที่เป็นทางเลือก ขั้นตอนการวิเคราะห์ประกอบด้วย การระบุผลที่เกิดจากการดำเนินการตามนโยบาย การประมาณค่าผลประโยชน์และต้นทุนในการดำเนินงานเป็นตัวเงิน การคิดค่าปัจจุบัน (present values) หรือการประมาณค่าผลประโยชน์และต้นทุนเมื่อหักค่าความเสี่ยง และอัตราเงินเพื่อ ส่วนขั้นตอนสุดท้ายเป็นการเปรียบเทียบอัตราผลตอบแทน (rate of return) ที่ได้รับจากการดำเนินงานแต่ละแบบเพื่อให้ได้รูปแบบที่มีประสิทธิภาพ ในกรณีที่นักวิจัยไม่สามารถประมาณค่าผลประโยชน์ของโครงการในรูปตัวเงินได้ นักวิจัยอาจใช้การวิเคราะห์ต้นทุน-ประสิทธิผลแทน คือศึกษาผลประโยชน์ในรูปของอัตราส่วนของผลผลิตต่อหน่วยปัจจัยการผลิตแทน (Bruno, J.E., 1976; Weimer, D.L. and Vining, A.R., 1992)

ในปัจจุบันเทคนิคการวิเคราะห์ผลประโยชน์-ต้นทุน และการวิเคราะห์ต้นทุน-ประสิทธิผล ได้รับการพัฒนาให้มีประสิทธิภาพมากขึ้น Weimer, D.L. และ Vining, A.R. (1992) เสนอแนะวิธีวิทยาใหม่ว่าในกรณีที่ไม่สามารถวัดผลประโยชน์เป็นตัวเงินได้ ให้ทำการวิเคราะห์ผลประโยชน์คุณภาพ-ต้นทุน (qualitative benefit-cost analysis) แทน โดยพิจารณาจากวัตถุประสงค์ของโครงการและในกรณีที่โครงการมีเป้าหมายหลายอย่างอาจใช้การวิเคราะห์แบบ multigoal analysis และ modified benefit-cost analysis ได้

### 4.3.3 โปรแกรมเชิงเส้น (Linear Programming)

เทคนิคการวิเคราะห์โปรแกรมเชิงเส้นเป็นการสร้างแบบจำลองหรือโมเดลของปรากฏการณ์ในรูปสมการเชิงเส้นทางคณิตศาสตร์จากเป้าหมาย และข้อจำกัดในการดำเนินงาน จากนั้นจึงตรวจสอบ โมเดลและแก้สมการหาคำตอบที่เป็นไปได้ คำตอบพื้นฐาน และคำตอบที่ดีที่สุดภายใต้เงื่อนไขที่เป็นข้อจำกัด (constraints) เทคนิคการวิเคราะห์โปรแกรมเชิงเส้นนี้นอกจากจะสามารถประยุกต์ใช้กับการวางแผนกำลังคน และการวางแผนการศึกษาแล้วยังสามารถใช้กับปัญหาการผลิต ปัญหาการใช้ทรัพยากรต่างชนิด ปัญหาการมอบหมายงาน และปัญหาเกี่ยวกับการลงทุนด้วย (Bruno, J.E., 1976) นอกจากนี้ Marsden, P.V. (1981) ยังเสนอให้ใช้เทคนิคการวิเคราะห์โปรแกรมเชิงเส้นในการวิเคราะห์ความไว (sensitivity analysis) พร้อมทั้งยกตัวอย่างให้เห็นตัวอย่างความคล้ายคลึงระหว่างการวิเคราะห์โมเดลลิสเรลกับการวิเคราะห์โปรแกรมเชิงเส้นด้วย

### 4.3.4 การพยากรณ์ และการวิเคราะห์อนุกรมเวลา (Forecasting and Time Series Analysis)

เทคนิคการพยากรณ์และการวิเคราะห์อนุกรมเวลาแบ่งออกเป็น 2 กลุ่ม กลุ่มแรกเป็นเทคนิคที่มีข้อตกลงเบื้องต้นว่าปรากฏการณ์ในอดีตมีสารสนเทศเพียงพอที่จะพยากรณ์ปรากฏการณ์ในอนาคตได้ เทคนิคการพยากรณ์ในกลุ่มนี้ได้แก่ เทคนิคการทำให้เรียบ (smoothing techniques) ประกอบด้วย เทคนิค moving average, exponential smoothing, autoregressive integrated moving average (ARIMA), integrated moving average (IMA), adaptive filtering multiple time series models ทั้งที่เป็นโมเดลอนุกรมเวลาแบบดั้งเดิมที่ต้องมีการกำหนดรูปแบบโมเดลก่อน และโมเดลอนุกรมเวลาแบบ Box-Jenkins ซึ่งไม่ต้องกำหนดรูปแบบของโมเดลไว้ก่อน กลุ่มที่สองเป็นเทคนิคที่มีข้อตกลงเบื้องต้นว่าปรากฏการณ์ที่จะเกิดขึ้นในอนาคตมีความสัมพันธ์เชื่อมโยงกับปรากฏการณ์อื่น ๆ เทคนิคการพยากรณ์ในกลุ่มนี้ได้แก่ เทคนิคการพยากรณ์ด้วยโมเดลการถดถอย และ โมเดลเศรษฐกิจมิติ (Montgomery, D.C., Johnson, L.A. and Gardiner, J.S., 1990; Makridakis, S. and Wheelwright, S., 1989; Box, G.E.P., Jenkins, G. and Reinsel, G.C., 1994)

การวิเคราะห์อนุกรมเวลาเหมาะสำหรับข้อมูลที่มีการรวบรวมข้อมูลระยะยาวหลายช่วงเวลา แต่ในบางกรณีที่มีการเก็บรวบรวมข้อมูลเพียง 4-5 ครั้งจากกลุ่มตัวอย่างหลายหน่วย

ลักษณะข้อมูลแบบนี้เรียกว่า panel data จำเป็นต้องมีเทคนิคการวิเคราะห์พิเศษ Markus (1979) ได้เสนอให้ใช้โมเดลหลายแบบ กรณีข้อมูลจัดประเภท ใช้ Markov model, Coleman model, Wiggins model และ log-linear model กรณีข้อมูลเป็นตัวแปรต่อเนื่อง ใช้การวิเคราะห์การถดถอย Wiley and Wiley model และ LISREL model

เทคนิควิธีต่าง ๆ ที่ใช้ในการวิจัยนโยบาย และการวางแผนการศึกษาที่กล่าวมาข้างต้นนี้ มีที่ใช้ในการวิจัยของต่างประเทศ ซึ่งมีผู้พัฒนาโปรแกรมคอมพิวเตอร์ใช้ในการวิเคราะห์ข้อมูล และมีผู้เชี่ยวชาญการวิเคราะห์ข้อมูลโดยใช้โปรแกรมคอมพิวเตอร์พื้นฐานเช่น โปรแกรมโลดัส (Judge, G., 1990) แต่เป็นที่น่าเสียดายว่าเทคนิคดังกล่าวมีใช้ในการวิจัยการศึกษาในประเทศไทยไม่มากนัก เหตุผลที่ยังไม่มีใช้แพร่หลายอาจเป็นเพราะเทคนิควิธีดังกล่าวต้องการความรู้พื้นฐานทางคณิตศาสตร์ประการหนึ่ง อีกประการหนึ่งอาจเป็นเพราะนักวิจัยยังอยู่ในระยะการพัฒนาทักษะวิธีวิทยาการวิจัยนโยบาย ซึ่งเป็นเทคนิคในสาขาเศรษฐศาสตร์และบริหารธุรกิจเป็นส่วนใหญ่ก็เป็นไปได้

#### 4.4 การวิเคราะห์ข้อมูลเชิงคุณภาพ (Analyzing Qualitative Data)

ข้อมูลสำหรับการวิจัยทางการศึกษา และการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์ ที่ใช้วิธีการวิเคราะห์ทางสถิติขั้นสูงส่วนใหญ่ เป็นข้อมูลเชิงปริมาณที่มีระดับการวัดตั้งแต่ระดับอันตรภาคขึ้นไป แต่ในความเป็นจริงข้อมูลดังกล่าวมีทั้งข้อมูลเชิงปริมาณและเชิงคุณภาพ นักวิจัยมักจะใช้วิธีสร้างตัวแปรดัมมี่ (dummy variables) ขึ้นแทนตัวแปรเชิงคุณภาพ แล้ววิเคราะห์การถดถอย หรือวิเคราะห์โดยใช้สถิติวิเคราะห์ขั้นสูงอื่น ๆ โดยใช้ข้อตกลงเบื้องต้นว่าตัวแปรดัมมี่เหล่านั้นสอดคล้องกับข้อตกลงเบื้องต้นของสถิติ วิธีการวิเคราะห์ดังกล่าวมีข้อจำกัด และนักสถิติได้พยายามพัฒนาวิธีการวิเคราะห์ข้อมูลเชิงคุณภาพให้ดีขึ้น ผลการพัฒนาทำให้เกิดวิธีวิทยาด้านสถิติใหม่ ๆ ที่เป็นประโยชน์ วิธีวิทยาดังกล่าวมีดังต่อไปนี้

##### 4.4.1 โมเดลล็อก-ลิเนียร์ (Log-linear Model)

การวิเคราะห์โมเดลล็อก-ลิเนียร์ เป็นชื่อรวมเรียกการวิเคราะห์ข้อมูลเชิงคุณภาพด้วยวิธีการทางสถิติเพื่อพัฒนาโมเดลที่สอดคล้องกับข้อมูลเชิงประจักษ์ เพื่อทดสอบสมมุติฐาน และเพื่อประมาณค่าพารามิเตอร์ในโมเดล ลักษณะของโมเดลล็อก-ลิเนียร์จึงครอบคลุมโมเดลความน่าจะเป็นเชิงเส้น และเชิงเส้นโค้ง (linear and nonlinear probability model = LPM and NPM) โมเดลโลจิท (logit model) โมเดลโพรบิท (probit model) โมเดลโลจิสติก

(logistic model) นักสถิติประมาณค่าพารามิเตอร์ในโมเดลด้วยวิธีไลค์ลิตีสูงสุด (maximum likelihood) และวิธีไค-สแควร์ต่ำสุด (minimum chi-square) หลักการวิเคราะห์เป็นแบบเดียวกับ การวิเคราะห์การถดถอยพหุ (multiple regression analysis) มีส่วนที่ต่างกันคือ โมเดล สำหรับการวิเคราะห์การถดถอยเป็นโมเดลเชิงเส้น (linear model) เท่านั้น (Aldrich, J.H. and Nelson, F.D., 1984)

Kennedy, J.J. (1983) อธิบายว่า หลักการวิเคราะห์โมเดลล็อก-ลิเนียร์กรณีตัวแปร ทั้งตัวแปรต้นและตัวแปรตามเป็นตัวแปรจัดประเภท (categorical variables) หรือเป็นข้อมูล เชิงคุณภาพนั้น เปรียบได้กับการวิเคราะห์ด้วยการวิเคราะห์ความแปรปรวน (anova) เมื่อใช้ วิธีการวิเคราะห์การถดถอยผสมกับการตรวจสอบความกลมกลืนด้วยไค-สแควร์ (chi-square goodness of fit) นั้นเอง นงลักษณ์ วิรัชชัย (2537) สรุปว่าจุดเด่นของการวิเคราะห์โมเดล ล็อกลิเนียร์สำหรับข้อมูลเชิงคุณภาพมี 2 ประการคือ ประการแรกผ่อนคลายข้อตกลงเบื้องต้น เกี่ยวกับโมเดลแบบบวก และโมเดลเชิงเส้น ประการที่สองสามารถศึกษาอิทธิพลจากปฏิสัมพันธ์ได้

วิธีวิทยาด้านสถิติเกี่ยวกับโมเดลล็อก-ลิเนียร์ ได้ถูกพัฒนาให้ดีขึ้น และถูกนำไปใช้ ประโยชน์ในหลายด้าน เช่น โมเดลโลจิสติก ได้ถูกนำไปใช้เป็นพื้นฐานในการพัฒนาโมเดลการวัด ที่ใช้กันอยู่อย่างแพร่หลายตามทฤษฎีการตอบสนองรายข้อ โมเดลการวัดโลจิสติกหนึ่งพารา- มิเตอร์ คือ โมเดลที่เรียกว่า Rasch model นั้นเอง (Linn, R.L., 1989) ส่วนโมเดลความน่าจะเป็นเชิงเส้น (LPM) ถูกนำไปใช้ในการวิเคราะห์ข้อมูลเกี่ยวกับความน่าจะเป็นของการเกิด เหตุการณ์ต่าง ๆ (Gujarati, D., 1978)

#### 4.4.2 การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression Analysis)

โมเดลการถดถอยโลจิสติกเป็นที่รู้จักกันอย่างดีในหมู่นักสถิติ แต่เพิ่งจะถูกนำมาใช้ ในการวิจัยสาขาต่าง ๆ ในช่วงคริสต์ทศวรรษ ที่ 1960 โดยมีการเริ่มใช้ในสาขาสาธารณสุข หรือวิทยาศาสตร์สุขภาพก่อน ในระยะหลังเมื่อโปรแกรมคอมพิวเตอร์ SPSS และ SAS มี subroutine ให้เลือกใช้ได้จึงเริ่มมีใช้มากขึ้นในสาขาสังคมศาสตร์และพฤติกรรมศาสตร์ การวิเคราะห์การถดถอยโลจิสติกมีหลักการเหมือนการวิเคราะห์การถดถอย ต่างกันตรงที่มีการ เปลี่ยนรูปตัวแปรตามซึ่งเป็นตัวแปรจัดประเภทโดยใช้โลจิท (logit transformation) และต่างกัน ที่การแปลความหมายสัมประสิทธิ์การถดถอยซึ่งต้องแปลความหมายในรูปของอัตราส่วนของ

แอดัมต่อ (Odds ratio) การวิเคราะห์โมเดลการถดถอยโลจิสติกในระยะหลังสามารถศึกษาอิทธิพลจากปฏิสัมพันธ์ของตัวแปรต้นที่มีต่อตัวแปรตามได้ด้วย (Hosmer, D.W. and Lemeshow, S., 1989; Daniel, W.W., 1995) เนื่องจากการวิเคราะห์โมเดลการถดถอยโลจิสติกสามารถใช้วิเคราะห์ข้อมูลได้เช่นเดียวกับการวิเคราะห์จำแนก (discriminant analysis) การวิเคราะห์การถดถอยโลจิสติกจึงเป็นทางเลือกใหม่อีกทางหนึ่งในการวิเคราะห์ข้อมูล

#### 4.5 สถิติแกร่ง (Robust Statistics)

วิธีวิทยาชั้นสูงด้านสถิติที่น่าสนใจอีกเรื่องหนึ่งคือ สถิติแกร่ง วิธีวิทยานี้เกิดขึ้นเนื่องจากข้อจำกัดของสถิติอนุมานที่ใช้กันอยู่นั้นขึ้นอยู่กับข้อตกลงเบื้องต้นของสถิติเกี่ยวกับลักษณะการแจกแจงประเภทของพารามิเตอร์ สภาพการสุ่ม และความเป็นอิสระ เมื่อข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้นย่อมให้ผลการวิเคราะห์คลาดเคลื่อน และนักสถิติได้พัฒนาสถิติที่แกร่งสำหรับการวิเคราะห์ข้อมูลที่มีลักษณะไม่เป็นไปตามข้อตกลงเบื้องต้น Huber, P.J. (1981) ได้เขียนตำราว่าด้วยสถิติแกร่งแบบต่าง ๆ เช่น การประมาณค่าพารามิเตอร์ด้วยวิธีไลค์ลิวสูงสุด (maximum likelihood type estimates = M-estimates) วิธีผลรวมเชิงเส้นของค่าสถิติแบบเรียงอันดับ (linear combinations of ordered statistics = L estimates) และวิธีประมาณค่าจากการทดสอบเกี่ยวกับตำแหน่ง (estimates derived from rank tests = R estimates) การทำให้เทคนิคกำลังสองน้อยที่สุดมีความแกร่ง (robustizing the least square approach) การหาค่าสหสัมพันธ์แบบแกร่ง และวิธีอื่น ๆ อีกมาก

ผู้อ่านจะเห็นได้ว่าวิธีวิทยาชั้นสูงด้านสถิติที่กล่าวมานี้มีข้อที่น่าสังเกตสองประการ ประการแรก แม้ว่าผู้เขียนจะแยกเสนอวิธีวิทยาด้านสถิติ ด้านการวิจัย ด้านการสุ่มตัวอย่าง ด้านการวัดและประเมินเป็นคนละหัวข้อ แต่วิธีวิทยาด้านสถิติก็มีส่วนเกี่ยวข้องกับวิธีวิทยาด้านการวิจัย การวัดและการประเมินผลด้วย ตัวอย่างเช่น การวิจัยที่มีคำถามวิจัยเกี่ยวกับความสัมพันธ์เชิงสาเหตุ ต้องการการวิเคราะห์อิทธิพลด้วยการวิเคราะห์โมเดลลิสเรล จำเป็นต้องมีแผนแบบการวิจัยที่มีกรอบความคิดในการวิจัยที่มีทฤษฎีสนับสนุน และมีการวัดตัวแปรที่สอดคล้องกับโมเดลลิสเรล จึงจะให้ผลการวิจัยที่มีความตรงสูง วิธีวิทยาด้านสถิติที่มีคุณลักษณะความทั่วไปมากเท่าไรก็สามารถนำไปใช้ในด้านอื่น ๆ ได้มากเท่านั้น เช่น การวิเคราะห์โมเดลลิสเรลก็ถูกนำไปใช้ทั้งในโมเดลการวัด การวิเคราะห์เนื้อหาสำหรับการวิจัยเชิงคุณภาพ และใช้ในการวิเคราะห์โปรแกรมเชิงเส้นด้วย ประการที่สอง วิธีวิทยาด้านสถิติที่น่าสนใจนี้ แม้ว่าจะมีรากฐานมาจากสถิติวิเคราะห์แบบเดิม แต่ได้มีการพัฒนาผสมผสานกับเทคนิควิธีอื่นจนดูเหมือน

เป็นเทคนิคใหม่ เหตุที่เป็นเช่นนี้เพราะผู้เขียนมิได้นำเสนอเทคนิคเดิมที่มีการปรับปรุงใหม่ เช่น การใช้ alternating conditional expectation (ACE) algorithm ในการวิเคราะห์การถดถอย เพื่อให้ได้ค่าสหสัมพันธ์พหุคูณมีค่าสูงสุด ของ De Veaux, R.D. (1990) การวิเคราะห์ความแปรปรวนกรณีพารามิเตอร์ความแปรปรวนวิวิธพันธ์ ของ Alexander, R.A. และ Govern, D.M. (1994) ไว้ในบทความนี้ด้วย ผู้อ่านที่สนใจอาจติดตามอ่านได้จากบทความวิชาการในวารสารเกี่ยวกับวิธีวิทยา

### อนาคตของวิธีวิทยาระดับสูงการวิจัยและสถิติ

เมื่อพิจารณาแนวโน้มการพัฒนาวิธีวิทยาระดับสูงด้านการวิจัยและสถิติที่นำเสนอมาทั้งหมด จะคาดการณ์ได้ว่าในอนาคตจะมีวิธีวิทยาระดับสูงที่มีคุณลักษณะความทั่วไป ใช้วิเคราะห์ได้อย่างกว้างขวางเพิ่มมากขึ้น ความก้าวหน้าของเทคโนโลยีคอมพิวเตอร์ช่วยให้การวิเคราะห์ข้อมูลเป็นไปอย่างสะดวก รวดเร็วและถูกต้องแม่นยำมากขึ้นตามเวลาที่ผ่านไป นักวิจัยจึงควรใช้ประโยชน์จากวิธีวิทยาระดับสูงที่ได้รับการพัฒนาใหม่ ๆ ในการตอบปัญหาวิจัยเพื่อให้ได้งานวิจัยที่มีประสิทธิภาพ และผลการวิจัยเป็นประโยชน์ต่อมวลมนุษยชาติ รวมทั้งช่วยสร้างและพัฒนาองค์ความรู้ทางวิชาการด้วย

จากภาพอนาคตของวิธีวิทยาระดับสูงด้านการวิจัยและสถิติที่กล่าวข้างต้นนี้ แสดงให้เห็นว่าในปัจจุบันมีวิธีวิทยาระดับสูงใหม่ ๆ จำนวนมากที่ยังมิได้ถูกบรรจุลงในหลักสูตรการเรียนการสอนระดับบัณฑิตศึกษา และในอนาคตซึ่งจะยังมีวิธีวิทยาใหม่ เกิดขึ้นอีกมากมาย ถ้าไม่มีการปรับปรุงและพัฒนาหลักสูตรให้ทันกับสภาพการณ์ ย่อมทำให้บัณฑิตจากสถาบันนั้นไม่ทันสมัย ดังนั้นสถาบันการศึกษาระดับอุดมศึกษาจึงควรปรับปรุงหลักสูตรโดยบรรจุวิธีวิทยาใหม่ ๆ นี้ลงในหลักสูตร เช่นอาจบรรจุลงในวิชาเลือก หรือวิชาที่เป็นการศึกษาอิสระ (independent study) หรืออาจจัดกิจกรรมการสนทนาเชิงวิชาการเพื่ออภิปรายถึงประโยชน์ จุดเด่น จุดด้อย ของวิธีวิทยาใหม่ ๆ นอกจากนี้หลักสูตรการศึกษาที่เกี่ยวข้องกับวิธีวิทยาโดยตรงควรจะต้องมีการกระตุ้นให้นักศึกษาระดับบัณฑิตศึกษาได้ทำวิทยานิพนธ์ที่เป็นงานวิจัยเกี่ยวกับวิธีวิทยา เพื่อพัฒนาวิธีวิทยาใหม่ ๆ มาใช้ประโยชน์ด้วย

## เอกสารอ้างอิง

- จุมพล พุณท์ทรวิน (2530). “เทคนิคการวิจัยแบบ EDFR.” *วารสารวิจัยสังคมศาสตร์*, มีนาคม: 34-57.
- นงลักษณ์ วิรัชชัย (2530). “การวิเคราะห์ห่อภิมาณ.” *สารานุกรมศึกษาศาสตร์*. (7, กรกฎาคม-พฤศจิกายน) 110-115.
- นงลักษณ์ วิรัชชัย (2535). “การวิเคราะห์ประมาณค่าส่วนประกอบความแปรปรวน (Analysis of Variance Component Estimation)”, *ข่าวสารวิจัยการศึกษา*. 15(4): 9-14.
- นงลักษณ์ วิรัชชัย (2537). “วิจัยปริทัศน์”, *วิธีวิทยาการวิจัย*. 6(2): 100-106.
- นงลักษณ์ วิรัชชัย (2537). *ความสัมพันธ์โครงสร้างเชิงเส้น (LISREL): สถิติวิเคราะห์สำหรับการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์*. กรุงเทพฯ: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- นภาพรณ์ หะวานนท์ (2535). “การศึกษาความคิดเห็นและทัศนะของคนจากการจัดกลุ่มสนทนา”, *ข่าวสารวิจัยการศึกษา*. 15(4): 15-21.
- พรชูลี อาชาวอรุณ (2537). “กลวิธีวิจัยอนาคต: กระบวนการอนาคตปริทัศน์ (A Future Research Strategy: Future Scanning Proces)”. *วิธีวิทยาการวิจัย*. 6(2): 76-91.
- อลิศรา ชูชาติ (2538). *ชุดวิชาวิจัยทางการศึกษานอกโรงเรียนเล่มที่ 7: การวิจัยเชิงปฏิบัติการอย่างมีส่วนร่วม*. กรุงเทพฯ: กรมการศึกษานอกโรงเรียน กระทรวงศึกษาธิการ.
- Aldrich, J.H. and Nelson, F.D. (1984). *Linear Probability, Logit and Probit Models*. Beverly Hills: Sage Publications.
- Alexander, R.A. and Govern, D.M. (1994). “A New and Simpler Approximation for ANOVA Under Variance Heterogeneity”, *Journal of Educational Statistics*. 19(2): 91-101.
- Andrich, D. (1995). “Models for Measurement, Precision, and the Nondichotomization of Graded Responses”, *Psychometrika*. 60(1): 7-26.
- Andrich, D. (1995). “Further Remarks on Nondichotomization of Graded Responses”, *Psychometrika*. 60(1): 37-46.

- Andrich, D. and Luo, G. (1993). "A Hyperbolic Cosine Latent Trait Model for Unfolding Dichotomous Single-Stimulus Responses", *Applied Psychological Measurement*. 17(3): 253-276.
- Baker, F.B. (1989). "Computer Technology in Test Construction and Processing", *Educational Measurement* (Third Edition). R.L. Linn (ed.). New York: Macmillan Publishing Company
- Bock, R.D., Mislevy, R. and Woodson, C. (1982). "The Next Stage in Educational Measurement", *Educational Researcher*. 11: 4-11.
- Box, G.E.P., Jenkins, G. and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control*. Englewood Cliffs, New Jersey: Prentice-Hall International, Inc.
- Bogdan, R.C. and Biklen, S.K. (1992). *Qualitative Research for Education*. Boston: Allyn and Bacon.
- Bruno, J.E. (1976). *Educational Policy Analysis: A Quantitative Approach*. New York: Crane, Russak & Company, Inc.
- Bryk, A., Raudenbush, S.W. and Congdon, R.T. (1994). *HLM: Hierarchical Linear Modelling with the HLM/2L and HLM/3L Programs*. Mooresville: Scientific Software International, Inc.
- Bunderson, C.V., Inouye, D.K. and Olsen, J.B. (1989). "The Four Generations of Computerized Educational Measurement", *Educational Measurement* (Third Edition). R.L. Linn (ed.) New York: Macmillan Publishing Company.
- Cheung, K.C. and Mooi, L.C. (1994). "A Comparison Between the Rating Scale Model and Dual Scaling for Likert Scales", *Applied Psychological Measurement*. 18(1): 1-13.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, New Jersey: Erlbaum.
- Crano, W.D. and Brewer, M.B. (1973). *Principles of Research in Social Psychology*. New York: McGraw-Hill Book Company.
- Daniel, W.W. (1995). *Biostatistics: A Foundation for Analysis in the Health Sciences*. New York: John Wiley & Sons, Inc.

- De Veaux, R.D. (1990). "Finding Transformations for Regression Using the ACE Algorithm" in *Modern Methods of Data Analysis*. edited by Fox, J. and Long, J.S. Newbury Park: Sage Publications.
- Draper, D. (1995). "Inference and Hierarchical Modelling in the Social Sciences", *Journal of Educational and Behavioral Statistics*. 20(2): 1995.
- Duke, R.D. (1978). "Simulation Gaming." in Fowles, J. (ed.) *Handbooks of Futures Research*. Westport, Connecticut: Greenwood Press.
- Fischer, G.H. and Parzer, P. (1991). "An Extension of the Rating Scale Model with an Application to the Measurement of Change", *Psychometrika*. 56(4): 637-651.
- Fowles, J. ed. (1978). *Handbooks of Futures Research*. Westport, Connecticut: Greenwood Press.
- Gujarati, D. (1978). *Basic Econometric Methods*. New York: McGraw-Hill, Inc.
- Hambleton, R.K. (1989). "Principles and Selected Applications of Item Response Theory", *Educational Measurement*. (Third Edition). R.L. Linn (ed.). New York: Macmillan Publishing Company.
- Hosmer, D.W. and Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley & Sons.
- Huber, P.J. (1981). *Robust Statistics*. New York Wiley & Sons.
- Joreskog, K.G. and Sorbom, D. (1988). *PRELIS: A Preprocessor for LISREL*. Mooresville: Scientific Software, Inc.
- Journal of Educational and Behavioral Statistics*. (1995). 20(2): 109-240.
- Judge, G. (1990). *Quantitative Analysis for Economics and Business: Using Lotus 1-2-3*. New York: Harvester Wheatsheaf.
- Kaplan, D. (1995). "The Impact of BIB Spiraling-Induced Missing Data Patterns on Goodness-of-Fit Tests in Factor Analysis." *Journal of Educational and Behavioral Statistics*. 20(1):69-82.
- Kennedy, J.J. (1983). *Analyzing Qualitative Data*. New York: Praeger Publishers.
- Kirk, R.E. (1995). *Experimental Design: Procedures for the Behavioral Sciences*, (3rd edition). New York: Brooks/Cole Publishing Company.

- Law, K.S., Schmidt, F.L. and Hunter, J.E. (1994). "Nonlinearity of Range Corrections in Meta-Analysis: Test of an Improved Procedure." *Journal of Applied Psychology*. 79(3): 425-38.
- Linn, R.L. (1989). "Current Perspectives and Future Directions", *Educational Measurement* (Third Edition). R.L. Linn (ed.). New York: Macmillan Publishing Company.
- Makridakis, S. and Wheelwright, S. (1989). *Forecasting Methods for Management*, (5th edition). New York: John Wiley & Sons.
- Markus, G.B. (1979). *Analyzing Panel Data*. Beverly Hills: Sage Publications.
- Marsden, P.V. (1981). *Linear Models in Social Research*. Beverly Hills: Sage Publications.
- Masters, G.N. and Wright, B.D. (1994). "The Essential Process in a Family of Measurement Models", *Psychometrika*. 49(4): 529-544.
- McLean, J.M. (1978). "Simulation Modelling." in Fowles, J. (ed.). *Handbooks of Futures Research*. Westport, Connecticut: Greenwood Press.
- Messick, S. (1984). "Response to Changing Assessment Needs: Redesign of the National Assessment of Educational Progress." *American Journal of Education*. 94(1): 90-105
- Montgomery, D.C., Johnson, L.A. and Gardiner, J.S. (1990). *Forecasting and Time Series Analysis* (2nd edition). New York: McGraw-Hill, Inc.
- Morgan, D. (1988). *Focus Group as Qualitative Research*. Beverly Hills, California: Sage.
- Muraki, E. (1993). "Information Functions of the Generalized Partial Credit Models", *Applied Psychological Measurement*. 17(4): 351-363.
- Pittenger, D.J. (1993). "The Utility of the Myer-Briggs Type Indicator", *Review of Educational Research*. 63(4) : 467-488.
- Rogosa, D. and Saner, H. (1995). "Longitudinal Data Analysis Examples With Random Coefficient Models", *Journal of Educational and Behavioral Statistics*. 20(2): 149-170.

- Scriven, M. (1994). "Evaluation as a Discipline", *Studies in Educational Evaluation*. 20: 147-166.
- Stecklein, J.E. (2532). "Policy Research: Nature and Approaches." ในสมหวัง พิธิยานุวัฒน์ และदनัย เทียนพุ่ม (บรรณาธิการ). *การวิจัยเชิงนโยบาย : รูปลักษณะใหม่ของการกำหนดอนาคต*. กรุงเทพฯ สมาคมวิจัยสังคมศาสตร์แห่งประเทศไทย
- Stover, J.G. and Gordon, T.J. (1978). "Cross-Impact Analysis." in Fowles, J. (ed.) *Handbooks of Futures Research*. Westport, Connecticut: Greenwood Press.
- Tierney, W.G. and Lincoln, Y.S. (1994). "At the Forefront: Teaching Qualitative Methods in Higher Education", *The Review of Higher Education*. 17(2): 107-124.
- Wang, M.C., Haertel, G.D. and Walberg, H.J. (1993). "Toward a Knowledge Base for School Learning", *Review of Educational Research*. 63: 249-294.
- Waller, N.G. (1993). "Software Review, Seven Confirmatory Factor Analysis Programs: EQS, EzPATH, LINCOS, LISCOMP, LISREL7, SIMPLIS and CALIS", *Applied Psychological Measurement*. 17(1): 73-100.
- Weber, R.P. (1985). *Basic Content Analysis*. Beverly Hills, California: Sage.
- Webster's Ninth New Collegiate Dictionary*, (1991), Springfield, Massachusetts.
- Webster, W.J., Mendro, R.L. and Almaguer, T.O. (1994). "Effectiveness indices: A 'Value Added' Approach to Measuring School Effect", *Studies in Educational Evaluation*. 20: 113-145.
- Weimer, D.L. and Vining, A.R. (1992). *Policy Analysis: Concepts and Practice*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Wilson, I.H. (1978). "Scenarios." in Fowles, J. (ed.) *Handbooks of Futures Research*. Westport, Connecticut: Greenwood Press.
- Wilson, D.T., Wood, R. and Gibbons, R. (1991). *TESTFAC*. Mooresville: Scientific Software.
- Wilson, M. and Adams, R. (1995). "Rasch Models for Item Bundles", *Psychometrika*. 60(2): 181-198.

# วิธีวิทยาขั้นสูงด้านการวัดและประเมินผลการศึกษา

ศิริชัย กาญจนวาสี

การพัฒนาทำให้เกิดความเจริญงอกงาม ความมั่งคั่งที่เกิดขึ้นเราทุกคนคาดหวังให้เป็น การเปลี่ยนแปลงไปในทิศทางที่พึงปรารถนา แต่ในขณะที่เดียวกันการเปลี่ยนแปลงย่อมมีผลกระทบทั้งทางบวกและทางลบต่อสังคม การเปลี่ยนแปลงที่เกิดจากการวางแผนและปฏิบัติชอบ ย่อมก่อให้เกิดความเจริญก้าวหน้า แต่ผลที่เกิดจากการเปลี่ยนแปลงอาจทำให้ปัญหาเปลี่ยนรูปแบบใหม่ จึงจำเป็นต้องการแนวคิดและวิธีการแก้ปัญหาแบบใหม่ที่แตกต่างไปจากเดิม

ขณะนี้เราเข้าสู่สังคมโลกยุคใหม่ ที่แต่ละสังคมถูกเชื่อมโยงเป็นสังคมโลกเดียวกันใน ด้านสารสนเทศ กระแสของข่าวสารและเทคโนโลยีเข้ามามีส่วนในการผลักดันการเคลื่อนไหว ของโลกอย่างรวดเร็ว และรุนแรง ซึ่งจะช่วยย่นย่อระยะเวลาของการสร้างสมองค์ความรู้ของ ศาสตร์ทุกแขนง ให้มีอัตราเร่งของความก้าวหน้าอย่างไม่เคยเป็นมาก่อน การพัฒนาองค์ความรู้ ของศาสตร์สาขาต่าง ๆ ได้เข้าสู่กระแสของการสื่อสารและใช้เทคโนโลยีเข้ามาเป็นเครื่องมือใน การศึกษาค้นคว้าวิจัย ซึ่งเป็นสิ่งที่เราจะต้องรับรู้เร่งความรู้เท่าทัน ปรับตัวและใช้ประโยชน์ใน ขณะที่ของเก่าที่ดีทรงคุณค่า จะต้องรักษาไว้ให้ได้และพัฒนาให้เหมาะสมกับสภาพการ

วิธีวิทยาทางด้านการวัดและประเมินผลการศึกษา ได้มีการศึกษามาเป็นเวลานาน และ เริ่มศึกษากันมาตั้งแต่ศตวรรษที่ 19 ซึ่งในยุคนั้นได้มีการเสนอทฤษฎีการทดสอบแบบคลาสสิก (classical test theory)  $(X=T+E)$  ก็ยังเป็นที่ยอมรับและยังใช้อย่างกว้างขวาง นับถึงปัจจุบัน นี้ ทฤษฎี classical มีอายุเกือบจะครบหนึ่งร้อยปีในอีกไม่กี่ปีข้างหน้า

บทความที่ถอดคำบรรยายจากการนำเสนอในการสัมมนาคุษฎีบัณฑิตสาขาการวัดและการประเมินผลการศึกษา ภาค วิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย “เรื่องวิธีวิทยาขั้นสูงด้านการวิจัย สถิติ การวัดและประเมินผลการศึกษา” วันจันทร์ที่ 21 สิงหาคม 2538 ณ อาคารศปฐศาลา จุฬาลงกรณ์มหาวิทยาลัย

แต่เนื่องจากข้อจำกัดบางประการของทฤษฎี classical จึงทำให้มีการพัฒนาปรับปรุงวิธีวิทยาด้านการวัดและประเมินผลการศึกษาเรื่อยมา รูปธรรมที่ปรากฏชัดคือในปี ค.ศ. 1980 ได้มีการประยุกต์ทฤษฎีการทดสอบแนวใหม่ ซึ่งเป็นทฤษฎีการวัดที่สามารถอธิบายความสัมพันธ์ระหว่างคุณลักษณะภายใน (trait) หรือความสามารถที่มีอยู่ภายในตัวบุคคล (true ability) กับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงใด

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) นี้มีพื้นฐานความเชื่อว่า พฤติกรรมการตอบสนองต่อข้อสอบของผู้สอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรงว่าถูกหรือผิด จะถูกกำหนดโดยคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคล ซึ่งเป็นสิ่งที่ไม่สามารถสังเกตได้โดยตรง

ทฤษฎี IRT นี้ จะอธิบายความสัมพันธ์ของคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคล กับพฤติกรรมการตอบสนองข้อสอบในรูปของฟังก์ชันทางคณิตศาสตร์ (ฟังก์ชันโลจิสติก ฟังก์ชันปกติสะสม) ซึ่งเรียกว่า โมเดลการตอบสนองข้อสอบซึ่งมีหลายรูปแบบ เช่น โมเดล 1, 2, 3 พารามิเตอร์ เป็นต้น

ทฤษฎี IRT นับได้ว่าเป็นวิธีวิทยาขั้นสูงของการวัดผล ทั้งนี้เพราะการวัดตามทฤษฎี IRT นี้สามารถประมาณค่าคุณลักษณะภายในของมนุษย์ หรือความสามารถของมนุษย์ได้อย่างน่าเชื่อถือ และสามารถอธิบายเชื่อมโยงกับพฤติกรรมภายนอกของมนุษย์ได้ ซึ่งเป็นสิ่งที่นักวัดผลทั้งหลายควรจะทำการศึกษาทฤษฎีนี้ให้เข้าใจอย่างลึกซึ้ง ทั้งนี้เพราะทฤษฎีทางการทดสอบที่จะเกิดขึ้นใหม่ ๆ ก็จะต้องเข้ามาเชื่อมโยงกับทฤษฎีดังกล่าวนี้ รวมทั้งการประยุกต์ IRT ในการศึกษา การทำหน้าที่ต่างกันของผู้สอบ (Differential Item Functioning; DIF) การเปรียบเทียบประสิทธิภาพระหว่างข้อสอบและแบบสอบ เป็นต้น ประกอบกับความก้าวหน้าของเทคโนโลยีด้านคอมพิวเตอร์ ก็จะมีส่วนสำคัญอย่างยิ่งต่อการพัฒนาและประยุกต์ทฤษฎีการวัดและประเมินผลทางด้านสังคมศาสตร์ คอมพิวเตอร์จะเข้าไปมีบทบาทในการประมาณสิ่งที่เป็นคุณลักษณะภายในของมนุษย์ได้ใกล้เคียงกับความเป็นจริงมากที่สุด ถึงแม้ว่าปัจจุบันจะยังไม่สามารถบอกได้ว่าถูกต้องที่สุดก็ตาม

ปัญหาที่เป็นพื้นฐานที่สุดที่ทำให้ศาสตร์ด้านการวัดและประเมินผลทางสังคมศาสตร์ ก้าวหน้าไปได้ช้ากว่าการวัดทางวิทยาศาสตร์ ซึ่งส่วนใหญ่เป็นการวัดทางด้านกายภาพ ก็คือปัญหาในเรื่อง validity ของการวัด trait ที่เราสนใจ ปัญหา validity ของการวัด เป็นปัญหา

ที่พื้นฐานที่สุด สำคัญที่สุด และยากที่สุด การวัดที่ดีจะต้องมีทฤษฎีที่ดีที่เป็นตัวนำว่าสิ่งที่เราต้องการวัดคืออะไร ประกอบด้วย component อะไรบ้าง และแต่ละ component มีอะไรเป็นตัวบ่งชี้ เพื่อนำไปสู่การสร้างข้อคำถามของแต่ละตัวบ่งชี้ เมื่อนำมาสร้างเป็นเครื่องมือแล้ว จะทำการทดสอบ validity ของเครื่องมือนั้นได้อย่างไร

ในการทดสอบ validity นั้น ในแนวของ classical validity จะทำการตรวจสอบ content validity, concurrent validity, predictive validity และ construct validity การตรวจสอบ validity ที่สำคัญที่สุดก็คือ construct validity หรือการตรวจสอบว่ามีความตรงตามทฤษฎีหรือไม่ เทคนิคที่นำมาใช้ในการตรวจสอบ construct validity ทาง classical ใช้วิธี Multitrait Multimethod (MTMM) หรือใช้ Exploratory Factor Analysis (EFA) สำหรับการตรวจสอบขั้นสูงควรใช้ confirmatory factor analysis (CFA) software ที่มีใช้อยู่ในปัจจุบัน ได้แก่ LISREL, LISCOM, EQS เป็นต้น

เทคนิค confirmatory factor analysis นี้ เป็นเทคโนโลยีของการวิเคราะห์ทางสถิติที่ดีที่สุดในปัจจุบัน ที่สามารถยืนยันได้ว่าเป็น construct validity ตามทฤษฎีหรือไม่ ทั้งนี้เพราะ construct validity ตามวิธีการนี้สามารถเชื่อมโยงไปถึงคุณลักษณะภายในของมนุษย์ได้ ถึงแม้ว่าปัจจุบันเรายังไม่แน่ใจว่าเป็นคุณลักษณะที่มุ่งวัดแท้จริงหรือไม่ แต่ถ้ามีการศึกษาและ confirm ด้วยเทคนิคนี้บ่อย ๆ ก็น่าจะยืนยันความเชื่อในความตรงของการวัดนั้นน่าจะมีความถูกต้องเหมาะสมมากขึ้น

นอกจากวิธีวิทยาการตรวจสอบความตรงแล้ว วิธีการศึกษาความเที่ยงทั่วไปเพื่อระบุถึงค่าความเที่ยงภายใต้สภาพการวัดต่าง ๆ กัน ก็เป็นสิ่งที่น่าสนใจมิใช่น้อย จากทฤษฎีการสรุปอ้างอิงทางการทดสอบ (Generalizability Theory; GT) สามารถที่จะช่วยให้นักวัดผลใช้ข้อมูลจากกลุ่มตัวอย่าง สรุปอ้างอิงถึงนัยทั่วไปของค่าความเที่ยงของการวัดภายใต้เงื่อนไขของการวัดต่าง ๆ กัน อันเป็นจุดอ่อนของทฤษฎี Classical Reliability เพื่อตอบคำถามว่าความเที่ยงควรจะเป็นเท่าไรเมื่อสภาพการวัดมีความหลากหลายในระดับต่าง ๆ กันของเงื่อนไขการวัด เช่น ความยาวของข้อสอบ ลักษณะเนื้อหาที่มุ่งวัด จำนวนครั้งของการสอน วิธีการวัด จำนวนผู้ตรวจให้คะแนน ระดับการวัด เป็นต้น ซึ่งจะช่วยนักวัดผลในการตัดสินใจทางการทดสอบว่า ถ้าต้องการใช้เครื่องมืออย่างมีคุณภาพ ควรจะใช้ภายใต้สถานการณ์การวัดลักษณะใด แนวทางของทฤษฎีนี้จะนำไปสู่การสรุปอ้างอิงนัยทั่วไปของความตรงต่อไปได้

แนวโน้มของการวัดผลในอนาคต สังคมเราจะต้องมีการสอบมากขึ้น และบ่อยขึ้น โดยจะยังคงใช้ concept ของ formative evaluation และ summative evaluation ปัญหาสำคัญอีกอย่างหนึ่งของการวัดผลจึงอยู่ที่การสร้างแบบสอบ ซึ่งปัญหาที่เป็นอยู่ในปัจจุบันก็คือ “การสร้างที่ซ้ำที่” และไม่ค่อยจะได้วิเคราะห์ข้อสอบกัน แนวทางการแก้ไขปัญหานี้ก็คือการนำเทคโนโลยีมาช่วยในการสร้างข้อสอบที่ดีขึ้นมาครั้งเดียว แล้วสามารถทำการ generate ข้อสอบที่ดีมาใช้ได้บ่อย ๆ ขึ้น

ในการสร้าง item generator นั้น จะต้องทำการวิเคราะห์ construct ของสิ่งที่ต้องการวัดออกมาก่อนว่าประกอบด้วย component อะไรบ้าง และแต่ละ component จะประกอบด้วยตัวบ่งชี้อะไรบ้าง เมื่อกำหนด Test Specification และ Item Specification และสร้าง Item Forms แล้วใช้คอมพิวเตอร์ช่วยในการ generate item ออกมา ซึ่งก็จะได้ข้อสอบที่มีลักษณะคล้าย ๆ กัน จัดเป็นชุดของแบบสอบเพื่อวัดตามลักษณะของตัวแปรที่ต้องการได้

นอกจากนี้ยังสามารถกำหนดเป็น table of specification ขึ้น เมื่อต้องการสร้างแบบสอบเพื่อวัดคุณลักษณะนี้มี test spec. อย่างนี้ ก็สามารถ generate ออกมาเป็นฉบับได้เลย สร้างให้มีแบบสอบคู่ขนาน A, B, C ก็สามารทำได้ ซึ่งขณะนี้ก็มีเทคโนโลยีด้านคอมพิวเตอร์รองรับอยู่แล้ว เพียงแต่คนที่สร้างระบบนี้จะต้องมีพื้นฐานด้าน programming พอสมควร

เมื่อมีการ generate ข้อสอบและนำไปใช้ ทุกครั้งของการใช้ควรมีการวิเคราะห์ผลเพื่อเก็บเข้าคลังข้อสอบ (item bank) จากพื้นฐานความรู้เกี่ยวกับ Item Information ตามทฤษฎี IRT สามารถนำไปสู่ระบบการคัดเลือกข้อสอบเพื่อสร้างเป็นแบบสอบคู่ขนานที่น่าเชื่อถือ รวมทั้งระบบการคัดเลือก

ปัญหาประการต่อมาของการวัดผลก็คือ ในการสอบวัดผลสิ่งเดียวกัน แต่ทำการสอบต่างเวลา ต่างสถานที่ และต่างกลุ่มกัน คะแนนที่ได้จะมีการปรับเทียบกันได้อย่างไรจึงจะทำให้เกิดความยุติธรรมสำหรับผู้สอบ ซึ่งปัญหานี้สามารถแก้ไขได้โดยใช้เทคนิค test equating ซึ่งจะต้องศึกษาและกำหนด design ของการ equate ให้เป็นไปอย่างยุติธรรม มีสิ่งที่ต้องศึกษาอยู่ 3 ประเด็น คือ การสร้างแบบสอบให้มีความทัดเทียมกัน วิธีการเก็บข้อมูลและเทคนิคการปรับแก้คะแนนระหว่างแบบสอบ ซึ่งมีสูตร model ต่าง ๆ ที่ต้องศึกษาในรายละเอียดทั้ง Horizontal Equating และ Vertical Equating เพื่อให้ Standard Error of Equating มีค่าจำกัดอยู่ใน

## ระดับที่ยอมรับได้

ปัญหาอีกประการหนึ่งของการวัดและประเมินผล ก็คือ การวัดผลในสิ่งที่อยู่ต่าง level กัน จะวัดอย่างไร และจะทำการวิเคราะห์ความสัมพันธ์ของผลการวัดต่างระดับกันอย่างไร เช่น ภาวะผู้นำของผู้บริหารโรงเรียนและพฤติกรรมการสอนของครูในชั้น ไปมีผลกระทบต่อผลสัมฤทธิ์ทางการเรียนของผู้เรียนได้อย่างไร เป็นต้น การวัดในลักษณะนี้เป็นลักษณะของการวัดที่เป็น multilevel จำเป็นต้องใช้ multilevel analysis ซึ่งมีเทคนิคและ software เช่น HLM ที่จะต้องทำการศึกษาในรายละเอียดต่อไป

สรุปประเด็นในเรื่องของวิธีวิทยาขั้นสูงของการวัดและประเมินผลทางการศึกษาที่น่าสนใจ และควรที่จะศึกษาวิธีการและนำมาใช้อย่างจริงจังได้แก่

1. การศึกษาการพัฒนาและการประยุกต์ทฤษฎีการตอบสนองข้อสอบ (IRT)
2. การศึกษาการพัฒนาและการประยุกต์ทฤษฎีการสรุปอ้างอิงทางการทดสอบ (GT)
3. concept และการตรวจสอบในเรื่องของ validity
4. การสร้างและการพัฒนาระบบ item generator/item bank
5. การสร้างและพัฒนาระบบ test equating
6. การวัดและการวิเคราะห์ multilevel analysis

# ความรู้ชายแดนด้านการวัดผลการศึกษา

อุทุมพร จามรมาน

## 1. แหล่งและขอบเขต

ความรู้ชายแดนที่ปรากฏในบทความนี้ได้มาจากการอ่านวารสาร 4 ฉบับ คือ (1) Journal of Educational Measurement (2) Educational Measurement : Issues and Practice (3) Applied Psychological Measurement และ (4) Psychometicka ในช่วง ค.ศ. 1993-1995

## 2. หัวข้อที่เกี่ยวข้อง

จากการอ่านวารสารดังกล่าวในช่วงเวลาที่ระบุ พบว่า มีการพูดถึงหัวข้อที่น่าจะยังคงต้องการการศึกษาค้นคว้าและวิจัยเพิ่มเติมอยู่ 7 หัวข้อ คือ

1. พื้นฐานและปัญหาในการวัด
2. การวัดแบบเดิม (Classical Test)
3. การวัดการตอบสนองรายข้อ (Item Response)
4. การทำคลัง
5. ทฤษฎีเปิดของคูล์มส์ (Coomb's unfolding)
6. การวัดความลำเอียง
7. การอ้างอิงสรุป

ดังมีรายละเอียดต่อไป

---

บทความนำเสนอในการสัมมนาคุชฎีบัณฑิตสาขาการวัดและประเมินผลการศึกษา เรื่องวิธีวิทยาขั้นสูงด้านการวิจัย สถิติ การวัดและประเมินผลการศึกษา วันจันทร์ที่ 21 สิงหาคม 2538 ณ ศปฐรศาลา จุฬาลงกรณ์มหาวิทยาลัย

### 3. พื้นฐานและปัญหาในการวัด

ประเด็นสำคัญของการวัดมี 2 ประเด็น คือ วัดอะไร วัดอย่างไร จึงจะตรงและเชื่อถือได้

**3.1 การวัดอะไรนั้น** อุทุมพร จามรมาน (2538) ได้ระบุว่า มีการวัดอยู่ 6 ประเภท คือ

1. วัดลักษณะทางร่างกาย เช่น อวัยวะ สุขภาพอนามัย
2. วัดลักษณะทางจิต และบุคลิกภาพ เช่น ทศนคติ ค่านิยม บุคลิกภาพ ความรู้สึก
3. วัดความสามารถทางสมอง คือ สถิติปัญญา ความถนัด และผลสัมฤทธิ์ทางการเรียน
4. วัดลักษณะความสัมพันธ์ระหว่างคนและกลุ่มคน
5. วัดทักษะทางกล้ามเนื้อ
6. วัดทักษะทางปัญญา เช่น ทักษะทางภาษาและการแก้ปัญหาคณิตศาสตร์

ความรู้ชายแดนในประเด็นนี้ได้แก่

1. การวัดสติปัญญา

ทฤษฎีกระบวนการประมวลผลทางปัญญา (Cognitive Information Processing Theories) ได้ระบุว่าสติปัญญามี 3 ประการ คือ สถิติปัญญาที่เกี่ยวกับพันธุกรรม ประสบการณ์ และสิ่งแวดล้อม

2. การวัดภาคปฏิบัติ (Performance) ในโรงเรียนให้เน้นการวัดตามสภาพที่แท้จริงและการวัดประเด็นที่เกี่ยวข้องกับการดำรงชีวิตจริง โครงการวิจัยใหญ่ ๆ ที่กำลังดำเนินอยู่ ได้แก่ โครงการวัดแบบ Authentic Assessment Portfolio Assessment และ Performance Based Assessment โครงการดังกล่าวเน้นที่สาระที่วัด วิธีการวัด ตัวชี้ของผลภาคปฏิบัติ และความถี่ของการวัดหลาย ๆ วิธี

**3.2 การวัดอย่างไรจึงจะได้ข้อมูลที่เชื่อถือได้** เป็นประเด็นสำคัญในสาขาการวัดผลมานาน และยังคงเป็นประเด็นที่อภิปรายถกเถียงกันอยู่ว่า “ข้อมูล” ที่ได้ นั้น สะท้อนธรรมชาติที่แท้จริงของผลการวัดหรือไม่ และมากน้อยเพียงใด

อุทุมพร จามรมาน (2538) ได้ระบุวิธีการวัดโดยอาศัยประสาทสัมผัสของคนไว้ 6 วิธี คือ (1) การสังเกตด้วยตา (2) การฟังด้วยหู (3) การถามด้วยปาก (4) การดมด้วยจมูก

(5) การจับต้องด้วยมือ (6) การผสมหลาย ๆ วิธี ซึ่งวิธีการดังกล่าวนำมาสู่การวัดในชื่อเรียกต่าง ๆ เช่น การสัมภาษณ์ การสังเกต การทดสอบ การสอบถาม

การวัดที่ให้ข้อมูลที่เชื่อถือได้ ต้องอาศัยความชัดเจนของสาระที่วัด (วัดอะไร) กับความสอดคล้องของวิธีการกับสาระที่วัด ดังนั้น การวัด (Measurement) จึงเน้นที่ข้อมูลที่ได้ว่ามีความเที่ยง (Reliability) และความตรง (Validity) มากกว่าเครื่องมือ (Measure) หรืออีกนัยหนึ่งก็คือการเน้นที่ความเที่ยงและความตรงของข้อมูลมากกว่าความเที่ยงและความตรงของเครื่องมือ

ความรู้ชายแดนในประเด็นนี้คือ

1. ทำอย่างไรจึงจะจัดวิธีการวัดให้ตรงกับสาระที่วัด แล้วให้ได้ข้อมูลที่เชื่อถือได้ เช่น การวัดความสัมพันธ์ของคน 2 คน จะวัดอย่างไรจึงจะได้ข้อมูลที่สะท้อนความสัมพันธ์ของคน 2 คนจริง

2. สูตร KR-20,21 เป็นสูตรที่วัดความสอดคล้องภายในได้จริงหรือไม่

3. การหาค่าความเที่ยง (Reliability) ของข้อมูลที่เป็นภาคบรรยาย (Qualitative หรือ Polytomous) จะทำอย่างไร

4. การหาค่าความเที่ยงของข้อมูลที่ได้จากวิธีการวัดแบบ Complex Multiple Choice (CMC) หรือ Multiple True False (MTF) จะทำอย่างไร (หมายเหตุ ข้อความแบบ MTE หมายถึง ตัวเลือกทั้งหลายไม่เป็นอิสระต่อกัน)

#### 4. การวัดแบบเดิม (Classical Test Theory)

การวัดแบบเดิมตั้งอยู่บนฐานของการพิจารณาค่าคะแนนที่ได้รับ (X) คะแนนจริง (T) และคะแนนความคลาดเคลื่อน (E) ภายใต้สมการเชิงเส้นตรง คือ

$$X_i = T_i + E_i$$

โดยที่ i หมายถึงคนที่ i

ประเด็นที่เป็นความรู้ชายแดนได้แก่

1. สมการที่ไม่ใช่เส้นตรง (Non-Linear) น่าจะสะท้อนค่าที่ได้รับได้ดีกว่า

2. การวัดค่าที่แท้จริง (True) ทำอย่างไร ซึ่งเป็นปัญหาของการวัดในข้อ 3 ที่กล่าวมาแล้ว

3. สูตรความเที่ยงแบบ KR-20,21 ยังคงเป็นสูตรที่ระบุค่าความสอดคล้องภายใน  
ใช่หรือไม่

4. จะระบุความเชื่อถือได้ของผลการวัดได้อย่างไร

5. ข้อตกลงเบื้องต้นที่นำไปสู่การได้สูตรความเที่ยง เช่น ความคลาดเคลื่อนเป็นอิสระ  
ต่อกัน ค่าจริงมีความสัมพันธ์เป็น 0 ยังคงจำเป็น และสะท้อนความเป็นไปได้มากน้อยเพียงไร  
การฝ่าฝืนข้อตกลงเบื้องต้นดังกล่าว จะทำให้มีวิธีหาค่าความเที่ยงแบบอื่นได้หรือไม่

## 5. การวัดการตอบสนองรายข้อ

ทฤษฎี Item Response Theory ได้เน้นที่การวัดรายข้อ โดยมุ่งที่การประมาณ  
คุณลักษณะของข้อ (item) 3 ประการ คือ ความยาก อำนาจจำแนก และการเดาถูก ซึ่งถ้า  
ประมาณค่า parameter ของข้อได้ ก็สามารถคำนวณค่าความสามารถ ( $\theta$ ) ของผู้ตอบได้

ทฤษฎีนี้ตั้งอยู่บนข้อตกลงเบื้องต้น เช่น ลักษณะที่วัดเป็นมิติเดียว (Unidimension)  
และความเป็นอิสระภายในของข้อ (Local Independence) ซึ่งในปัจจุบันพบว่า ลักษณะที่วัด  
มิได้มีมิติเดียว หากแต่มีหลายมิติ เช่น ความสามารถทางสมอง ลักษณะทางจิตใจ ทักษะทาง  
ร่างกาย ดังนั้น การฝ่าฝืนข้อตกลงเบื้องต้น จึงเป็นสิ่งที่น่าสนใจว่าจะยังคงใช้ทฤษฎีได้หรือไม่

ความรู้ชายแดนที่เกี่ยวข้องได้แก่

1. การวัดมิติพหุ (Multidimension) โดยอิงทฤษฎี IRT จะทำอย่างไร

2. ความไม่เป็นอิสระภายใน (Local Dependence) จะกระทบทฤษฎีนี้มากน้อยเพียงใด

3. ทฤษฎีนี้ให้คำตอบที่เกี่ยวข้องกับลักษณะของข้อ 3 อย่างคือ ความยาก อำนาจ-  
จำแนก และการเดาถูก (3 parameter) ซึ่งในเชิงคณิตศาสตร์ parameter ที่เกี่ยวข้องกับเส้น  
โค้ง (ICC) มีมากกว่า 3 อย่าง การนำ parameter ตัวที่ 4,5 มาประยุกต์กับลักษณะของข้อ  
จะทำได้หรือไม่ และให้ความหมายถึงการวัดรายข้อหรือไม่ (parameter ตัวที่ 4,5 คืออะไรบ้าง)

4. ข้อความที่นำมาใช้กับทฤษฎี IRT ใช้กับ Dichotomous Item ซึ่งไม่เพียงพอต่อ  
การวัดลักษณะบางอย่าง ซึ่งไม่จำเป็นต้องเป็นข้อความประเภทนี้เท่านั้น การวัดลักษณะเชิง  
บรรยาย ซึ่งให้ค่าและคะแนนที่มีใช่ 0,1 ถ้านำมาใช้กับทฤษฎีนี้ได้ และควรทำอย่างไร  
(ข้อความเหล่านี้บางทีเรียกว่า logic item หรือ polytomous item)

## 6. การทำคลัง (Banking)

การทำคลังข้อสอบและคลังแบบทดสอบ โดยอิงคอมพิวเตอร์กับทฤษฎี IRT เป็นสิ่งที่เป็นไปได้ ได้มีโปรแกรมสำเร็จรูปที่สามารถจัดเก็บข้อสอบเพื่อการเรียกใช้และปรับเปลี่ยนค่าสถิติ หลังการใช้แล้ว

ความรู้ชายแดนในเรื่องนี้ ได้แก่

1. การทำคลังแบบทดสอบ (Test Bank) เพื่อการใช้ให้ตรงกับจุดประสงค์การวัด และเนื้อหาที่วัดในหลักสูตร ยังไม่ได้มีผู้ศึกษาค้นคว้าไว้ ผู้ที่ให้ความคิดในเรื่องนี้คือ Eva Baker แต่การดำเนินงานยังไม่ได้ทำจริงจัง

2. การสร้างข้อสอบ (Theory Based Item Construction) ตามโครงสร้างเดิม พบว่าทำได้ในเนื้อหาที่เกี่ยวข้องกับความสามารถทางสมอง (Cognitive) เช่น การสร้างข้อสอบภายใต้เนื้อหาเกี่ยวกับทรง (shape) จำนวน (number) ทำได้แล้ว การสร้างข้อสอบในเนื้อหาอื่นจะอย่างไร

3. การทดลองใช้ข้อสอบ แล้วนำข้อมูลมาคำนวณค่าพารามิเตอร์ของข้อสอบรายข้อ มักมีปัญหาในเรื่องจำนวนคน ความเป็นตัวแทนของประชากร วิธีสุ่มคน ซึ่งถ้าใช้ทฤษฎี CTT ในการวิเคราะห์ต้องระมัดระวังในเรื่องลักษณะกลุ่มที่เป็นตัวแทน (Norm group) และถ้าใช้ทฤษฎี IRT ในการวิเคราะห์ต้องเน้นความระมัดระวังในเรื่องจำนวนคนที่จะนำมาประมาณค่าพารามิเตอร์ของข้อ

ความแตกต่างระหว่างทฤษฎีทั้ง 2 ในการวิเคราะห์ข้อสอบ มีดังนี้

CTT	IRT
1. จำเป็นต้องใช้กลุ่มปกติวิสัย (norm group)	1. ไม่จำเป็น
2. ให้ค่า p,r	2. ให้ค่า a,b,c
3. ค่าที่ได้เปลี่ยนแปลงตามกลุ่ม	3. ค่าที่ได้ไม่เปลี่ยนแปลงตามกลุ่ม ถ้าได้จำนวนคนมากพอ
4. คำนวณง่ายและคำนวณจากเครื่องคิดเลขได้	4. คำนวณจากคอมพิวเตอร์

CTT	IRT
<p>5. ค่าความสามารถของคนเป็นค่าที่ได้จากการสังเกต (observed score)</p>	<p>5. ค่าความสามารถของคน คำนวณจากคุณลักษณะของข้อ</p> <p>6. สามารถสร้างกลุ่มปรกติวิสัยได้ หลังจากการทดลองใช้หลาย ๆ ครั้ง (สามารถ build up กลุ่มปรกติวิสัยได้)</p> <p>7. สามารถสร้างลักษณะการแจกแจงข้อมูลตามลักษณะที่กำหนดได้ โดยการ Simulate data</p> <p>8. ใช้ได้กับข้อความแบบ polytomous</p> <p>9. สามารถทำ item calibration ได้โดยวิธี (ก) link (ข) pool data</p> <p>10. ใช้ได้กับลักษณะมิติพหุ</p> <p>11. ให้สารสนเทศของข้อ (item information)</p> <p>12. สามารถระบุความลำเอียงของข้อ (item bias) ได้จาก Differential Item Function (DIF) โดยวิธี SIBTEST ของ Shealy and Stout</p>

### 6. ทฤษฎีเปิดของคัมส์ (Coombs' Theory of Unfolding)

คัมส์ (1950) ได้เสนอว่า คนทุกคนมีความพึงพอใจ (Preference) ของตนเองเมื่อให้ตัดสินสิ่งเร้าที่ละคู่ เช่น เขาจะพึงพอใจ x มากกว่า y จุดของความพึงพอใจของเขาบนเส้นต่อเนื่องจะอยู่ใกล้ x มากกว่า y

ดังนั้นทฤษฎีนี้ระบุว่า

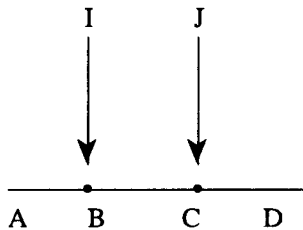
1. ตัวแปรสิ่งเร้าตัวหนึ่งที่เกี่ยวข้องกับการตัดสินเรื่องความพึงพอใจ
2. ตัวแปรนี้เป็นตัวแปรเชิงปริมาณ
3. คนแต่ละคนจะมีค่าของความพึงพอใจสูงสุงบนตัวแปรนั้น

4. ความพึงพอใจในตัวแปรสิ่งเร้าลดลง ระยะห่างระหว่างสิ่งเร้ากับจุดของความพึงพอใจจะเพิ่มขึ้น

5. คนทั้งหมดเห็นด้วยกับตำแหน่งของสิ่งเร้าบนมิติที่เกี่ยวข้อง

คুমส์เรียกจุดของความพึงพอใจสูงสุดของคนว่า จุดอุดมคติของเขาคนนั้น และเรียกความพึงพอใจในการจัดลำดับสิ่งเร้าว่า มาตรฐาน I (I คือ ลำดับความพึงพอใจของเขา)

ตัวอย่างเช่น มีสิ่งเร้า A,B,C,D และมีค่าความพึงพอใจ ซึ่งแสดงด้วยจุดบนเส้นตรงเส้นหนึ่ง



คน I มีจุดบนเส้นตรงต่อเนื่องที่แสดงถึงความพึงพอใจสูงสุดของเขา ดังนั้นลำดับที่ของความพึงพอใจสิ่งเร้าจึงเป็น A B C D อีกคนหนึ่ง มีความพึงพอใจสูงสุดอยู่ที่ J ลำดับความพึงพอใจของเขาจึงเป็น C D A B

ทฤษฎีของคুমส์ มีจุดอ่อนคือ

1. ดูเหมือนว่าจะเป็นทฤษฎีที่เกี่ยวกับมิติเดียว ซึ่งในกรณีที่สิ่งเร้ามีหลายมิติ ทฤษฎีนี้จะใช้ไม่ได้

2. การระบุว่า ทุกคนมีความพึงพอใจสูงสุดเพียง 1 ค่า อาจไม่จริง เพราะบางคนพอใจเลือก x แทนที่จะเลือก y เพราะเงื่อนไขที่ต่างกัน เช่น คนที่เลือกกาแฟไม่ใส่น้ำตาลและนม แทนกาแฟที่ใส่น้ำตาลและนมอาจเป็นเพราะเขาอยู่ในช่วงลดความอ้วน หรือเขาอาจชอบกาแฟแบบนี้จริงก็ได้

3. การระบุให้ความมากน้อยของความพอใจเกี่ยวข้องกับระยะจากจุดที่เขาพอใจสูงสุด เป็นการระบุที่ไม่มีทิศทาง (direction) ว่าข้างซ้ายหรือข้างขวาของจุด และอัตราการเพิ่มหรือลดจะเท่ากันหรือไม่

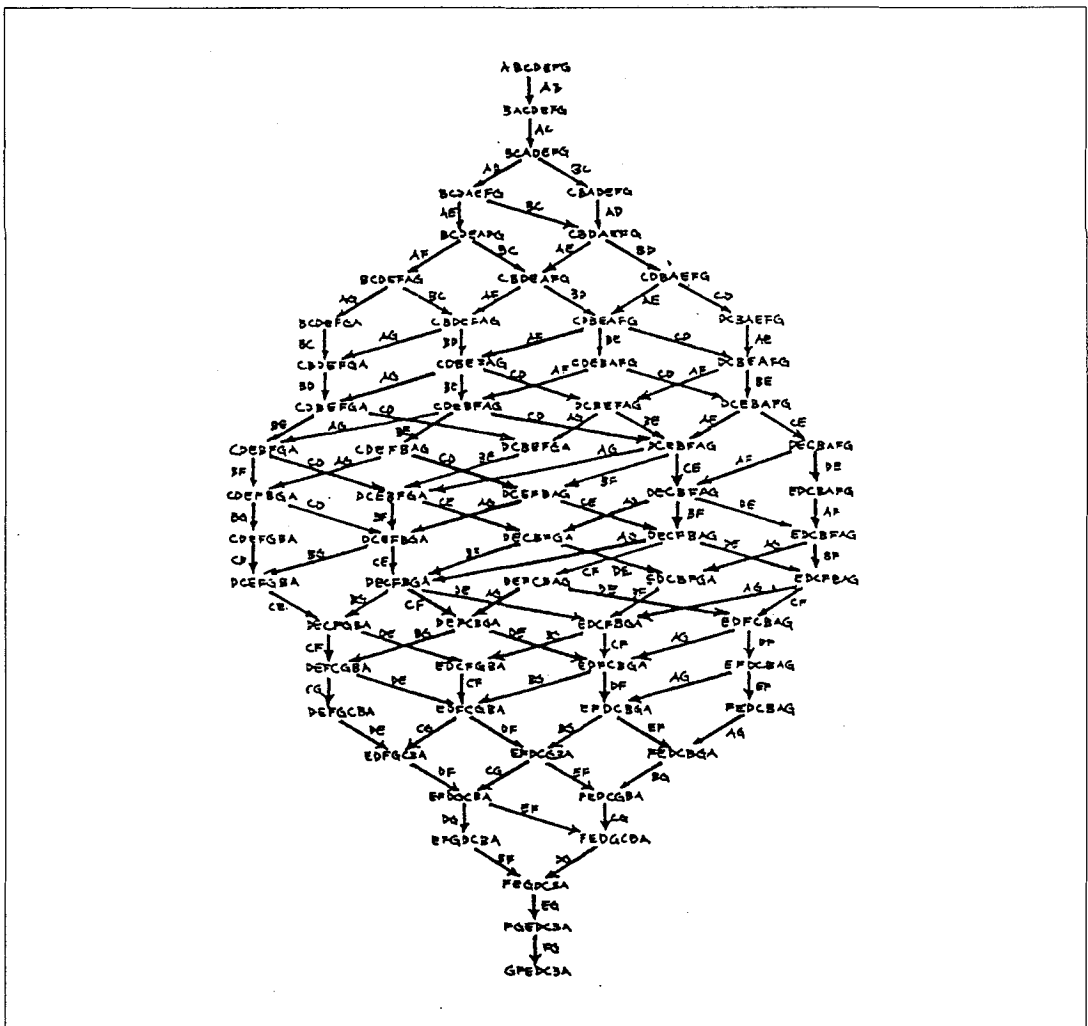
4. ตำแหน่งของสิ่งเร้าบนมิติที่เกี่ยวข้องได้มาจากการตัดสินใจของคนกลุ่มหนึ่ง ซึ่งคนในกลุ่มอาจมีความเห็นที่ไม่สอดคล้องกัน ทำให้ตำแหน่งของสิ่งเร้าไม่เป็นตัวแทนของกลุ่ม

ตัวอย่าง คูมส์ได้ให้นักศึกษา 62 คนตัดสินเกี่ยวกับเกรด ( $A^+$ ,  $A$ ,  $A^-$ ,  $B^+$ ,  $B$ ,  $B^-$ ,  $C^+$ ) โดยให้พิจารณาเกรดทีละคู่แล้วระบุว่าพอใจเกรดใดมากกว่า

เพื่อจะได้ไม่ต้องเกี่ยวข้องกับเครื่องหมายจะใช้สัญลักษณ์  $A, B, C, D, E, F, G$  แทนเกรดดังกล่าวในการจัดลำดับความพึงพอใจต่อเกรด 7 ตัวนี้จะมี  $2^6 = 64$  แบบ

การวิเคราะห์ให้พิจารณาคำตอบที่สอดคล้องในแนวเดียวกันกับลำดับของ  $A, B, C, D, E, F, G$  ของคนได้ 58 คน จาก 62 คน ที่จัดลำดับไปในทางเดียว (Transitive Order)

ผลการวิเคราะห์ของคูมส์เปิดเผยให้เห็นลำดับที่ของความพึงพอใจในเกรดทั้ง 7 ตัว ดังนี้



## 7. การวัดความลำเอียงของข้อ (Item Bias)

การที่ข้อสอบเดียวกันวัดกลุ่มคน 2 กลุ่ม แล้วให้ลักษณะของข้อที่แตกต่างกัน อาจมาจากการลำเอียงของผู้ตอบ ซึ่งสามารถตรวจสอบได้ โดยอิงความรู้เรื่อง Differential Item Information (DIF) กับข้อมูล Dichotomous เพื่อเทียบ (equate) ค่า/คะแนนที่ได้

ความรู้ชายแดนที่เกี่ยวข้องคือ

1. วิธีการของ Shealy and Stout ชื่อ SIBTEST เป็นวิธีที่ได้รับการแนะนำว่าเหมาะสมกว่าวิธีของคนอื่น การหาวิธี detect ความลำเอียงของข้อมูล Polytomous ยังเป็นสิ่งที่ต้องกระทำต่อไป

2. ปัจจัยที่มีผลต่อความลำเอียงของข้อ เช่น เพศ ภาควิชาศาสตร์ เขต ภูมิภาค เป็นปัจจัยที่ก่อให้เกิดความลำเอียงจริงหรือไม่ และยังมีปัจจัยอื่นอีกหรือไม่

## 8. การอ้างอิงสรุป (Generalizability Theory)

ได้มีการระบุว่า แหล่งที่เกี่ยวข้องกับผลสัมฤทธิ์ทางการเรียนมี 5 แหล่ง คือ

1. all possible tasks
2. all possible occasions
3. all possible raters
4. all possible measurement methods
5. all possible level units (คน โรงเรียน)

ซึ่งเมื่อประยุกต์กับทฤษฎีอ้างอิงสรุป (Generalizability Theory) พบว่าเป็นปัจจัยที่มีผลต่อค่าที่สะท้อนผลสัมฤทธิ์ทางการเรียน

ความรู้ชายแดนในที่นี้ได้แก่

1. ยังคงมีปัจจัยอื่นที่เกี่ยวข้องอีกหรือไม่ เช่น item formats, variable types, sample size
2. ปัจจัยดังกล่าวส่งผลกระทบต่อค่า effect size มากน้อยเพียงใด

## สรุป

จากการระบุนความรู้ชายแดน (Borderline of knowledge) ในสาขาการวัดผล พบว่า ยังมีประเด็นที่ต้องการการศึกษาค้นคว้าอีกมาก บางประเด็นที่เคยคิดว่าเป็นความรู้จริงก็กลับมาสู่ข้อสงสัย เช่น สูตรคำนวณค่าความเที่ยงแบบ KR-20,21 หรือความเข้าใจพื้นฐานที่เกี่ยวกับความเที่ยงตรงของเครื่องมือหรือของข้อมูล

บทความนี้ต้องการกระตุ้นให้นักวิชาการวัดผลได้เกิดการตื่นตัว และเพิ่มการคิดวิเคราะห์ในสาระที่เกี่ยวข้อง เพื่อสร้างศาสตร์การวัดผลให้เป็นศาสตร์ที่แข็งแกร่งกว่านี้

## ความรู้ชายแดนด้านการประเมินผลการศึกษา

สุวิมล ว่องวานิช

บทความนี้ต้องการนำเสนอความรู้ชายแดนด้านการประเมินผลการศึกษาซึ่งได้พัฒนาอย่างมากในช่วง 20-30 ปี ที่ผ่านมา โดยมีเนื้อหาครอบคลุมแนวคิดและการเปลี่ยนแปลงของแนวคิดด้านการประเมิน ประเด็นการประเมินที่เหล่านักคิด นักประเมินกำลังให้ความสนใจ ทั้งนี้ผู้เขียนได้ทำการรวบรวม วิเคราะห์และสังเคราะห์จากหนังสือ บทความ รายงานการวิจัย และเอกสารอื่น ๆ ที่เกี่ยวข้องกับการประเมิน

### แนวคิดการประเมิน: อดีตสู่ปัจจุบัน

นับแต่ Ralph W. Tyler (1930-1945) ผู้ซึ่งได้รับการยกย่องให้เป็น บิดาแห่งการประเมินทางการศึกษา (father of educational evaluation) ได้ให้นิยามของการประเมินว่าเป็นการศึกษาความสอดคล้องระหว่างผลที่ได้รับกับจุดมุ่งหมายที่กำหนดอันเป็นที่มาของการประเมินที่อิงวัตถุประสงค์ (objective-based evaluation) การประเมินผลเริ่มเข้าสู่ยุคที่มีความชัดเจนยิ่งขึ้น โดยเฉพาะความชัดเจนในสิ่งที่ต้องประเมิน (object of evaluation) ความหมายของ “การประเมิน” เริ่มมีการจำแนกให้เห็นความแตกต่างในความหมายของ “การวัดผล” หรือ “การวิจัย” และการประเมินผลตามแนวคิดนี้ก็ยังคงใช้ได้และหลายคนยังนิยมใช้กันอยู่จนทุกวันนี้ โดยเฉพาะการประเมินผลในห้องเรียน

---

บทความนำเสนอในการสัมมนาชุมชนนักคิดสาขาการวัดและประเมินผลการศึกษา เรื่องวิธีวิทยาขั้นสูงด้านการวิจัย สถิติ การวัดและประเมินผลการศึกษา วันจันทร์ที่ 21 สิงหาคม 2538 ณ ศศปฐศาลา จุฬาลงกรณ์มหาวิทยาลัย ขอขอบพระคุณ ศ.ดร.อุทุมพร จามรมาน และ ผศ.ดร.นงลักษณ์ วิรัชชัย ที่ช่วยเหลือจัดหาเอกสารที่เป็นประโยชน์ต่อการเขียนบทความนี้

มโนทัศน์เกี่ยวกับการประเมินได้รับการพัฒนามาเรื่อย จนมาถึงยุคแห่งการขยายความคิดเกี่ยวกับการประเมิน (The Age of Expansion 1958-1972) ในยุคนี้มีแนวคิดเกี่ยวกับการประเมิน ตลอดจนการนำเสนอโมเดลการประเมินอย่างมากมาย นิยามของการประเมินเริ่มเปลี่ยนจากการประเมินที่เน้นที่ผลผลิตอย่างเดียวมาเป็นการประเมินกระบวนการด้วยเพื่อให้ได้ข้อมูลที่ช่วยในการตัดสินใจคุณค่าหรืออนาคตของโครงการ และ/หรือข้อมูลที่ช่วยในการปรับปรุงโครงการ ในช่วงนี้นับว่าเป็นยุคของความพยายามที่จะทำความเข้าใจให้ตรงกัน ในหมู่นักประเมินว่า จริง ๆ แล้ว การประเมินคืออะไรกันแน่ และท้ายที่สุดก็ค่อนข้างจะทำให้ได้แนวคิดเกี่ยวกับการประเมินที่แบ่งได้เป็น 2 กลุ่มใหญ่ ๆ คือ การประเมินเพื่อการตัดสินใจคุณค่าของโครงการ (judgmental evaluation) และการประเมินเพื่อหาข้อมูลสนับสนุนการตัดสินใจโครงการ (decision-based evaluation) ทั้งนี้วิธีดำเนินการประเมินที่นักประเมินแต่ละคนเสนออาจจะแตกต่างกัน หรือมีจุดเน้นต่างกัน อย่างไรก็ตาม บทบาทของการประเมินในช่วงนี้มีการจำแนกอย่างเห็นได้ชัดเกี่ยวกับการประเมินความก้าวหน้า (formative evaluation) และการประเมินผลสรุป (summative evaluation) ตามแนวคิดของ Scriven ซึ่งเป็นที่ยอมรับกันมาก และแนวคิดนี้ก็สอดคล้องกับการประเมินกระบวนการ (process evaluation) และการประเมินผลผลิต (product evaluation) ซึ่งใช้กันมากในปัจจุบัน

ความคิดเกี่ยวกับการประเมินในยุคที่เรียกว่า The Age of Expansion (1958-1972) นี้อาจกล่าวได้ว่าประเด็นที่ยังเห็นไม่ตรงกันคือเรื่องของ “การตัดสินใจคุณค่า” (value judgment) Scriven เป็นผู้หนึ่งที่ทำให้ความสำคัญกับ “คุณค่า” อย่างมาก แม้จนถึงทุกวันนี้จุดยืนทางความคิดในเรื่องนี้ของเขาก็ยังไม่เปลี่ยนแปลงเลย นั่นคือ หากนักประเมินจะต้องทำการประเมินอะไรก็ตาม สิ่งที่นักประเมินต้องทำคือการให้ข้อสรุปเชิงประเมิน (evaluative conclusion หรือ evaluative claim) และถ้าผลการประเมินไม่สามารถสรุปในลักษณะนี้ได้ ก็ถือว่านักประเมินไม่ได้ทำหน้าที่ของตนเองอย่างสมบูรณ์ ความคิดของ Scriven ไม่ค่อยสอดคล้องกับแนวคิดของ Stufflebeam ซึ่งเสนอ CIPP Model ในช่วงแรก (1969, 1971) และนักประเมินอื่น ๆ เช่น Alkin (1969) คนกลุ่มหลังเห็นว่าการประเมินเป็นการจัดหาข้อมูลที่เป็นประโยชน์ที่ช่วยในการตัดสินใจของผู้ทำหน้าที่ตัดสินใจโครงการ นั่นคือ ผู้ประเมินไม่ได้ตัดสินใจคุณค่าของโครงการ แต่ให้ข้อมูลเพื่อให้ผู้มีอำนาจตัดสินใจคุณค่า อย่างไรก็ตาม แนวคิดในช่วงหลังของ Stufflebeam ก็ได้เปลี่ยนแปลงไป โดยเห็นว่าผู้ประเมินควรให้ข้อมูลเชิงประเมินด้วย

การพัฒนาแนวคิดเกี่ยวกับการประเมินในช่วงปี 1973 เป็นต้นมา นับเป็นความพยายามที่จะทำให้การประเมินมีความเป็นมาตรฐานทางวิชาชีพมากขึ้น (The Age of Professionalization) การประเมินจะให้ข้อมูลที่ เป็นประโยชน์ต่อการตัดสินใจต่อเมื่อได้ให้ข้อมูลที่สอดคล้องกับความต้องการของผู้เกี่ยวข้อง โดยเห็นได้ชัดจากการนำเสนอแนวคิดของ Stake ที่เรียกว่า responsive evaluation (1973) และแนวคิดของ Guba & Lincoln (1981) เกี่ยวกับการประเมินผล ซึ่งปรากฏในหนังสือที่ชื่อว่า effective evaluation ไม่นานหลังจากนั้น Patton ก็นำเสนอการประเมินที่เป็นแบบ utilization-focused evaluation เกือบทุกแนวคิดมีความเห็นว่าผู้ประเมินต้องให้ข้อมูลที่สอดคล้องกับความต้องการของผู้ใช้ จึงจะทำให้เกิดการนำผลการประเมินไปใช้ ความคิดเกี่ยวกับการประเมินของคนในกลุ่มนี้ส่วนใหญ่เป็นวิธีการที่จัดอยู่ในกลุ่มของ qualitative evaluation หรือ responsive evaluation

ความพยายามในการพัฒนาการประเมินให้เป็นวิชาชีพมากขึ้นในช่วงนี้สามารถแบ่งได้เป็น 3 กลุ่มใหญ่ ๆ คือ (1) การพัฒนาแนวคิดในเชิงการทฤษฎีที่ช่วยชี้ว่าการประเมินให้สนองความต้องการของผู้ใช้ (2) การพัฒนาในเชิงการกำหนดมาตรฐานการประเมินเพื่อให้การประเมินเป็นวิชาชีพที่ได้รับการยอมรับ และ (3) การพัฒนาในเชิงเทคนิคต่าง ๆ ที่ใช้ในการทำการประเมิน

การพัฒนาแนวคิดเชิงทฤษฎีในปัจจุบันเปลี่ยนจากช่วงก่อนปี 1973 แต่เดิม ประเด็นการประเมินเป็นไปตามความคิดของนักประเมินที่เห็นว่าสำคัญและเป็นประโยชน์ต่อผู้ใช้ แต่ในช่วงหลังพัฒนาการของการประเมินเปลี่ยนไปเป็นว่า ประเด็นการประเมินควรขึ้นอยู่กับความต้องการของผู้ใช้ นั่นคือ ต้องศึกษาก่อนว่าผู้ใช้ผลการประเมินต้องการอะไร แล้วจึงเอาตัวนั้นมากำหนดแนวทางในการประเมิน ดังนั้น ช่วงหลังจึงมีการเสนอแนวคิดที่เป็นแบบ utilization-focused evaluation หรือ need-based evaluation กันมากขึ้น

ในขณะที่กลุ่มหนึ่งยังคงพัฒนาความคิดเกี่ยวกับการประเมินเชิงทฤษฎี อีกกลุ่มหนึ่งก็เริ่มพัฒนาความคิดเกี่ยวกับการทำให้การประเมินเป็นกระบวนการที่เป็นมาตรฐาน มีการตั้งคณะกรรมการทำงานเพื่อกำหนดมาตรฐานของการประเมิน (standard for evaluation) มีการนำเสนอความคิดที่เรียกว่า meta evaluation (โดย Scriven) เพื่อตรวจสอบการประเมินอีกครั้ง (evaluation of evaluation) และมีการรวมกลุ่มกันเพื่อพัฒนาความก้าวหน้าให้กับศาสตร์ของการประเมินอย่างต่อเนื่อง ไม่ว่าจะเป็นการก่อตั้งสมาคมการประเมิน การจัดทำวารสาร

ทางวิชาการเพื่อแลกเปลี่ยนความรู้และข่าวสารด้านการประเมิน หรือการตั้งกลุ่มผู้สนใจศึกษา ภายใต้หัวข้อที่สมาชิกในกลุ่มสนใจ (เช่น Topical Interest Group) ก็ตาม นับเป็นการรวมกลุ่มของนักประเมินในการพัฒนาศาสตร์ของการประเมินให้แก่งยิ่งขึ้น

นอกจากนี้ในส่วนเทคนิคที่ใช้ในการประเมินก็มีการพัฒนาอย่างมาก นักประเมินเห็น ว่าบทบาทของตนเองไม่ได้จำกัดอยู่เพียงการประเมินผลของโครงการเท่านั้น แต่เห็นว่าจะต้อง เป็นผู้มีส่วนในการพัฒนาปรับปรุงโครงการได้ด้วย (to improve program) และการพัฒนา โครงการให้ดีขึ้นไม่ได้หมายความว่า ต้องมีโอกาสเข้าไปทำการประเมินระหว่างที่โครงการ กำลังดำเนินอยู่นั้น แต่ควร จะต้องมีโอกาสเข้าไปมีส่วนร่วมตั้งแต่การพัฒนาหรือจัดตั้ง โครงการ ผู้ที่เชื่อในแนวคิดเกี่ยวกับ need-based evaluation หรือ consumer-oriented evaluation น่าจะเห็นด้วยกับแนวคิดของการมีส่วนร่วมในการจัดทำโครงการตั้งแต่ต้น เพราะ ทำให้ได้โครงการที่มีลักษณะตอบสนองความต้องการของผู้ที่ได้รับผลกระทบจากโครงการมากที่สุด แต่เดิมการประเมินเกิดขึ้นหลังจากโครงการสิ้นสุด ซึ่งทำได้เพียงการประเมิน “ผล” (product evaluation) ต่อมาแนวคิดเกี่ยวกับการประเมินระหว่างโครงการกำลังดำเนิน อยู่ด้วย ทำให้เกิดการประเมิน “กระบวนการ” (process evaluation) ขณะนี้มีแนวคิดเกี่ยวกับการประเมินก่อนมีโครงการ ทำให้เกิดการประเมิน “ความต้องการ” (needs assessment) บทบาทของผู้ประเมินเริ่มไปผูกติดอยู่กับวงจรการบริหารจัดการโครงการมากขึ้น และผูกติด กับกลุ่มผู้บริหารจัดการด้วย

ความคิดเกี่ยวกับการประเมินไม่ว่าจะเป็นแนวใด ต่างมุ่งไปที่เป้าหมายเดียวกัน คือ การก่อให้เกิดการพัฒนาและดำเนินโครงการที่เอื้อประโยชน์สูงสุดต่อผู้เกี่ยวข้อง การทำการ ประเมินผลลัพธ์อย่างเดียวยังไม่ทำให้สามารถหาจุดบกพร่องมาปรับปรุงโครงการได้ และการทำ การประเมินแต่กระบวนการแม้จะทำให้ได้ข้อมูลที่ช่วยในการปรับปรุงโครงการ แต่ก็ทำให้ต้อง เสียเวลากับการดำเนินโครงการที่ไม่เกิดประโยชน์เท่าที่ควร หากพบภายหลังว่าโครงการนั้น ไม่ได้สนองตอบความต้องการของผู้เกี่ยวข้องอย่างแท้จริง บทบาทของผู้ประเมินจึงเป็นทั้งผู้ ตัดสินคุณค่าของสิ่งที่ถูกประเมิน (merit and worth evaluator) เป็นผู้อำนวยความสะดวก ในการประเมิน (evaluation facilitator) และขณะเดียวกันก็ดูเหมือนว่าจะทำหน้าที่เป็นนัก- พัฒนาโครงการด้วย (program developer)

โดยสรุป จะเห็นได้ว่าการประเมินมีพัฒนาการที่เปลี่ยนไปเป็นยุค ๆ ช่วงแรกเป็นยุคของความพยายามทำความเข้าใจเกี่ยวกับความหมายของการประเมิน และความพยายามสร้างรูปแบบหรือนำเสนอแนวคิดเชิงทฤษฎีเกี่ยวกับการประเมิน ต่อมาความพยายามจะสร้างความเป็นวิชาชีพให้กับศาสตร์ของการประเมิน โดยเน้นที่การทำให้การประเมินมีมาตรฐานเป็นที่ยอมรับ ณ จุดนี้ นิยามเกี่ยวกับการประเมินค่อนข้างเป็นที่ยุติ แม้จะมีความแตกต่างกันทางความคิดบ้าง แต่ก็ถือว่าเป็นเรื่องไม่สำคัญมากนัก สิ่งที่นักประเมินแต่ละคนต่างมีความเห็นสอดคล้องคือ การประเมินจะไม่มีประโยชน์เลยหากไม่ได้มีการนำไปใช้ประโยชน์ในการปรับปรุง เปลี่ยนแปลงหรือพัฒนาสิ่งที่ถูกประเมินให้ดีขึ้น ช่วงนี้จะพบว่ามีการศึกษาวิธีการที่จะนำไปสู่การใช้ประโยชน์จากการประเมิน และขณะเดียวกันความพยายามสร้างศาสตร์ของการประเมินก็ยังคงดำเนินต่อไป

ความรู้ชายแดนเกี่ยวกับการประเมินผล จึงน่าจะครอบคลุม 3 ประเด็นที่น่าสนใจ คือ การพัฒนาวิทยาการหรือศาสตร์ของการประเมินให้แกร่ง การพัฒนามาตรฐานของการประเมิน และการพัฒนาเทคนิควิธีที่ช่วยในการทำการประเมินให้ได้คำตอบที่มีคุณค่าเพียงพอและเหมาะสมต่อการตัดสินใจ

### วิทยาการของการประเมิน (Discipline of Evaluation)

ในขณะที่การวิจัยและการวัดผลการศึกษาต่างมีวิทยาการ (discipline) ของตนเองหลายคนอาจสงสัยว่า การประเมินมีวิทยาการของตนเองหรือไม่ เรื่องนี้มีผู้ถกเถียงกันมากบ้างก็ว่าไม่น่าจะมี เนื่องจากไม่ได้มีการสร้างทฤษฎีอะไรของตนเองเลย บ้างก็ว่ามีวิทยาการแต่เป็นแบบสหวิทยาการ (interdisciplinary) ในเรื่องนี้ Scriven (1994) ได้เสนอบทความชื่อว่า "Evaluation as a Discipline" ซึ่งมีรายละเอียดที่น่าสนใจจะศึกษาเป็นอย่างยิ่ง Scriven กล่าวว่า การประเมินถือเป็นวิทยาการหนึ่ง แต่เป็นวิทยาการที่เรียกว่า transdiscipline เนื่องจากเป็นวิทยาการที่เป็นเครื่องมือ (tool discipline) เหมือนวิทยาการด้านสถิติหรือการวัดผล คือมีเนื้อหา หรือวิธีการที่สามารถนำไปใช้ในศาสตร์อื่น และตัวการประเมินเองก็สามารถนำความรู้จากศาสตร์อื่นมาใช้ในกระบวนการประเมินได้ ลักษณะของ transdiscipline จะประกอบด้วย 2 ส่วน คือ ส่วนที่เป็น "วิทยาการที่เป็นแกน" (core discipline) และส่วนที่เป็น "วิทยาการประยุกต์" (applied discipline) วิทยาการที่เป็นแกน เป็นแนวคิดหรือทฤษฎีที่เกี่ยวกับการประเมิน ซึ่งเป็นเรื่องของ logic of evaluation ในขณะที่ส่วนที่เป็น

วิทยาการประยุกต์นั้น เป็นการนำความรู้ไปประยุกต์ใช้กับวิทยาการอื่น ๆ Scriven ยกตัวอย่างว่าวิทยาการประยุกต์ที่สำคัญมี 6 ประการ ซึ่งเขาเรียกว่า Big Six ได้แก่ program evaluation, personnel evaluation, performance evaluation, product evaluation, proposal evaluation, และ policy evaluation นอกจากนี้เขายังกล่าวว่า meta evaluation (evaluation of evaluation) ก็ถือว่าเป็นวิทยาการประยุกต์ เพราะนำวิทยาการด้านการประเมินมาใช้ในการประเมินตัวเอง (self-reference)

Scriven เห็นว่ามีความจำเป็นต้องพัฒนาวิทยาการของการประเมินขึ้นมา ที่ผ่านมามีคนส่วนหนึ่งจะไม่ยอมรับว่าการประเมินเป็นศาสตร์หนึ่ง (science of evaluation) เนื่องจากการประเมินเกี่ยวข้องกับ “การตัดสินคุณค่า” (value judgment) ในขณะที่ทางวิทยาศาสตร์ถือว่าสิ่งที่สนใจศึกษา “ไม่มีส่วนที่เป็นคุณค่า” (value-free science) ดังนั้น คนกลุ่มนั้นจึงไม่คิดว่าการประเมินเป็นวิทยาศาสตร์ และแม้หลายคนอาจจะยอมให้การประเมินเป็นวิทยาศาสตร์ แต่ต้องไม่มีส่วนที่เป็น “การตัดสินคุณค่า” รวมอยู่ด้วยโดยปล่อยให้การตัดสินคุณค่าของสิ่งที่ถูกประเมินเป็นหน้าที่ของผู้มีอำนาจในการตัดสินใจ ความคิดเหล่านี้เองที่ Scriven มองว่าทำให้ศาสตร์ของการประเมินไม่ก้าวหน้า และยังตั้งข้อสงสัยว่าทำไมศาสตร์ของการประเมินจึงต้องใช้เวลาอันยาวนานในการพัฒนา เขามองว่าส่วนหนึ่งเนื่องมาจากคนนอกศาสตร์ไม่ยอมรับความเป็นศาสตร์ และคนในแวดวงการประเมินเองก็ยังไม่มีความเห็นสอดคล้องกันทางความคิดของศาสตร์ด้านนี้ ดังนั้น จึงมีความจำเป็นต้องพัฒนาศาสตร์นี้ให้ได้

Scriven ได้เสนอความคิดว่า หากมองการประเมินว่าเป็น transdiscipline ก็ควรจะประกอบด้วยลักษณะสำคัญ 4 ประการ คือ (1) มีความเป็นปรนัย (objectivist view) โดยนักประเมินมีหน้าที่ให้ข้อสรุปเชิงประเมิน (evaluative conclusions) ในการประเมินสิ่งต่าง ๆ (2) มีลักษณะที่เป็นการอิงผู้ได้รับผลกระทบจากสิ่งที่ถูกประเมิน (consumer-oriented view) (3) มีลักษณะที่สรุปได้กว้าง (generalized view) โดยสามารถสรุปโมทัศน์เกี่ยวกับการประเมินไปได้ทุก ๆ สาขาของการประเมิน กล่าวคือ ต้องให้แนวคิดเกี่ยวกับการประเมินที่ใช้ได้กับการประเมินเรื่องใด ๆ ก็ตาม ไม่เจาะจงว่าเป็นการประเมินโครงการหรือการประเมินผลผลิต เป็นต้น และ (4) มีลักษณะที่เป็นความรู้เชิงเทคนิควิธี (technical view) ที่เป็นวิธีการของวิทยาการของการประเมินเอง

หากพิจารณาโมเดลหรือแนวคิดเกี่ยวกับการประเมินที่ผ่านมา Scriven วิจารณ์ว่าหลายแนวคิดไม่คำนึงถึงบทบาทหน้าที่ของผู้ประเมินในการตัดสินคุณค่า ในที่นี้ต้องทำความเข้าใจให้ตรงกันว่า คำว่า “ตัดสินคุณค่า” นั้น หมายถึง การให้ข้อสรุปเชิงประเมินเกี่ยวกับคุณค่าของสิ่งที่ถูกประเมินซึ่งกระทำโดยผู้ประเมิน ในบางแนวคิดข้อมูลที่ได้จากการประเมินเป็นเพียงข้อมูลสนับสนุนการตัดสินใจ แต่ไม่เป็นข้อสรุปเชิงประเมิน (เช่น แนวคิดของ Alkin) แต่บางแนวคิดสามารถให้ข้อมูลเชิงประเมินได้ (เช่น แนวคิดของ Stake, Provus, Rossi & Freeman) แต่ปล่อยให้ผู้มีอำนาจในโครงการเป็นผู้ตัดสินคุณค่าเอง

ในช่วงต้นปี 1990s มีการรวมกลุ่มกันของผู้สนใจที่จะพัฒนาศาสตร์ทางการประเมิน เช่น มีกลุ่ม TIG (Topical Interest Group) ซึ่งสนใจศึกษาเกี่ยวกับ Theories of Evaluation นำโดย Fournier และคณะ (1993) คนกลุ่มนี้เห็นว่า ทฤษฎีการประเมินน่าจะจำแนกออกเป็น 2 ประเภท คือ **Normative Theories** เป็นทฤษฎีที่เกี่ยวกับตรรกะหรือความสมเหตุสมผลของการประเมิน (logic and justification of evaluation) และ **Descriptive Theories** เป็นทฤษฎีที่เกี่ยวกับปฏิบัติการการประเมิน (practice of evaluation) การทำงานของคนกลุ่มนี้ยังไม่มีข้อสรุปที่แน่นอน สิ่งที่พวกเขากำลังคิดต่อไป คือ ความพยายามหากระบวนการที่มีเหตุผลที่จะให้ข้อสรุปเชิงประเมิน (reasoning process in coming to evaluative conclusions) และการจัดประเภทของการประเมินแบบนี้ Scriven (1994) ได้วิจารณ์ว่า เป็นการกล่าวถึงการประเมินใน 2 ประเด็น คือ ทฤษฎีการประเมิน (Theories of Evaluation) และทฤษฎีเกี่ยวกับการประเมิน (Theories about Evaluation) โดยเขากล่าวว่า ถ้าเป็น theories of evaluation หมายถึง logic of evaluation ส่วน theories about evaluation หมายถึง practice of evaluation

จากผลงานด้านความคิดของคนกลุ่มนี้ประกอบกับการวิจารณ์ของ Scriven และบทความ โดย Scriven ให้ข้อสรุปว่าปัจจุบันประเด็นของความพยายามสร้างทฤษฎีของการประเมินหรือวิทยาการของการประเมินกำลังได้รับความสนใจจากกลุ่มนักคิดทั้งหลาย สิ่งที่ Scriven เสนอประกอบกับการวิเคราะห์วิจารณ์จากแนวคิดต่าง ๆ ผู้เขียนคิดว่าศาสตร์ของการประเมินต้องให้ความสนใจในประเด็นต่อไปนี้

- (1) ข้อมูลจากการประเมินควรเป็น value judgment หรือเป็น decision support
- (2) การประเมินต้องให้ได้ข้อสรุปเชิงประเมินหรือไม่ (evaluative conclusions)

- (3) นักประเมินมีบทบาทหน้าที่ต้องให้ข้อสรุปเชิงประเมินหรือไม่ (draw evaluative conclusions) หรือให้ผู้บริหารโครงการเป็นผู้ตัดสินใจ
- (4) logic of evaluation หรือที่เป็น core discipline ซึ่ง Scriven จัดอยู่ใน theories of evaluation นั้นควรประกอบด้วย ทฤษฎี แนวคิด หรือมโนทัศน์อะไรบ้าง

### การพัฒนามาตรฐานของการประเมิน

ในปี 1981 Stufflebeam และคณะ ได้พัฒนามาตรฐานของการประเมินเพื่อเป็นแนวทางในการดำเนินการประเมินให้เป็นที่เชื่อถือ นับเป็นความพยายามในการทำให้ศาสตร์ของการประเมินเป็นวิชาชีพที่ยอมรับได้ อย่างไรก็ตาม ในช่วงหลังนับแต่ Evaluation Network (ENet) และ Evaluation Research Society (ERS) ได้มารวมกันเป็นสมาคมชื่อว่า American Evaluation Association (AEA) ทางสมาคมก็เริ่มตั้งกลุ่มคณะทำงาน (Task Force) ในปี 1992 เพื่อหาหลักปฏิบัติสำหรับนักประเมิน (Guiding Principles for Evaluators) เพื่อให้เกิดความเป็นมาตรฐานในวิชาชีพของการประเมิน ในปี 1993 คณะทำงานโดย Dianna Newman และคณะ ได้เสนอข้อสรุปเกี่ยวกับแนวปฏิบัติสำหรับนักประเมินให้กับคณะกรรมการบริหารสมาคม ซึ่งประกอบด้วยหลักใหญ่ 5 ประการ และข้อความย่อย ๆ จำนวน 23 ข้อ และข้อเสนอดังกล่าว มีการแจกให้สมาชิกของสมาคม AEA ลงคะแนนว่าจะยอมรับในหลักปฏิบัติ 5 ประการและข้อความย่อยหรือไม่ สำหรับหลักการที่กลุ่มเสนอและยังไม่ได้มีการยอมรับอย่างเป็นทางการนั้น มีดังต่อไปนี้ โดยทั้ง 5 หลักการนี้ไม่ได้มีการเรียงตามลำดับความสำคัญ

1. **Systematic Inquiry:** Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.
2. **Competence:** Evaluators provide competent performance to stakeholders.
3. **Integrity/Honesty:** Evaluators ensure the honesty and integrity of the entire evaluation process.
4. **Respect for People:** Evaluators respect the security, dignity and self-worth of the respondent, program participants, clients, and other stakeholders with whom they interact.

5. **Responsibilities for General and Public Welfare:** Evaluator articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

จะเห็นว่าแม้จะมีกลุ่มการทำงานที่นำโดย Stufflebeam ได้เสนอมาตรฐานของการประเมินมาก่อนหน้านี้แล้วก็ตาม การพัฒนาวิชาการประเมินให้เป็นที่ยอมรับก็ไม่ได้หยุดนิ่งอยู่กับที่ สมาคม AEA ได้ตั้งคณะทำงานภายใต้สมาคม AEA เพื่อศึกษาเรื่องนี้ดังกล่าวข้างต้น คณะทำงานซึ่งเสนอหลักปฏิบัติ 5 ประการนั้น ได้ให้ข้อคิดเห็นเพิ่มเติมว่า การกำหนดหลักการดังกล่าวไม่ได้หมายความว่า จะนำมาใช้แทนมาตรฐานการประเมินที่มีการกำหนดก่อนหน้านี้ แต่ต้องการให้เกิดการพัฒนาวิชาชีพให้นักประเมินได้ทำงานอย่างถูกต้องเหมาะสมและท้ายที่สุดได้ให้ข้อคิดเพิ่มเติมว่าหลักการที่กำหนดอาจจะยึดติดกับวัฒนธรรมทางตะวันตกโดยเฉพาะประเทศสหรัฐอเมริกา หากวัฒนธรรมหรือประเทศแตกต่างออกไป หลักปฏิบัติเกี่ยวกับการประเมินก็อาจเปลี่ยนแปลงไปจากที่เสนอก็ได้

## แนวคิดทฤษฎี วิธีการด้านการประเมินที่นำศึกษาพัฒนาต่อไป

### แบบของการประเมิน

#### Qualitative Evaluation, Participatory Evaluation

วิธีการที่ใช้ในการประเมินถ้าศึกษาให้ดีจะเห็นว่าค่อนข้างจะคล้อยตามวิธีวิทยาการวิจัย ในขณะที่การวิจัยมีการทำวิจัยแบบ quantitative research ทางการประเมินก็มีการทำ quantitative evaluation หรือเมื่อมีการวิจัยแบบ qualitative research ทางการประเมินก็มี qualitative evaluation วิจัยมีการทำ meta-analysis ทางการประเมินมีการทำ meta-evaluation ที่เป็นเช่นนี้อาจเนื่องจากกลุ่มนักวิชาการใน 2 ศาสตร์นี้ได้รับพื้นฐานการพัฒนาความรู้มาใกล้เคียงกัน นักวิจัย นักวัดผล และนักประเมินอาจจะเป็นคนกลุ่มเดียวกัน จึงเกิดการแลกเปลี่ยนวิทยาการที่พัฒนาขึ้นมาใหม่ ดังนั้น ถ้าพิจารณาการออกแบบการประเมินในช่วงหลังจะเห็นว่าเริ่มมีการนำเสนอแนวคิดเกี่ยวกับการประเมินเชิงคุณภาพ (qualitative evaluation) มาใช้กันมากในการประเมินงานต่าง ๆ

ในขณะที่ศาสตร์ด้านการวิจัยมีการกล่าวถึง การทำวิจัยเชิงคุณภาพที่เป็นแบบ Participatory Research หรือ Participatory Action Research (PAR) ในการประเมินก็มีการ

กล่าวถึง การประเมินแบบ Participatory Evaluation เมื่อเร็ว ๆ นี้ Garaway (1995) ได้เขียนบทความเกี่ยวกับการประเมินที่นำผู้ที่เกี่ยวข้องกับการประเมินเข้ามามีส่วนในการประเมินด้วย ทั้งนี้กลุ่มของผู้ประเมินมาจากผู้มีส่วนร่วมในโครงการ (project participants) และนักประเมินอาชีพ (professional evaluator) คนกลุ่มนี้จะร่วมกันทำงานตั้งแต่การจัดการ การลงมือทำงานด้านการประเมินและการเผยแพร่ผลการประเมินไปยังกลุ่มสมาชิกต่าง ๆ ที่เกี่ยวข้อง คำถามที่น่าสนใจคือกลุ่มของผู้มีส่วนร่วมในโครงการที่จะนำมามีส่วนร่วมในการประเมินคือใคร และจะทำการประเมินโดยวิธีนี้อย่างไร Garaway ได้นำเสนอกรอบความคิดในการทำการประเมินโดยวิธีนี้ และให้ข้อคิดเห็นไว้ว่า สิ่งที่สำคัญของการประเมินแบบนี้คือการมีผู้อำนวยความสะดวกด้านการประเมินที่เหมาะสม (evaluation facilitator) การกำหนดวิธีการประเมินที่เหมาะสม การคำนึงถึงความตรง และค่าใช้จ่ายที่ใช้ในการประเมิน ตลอดจนเวลาที่ใช้ในการประเมิน การประเมินแบบนี้ต้องมีค่าใช้จ่ายรวมทั้งสิ้นของผู้ประเมินที่ทำหน้าที่เป็น facilitator และผู้มีส่วนร่วมในการประเมิน Garaway มองว่าการประเมินแบบนี้มีข้อดี เพราะเป็นการดึงเอาบุคลากรที่มีประสบการณ์ในงานของตนเองเข้ามามีส่วนร่วมในการประเมิน และถือว่าคนกลุ่มนี้เป็นผู้ตัดสินใจโครงการได้ดีกว่าผู้อื่น และยังเป็นการสร้างบรรยากาศของการเรียนรู้ที่ดี โดยการนำผู้ใช้ผลการประเมิน (end-users) มาทำการประเมินร่วมกันกับนักประเมินอาชีพ

การประเมินแบบ participatory evaluation นี้ไม่เหมือนการประเมินภายใน (internal evaluation) ซึ่งทำการประเมินโดยบุคคลในหน่วยงาน ไม่มีนักประเมินอาชีพมาช่วยในการประเมิน หากจะใช้นักประเมินอาชีพต้องอยู่ในรูปแบบที่ให้นักประเมินอาชีพทำการประเมินที่เป็นอิสระจากนักประเมินภายใน ซึ่งเรียกว่าเป็น การประเมินภายนอก (external evaluation) นั่นคือ นักประเมินภายในและนักประเมินภายนอกมิได้ทำงานการประเมินร่วมกันเหมือน participatory evaluation

#### Fourth Generation Evaluation Theory

ระเบียบวิธีที่ใช้ในการประเมินที่กำลังกล่าวถึงมากและถือเป็นคลื่นลูกใหม่ของทฤษฎีการประเมินในขณะนี้ คือ Fourth Generation Evaluation Theory ของ Guba & Lincoln (1989) แนวคิดนี้จัดอยู่ในกลุ่มของ naturalistic evaluation ลักษณะเด่นที่ทำให้คนส่วนหนึ่งชอบในแนวคิดนี้ คือการให้อำนาจในการตัดสินใจกับผู้ได้รับผลกระทบจากโครงการมาก (stakeholders) มีการให้คุณค่าเท่า ๆ กันในการสร้างกลุ่มผู้ประเมินซึ่งเป็นผู้ได้รับผลกระทบจากการประเมิน ทำให้เกิดการยอมรับร่วมกันในข้อสรุปหรือประเด็นต่าง ๆ ที่ได้จากการประเมิน

และทำให้เกิดความมั่นใจในการใช้กระบวนการประเมินและผลการประเมิน (O'Neill, 1995) อย่างไรก็ตาม Scriven (1994) ได้วิจารณ์แนวคิดที่ว่าจัดอยู่ในกลุ่ม constructivism และไม่ค่อยเห็นด้วยกับแนวคิดนี้ เนื่องจากการประเมินตามแนวนี้จะไม่ได้ถือว่าการประเมินเป็นการค้นหาคุณภาพ คุณค่าหรือความคุ้มค่าของสิ่งที่ถูกประเมิน แต่ถือว่าการประเมินเป็นผลจากการทำงานของคนแต่ละกลุ่มแล้วมาหาข้อสรุปร่วมกัน บทความช่วงหลัง ๆ มีการวิพากษ์วิจารณ์แนวคิดการประเมินแบบนี้กันมาก และน่าจะเป็นแนวคิดที่ยังต้องศึกษาและวิเคราะห์วิจารณ์กันอีกนาน

### **Need-based Evaluation, Consumer-Oriented Evaluation**

Scriven กล่าวถึงรูปแบบการประเมินที่เป็น consumer-oriented evaluation ไว้นานแล้ว แต่ปรากฏว่าความคิดนี้เริ่มมีการให้ความสำคัญกันมากขึ้น ในขณะที่ Stake (1972) เสนอความคิดเกี่ยวกับ responsive evaluation พบว่าความสำคัญอยู่ที่ clients หรือ stakeholders ซึ่งต้องทำความเข้าใจให้ตรงกันว่า คนกลุ่มนี้หมายถึงใคร ถ้าหมายถึงเฉพาะกลุ่มที่เป็นผู้ตัดสินใจ (decision makers) ซึ่งเป็นคนขอให้ทำการประเมิน ก็ไม่น่าจะสอดคล้องกับความคิดของ Scriven เพราะแนวคิดของ Scriven นั้น เป้าหมายของการประเมินเน้นที่การนำผลการประเมินไปใช้ กลุ่มผู้ใช้ (users) เป็นกลุ่มที่ไม่ได้หมายถึงผู้มีอำนาจตัดสินใจเพียงกลุ่มเดียว เพราะการประเมินเกี่ยวข้องกับคนหลายกลุ่ม คนที่ได้รับผลกระทบเหล่านี้ (impacted population) ควรได้มีส่วนร่วมรับรู้ผลการประเมินด้วย แนวคิดการประเมินแบบ consumer-oriented evaluation จึงควรให้ความสำคัญมากกว่า manager-oriented evaluation (Scriven, 1989) เพราะการประเมินต้องดูที่ผลที่เกิดขึ้นว่าเป็นไปตามความต้องการของกลุ่มคนที่ได้รับผลกระทบหรือไม่ โดยเหตุนี้ การประเมินจึงไม่ควรยึดติดอยู่กับวัตถุประสงค์อย่างเดียว แต่ควรเป็นแบบ goal-free evaluation ดังนั้น การประเมินจึงต้องคำนึงถึง needs ของคนที่ได้รับผลกระทบด้วย และเพื่อให้ชัดเจนยิ่งขึ้น ก็ควรเรียกการประเมินให้เจาะจงว่าเป็น need-based evaluation ซึ่งเป็นแนวคิดเดียวกับ consumer-oriented evaluation

### **Individualized Objective-Based Evaluation**

ในการประเมินโครงการทางการศึกษา หากโครงการดังกล่าวให้บริการกลุ่มคนที่มีระดับต่าง ๆ กัน และบริการที่ให้แตกต่างกัน วัตถุประสงค์ของโครงการจะขึ้นอยู่กับกลุ่มคนที่เข้ามารับบริการ (program participants) ระดับของกลุ่มคนที่ต่างกัน จะมีความต้องการไม่เหมือนกัน ดังนั้นจึงมีผู้เสนอแนวคิดว่าการประเมินที่อิงวัตถุประสงค์ ควรพิจารณาวัตถุประสงค์แต่ละด้าน

หากการประเมินใด ๆ อิงการประเมินกระบวนการหรือผลงานที่เป็นมาตรฐานเดียวกัน น่าจะให้ ข้อมูลที่ไม่ถูกต้อง แนวคิดการประเมินแบบ individualized objective-based evaluation จึงเกิดขึ้น (Lam, 1995) และเป็นรูปแบบการประเมินที่น่าสนใจซึ่งแตกต่างไปจาก objective-based evaluation

### Meta-Evaluation

วิทยาการที่แข็งแกร่งต้องมีการตรวจสอบและบ่งบอกถึงจุดบกพร่องของตนเองได้ การทำ meta-evaluation จึงเป็นการประเมินคุณภาพของการประเมินอีกครั้ง อย่างไรก็ตาม ตัวอย่างของการประเมินแบบนี้มีไม่มากนัก คนที่อยู่ในสาขาวิจัยแต่ไม่มีความชำนาญในสาขา การประเมิน มักจะสงสัยว่า meta-analysis แตกต่างจาก meta-evaluation อย่างไร การทำ meta-analysis คือการสังเคราะห์ผลการวิจัย นั่นคือ การนำผลการวิจัยที่มีประเด็นวิจัย เดียวกันมาสังเคราะห์เพื่อหาข้อสรุปรวม ในขณะที่ meta-evaluation เป็นการประเมินของ การประเมินอีกครั้ง หากพิจารณาบริบทของการประเมินแล้วจะเห็นว่าไม่น่าจะมีการสังเคราะห์ ผลการประเมินได้ เนื่องจากการประเมินอะไรก็ตาม ผลการประเมินที่ได้ขึ้นอยู่กับบริบทของ หน่วยงานนั้น ๆ (context-dependent) โครงการหนึ่งดำเนินการภายใต้หน่วยงานหนึ่ง ด้วย เงื่อนไขหนึ่ง ผลการประเมินจะออกมาเป็นแบบหนึ่ง จึงเป็นเรื่องที่ไม่สมเหตุสมผลและไม่น่าจะ เกิดประโยชน์ด้วย ที่จะนำเอาผลการประเมินโครงการในลักษณะเดียวกันของหน่วยงานต่าง ๆ หรือผลการดำเนินงานของโครงการในหน่วยงานเดิม แต่ทำมาแล้วหลาย ๆ ครั้ง มาสังเคราะห์ เพื่อหาข้อสรุปรวมอีกครั้ง เพราะเป้าหมายของการประเมินไม่ได้เน้นที่การสรุปอ้างอิงผลการ ประเมินไปยังประชากรหรือกลุ่มเป้าหมายอื่นเหมือนการวิจัย เรื่องของ generalizability จึง ไม่ใช่ประเด็นสำคัญของการประเมิน

อย่างไรก็ตาม จากการศึกษารายงานผลการประเมินของ Gallegos (1994) ซึ่งทำการ ประเมินเรื่อง Meta-Evaluation of School Evaluation Models โดยผู้ประเมินนำโมเดลที่ ใช้ในการประเมินโรงเรียนจากที่ต่าง ๆ ในสหรัฐอเมริกาจำนวน 51 โมเดล มาทำการสังเคราะห์ โมเดลที่ใช้ในการประเมิน โดยตรวจสอบกระบวนการ มาตรฐานและเกณฑ์ที่ใช้ในการประเมิน ทั้งนี้มีการใช้คณะผู้ประเมินเข้าไปทำการประเมินโมเดลเหล่านั้นอย่างอิสระจากกัน บทความนี้ ทำให้ผู้เขียนเริ่มเปลี่ยนความคิดใหม่เกี่ยวกับการทำ meta-evaluation นั่นคือ การทำ meta-evaluation ไม่น่าจะหมายความว่าเพียงการประเมินวิธีการหรือคุณภาพของการประเมินใน โครงการใดโครงการหนึ่งโดยเฉพาะ แต่สามารถทำได้โดยการสังเคราะห์วิธีการที่ใช้ในการประเมิน

(synthesis of evaluation process) ซึ่งตรงจุดนี้จะคล้าย meta-analysis แล้วทำการประเมินคุณภาพของวิธีการที่ได้จากการสังเคราะห์เหล่านั้น อันทำให้ได้ข้อมูลที่สามารถสรุปรวมเกี่ยวกับจุดดีจุดด้อยของกระบวนการหรือวิธีการทั้งหลายที่ใช้ในการประเมินสิ่งใดสิ่งหนึ่งได้ดี เป็นการผสมผสานความคิดของการสังเคราะห์กับการตรวจสอบคุณภาพของการประเมินเข้าด้วยกัน

### Needs Assessment

นักประเมินหากต้องทำหน้าที่เป็นผู้พัฒนาโครงการ (program developer) ก็ต้องมีความรู้ด้านการวัดความต้องการ (needs assessment) เทคนิคการวิเคราะห์ความต้องการได้รับการพัฒนาอย่างมากในช่วงปี 1980 เป็นต้นมา หลักของการวัดความต้องการอยู่ที่ความพยายามหาความแตกต่างระหว่างสิ่งที่มุ่งหวังกับสิ่งที่ป็นจริง ช่องว่างที่เกิดขึ้นบอกถึงสภาพปัญหาที่เกิดขึ้นและต้องการแก้ไข และที่สำคัญต้องมีการจัดเรียงลำดับความสำคัญของปัญหาดังกล่าวที่เกิดขึ้น เทคนิคการวัดความต้องการจึงแบ่งออกได้เป็นหลายประเภท ที่ใช้กันมากได้แก่ (1) กระบวนการกลุ่ม (group process) เช่น brainstorming, focus-group, technique, nominal group technique (2) เทคนิคที่เกี่ยวกับการจัดเรียงลำดับความสำคัญของปัญหาถือเป็นการวัดความต้องการที่อยู่ในกลุ่ม setting priority (3) เทคนิคเชิงอนาคต (future technique) เช่น Delphi Technique, Forecasting Technique, Fault Tree Analysis, Cross-Impact Analysis และ (4) การใช้วิธีการสำรวจ (needs survey) จากการศึกษาวิธีการวัดความต้องการที่มีการทำกันมาตั้งแต่ปี 1981 ของ Witkin (1994) พบว่าส่วนใหญ่มีการทำในวงการศึกษา และกลุ่มเป้าหมายเป็นนักเรียน ครู ซึ่งถือเป็นกลุ่มผู้รับบริการ (service provider หรือ primary needs) นอกจากนี้ยังพบว่าวิธีการที่ใช้ส่วนใหญ่เป็นวิธีการแบบเดี่ยว (single-method study) แทนที่จะเป็นพหุวิธี (multiple methods) เทคนิคที่เป็นแบบการจัดเรียงลำดับความสำคัญไม่ค่อยได้มีการใช้ ผลการวิจัยนี้ชี้ให้เห็นว่าเทคนิคการวัดความต้องการที่ยังใช้กันยังมีขีดจำกัด แม้จะมีหลายเทคนิคแต่ก็ยังใช้กันอย่างไม่กว้างขวาง

เรื่องของการทำ needs assessment เป็นเรื่องที่มีผู้ให้ความสนใจกันมาก เห็นได้ชัดจากรายงานการวิจัยในช่วงปี 1990 เป็นต้นมา มีประเด็นวิจัยใหม่ทั้งการพัฒนา needs assessment models เพื่อใช้กับงานด้านต่าง ๆ การศึกษาในเชิงการประยุกต์ใช้ และการพัฒนาเชิงเทคนิควิธี เพื่อหาวิธีการวัดความต้องการที่เหมาะสมและให้ข้อมูลที่น่าเชื่อถือ ที่ผ่านมามีเทคนิคส่วนใหญ่ที่ใช้กันมักนำมาจากศาสตร์หรือสาขาวิชาอื่น เช่น การทำ cost-analysis, forecasting technique เป็นต้น น่าจะถึงเวลาแล้วที่นักประเมินควรให้ความสนใจกับการพัฒนาเทคนิคเหล่านี้ให้เป็นวิทยาการของตนเอง

### Evaluability assessment

การประเมินก็เหมือนกับการดำเนินโครงการ ต้องมีการตรวจสอบก่อนเริ่มลงมือปฏิบัติงาน ว่ามีความเป็นไปได้ที่จะทำการประเมินมากน้อยเพียงใด โครงการนั้น ๆ สามารถทำการประเมินได้หรือไม่ มโนทัศน์เกี่ยวกับเรื่องนี้มีผู้เสนอมานานแล้ว (เช่น Smith, 1989) โดยเห็นว่าการประเมินต้องเป็นเครื่องมือที่ช่วยพัฒนาโครงการและทำให้ได้ผลการประเมินที่เป็นประโยชน์ การทำ evaluability assessment จึงเป็นกระบวนการเพื่ออธิบายโครงสร้างของโครงการ เช่น วัตถุประสงค์ของโครงการ ตระกาะของโครงการ กิจกรรม ตัวบ่งชี้ความสำเร็จของโครงการ นอกจากนี้ยังเป็นการวิเคราะห์ความเป็นไปได้ในการทำให้โครงการบรรลุวัตถุประสงค์ ทำให้การประเมินให้ข้อมูลที่ลึกซึ้ง และเป็นที่ยอมรับของผู้บริหาร ผู้ตัดสินใจนโยบาย ผู้จัดทำโครงการ กล่าวโดยสรุปเป็นเสมือนการวิเคราะห์โครงการก่อนการประเมิน อย่างไรก็ตาม แม้แนวคิดนี้จะไม่ค่อยใหม่ แต่พบว่าการประเมินเองไม่ค่อยมีการดำเนินการในเรื่องนี้อย่างเป็นระบบมากนัก จึงเป็นอีกประเด็นหนึ่งที่น่าจะมีการพัฒนาเทคนิควิธีหรือประยุกต์เทคนิคไปใช้ให้มากขึ้น

### Teacher Evaluation

การประเมินครูอาจารย์นับเป็นตัวอย่างหนึ่งของการประยุกต์วิธีการประเมินไปใช้โมเดล การประเมินครูอาจารย์มีมากมาย แต่ละแบบต่างมีลักษณะสำคัญเฉพาะ Scriven (1995) ได้เสนอรูปแบบการประเมินถึง 12 รูปแบบ โดยได้แยกลักษณะของโมเดลออกเป็น 2 ประเภท คือ ประเภทที่เป็น criterion model ซึ่งจะให้แนวคิดเกี่ยวกับเกณฑ์ที่ใช้ในการประเมิน (เช่น การประเมินครูอาจารย์โดยอิงความสามารถในการสอน อิงผลการปฏิบัติงาน เป็นต้น) กับประเภทที่เป็น method model ซึ่งให้แนวคิดเกี่ยวกับวิธีการที่ใช้ในการประเมิน (เช่น การใช้เพื่อนร่วมงานเป็นผู้ให้ข้อมูล การให้ผู้เชี่ยวชาญเป็นผู้ประเมิน การใช้วิธีการสัมภาษณ์ในการเก็บข้อมูล) Scriven แสดงความเห็นว่าการประเมินบุคคลต้องเน้นที่ merit และเพื่อให้การประเมินยุติธรรมควรประเมินครูจากงานที่มอบหมายให้ทำ หรืองานที่มีการตกลงกันตอนว่าจ้าง ดังนั้น Scriven จึงไม่เห็นด้วยกับการประเมินครูที่ใช้ Research-Based Teacher Evaluation (RBTE) แต่เห็นว่า Duties-Based Teacher Evaluation (DBTE) หรือบางคนเรียก Duties of the Teacher (DOTT) มีความเหมาะสมกว่าเนื่องจากครูอาจารย์ควรถูกประเมินโดยเกณฑ์ที่มีความชอบธรรม การรับครูอาจารย์เข้าทำงานมีการตกลงกันว่าให้ทำหน้าที่สอนหนังสือ บทบาทส่วนใหญ่จึงอยู่ที่ห้องเรียน ดังนั้น หากจะประเมินครูก็ควรประเมินผลการปฏิบัติงานในห้องเรียนซึ่งถือเป็นหน้าที่โดยตรง การเสนอแนวคิดนี้นับว่าเป็นประโยชน์

ต่อการประเมินบุคลากร (personnel evaluation) และมีหลายประเด็นที่น่าจะนำมาพิจารณาในการพัฒนาโมเดลต่อไป

## บทสรุป

บทความนี้ได้รับอิทธิพลทางความคิดอย่างมากจาก Scriven ซึ่งถือเป็นนักคิด นักปรัชญาที่สำคัญคนหนึ่งของวงการประเมิน ในช่วง 2-3 ปีนี้ จะพบแนวคิดของ Scriven ในบทความต่าง ๆ ตามวารสารการประเมินเป็นจำนวนมาก ถ้าวิเคราะห์ดูแล้ว จะเห็นว่าหลักการทางความคิดของนักประเมินผู้นี้ไม่ค่อยเปลี่ยนแปลงไปจากเดิมเลย ความคิดที่เขาเสนอส่วนใหญ่จะเป็นแนวคิดที่เป็นหลักการ ไม่เจาะจงว่าจะเขียนขึ้นมาเพื่อใช้ในการประเมินสิ่งใดสิ่งหนึ่งโดยเฉพาะ เพราะเขามีความเชื่อว่าศาสตร์ของการประเมินจะต้องแสดงความคิดเชิงทฤษฎีที่มีความเป็นทั่วไปสูง เขาเห็นว่าบทความหรือหนังสือด้านการประเมินผลส่วนใหญ่จะใช้คำว่า “การประเมิน” ไม่ถูกต้อง นั่นคือ สิ่งที่น่าเสนอมักจำกัดอยู่ที่การประเมินโครงการหรือการประเมินบุคลากร หรือการประเมินทางการศึกษา (ซึ่งมักหมายถึงการประเมินนักเรียนเป็นส่วนใหญ่) จริง ๆ แล้ว แนวคิดการประเมินต้องพัฒนาให้ใช้ได้กว้าง จึงจะเกิดประโยชน์ต่อการประเมินอย่างแท้จริง

หากบทบาทของผู้ประเมินต้องมีถึง 3 ประการ คือ การเป็นผู้ตัดสินคุณค่า การเป็นผู้อำนวยความสะดวกในการทำการประเมิน และการเป็นผู้มีส่วนในการพัฒนาโครงการ ศาสตร์ของการประเมินต้องมีลักษณะที่ผสมผสานวิทยาการต่าง ๆ เข้าด้วยกันให้เป็นวิทยาการของตนเอง สิ่งที่น่าพัฒนาในเชิงแนวคิด คือ ขอบเขตการประเมินควรครอบคลุมอะไรบ้างมากน้อยเพียงใด พัฒนาการของแนวคิดการประเมินในขณะนี้เป็นที่ยอมรับร่วมกันว่า ควรมีการเปลี่ยนแปลงความคิดจากที่เคยเน้นแต่การประเมินผลลัพธ์สุดท้ายอย่างเดียว มาเป็นการประเมินเพื่อก่อให้เกิดการพัฒนาอันเป็นบทบาทของการประเมินความก้าวหน้า แนวคิดที่เริ่มเปลี่ยนแปลงไป คือ การทำให้การประเมินสนองความต้องการของผู้ใช้ และประเด็นนี้เองที่ยังไม่มีข้อสรุปชัดเจนว่าผู้ใช้ควรครอบคลุมกลุ่มใดบ้าง และจำเป็นหรือไม่ที่ผู้ประเมินต้องเป็นผู้ตัดสินคุณค่าของสิ่งที่ถูกประเมิน จะปล่อยให้เป็นที่ของผู้มีอำนาจในการตัดสินหรือผู้ได้รับผลกระทบจากโครงการได้หรือไม่ และหากผู้ประเมินต้องเป็นผู้ตัดสินคุณค่าเอง จะใช้กระบวนการแบบใดเพื่อให้ได้ข้อสรุปดังกล่าว ทั้งหมดนี้เป็นงานที่ทำทลายความคิดของกลุ่มคนที่เกี่ยวข้องกับการประเมินเป็นอย่างยิ่ง

## หนังสืออ้างอิง

- Dissertation Abstracts International.** (1989-1994). Topics on Needs Assessment. University Microfilms, Inc.
- Evaluation Practice.** American Evaluation Association. Jai Press Inc. (Vol,15 No. 1, 1994).
- Evaluation Practice.** American Evaluation Association. Jai Press Inc. (Vol,15 No. 2, 1994).
- Evaluation Practice.** American Evaluation Association. Jai Press Inc. (Vol,15 No. 3, 1994).
- Evaluation Practice.** American Evaluation Association. Jai Press Inc. (Vol,16 No.1, 1995).
- Evaluation Practice News.** American Evaluation Association. (June, 1994).
- Gallegos, A. (1994). Meta-Evaluation of School Evaluation Models. **Studies in Educational Evaluation.** Vol.20, pp.41-54.
- Garaway, G.B. (1995). Participatory Evaluation. **Studies in Educational Evaluation.** Vol.21, pp.85-102.
- Madaus, G.F.; Scriven.M.S.; and Stufflebeam, D.1. (1989). **Evaluation Models: Viewpoints on Educational and Humen Services Evaluation.** Boston: Kluwer Academic Publishers.
- O'neil, T. (1995). Implementation frailties of Guda and Lincoln's Fourth Generation Evaluation Theory. **Studies in Educational Evaluation.** Vol.21, pp.5-21.
- Personal letter from Fournier, D.M.to Utumporn Jamornmann, dated April 29, 1993. Re. Theories of Evaluation.
- Report from the AEA Task Force on Guiding Principles for Evaluators (Draft). (Revised 18 August 1993).
- Scriven,M. (1991). Pros and Cons about Goal-Free Evaluation. **Evaluation Practice.** Vol. 12, No.1, pp.55-76
- Scriven, M. (1995). A Unified Theory Approach to Teacher Evaluation. **Studies in Educational Evaluation.** Vol.21, pp.111-129.
- Scriven, M. (1995). **Theories of Evaluation.** Topical Interest Group of the American Evaluation Association. Vol. 2, No. 1.
- Scriven, M. (1995). Evaluation as a Discipline. **Studies in Educational Evaluation.** Vol.20, pp.147-166.
- Smith, M.F. (1989). **Evaluability Assessment: A Practical Approach.** Boston: Kluwer Academic Publishers.

## AN APPLICATION OF RULE ASSESSMENT APPROACH FOR DIAGNOSING MISCONCEPTIONS\*

SIRIDEJ SUSHEVA

### ABSTRACT\*\*

This paper introduced the new diagnostic method for detecting misconceptions. The method was based on the rule assessment approach which stated that each testee had various rules of operation, and that the same responses might be obtained from the same or different rules. Therefore, the rule assessment of the whole set of testee's responses would make possible the tracing back of the rule and the diagnosis of the testee's misconceptions. This diagnostic process consisted of 3 steps : the first one was the developing of cognitive patterns in accordance with the concept of the content wanted to be diagnosed; the second one was the construction of item forms and the test in such a way that the response pattern could be traced back to get the testee's cognitive pattern; the last one was the two-step diagnostic process : exploratory diagnosis and confirmatory diagnosis. From the diagnostic method for misconceptions in mathematics, the negative signed-number addition, of a sample of 940 grade 8 students, it was found that the diagnostic result of the developed method was stable and had criterion-related validity when the teachers' diagnostic

---

\* Paper presented at the International Conference on Teacher Education: Innovative Alternatives for the 21st Century, July 11-14, 1995. Bangkok Thailand.

\*\* Doctoral Dissertation in Educational Measurement and Evaluation, Department of Educational Research, Chulalongkorn University, 1995.

result were used as a criterion. This method provided diagnostic result consistently with that of the Rule Space model's, and was able to avoid the limitation of the Rule Space model; i.e. - to discriminate misconceptions of the testees whose item score response patterns were the same or all zero's. In addition, the computer program written from this diagnostic method functioned as planned.

## INTRODUCTION

Recently several researchers from various disciplines such as cognitive psychology and psychometrics have developed rule assessment methods for diagnosing students' misconceptions. Researchers from each of these disciplines were motivated by an interest in knowledge structure and development in human cognition (i.e. Anderson (1981) and Siegler (1976) from cognitive psychology, and Tatsuoka (1982) from cognitive psychometrics) Tatsuoka and associates (Tatsuoka, 1983a, 1983b, 1985, 1987; Tatsuoka & Tatsuoka, 1985, 1987) have developed a systematic approach to the study of psychometric properties of misconceptions such as the stability of misconceptions committed by a student throughout a test.

Diagnosis of misconceptions can provide useful information in evaluating instruction and instructional materials and can suggest specific prescription for remediating a student. For example, the source of many of the misconceptions constructed by students is the ambiguity of explanation or the lack of precise, accurate instruction (Harnisch & Tatsuoka, 1991). Multiple-choice type diagnostic test is more widely popular in Thailand than other methods. This kind of test can diagnose student's misconception by the choice he/she chooses which is in accordance with the provided misconceptions. In fact there are a lot of misconceptions committed by the students in a specific content, but each item in the test can contain only four or five choices. The students might have other misconceptions which are not in the choices but they must choose one of the provided choices. In addition, the same choices could be obtained from various misconceptions that are not always the specific misconception provided by the test-maker and either the right choices could be obtained from misconceptions. To avoid these limitations of the multiple-choice type diagnostic test, personal indices (Drasgow, 1982; Harnisch, 1981, 1983; Harnisch and Linn, 1981, 1982; Harnisch and Tatsuoka, 1983; Sato, 1975; Wright and Stone, 1979) designing to summarized response patterns are used for detecting anomalous patterns which may result from application of some erroneous rules, but they have only limited power. For instance, these indices cannot diagnose sources of misconceptions or provide prescriptive information for

remediating students. Nevertheless, the use of personal indices is economical and applicable to general areas because such indices can first be used to spot candidates among students who may possibly possess misconceptions and may hence need personal attention from teachers. This procedure seem to be widely used in Japan (Tatsuoka, 1983). To get the more information for the misconceptions diagnosis beyond the personal indices' capability, Tatsuoka and associates (Tatsuoka, 1983a, 1983b, 1985, 1986; Tatsuoka & Tatsuoka, 1983, 1987) have developed a new probabilistic model for handling actual response patterns together with theoretically estimated response curves. This model has been named the Rule Space model. The model begins by mapping all response patterns in a cartesian product space  $\{(\theta, f(x))\}$ , called rule space. Mapping is accomplished using the Standardized Extended Caution Index and the true score ( $\theta$ ) following the Item Response Theory. This approach facilitated the discovery that a large number of actual response patterns match response patterns associated with some erroneous rule of operation by examining the proximity of a student's position in the rule space to the position represented by the erroneous rules. The Rule Space model has the limitation to discriminate misconceptions of the testees whose item score response patterns were the same or all zero's and it does not suitable to use in the classroom level because this model, applying Item Response Theory, has to take a large number of student to estimate the parameters in the model. The developed diagnostic method which is based on rule assessment approach was designed to handle the limitations of the previous approaches and provide the exact prescriptive information for remediating each student who has misconception.

### **RULE ASSESSMENT APPROACH**

The rule assessment approach (Seigler, 1976) focused upon conceptual development. It can be summarized in terms of five core assumptions.

1. Children's conceptual understanding progresses through a regular sequence of qualitatively discrete rules.
2. These rules are ordered in terms of increasing correlation with the correct rule in environments that the children encounter. Children will not adopt a new rule that is less highly correlated with the correct rule than the rule that they are already using.
3. The effectiveness of a learning experience is largely determined by whether the learning experience distinguished the child's existing rule from the correct rule. If the learning experience does not discriminate between the two rules, the child will continue to use the original rule. If the learning experience does discriminate between the two rules, the child will either adopt a more advance rule or will enter into a "rule search" state character-

ized by uncertainty as to which rule is correct.

4. A major reason why children do not immediately adopt the correct rule for all concepts and why they have difficulty in learning it consists in the limited encoding of the correct rules' component dimensions. Understanding of concepts frequently requires integration of several component dimensions. Failure to encode one or more of them can restrict children's ability to learn the concepts.

5. Failure to encode a dimension may be caused by lack of knowledge of the dimension's importance or lack of perceptual salience relative to other dimensions in the situation in which the concept is to be applied. Increasing a child's knowledge of the dimension's importance or increasing the dimension's salience can lead to improve ability to learn, which in turn can lead to improve knowledge. Central to this approach is a methodology for determining which rule an individual child is using. According to this approach, each testee had various rules of operation, and that the same responses might be obtained from the same or different rules. Therefore, the rule assessment of the whole set of testee's responses would make possible the tracing back of the rule and the diagnosis of the testee's misconceptions.

### **THE DEVELOPED DIAGNOSTIC METHOD**

The developed diagnostic method consisted of 3 steps : the development of the rules of operation; the construction of item forms and the test; and the two-step diagnostic process. The first step was the development of cognitive patterns or the rules of operation in accordance with the concept of the content wanted to be diagnosed. A rule was a description of a set of procedures or operations that one used in solving a problem in some well-defined domain such as arithmetic. A right rule was defined as a rule that produced the right answer to every item in a test, but an erroneous rule might fortuitously yeild the right answer for some subset of the items (Tatsuoka & Tatsuoka, 1987). The rules were developed on the basis of logical analysis of the task of solving signed-number addition problems and on analysis of actual responses given by students on a test, followed by clinical interviews. In this study 126 rules of the sign-number addition were developed. Table I describes three examples of the signed-number addition rule.

**TABLE I**  
**The Sets of Response Generated by Three Rules**

Items	Rule 1-2		Rule 2-1		Rule 2-4	
	Responses	Scores	Responses	Scores	Responses	Scores
1) $-9 + 3 = -6$	6	0	-12	0	-12	0
2) $-6 + (-2) = -8$	-8	1	-8	1	-8	1
3) $2 + (-7) = -5$	5	0	-9	0	-9	0
4) $-4 + 8 = 4$	-4	0	12	0	-12	0
5) $-3 + (-9) = -12$	-12	1	-12	1	-12	1
6) $5 + (-3) = 2$	-2	0	8	0	-8	0

Rule 1-2 : If the signs of the two numbers are same then the student adds the two absolute values together and takes the sign of the number with the smaller absolute value in his/her answer.

Rule 2-1 : The two numbers are always added and the sign of the number with the larger absolute value is always taken in the answers.

Rule 2-4 : The two numbers are always added as seen in Rule 2-1 but the negative sign is always taken in the answers.

The second step was the construction of item forms and the test in such a way that the response pattern could be traced back to get the testee's cognitive pattern. The basic strategy that has been used involves the formulation of problem types that yield distinct patterns of correct answers and errors for students using different rules. The six item forms were formulated for the content of the signed-number addition and used to construct the test as shown in figure I

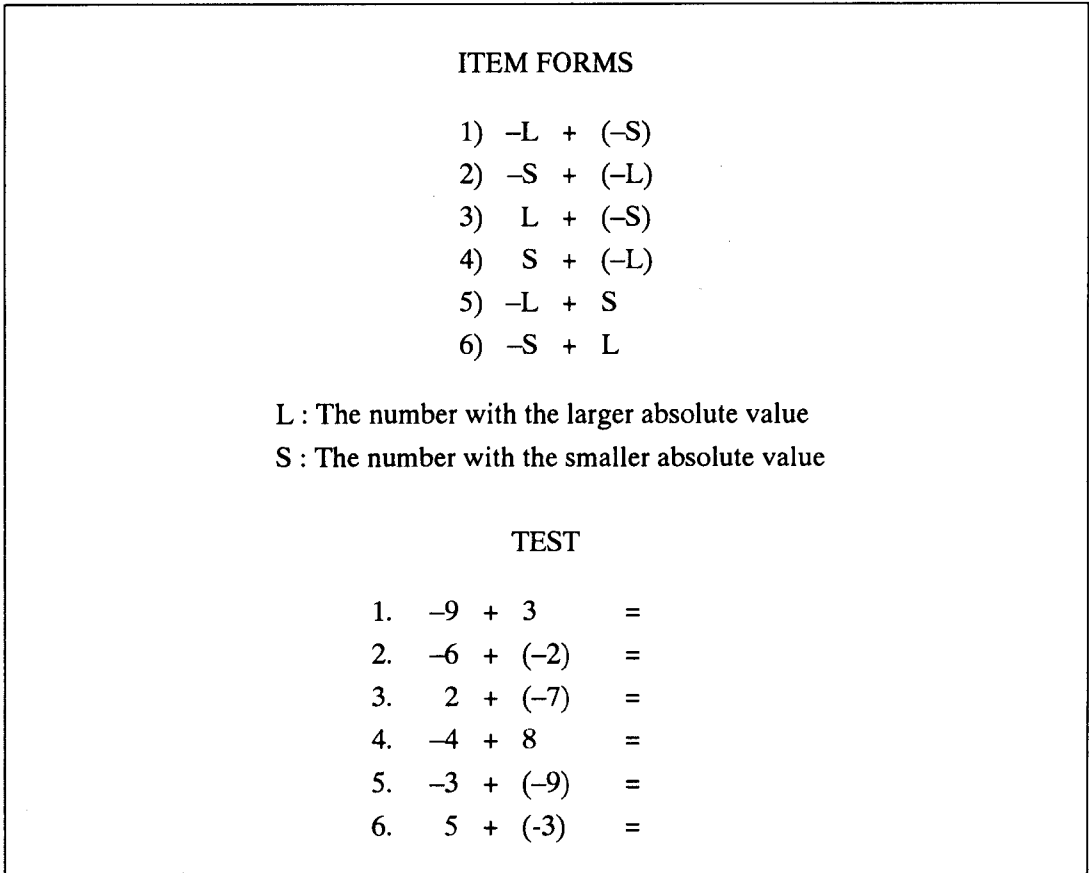


FIGURE I The six item forms for the content of the signed-number addition and the test.

The last step was the two-step diagnostic process : exploratory diagnosis and confirmatory diagnosis.

**Exploratory Diagnosis** : After formulating the response sets of each rule, the student's response set was compared with the response set of the rule one by one. All of 126 rules generated unique response sets. Therefore, the results of the comparisons were classified into three cases.

**CASE I** : Student's response set was the same as the response set of a rule. It will be diagnosed that he/she used that rule.

**EXAMPLE**

	Student's response	Rule produced the same response
1)	$-9 + 3 = \underline{-6}$	#1-5
2)	$-6 + (-2) = \underline{8}$	#1-5
3)	$2 + (-7) = \underline{-5}$	#1-5
4)	$-4 + 8 = \underline{-4}$	#1-5
5)	$-3 + (-9) = \underline{12}$	#1-5
6)	$5 + (-3) = \underline{2}$	#1-5

**CASE II** : Student's response set was not absolutely the same as the response set of any rules, but it was mostly like the response sets of only one rule. It will be diagnosed that he/she used that rule.

**EXAMPLE**

	Student's response	Rule produced the same response
1)	$-9 + 3 = \underline{-6}$	#1-7
2)	$-6 + (-2) = \underline{9}$	
3)	$2 + (-7) = \underline{-5}$	#1-7
4)	$-4 + 8 = \underline{4}$	#1-7
5)	$-3 + (-9) = \underline{12}$	#1-7
6)	$5 + (-3) = \underline{2}$	#1-7

(The response set of Rule 1-7 was -6, 8, -5, 4, 12, 2).

**CASE III** : Student's response set was not absolutely the same as the response set of any rules, but it was mostly like the response sets of more than one rule. This student had to be diagnosed in the confirmatory diagnosis process.

**EXAMPLE**

	Student's response	Rule produced the same response
1)	$-9 + 3 = \underline{-6}$	#1-1, #1-7
2)	$-6 + (-2) = \underline{-8}$	#1-1
3)	$2 + (-7) = \underline{-5}$	#1-1, #1-7
4)	$-4 + 8 = \underline{4}$	#1-1, #1-7
5)	$-3 + (-9) = \underline{12}$	#1-7
6)	$5 + (-3) = \underline{2}$	#1-1, #1-7

(The response set of Rule 1-1 was -6, -8, -5, 4 -12, 2

The response set of Rule 1-7 was -6, 8, -5, 4 12, 2)

**Confirmatory Diagnosis** : The students in case III had to be diagnosed by taking the confirmatory test. It is a true or false type test constructed from the previous test by selecting the items that are not be found all the suspect rules. From case III item I and item V are selected to construct the true or false type test which the provided answer of each item is generated from the rule that cannot be found in the his/her answer in the previous test.

### CONFIRMATORY TEST

$$1) \quad \quad \quad -6 + (-2) = 8$$

$$2) \quad \underline{\quad\quad} -3 + (-9) = -12$$

The provided answer of item 1 is generated from the rule 1-7 and the answer of item 2 is generated from the rule 1-1. The students' response can be generated into two cases.

**CASE I** : True and false or false and true response.

$$1) \quad \underline{/} -6 + (-2) = 8$$

$$2) \quad \underline{x} -3 + (-9) = -12$$

The student with these responses is diagnosed to have rule 1-7.

$$1) \quad \underline{x} -6 + (-2) = 8$$

$$2) \quad \underline{/} -3 + (-9) = -12$$

The student with these responses is diagnosed to have rule 1-1.

**CASE II** : All true or false response.

The student with all true or all false responses is diagnosed to have no consistent rules or use guessing.

## **VALIDITY AND RELIABILITY OF THE DEVELOPED METHOD**

The developed method was used to diagnose misconceptions of a sample of 940 grade 8 students. It was found that the diagnostic result of the developed method was stable and had criterion-related validity when the teachers' diagnostic result were used as a criterion. This method provided diagnostic result consistently with that of Rule Space model's, and was able to avoid the limitation of Rule Space model; i.e. – to discriminate misconceptions of the testees whose item score response patterns were the same or all zero's. In addition, the computer program written from this diagnostic method functioned as planned.

## **CONCLUSION**

The developed diagnostic method, by comparing the students' response sets to the rules' response sets with the confirmatory test for some students, facilitates diagnosis of the misconception that resulted in the erroneous rule. Students' responses are often affected by random errors produced by uncontrollable factors (Harnisch & Tatsuoka, 1991). The developed method enables us to take into account the careless, the erroneous rules, and the guessing. This approach for classifying students' error patterns avoid some of the weakness of personal indices, such as the problems of setting cut-off scores for determining atypicality of response patterns, dependence on a baseline ordering of items, and the lack of prescriptive information for remediating. In addition, it decreases the limitation of the Rule Space model; i.e. – to discriminate misconceptions of the testees whose item score response patterns were the same or all zero's.

## **ACKNOWLEDGEMENTS**

The author would like to gratefully acknowledge and thank thesis advisors, Asst. Prof. Nonglak Wiratchai, Ph.D. and Prof. Suwimon Wongwanich, Ph.D. Special thanks is extended to Dr. Harnisch and Mr. Choi from the University of Illinois at Urbana-Champaign for their assistance with the data analysis.

## REFERENCES

- Anderson, J.R. 1983. *The architecture of cognition*. Cambridge, MA : Harvard University Press.
- Birenbaum, M., et al. 1992. *Toward a stable diagnostic representation of students' errors in algebra* (Research Rep.). Princeton NJ : Educational Testing Service.
- Birenbaum, M. & Tatsuoka, K.K. 1987. Effects of on-line test feedback on the seriousness of subsequent errors. *Journal of Educational Measurement* 2 : 145-155.
- \_\_\_\_\_. 1992. Effect of response format on diagnostic assessment of scholastic achievement. *Applied Psychological Measurement* 16 : 353-363.
- Bloom, B. S. 1971 *Handbook on formative and summative of student learning*. New York : McGraw-Hill.
- Brown, F. G. 1978. 2nd ed. *Principle of education and psychological testing* New York : Holt, Rinehart and Winston.
- Drasgow, F. 1982. Choice of test model for appropriateness measure. *Applied Psychological Measurement* 6 : 297-308.
- Fisher, K. M. 1985. A misconceptions in biology : amino acid and transition. *Journal of Research in Science and Teaching* 22 : 53-62.
- Flavell, J. H. 1985. *Cognitive psychology*. New Jersey : Prentice-Hall.
- Gabel, D. L., et al. 1987. Understanding the particular nature of matter. *Journal of Chemical Education* 64 : 695-697.
- Good, C. V. 1973. *Dictionary of education*. New York : McGraw-Hill Book, Co.
- Gronlund, N. E. 1981. 4 th ed. *Measurement and evaluation in teaching* New York : Mcmillan Publishing Co, Inc.
- Halloun, I.A. and Hestenes D. 1985. Common sense concepts about motion. *The American Journal of Physics* 53 : 1056-1065.
- Harnisch D. L., & Linn, L. R. 1981. Analysis of item response pattern: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement* 3 : 39-87.
- Harnisch D.L., & Tatsuoka, K.K. 1983. A comparison of appropriateness indices based on item response theory. In R. Hambleton (Ed.) *Application of item response theory* Vancouver : British Columbia, 104-112.
- \_\_\_\_\_. 1991. *An application of IRT for diagnosing misconceptions* a paper presented in the proceedings of international educational statistics and measurement symposium. Taiwan : Educational Measurement Center, 12-20.

- Helm, H. 1980. Misconceptions in physics among south africa students. *Physics Educaton* 15 : 92-105.
- Hills, J. R. 1981. 2nd ed. *Measurement and evaluation in the classroom* Ohio : Bell & Howell Company.
- Kiokaew, S. 1988. Comparing college freshmen's concept of covalent bonding and the college of science and the college of education at Prince of Songkhla univesity, Thailand. Doctoral Dissertation, University of Missouri-Columbia.
- Levine, M. V., Rubin, D. B. 1979. Measuring the appropriateness of multiple chioce test scores. *Journal of Educational Statistics* 4 : 269-290.
- Lindquist, Franklin. 1956. *Educational measurement* Washington : American Council of Education.
- Mehrens, W. A. and Lehman, I. J. 1984. *Measurement and evaluation in education and psychology* New York : Holt, Rinehart and Winston.
- \_\_\_\_\_. 1980. 3rd ed. *Standardized tests in ecucation*. New York : Holt, Rinehart and Winston.
- Noll, V. H. and Scannell, D. P. 1972. 3rd ed. *Inrtoduction to educational measurement* Boston : Houghton Mifflin.
- Page, G. T. & Thomson, J. B. with Marshall, A. R. 1977. *International dictionary of education*. Great Britain : The Anchor Press Ltd.
- Peterson, R. F. and Treagust, D. F. 1989. Dvelopment and application of a diagnostic instrument to evaluate grade 11 and 12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science and Teaching* 26 : 301-314.
- Piaget, J. 1964. Cognitive development in the Child. *Journal of Research in Science and Teaching* 2 : 176-186.
- Pines, A. L. and West L. H. T. 1983. A framework for conceptual change special reference to misconceptions. In Novak, J. D. (ed.), *Proceedings of the international seminar misconceptions in science and mathematics* New York : Cornell University, 47-55.
- Ron, H. 1983. Enhancement and assessment of the reliability of instruments of the measurement of conceptual framework. In Novak, J. D. (ed.), *Proceedings of the international seminar misconceptions in science and mathematics* New York : Cornell University, 153-161.
- Sato, T. 1975. *The construction and interpretation of S-P tables* Tokyo : eiji Tosho.
- \_\_\_\_\_. 1982. *Application of S-P curve theory*. Tokyo : Meiji Tosho.

- Schmidt, H. J. 1987. Secondary school students' learning difficulties in stoichiometry. In Novak, J. D. (ed.), *Misconceptions and educational strategies in science and mathematics* New York : Cornell University, 396-404.
- Seigler, R. S. 1976. Three aspects of cognitive development. *Cognitive psychology* 8 : 481-502.
- \_\_\_\_\_. 1978. The origins of scientific reasoning. In R. S. Seigler (Ed.), *Children's thinking : What develops?* Hillsdale, N.J. : Erlbaum.
- Simson, W. W. and Marek A. E. 1988. Understanding and misconceptions of biology concepts holds by students attending small high school. *Journal of Research in Science and Teaching* 25 : 361-374.
- Singha, H. S. 1974. *Modern educational testing*. New Delhi : Sterling Pub.
- Sheehan, K. & Tatsuoka, K. K. 1992. *A diagnostic classification model for problem solving skill* (Research Rep.). Princeton NJ : Educational Testing Service.
- Sternburg, J.R. 1982. *Handbook of human intelligence*. London : Cambridge University Press.
- Tatsuoka, K.K. 1983a. Rule space : An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement* 20 : 345-354.
- \_\_\_\_\_. 1987. Validation of cognitive sensitivity for item response curves. *Journal of Educational Measurement* 24 : 233-245.
- \_\_\_\_\_. 1983b. Changes in error type over learning stage. *Journal of Educational Psychology* 76 : 120-129.
- \_\_\_\_\_. 1984. Caution indices based on item response theory. *Psychometrika* 9 : 95-110.
- \_\_\_\_\_. 1985. A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics* 12 : 55-73.
- \_\_\_\_\_. 1986. Diagnosing cognitive error : Statistical pattern classification and recognition approach. *Behaviormetrika* 19 : 193-206.
- \_\_\_\_\_. 1991. *Item construction and psychometric model appropriate for constructed responses* (Research Rep. No. RR-91-49-ONR) Princeton NJ : Educational Testing Service, 73-86.
- Tatsuoka, K. K. & Birenbaum, M. 1992. *Proficiency scaling based on attribute characteristic curves* (Research Rep.) Princeton NJ : Educational Testing Service.
- Tatsuoka, K. K. & Linn, L. R. 1983. Indices for detecting unusual patterns : Links between two general approaches and potential application. *Applied Psychological Measurement* 7 : 81-96.
- Tatsuoka, K. K., & Tatsuoka, M. M. 1982. Detection of aberrant response patterns. *Journal of Educational Statistics* 7 : 215-231.

- Tatsuoka, K.K., & Tatsuoka, M.M. 1983. Spotting erroneous rules of operation by individual consistency index. *Journal of Educational Measurement*. 20 : 221-230.
- Tatsuoka, K. K., & Tatsuoka, M. M. 1987. Bug distribution and pattern classification. *Psychometrika* 52 : 193-206.
- Treagust, D. 1986. Evaluating students' misconceptions by means of diagnostic multiple choice items. *Research in Science Education* 16 : 199-207.
- Wright, B. D., & Stone, M. H. 1979. *Best design, Rasch measurement*. Chicago : MESA Press.

# ผลของการแจ้งให้ทราบล่วงหน้าต่างวิธี ตำแหน่งหน้าที่ ของผู้ตอบและการติดตามทางไปรษณีย์ที่มีต่ออัตรา การตอบกลับและความจริงใจ ในการตอบแบบ สอบถามที่ส่งทางไปรษณีย์\*

วิระยุทธ ชატะกาญจน์

## ABSTRACT

The purposes of this study were to compare the effects of differrent pre-notification methods, respondents' positions, and follow-up mailings on response rates and sincere responses of mailed qestionnaires. The subjects were 1,500 teachers in secondary schools under the jurisdiction of the Department of General Education, Ministry of Education, located in Bangkok metropolis, equally classified by the positions of school administrationtors, heads of department, and instructors. The 5×3×2 factorial design was used with the analysis of 3-way ANOVA of three factors with one percentage per cell followed by post hoc of Z-Proportion. The 3-way ANOVA followed by post hoc of Student-Newman-Keuls was used for the analysis of sincere responses. The findings were as follows;

1. All methods of pre-notifications significantly effected on response rates, and pre-notification by telephone had higher response rates than using cards, postcards, and letters.
2. The response rates of the instructors were found higher than those of heads of departments and adminitrators.
3. The follow-up mailings effected higher response than nonfollow-ups, but sincere responses were not differrent.
4. There was an interaction between pre-notification methods and respondents' positions on sincere responses, those instructors with pre-notification by telephone had highest sincere responses.

---

\*วิทยานิพนธ์ครุศาสตรบัณฑิตสาขาวิจัยการศึกษา ปีการศึกษา 2537 โดยมีศาสตราจารย์ ดร.สมหวัง พิธิยานุวัฒน์ เป็นอาจารย์ที่ปรึกษา

## บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบผลของวิธีการติดต่อแจ้งให้ทราบล่วงหน้า ตำแหน่งหน้าที่ของผู้ตอบและการติดตาม ที่มีต่ออัตราการตอบกลับและความจริงใจในการตอบแบบสอบถามที่ส่งทางไปรษณีย์ โดยใช้กลุ่มตัวอย่างเป็นครู-อาจารย์ โรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ในเขตกรุงเทพมหานคร 1,500 คน ซึ่งมีตำแหน่งเป็นผู้บริหาร หัวหน้าหมวดวิชา และครูผู้สอน ตำแหน่งละ 500 คน ได้ออกแบบการทดลองโดยใช้วิธีการ Factorial Design แบบ  $5 \times 3 \times 2$  ระดับ วิเคราะห์ผลที่มีต่ออัตราการตอบกลับ โดยการใช้วิธีการวิเคราะห์ 3-Way ANOVA ชนิดที่การวัดเป็นร้อยละ 1 จำนวน ใน 1 รายการ และทดสอบภายหลังด้วยวิธีการทดสอบค่าสัดส่วนซี (Z-proportion) สำหรับความจริงใจในการตอบแบบสอบถามได้ใช้วิธีการวิเคราะห์ 3-Way ANOVA และทดสอบภายหลังด้วยวิธีการของสตีเวนสัน-นิวแมน-คูลส์ ผลการวิจัยพบว่า

1. การแจ้งให้ทราบล่วงหน้าทุกวิธีมีอัตราการตอบกลับของแบบสอบถามสูงกว่าการไม่แจ้งให้ทราบล่วงหน้าและการติดต่อแจ้งให้ทราบล่วงหน้าโดยการใช้โทรศัพท์มีอัตราการตอบกลับสูงกว่าการติดต่อโดยการส่ง การ์ด ไปรษณีย์บัตรและจดหมาย
2. ผู้ตอบแบบสอบถามที่มีตำแหน่งเป็นครูผู้สอนมีอัตราการตอบกลับแบบสอบถามสูงกว่าผู้ตอบแบบสอบถามที่มีตำแหน่งเป็น หัวหน้าหมวดวิชา และผู้บริหาร
3. การติดตามแบบสอบถามทำให้มีอัตราการตอบกลับของแบบสอบถามสูงกว่าการไม่มีการติดตามแบบสอบถามแต่ผู้ตอบมีความจริงใจในการตอบไม่แตกต่างกัน
4. มีปฏิสัมพันธ์ระหว่างวิธีการแจ้งให้ทราบล่วงหน้าและตำแหน่งหน้าที่ของผู้ตอบต่อความจริงใจในการตอบแบบสอบถามโดยผู้ตอบที่มีตำแหน่งเป็นครูผู้สอนและได้รับแจ้งให้ทราบล่วงหน้าโดยการใช้โทรศัพท์มีความจริงใจในการตอบสูงสุด

## ความเป็นมาและความสำคัญของปัญหา

การวิจัยเป็นการประดิษฐ์และการคิดค้นหาความจริงที่เกิดจากความอยากรู้หรือข้อสงสัยของมนุษย์ต่อปรากฏการณ์ของชีวิตและธรรมชาติ โดยมีวิธีการแสวงหาคำตอบของปัญหาอย่างมีระบบ ใช้วิธีการทางวิทยาศาสตร์มีระเบียบแบบแผนหรือขั้นตอนในการดำเนินงานที่เฉพาะ แต่ละขั้นตอนมีความสัมพันธ์ต่อเนื่องกันอย่างใกล้ชิดและส่งผลกระทบซึ่งกันและกัน ขั้นตอนที่สำคัญมากขั้นตอนหนึ่ง คือ ขั้นตอนการเก็บรวบรวมข้อมูลซึ่งในปัจจุบันการวิจัยทาง

พฤติกรรมศาสตร์เกือบร้อยละ 80 นิยมใช้แบบสอบถามเป็นเครื่องมือในการเก็บรวบรวมข้อมูล (อุทุมพร จามรมาน, 2530) และจากงานวิจัยของ วีระศักดิ์ บรรณารธรรม (2523) และรัตติกร ทิมเจริญ (2531) พบว่าแบบสอบถามที่ส่งทางไปรษณีย์เป็นเครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลมากที่สุด เนื่องจากการส่งและคืนแบบสอบถามทางไปรษณีย์เป็นวิธีที่สะดวก ประหยัด ทรัพยากร ทั้งกำลังคนและเงิน นอกจากนี้ยังสามารถใช้ครอบคลุมกลุ่มตัวอย่างจำนวนมากที่กระจายตามพื้นที่ต่าง ๆ ได้

แต่การเก็บรวบรวมข้อมูลด้วยแบบสอบถามที่ส่งทางไปรษณีย์มีปัญหาที่สำคัญ คือ กลุ่มตัวอย่างจะส่งแบบสอบถามกลับคืนมาในอัตราที่ค่อนข้างต่ำและไม่สามารถที่จะแน่ใจได้ว่าคำตอบที่ได้รับกลับคืนมา ผู้ตอบตั้งใจตอบอย่างจริงจังหรือตอบแบบสุ่มโดยไม่ได้อ่านข้อคำถามหรือคิดอะไรเลย (SAX, 1979) และสมาคมการศึกษาของอเมริกา (NEA) พบว่าข้อมูลของแบบสอบถามที่จะถือได้ว่าเป็นตัวแทนของประชากรได้ นักวิจัยต้องได้รับแบบสอบถามกลับคืนมามากกว่าร้อยละ 90 (Engelhart, 1972) และอัตราการตอบกลับแบบสอบถามที่ต่ำที่สุด ซึ่งทำให้ตัวประมาณค่าไม่มีความลำเอียง คือ อัตราการตอบกลับตั้งแต่ร้อยละ 95 (สุณี พนิตประชา)

เทคนิควิธีการเพิ่มอัตราการตอบกลับที่สามารถเพิ่มอัตราการตอบกลับให้สูงขึ้นอย่างได้ผล คือวิธีการแจ้งให้ทราบล่วงหน้าก่อนที่จะส่งแบบสอบถามไปให้ หรือการติดตามหลังการส่งแบบสอบถาม (Fox และคณะ, 1988) แต่ในปัจจุบันนักวิจัยสามารถทำการติดต่อกับผู้ตอบแบบสอบถามได้หลายวิธี เช่น โทรศัพท์ การ์ด จดหมาย และไปรษณียบัตร ซึ่งยังไม่สามารถทราบได้ว่าการติดต่อกับผู้ตอบโดยการวิธีการใดจะสามารถเพิ่มอัตราการตอบกลับได้สูงสุด ประกอบกับสังคมไทยในปัจจุบันยังให้ความสำคัญกับบทบาท ตำแหน่งหน้าที่ของบุคคลอยู่มาก โดยที่ในต่างประเทศ Berdie และ Anderson พบว่าการติดตามแต่ละวิธีมีประสิทธิภาพแตกต่างกันตามตำแหน่งหน้าที่ของผู้ตอบ ดังนั้นตำแหน่งหน้าที่ของผู้ตอบจึงเป็นตัวแปรที่น่าจะศึกษาว่า ในสังคมไทยนั้นตำแหน่งหน้าที่ของผู้ตอบจะมีผลกระทบต่ออัตราการตอบกลับของแบบสอบถามที่ส่งทางไปรษณีย์หรือไม่ และงานวิจัยที่จะก่อให้เกิดประโยชน์อย่างแท้จริงนั้น ต้องเป็นงานวิจัยที่ให้ความรู้หรือข้อเท็จจริงใหม่ ๆ ได้อย่างถูกต้อง ดังนั้นสิ่งที่ควรคำนึงถึงนอกจากอัตราการตอบกลับที่สูงแล้ว คุณภาพของข้อมูลที่ได้รับคืนมายังเป็นสิ่งที่สำคัญอีกประการหนึ่งด้วย จึงมีความจำเป็นที่จะต้องศึกษาเรื่องนี้ อันจะเป็นประโยชน์ต่อการวิจัยทางสังคมศาสตร์และพฤติกรรมศาสตร์ สามารถที่จะเพิ่มศักยภาพและความเชื่อถือได้ของการเก็บรวบรวมข้อมูลโดยใช้แบบสอบถามที่ส่งทางไปรษณีย์ ทำให้องค์ความรู้ของวิธีวิทยาการวิจัยมีความแข็งแกร่งยิ่งขึ้น

## วัตถุประสงค์ในการวิจัย

1. เพื่อเปรียบเทียบอัตราการตอบกลับและความจริงใจในการตอบแบบสอบถามเมื่อได้รับการแจ้งให้ทราบล่วงหน้าโดยการติดต่อที่แตกต่างกัน 5 วิธี คือ การติดต่อโดยใช้โทรศัพท์ การจด ไปรษณียบัตร จดหมาย และการไม่ติดต่อ
2. เพื่อเปรียบเทียบอัตราการตอบกลับและความจริงใจในการตอบแบบสอบถาม เมื่อผู้ตอบมีตำแหน่งหน้าที่แตกต่างกัน คือ ตำแหน่งผู้บริหาร หัวหน้าหมวดวิชา และครูผู้สอน
3. เพื่อเปรียบเทียบอัตราการตอบกลับและความจริงใจในการตอบแบบสอบถาม เมื่อผู้ตอบได้รับการติดตามกับการไม่ได้รับการติดตาม
4. เพื่อศึกษาปฏิสัมพันธ์ระหว่างวิธีการแจ้งให้ทราบล่วงหน้า ตำแหน่งหน้าที่ของผู้ตอบ และการติดตามที่มีต่ออัตราการตอบกลับและความจริงใจในการตอบแบบสอบถาม

## สมมุติฐานการวิจัย

1. กลุ่มที่ได้รับการแจ้งให้ทราบล่วงหน้าด้วยโทรศัพท์ น่าจะมีอัตราการตอบกลับและความจริงใจในการตอบแบบสอบถาม สูงกว่า กลุ่มที่ได้รับการแจ้งให้ทราบล่วงหน้าด้วยการจดไปรษณียบัตร จดหมาย และกลุ่มที่ไม่ได้รับการแจ้งให้ทราบล่วงหน้า และทุกกลุ่มน่าจะมีอัตราการตอบกลับและความจริงใจในการตอบแบบสอบถามสูงกว่ากลุ่มที่ไม่ได้รับการแจ้งให้ทราบล่วงหน้า
2. กลุ่มครูผู้สอนน่าจะมีอัตราการตอบกลับและความจริงใจในการตอบแบบสอบถาม สูงกว่ากลุ่มผู้บริหารและหัวหน้าหมวดวิชา
3. กลุ่มที่มีการติดตามแบบสอบถามน่าจะมีอัตราการตอบกลับแบบสอบถามสูงกว่ากลุ่มที่ไม่มีการติดตาม
4. กลุ่มที่ไม่มีการติดตามแบบสอบถามน่าจะมีความจริงใจในการตอบแบบสอบถาม สูงกว่ากลุ่มที่มีการติดตาม
5. กลุ่มที่มีตำแหน่งหน้าที่แตกต่างกัน เมื่อได้รับการแจ้งให้ทราบล่วงหน้าและการติดตามที่แตกต่างกัน น่าจะมีอัตราการตอบกลับและความจริงใจในการตอบแบบสอบถามที่แตกต่างกัน

## วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้เป็นการวิจัยเชิงทดลองภาคสนาม (Field Experiment) ซึ่งได้วางแผนการวิจัยแบบแฟคตอเรียล 3 ปัจจัย  $5 \times 3 \times 2$  ระดับ และผู้วิจัยได้ดำเนินการวิจัยตามขั้นตอนดังนี้

## ประชากรและกลุ่มตัวอย่าง

ประชากรมีอยู่ 3 กลุ่ม คือ เป็นผู้บริหาร หัวหน้าหมวดวิชา และครูผู้สอนในโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ที่ปฏิบัติราชการในปีการศึกษา 2537 ในเขตกรุงเทพมหานคร

การวิจัยในครั้งนี้ใช้กลุ่มตัวอย่างทั้งหมด 1,500 คน ซึ่งได้มาจากวิธีการสุ่มแบบแบ่งชั้น 2 ขั้นตอน (Two-Stage Stratified Random Sampling) โดยที่กลุ่มตัวอย่างมีตำแหน่งเป็น ผู้บริหาร หัวหน้าหมวดวิชา และครูผู้สอน ตำแหน่งละ 500 คน

## แบบแผนการทดลองและจำนวนกลุ่มตัวอย่างในแต่ละกลุ่มทดลอง

วิธีการติดต่อ	ไม่มีการติดตาม			วิธีการติดต่อ	มีการติดตาม 2 ครั้ง		
	ตำแหน่งหน้าที่				ตำแหน่งหน้าที่		
แจ้งให้ทราบ	ผู้บริหาร	หัวหน้าหมวดวิชา	ครูผู้สอน	แจ้งให้ทราบ	ผู้บริหาร	หัวหน้าหมวดวิชา	ครูผู้สอน
ล่วงหน้า	50	50	50	ล่วงหน้า	50	50	50
โทรศัพท์	50	50	50	โทรศัพท์	50	50	50
การ์ด	50	50	50	การ์ด	50	50	50
ไปรษณีย์	50	50	50	ไปรษณีย์	50	50	50
จดหมาย	50	50	50	จดหมาย	50	50	50
ไม่แจ้ง	50	50	50	ไม่แจ้ง	50	50	50

----- กลุ่มควบคุม -----

## เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล

1. แบบสอบถามเรื่องคุณลักษณะที่พึงประสงค์ของผู้บริหารโรงเรียนมัธยมศึกษาสังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ในเขตกรุงเทพมหานคร ซึ่งผู้วิจัยได้พัฒนาขึ้นจากแบบสอบถามของ ทศนีย์ เอี่ยมผ่อง (2535) มีข้อกระทงคำถามทั้งหมด 70 ข้อ โดยมีข้อกระทงคู่ขนานที่ใช้วัดความจริงใจในการตอบแบบสอบถามแทรกปะปนอยู่ด้วย 9 คู่ และแบบสอบถามมีค่าความเที่ยงเท่ากับ .99
2. ข้อความที่แจ้งให้ทราบล่วงหน้า เป็นข้อความที่ใช้แนะนำตัวผู้วิจัย หัวข้อที่จะทำวิจัยและขอความร่วมมือในการตอบแบบสอบถามไว้ล่วงหน้า ซึ่งทุกวิธีการแจ้งให้ทราบล่วงหน้ามีข้อความที่เหมือนกัน
3. จดหมายนำ ในการวิจัยครั้งนี้ใช้จดหมายนำเพื่อขอความร่วมมือในการตอบแบบสอบถาม 2 ฉบับ ซึ่งออกโดยกองการมัธยมศึกษา กรมสามัญศึกษา และอีกฉบับออกโดยผู้วิจัยเอง
4. จดหมายติดตามครั้งที่ 1 และครั้งที่ 2 เป็นข้อความที่กล่าวอ้างถึงแบบสอบถามที่ได้จัดส่งไปให้และข้อความที่เกี่ยวกับการที่ผู้วิจัยยังไม่ได้รับแบบสอบถามกลับคืน การขอร้องให้กลุ่มตัวอย่างเห็นความสำคัญของการตอบกลับแบบสอบถาม

## การเก็บรวบรวมข้อมูล

ใช้เวลาในการเก็บรวบรวมข้อมูลทั้งหมด 7 สัปดาห์ คือ ตั้งแต่วันที่ 23 มกราคม - 10 มีนาคม 2538

### เริ่มการทดลอง

### สิ้นสุดการทดลอง

1 สัปดาห์	2 สัปดาห์	2 สัปดาห์	2 สัปดาห์
23-27 ม.ค. 38	30 ม.ค.-10 ก.พ.	13-24 ก.พ.	27 ก.พ.-10 มี.ค.
แจ้งให้ทราบ	ส่งแบบ	ติดตาม	ติดตาม
ล่วงหน้า	สอบถาม	ครั้งที่ 1	ครั้งที่ 2

## การวิเคราะห์ข้อมูล

ทำการวิเคราะห์ข้อมูลโดยการคำนวณอัตราการตอบกลับของแบบสอบถามในกลุ่มทดลองแต่ละกลุ่มเป็นค่าร้อยละ ส่วนความสนใจในการตอบแบบสอบถามจะตรวจให้คะแนนตามวิธีการตรวจความสนใจในการตอบของผู้ตอบแบบสอบถามโดยจำแนกตามระยะเวลาในการตอบกลับของแบบสอบถาม และวิเคราะห์ผลของเทคนิควิธีการแจ้งให้ทราบล่วงหน้า ตำแหน่งหน้าที่ของผู้ตอบและวิธีการติดตาม ที่มีต่ออัตราการตอบกลับของแบบสอบถาม โดยการใช้สถิติวิเคราะห์ความแปรปรวน 3 ทาง (3-Way ANOVA) ชนิดที่การวัดเป็นความถี่ร้อยละ 1 จำนวน ใน 1 รายการ และทดสอบเป็นรายคู่ ภายหลังจากวิเคราะห์ด้วยวิธีการทดสอบค่าสัดส่วนซี (Z-proportion) สำหรับความสนใจในการตอบแบบสอบถามได้ทำการวิเคราะห์ โดยใช้สถิติวิเคราะห์ความแปรปรวน 3 ทาง (3-Way ANOVA) และทำการทดสอบเป็นรายคู่ภายหลังจากวิเคราะห์ด้วยวิธีการของ สตีวเดนส์-นิวแมน-คูลส์ (SNK)

## สรุปผลการวิจัย

1. กลุ่มที่ได้รับการแจ้งให้ทราบล่วงหน้าด้วยการใช้โทรศัพท์มีอัตราการตอบกลับของแบบสอบถามสูงกว่ากลุ่มที่ได้รับการแจ้งให้ทราบล่วงหน้าด้วยการดี ไปรษณียบัตร จดหมาย และกลุ่มที่ไม่ได้รับการแจ้งให้ทราบล่วงหน้า และทุกกลุ่มที่ได้รับการแจ้งให้ทราบล่วงหน้ามีอัตราการตอบกลับสูงกว่ากลุ่มที่ไม่ได้รับการแจ้งให้ทราบล่วงหน้าอย่างมีนัยสำคัญที่ระดับ .05
2. กลุ่มที่มีตำแหน่งเป็นครูผู้สอนมีอัตราการตอบกลับของแบบสอบถามสูงกว่ากลุ่มผู้ตอบที่มีตำแหน่งเป็นผู้บริหารและหัวหน้าหมวดวิชา และกลุ่มผู้ตอบที่มีตำแหน่งเป็นหัวหน้าหมวดวิชา มีอัตราการตอบกลับของแบบสอบถามสูงกว่ากลุ่มผู้ตอบที่มีตำแหน่งเป็นผู้บริหารอย่างมีนัยสำคัญทางสถิติที่ระดับ .05
3. เฉพาะในระยะเวลาการตอบกลับหลังการติดตาม กลุ่มที่มีการติดตามแบบสอบถามจะมีอัตราการตอบกลับของแบบสอบถามสูงกว่ากลุ่มที่ไม่มีการติดตามแบบสอบถามอย่างมีนัยสำคัญทางสถิติที่ระดับ .05
4. กลุ่มที่มีการติดตามและไม่มีการติดตามแบบสอบถามมีความสนใจในการตอบแบบสอบถามไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05
5. มีปฏิสัมพันธ์ระหว่างวิธีการแจ้งให้ทราบล่วงหน้าและตำแหน่งหน้าที่ของผู้ตอบต่อความสนใจในการตอบแบบสอบถามอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

6. กลุ่มผู้ตอบที่มีตำแหน่งเป็นครูผู้สอนและได้รับการแจ้งให้ทราบล่วงหน้าโดยการใช้อีเมล มีความจริงใจในการตอบแบบสอบถามสูงกว่ากลุ่มอื่น ๆ อย่างมีนัยสำคัญที่ระดับ .05

## อภิปรายผลการวิจัย

1. การแจ้งให้ทราบล่วงหน้าด้วยการใช้อีเมลมีอัตราการตอบกลับของแบบสอบถามสูงกว่ากลุ่มที่ได้รับการแจ้งให้ทราบล่วงหน้าด้วยการจด ไปรษณียบัตร จดหมายและกลุ่มที่ไม่ได้รับการแจ้งให้ทราบล่วงหน้า และทุกกลุ่มที่ได้รับการแจ้งให้ทราบล่วงหน้ามีอัตราการตอบกลับสูงกว่ากลุ่มที่ไม่ได้รับการแจ้งให้ทราบล่วงหน้า ซึ่ง Dillman (1987) ได้กล่าวถึงเหตุผลที่การแจ้งให้ทราบล่วงหน้าสามารถเพิ่มอัตราการตอบกลับของแบบสอบถามให้สูงขึ้นเพราะ

1. เป็นการแจ้งให้ผู้ตอบทราบล่วงหน้าว่าจะมีการส่งแบบสอบถามไปให้ตอบ ช่วยลดการทิ้งแบบสอบถามโดยไม่ได้ตั้งใจของผู้ตอบที่มีความสนใจ
2. เป็นการให้ความไว้วางใจ เห็นความสำคัญของผู้ตอบภายใต้ทฤษฎีการแลกเปลี่ยนทางสังคม

ผลการวิจัยที่ได้สอดคล้องกับงานวิจัยหลาย ๆ เรื่อง ที่พบว่า การแจ้งให้ทราบล่วงหน้าโดยการใช้อีเมลหรือจดหมาย มีอัตราการตอบกลับของแบบสอบถามสูงกว่าการไม่แจ้งให้ทราบล่วงหน้า อีกทั้งการแจ้งให้ทราบล่วงหน้าโดยการใช้อีเมลเป็นการสื่อสาร 2 ทาง ซึ่งผู้ตอบแบบสอบถามเกิดความพึงพอใจในการสื่อสารมากที่สุดจึงทำให้ได้รับอัตราการตอบกลับสูงกว่าการแจ้งให้ทราบล่วงหน้าด้วยวิธีการอื่น ๆ

2. ครูผู้สอนมีอัตราการตอบกลับของแบบสอบถามสูงกว่ากลุ่มผู้ตอบที่มีตำแหน่งเป็นผู้บริหารและหัวหน้าหมวดวิชา และกลุ่มผู้ตอบที่มีตำแหน่งเป็นหัวหน้าหมวดวิชา มีอัตราการตอบกลับของแบบสอบถามสูงกว่ากลุ่มผู้ตอบที่มีตำแหน่งเป็นผู้บริหาร ทั้งนี้ Wofford และคณะ (1977) กล่าวว่า ระดับตำแหน่งของบุคคลภายในองค์กรเป็นปัจจัยสำคัญที่กำหนดความพึงพอใจในการรับและส่งข่าวสาร โดยผู้ที่มีตำแหน่งต่ำจะมีความพึงพอใจในการรับและส่งข่าวมากกว่าผู้ที่มีตำแหน่งสูง บุคคลที่มีตำแหน่งสูงมีแนวโน้มที่จะได้รับข่าวสารที่จำเป็นสำหรับการปฏิบัติงานได้มากกว่า บุคคลที่มีตำแหน่งในองค์กรต่ำ จึงทำให้บุคคลที่มีตำแหน่งสูงให้ความสำคัญหรือให้ความสนใจกับข่าวสารที่ไม่มีผลต่อการปฏิบัติงานโดยตรงน้อยกว่าผู้ที่มีตำแหน่งต่ำ และผู้ที่มีตำแหน่งสูงจำเป็นต้องมีภาระงานในความรับผิดชอบมากขึ้น จึงทำให้ผู้ที่มีตำแหน่งหน้าที่สูงกว่ามีอัตราการตอบแบบสอบถามน้อยกว่าผู้ที่มีตำแหน่งหน้าที่ในองค์กรต่ำกว่า

3. ในระยะเวลาการตอบกลับหลังการติดตามทั้ง 2 ครั้ง กลุ่มที่มีการติดตามแบบสอบถามมีอัตราการตอบกลับของแบบสอบถามสูงกว่ากลุ่มที่ไม่มีการติดตามแบบสอบถาม ผลการวิจัยที่ได้สอดคล้องกับงานวิจัยของ Fox และ คณะ (1988) ที่ได้วิเคราะห์เมทาดังงานวิจัยที่เกี่ยวกับเทคนิคการเพิ่มอัตราการตอบแบบสอบถามที่ส่งทางไปรษณีย์ พบว่าการติดตามสามารถเพิ่มอัตราการตอบกลับของแบบสอบถามให้สูงขึ้นได้ร้อยละ 11-35 และในการวิจัยครั้งนี้ พบว่าการติดตามแบบสอบถามสามารถเพิ่มอัตราการตอบกลับให้สูงขึ้นได้ร้อยละ 18.14 ทั้งนี้เป็นไปตามที่ Dillman (1978) ได้กล่าวถึงเหตุผลที่การติดตามสามารถเพิ่มอัตราการตอบกลับให้สูงขึ้นได้เพราะ

1. เป็นรูปแบบของการได้รับรางวัลตามทฤษฎีการแลกเปลี่ยนทางสังคมที่เกิดจากความตั้งใจ การรับรู้คุณค่าของการติดตาม

2. การติดตามเป็นการกระตุ้นเตือนผู้ตอบว่าเขาลืมตอบแบบสอบถาม

3. การติดตามและไม่มีการติดตามแบบสอบถามมีความจริงจังในการตอบแบบสอบถามไม่แตกต่างกัน จากการวิจัยเรื่องนี้ทำให้ได้ข้อมูลเพิ่มเติมว่าการติดตามและไม่ติดตามแบบสอบถามนั้นทำให้ผู้ตอบมีความจริงจังในการตอบแบบสอบถามไม่แตกต่างกัน ซึ่งเป็นการสนับสนุนผลการวิจัยในครั้งก่อน ๆ ที่พบว่า คำตอบจากแบบสอบถามที่ตอบกลับมาซ้ำมีคุณภาพลดลงนั้นไม่ได้มีสาเหตุมาจากการติดตาม แต่เป็นเพราะลักษณะเฉพาะตัวของกลุ่มตัวอย่างที่ตอบแบบสอบถามกลับคืนมาซ้ำ

4. มีปฏิสัมพันธ์ระหว่างวิธีการแจ้งให้ทราบล่วงหน้าและตำแหน่งหน้าที่ของผู้ตอบต่อความจริงจังในการตอบแบบสอบถาม ซึ่งผลการวิจัยที่ได้สอดคล้องกับงานของ Berdie และ Anderson (1975) ที่พบว่าการติดตามแต่ละวิธีมีประสิทธิภาพแตกต่างกันขึ้นอยู่กับตำแหน่งหน้าที่ของผู้ตอบ ดังนั้นการแจ้งให้ทราบล่วงหน้าก็น่าจะเหมือนกับการติดตามที่มีประสิทธิภาพแตกต่างกันไปตามตำแหน่งหน้าที่ของผู้ตอบแบบสอบถาม โดยที่ลักษณะของปฏิสัมพันธ์ที่พบมีทั้งเป็นแบบมีอันดับ (ordinal) และไร้อันดับ (disordinal) ส่วนเหตุผลที่ผู้ตอบที่มีตำแหน่งเป็นครูผู้สอนและได้รับการแจ้งให้ทราบล่วงหน้าโดยการใช้โทรศัพท์มีความจริงจังในการตอบแบบสอบถามสูงกว่ากลุ่มย่อยอื่น ๆ นั้น เพราะการแจ้งให้ทราบล่วงหน้าโดยการใช้โทรศัพท์ทำให้เกิดความเข้มของสิ่งเร้าในการเกิดพฤติกรรมมากกว่าวิธีการอื่น ๆ ทำให้เกิดแรงขับและแรงจูงใจใฝ่สัมพันธ์มากที่สุด และทำให้ผู้ตอบแบบสอบถามรู้สึกว่าคุณค่า มีความสำคัญ ผู้ตอบได้รับการสนองความต้องการในขั้นที่ 4 ตามทฤษฎีมานุษยนิยมของมาสโลว์ การติดต่อโดยการใช้โทรศัพท์ผู้ตอบเกิดความรู้สึกว่าอยู่ใกล้นักวิจัย ได้มีโอกาสสร้างความสัมพันธ์กับ

ผู้วิจัย ซึ่งตามแบบจำลองความพึงพอใจในการติดต่อสื่อสารกับระยะทางของการสื่อสารและทิศทางของการสื่อสาร จะทำให้ผู้ตอบแบบสอบถามเกิดความพึงพอใจในการติดต่อสื่อสารมากที่สุด ประกอบกับการที่ผู้ตอบแบบสอบถามมีตำแหน่งเป็นครูผู้สอนซึ่งเป็นระดับตำแหน่งที่ต่ำกว่าหัวหน้าหมวดวิชาและผู้บริหาร จึงทำให้มีความจริงใจในการตอบแบบสอบถามมากกว่ากลุ่มย่อยอื่น ๆ

## ข้อเสนอแนะในการนำไปใช้

1. ในการเก็บรวบรวมข้อมูลโดยใช้แบบสอบถามที่ส่งทางไปรษณีย์ก่อนที่จะจัดส่งแบบสอบถามไปให้กลุ่มตัวอย่าง ควรได้ติดต่อแจ้งให้ผู้ตอบทราบล่วงหน้าเสียก่อน และน่าจะติดต่อแจ้งให้ทราบล่วงหน้าโดยการใช้โทรศัพท์ ซึ่งจะทำให้ได้รับอัตราการตอบกลับของแบบสอบถามสูงสุด

2. งานวิจัยที่ใช้แบบสอบถามส่งทางไปรษณีย์เป็นเครื่องมือในการเก็บรวบรวมข้อมูลหากกลุ่มตัวอย่างเป็นผู้บริหารหรือเป็นบุคคลที่มีตำแหน่งหน้าที่ในองค์กรสูง ผู้วิจัยจำเป็นต้องใช้เทคนิควิธีต่าง ๆ เพื่อที่จะทำให้อัตราการตอบกลับของแบบสอบถามสูงขึ้น เช่น การแจ้งให้ทราบล่วงหน้าโดยการใช้โทรศัพท์ การติดตามแบบสอบถาม

3. ผู้วิจัยยังสามารถใช้วิธีการติดตามแบบสอบถาม สำหรับการทวงถามกลุ่มตัวอย่างที่ไม่ตอบกลับแบบสอบถามได้ เพราะการติดตามทำให้ได้รับอัตราการตอบกลับสูงขึ้นและความจริงใจในการตอบแบบสอบถามก็ไม่แตกต่างจากกลุ่มที่ไม่มีการติดตาม

4. หากผู้วิจัยต้องการที่จะให้ได้รับอัตราการตอบกลับที่สูงและผู้ตอบมีความจริงใจในการตอบสูงสุดเฉพาะผู้ตอบที่มีตำแหน่งเป็นครูผู้สอน ควรใช้เทคนิคการแจ้งให้ทราบล่วงหน้าโดยการใช้โทรศัพท์ก่อนที่จะส่งแบบสอบถามไปให้กลุ่มตัวอย่าง

## ข้อเสนอแนะสำหรับการทำวิจัยต่อไป

1. เนื่องจากมีงานวิจัยที่เกี่ยวกับอัตราการตอบกลับของแบบสอบถามอยู่เป็นจำนวนมากและผลการวิจัยบางเรื่องสนับสนุนและโต้แย้งกันอยู่ น่าจะได้ทำการสังเคราะห์งานวิจัย (Meta Analysis) เพื่อศึกษาค่าขนาดอิทธิพล (Effect size) ของตัวแปรต่าง ๆ ที่มีต่ออัตราการตอบกลับของแบบสอบถาม อันจะทำให้ข้อความรู้ที่เกี่ยวกับเรื่องนี้มีความถูกต้องน่าเชื่อถือมากขึ้น

2. ควรศึกษาเปรียบเทียบระหว่างวิธีการติดตามแบบสอบถามกับการปรับค่าของแบบสอบถามสำหรับกลุ่มตัวอย่างที่ไม่ตอบกลับแบบสอบถาม ว่าวิธีการใดสามารถประมาณค่าพารามิเตอร์ได้ถูกต้องมากกว่ากัน

### บรรณานุกรม

- จุมพล สวัสดิยากร. **หลักและวิธีการทางสังคมศาสตร์**. กรุงเทพมหานคร: โรงพิมพ์สุวรรณ-  
ภูมิ, 2520.
- บุญธรรม กิจปรีดาบริสุทธิ์. **ระเบียบวิธีการทางสังคมศาสตร์**. พิมพ์ครั้งที่ 6 กรุงเทพมหานคร: โรงพิมพ์สามเจริญพานิชย์ (กรุงเทพ) จำกัด, 2535.
- ภัสสร ลิมานนท์. การเก็บรวบรวมข้อมูลโดยการส่งและคืนแบบสอบถามทางไปรษณีย์ใน **จดหมายข่าวประชากร 36** (สิงหาคม 2529).
- รัตติกร ทิมเจริญ. **การวิเคราะห์ข้อความรู้ในวิทยานิพนธ์ครุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย ในช่วงปีการศึกษา 2525-2528**. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2523.
- วีระศักดิ์ บรรณาธรรม. **การวิเคราะห์วิทยานิพนธ์ระดับมหาบัณฑิตที่มีเนื้อหาเกี่ยวกับอุดมศึกษา**. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2528.
- สมยศ นาวิการ. **การสื่อสารขององค์การ**. กรุงเทพมหานคร: สำนักพิมพ์บรรณกิจ, 2527.
- สมหวัง พิธิยานุวัฒน์. “การเก็บรวบรวมข้อมูล” ในไพฑูริย์ สีนลารัตน์ และสำลี ทองธิว (บรรณาธิการ), **การวิจัยทางการศึกษา: หลักและวิธีการสำหรับนักวิจัย**. กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย, 2530.
- \_\_\_\_\_. **การวิจัยเชิงบรรยาย**. กรุงเทพมหานคร : สำนักพิมพ์โอเดียนสโตร์, 2525.
- สุณี พนิตประชา. **อัตราต่ำสุดของการตอบแบบสอบถามกลับคืนทางไปรษณีย์ที่ทำให้ตัวประมาณค่าไม่มีความลำเอียง**. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2531.
- อุทุมพร จามรมาน. **แบบสอบถาม: การสร้างและการใช้**. กรุงเทพมหานคร : ม.ป.ท. 2530.

- Alrect, P.L. and Robert B.S. **The Survey Research Handbook**. Illinois: Richard D. Irwin, Inc, 1985.
- Berdie , D.R. and Anderson,J.F. "Effect on response rate of formal and informal question-  
nair follow-up techniques". **Journal of Applied Psychology**. 60 (April 1975) :  
223-227.
- \_\_\_\_\_. **Questionnaire : Design And Use**. New Jersey : The Scareerow Press, 1974.
- Dillman D.A. **Mail and Telephone Survey : The Total Design Method**. New  
York: John wiley and sons, 1978.
- Engelhert, M.D. **Methods of Educational Research**. Chicago : Rand Menally, 1972.
- Fox, R.J. : Crask, M.R. and Kim, J. "Mail survey response rate: A meta analysis of selected  
techniques for inducing response". **Public Opinion Quartery**. 52 (1988) : 467-  
491.
- Winer, B.J. **Statistical Principle in Experimental Design**. 2nd. ed. Mcgraw-Hill  
Kogskuska, Ltd.,1971.
- Wofford, J.C. ,Gerloff, A.M. and Cummins, B.L. **Organnizational Communication**.  
Tokyo: Mc Graw-Hill Kogakuska, Ltd., 1977.

## การติดตามคุณภาพและประสิทธิภาพของ แบบประเมินตนเองสำหรับครูมัธยมศึกษา

เยาวดี วิบูลย์ศรี

### ความเป็นมาและความสำคัญของปัญหา

การประเมินตนเอง (Self Evaluation) เป็นวิธีการหนึ่งในการวัดบุคลิกภาพของแต่ละคน โดยผู้ที่ถูกวัดจะเป็นผู้ที่อธิบายหรือประเมินพฤติกรรมของตนเองด้วยการทำเครื่องหมายลงในแบบประเมินที่ตรงกับบุคลิกภาพของเขา หรือด้วยการจัดลำดับข้อความที่ตรงกับลักษณะนิสัยของเขา หรือด้วยการเขียนรายงานที่แสดงนิสัยประจำตัว ทักษะคติ ความสนใจ และค่านิยมของตนเอง การประเมินตนเองเป็นวิธีการที่สะดวก ประหยัดเวลาและทรัพยากร และที่สำคัญคือเป็นประโยชน์ต่อตนเองในการเพิ่มพูนและพัฒนาประสิทธิภาพการทำงานของตนให้ดีขึ้น (Campbell and Lee 1988:307)

การประเมินผลการปฏิบัติงานของครูนั้น กระทำเพื่อนำผลหรือข้อมูลที่ได้มาใช้เพื่อปรับปรุงการเรียนการสอนให้มีประสิทธิภาพได้ระดับมาตรฐานซึ่งผู้ที่จะทำหน้าที่ประเมินครูได้แก่นักเรียน เพื่อนครูด้วยกัน ผู้บังคับบัญชา และตัวครูเอง แต่ผู้ที่จะทำหน้าที่ประเมินการปฏิบัติงานของครูได้ดีที่สุดคือตัวครูเอง เพราะครูย่อมทราบดีว่าตนเองเตรียมวางแผน เตรียมการสอน และมีความรู้ในเนื้อหาเพียงใด นอกจากนั้น เมื่อพิจารณาในด้านการประเมินเพื่อพัฒนาตนเองแล้ว ยิ่งเป็นเหตุผลสำคัญในการส่งเสริมให้ใช้แบบประเมินตนเองสำหรับครู และมีข้อค้นพบจากงานวิจัยว่าผลการประเมินตนเองของครูจะมีความสอดคล้องกับผลการประเมินของผู้บริหารและเพื่อนครูในระดับที่ค่อนข้างสูง (ชรณี หมิ่นรักษ์ 2512:22-35 ; วิภา เกษตรศิริเจริญ 2521: 46-61) และมีค่าความเที่ยงสูงกว่าการประเมินประสิทธิภาพการสอนของครูโดยผู้เรียน

(Centra อ้างใน Kenenth, Doyle, and Webber 1972:72-73)

สำหรับแบบประเมินตนเองสำหรับครูมัธยมศึกษาที่มีเกณฑ์มาตรฐานนั้น ผู้วิจัยได้สร้างขึ้นใน พ.ศ. 2522 เรียกว่าแบบรายงานตนเองสำหรับครูมัธยมศึกษาตามที่ปรากฏในรายงานวิจัยฉบับปี พ.ศ. 2522 (เยาวดี วิบูลย์ศรี, 2522) แต่มาในยุคปัจจุบันที่แนวคิดเกี่ยวกับการประเมินตนเองได้รับการยอมรับกันอย่างแพร่หลาย จึงเห็นสมควรที่จะเปลี่ยนชื่อจากแบบรายงานตนเองเป็นแบบประเมินตนเอง ลักษณะของแบบประเมินในรายงานวิจัยฉบับปี 2522 นั้น ประกอบด้วยข้อรายการ 104 ข้อ ครอบคลุมลักษณะของครูที่ดี 6 ด้าน คือ ด้านวิชาการ 17 ข้อ ด้านการสอน 25 ข้อ ด้านคุณธรรม 17 ข้อ ด้านบุคลิกภาพ 20 ข้อ ด้านสังคม 15 ข้อ และด้านส่วนตัว 12 ข้อ โดยแบ่งการประเมินเป็น 7 ระดับ จากมากที่สุดไปจนถึงน้อยที่สุด เมื่อแบบประเมินฉบับนี้ผ่านการทดลองใช้แล้ว ผู้วิจัยได้นำแบบประเมินที่สร้างขึ้นมาหาค่าปกติเปอร์เซ็นต์ไทล์ (Percentile Norm) จากกลุ่มครูมัธยมศึกษาสายสามัญที่สังกัดกองการมัธยมศึกษา กรมสามัญศึกษา และสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน กระทรวงศึกษาธิการ จำนวน 1,491 คน และนำไปวิเคราะห์คุณภาพตามแนวทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) พบว่ามีค่าความเที่ยงประเภทความคงที่ภายใน (Cronbach's Alpha) เท่ากับ .97 มีค่าความตรงร่วมสมัย (Concurrent Validity) โดยการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ครูประเมินตนเองกับผู้บังคับบัญชาเป็นผู้ประเมินอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $r=.59$ ) และข้อรายการทั้ง 104 ข้อนี้ มีค่าอำนาจจำแนกตามลักษณะครูที่ดี โดยที่ผลการทดสอบด้วยสถิติทดสอบที (t-test) มีนัยสำคัญทางสถิติที่ระดับ .05

เนื่องจากการวิเคราะห์คุณภาพของแบบประเมินตนเองสำหรับครูมัธยมศึกษา ในรายงานวิจัยฉบับปี 2522 นั้น ได้อิงทฤษฎีการวัดแบบดั้งเดิม ซึ่งในปัจจุบันนี้มีผู้วิจารณ์ว่ามีจุดอ่อนหลายประการ เป็นต้นว่า ค่าสถิติของข้อกระทง (Item statistics) บางตัวได้แก่ ค่าความยาก (Difficulty) และค่าอำนาจจำแนก (Discrimination) ขึ้นอยู่กับกลุ่มตัวอย่างที่ถูกวัด ส่วนการเปรียบเทียบคุณลักษณะ (trait) ไต ๆ ในแต่ละบุคคลมีข้อจำกัดคือ ต้องเป็นคะแนนที่ได้จากแบบวัดเดียวกันหรือแบบวัดที่เป็นคู่ขนานกัน นอกจากนี้ คุณภาพของเครื่องมือ เช่น ค่าความเที่ยง (Reliability) ที่นิยามในลักษณะของแบบสอบคู่ขนานนั้น มีความเป็นไปได้ยากในทางปฏิบัติ และค่าความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement) ตามแนวทฤษฎีการวัดแบบดั้งเดิมนั้น กำหนดให้ผู้ตอบทุกคนมีค่าเท่ากัน ซึ่งภายหลังได้พบว่าผู้มีคะแนนปานกลางจะมีค่าความคลาดเคลื่อนมาตรฐานของการ

วัดน้อยกว่าผู้ที่มีคะแนนสูงหรือต่ำ (Hambleton and Swaminathan 1985: 2-3) ประการสุดท้าย คือการแปลผลจากคะแนนดิบ (Raw Score) ได้รับการวิจารณ์ว่าคะแนนดิบไม่ใช่คะแนนที่อยู่ในมาตราอันตรภาค (Interval Scale) เป็นผลให้ไม่สามารถแสดงความหมายใด ๆ บนมาตราเชิงเส้นได้ ซึ่งขัดกับหลักพื้นฐานของการวัดตัวแปร

เนื่องจากความสำคัญของแบบประเมินตนเองสำหรับมัธยมศึกษาที่ผู้วิจัยได้พัฒนาขึ้นในปี พ.ศ. 2522 จนถึงปัจจุบันเป็นเวลามากกว่า 10 ปีแล้ว และนับวันจะยังมีผู้นำไปใช้มากขึ้น ๆ ผู้วิจัยจึงเห็นสมควรที่จะได้ทำการติดตามผลด้านคุณภาพและประสิทธิภาพของแบบประเมินที่สร้างขึ้น โดยเน้นการวิเคราะห์คุณภาพตามวิธีมาตราส่วนประมาณค่าของแอนดริช และประสิทธิภาพในด้านความเป็นไปได้ และความเหมาะสมในการนำไปใช้ในเชิงปฏิบัติ เพื่อให้เกิดความมั่นใจในการนำไปใช้พัฒนาประสิทธิภาพของครูมัธยมศึกษาต่อไป

### วัตถุประสงค์ของการวิจัย

เพื่อติดตามคุณภาพและประสิทธิภาพของแบบประเมินตนเองสำหรับครูมัธยมศึกษาที่สร้างในปี พ.ศ. 2522 สำหรับการนำไปใช้ต่อไป โดยติดตามผลในด้านต่าง ๆ ดังนี้

1. คุณภาพของแบบประเมินติดตามด้วยวิธีการวิเคราะห์ตามมาตราส่วนประมาณค่าของแอนดริช ได้แก่การตรวจสอบด้วยวิธีต่อไปนี้

1.1 คุณภาพของข้อกระทงใช้วิธีการที่ประยุกต์มาจากแนวคิดของราสซ์ ซึ่งพิจารณาจากค่าความเหมาะสมของข้อกระทง (Item Fit)

1.2 ความเที่ยง (Reliability) วิเคราะห์ตามแบบความคงที่ภายในด้วยวิธีการที่ประยุกต์มาจากแนวคิดของราสซ์ เช่นกัน

1.3 ความตรงเชิงทฤษฎี (Construct Validity) โดยประเมินจากร้อยละของข้อกระทงในแบบประเมินฯ ที่มีความเหมาะสม (Fit) กับโมเดล

2. ประสิทธิภาพในด้านความเป็นไปได้ และความเหมาะสมในการนำแบบประเมินตนเองสำหรับครูมัธยมศึกษาไปใช้ในเชิงปฏิบัติ เพื่อให้เกิดการพัฒนาประสิทธิภาพของครู

## สมมติฐานของการวิจัย

เนื่องจากแบบประเมินตนเองสำหรับครูมัธยมศึกษาฉบับนี้ ได้สร้างขึ้นตามคุณลักษณะที่ดีของครู โดยคุณลักษณะดังกล่าวผู้วิจัยได้มาจากการสังเคราะห์แนวคิดของบุคคลสำคัญ ๆ ในวงการศึกษ และจากรายงานการวิจัยที่เกี่ยวข้องกับคุณลักษณะของครูที่ดี แม้ปัจจุบันด้านการศึกษามีการพัฒนาและก้าวหน้ามาก น่าจะยังคงเป็นศักยภาพที่จำเป็นสำหรับการเป็นครู จึงน่าจะคงมีประสิทธิภาพในการนำไปใช้ในปัจจุบันได้ ประกอบกับแบบประเมินฉบับนี้มีคุณภาพด้านความตรงและความเที่ยงในเกณฑ์ที่สูง รวมทั้งข้อกระทงทุกข้อสามารถจำแนกลักษณะของครูที่ดีได้ และเมื่อนำมาวิเคราะห์คุณภาพด้วยวิธีการของมาตราส่วนประมาณค่าของแอนดริช ซึ่งเป็นวิธีการวิเคราะห์ที่ละเอียดขึ้น ก็น่าจะมั่นใจได้ว่าแบบประเมินฉบับนี้จะยังคงมีคุณภาพด้านความตรงและความเที่ยงในเกณฑ์ที่สูงเช่นเดิม ดังนั้นผู้วิจัยจึงคาดหวังดังต่อไปนี้ว่า

1. แบบประเมินตนเองสำหรับครูมัธยมศึกษาที่ผู้วิจัยพัฒนาขึ้นน่าจะประกอบด้วยข้อกระทงที่มีคุณภาพตามวิธีการวิเคราะห์ด้วยวิธีการของมาตราส่วนประมาณค่าของแอนดริช และมีความเที่ยงและความตรงเชิงทฤษฎี เมื่อวิเคราะห์ด้วยวิธีการดังกล่าว
2. แบบประเมินตนเองสำหรับครูมัธยมศึกษาที่ผู้วิจัยพัฒนาขึ้น น่าจะมีประสิทธิภาพในด้านความเป็นไปได้ และความเหมาะสมในการนำไปใช้ในเชิงปฏิบัติเพื่อให้เกิดการพัฒนาประสิทธิภาพของครู

## เครื่องมือที่ใช้ในการวิจัย

1. แบบประเมินตนเองสำหรับครูมัธยมศึกษาที่ ผู้วิจัยได้สร้างขึ้น แล้วนำไปพัฒนาจนมีคุณภาพทั้ง 4 ด้าน คือ มีความตรงตามเนื้อเรื่อง (Content Validity) ความตรงร่วมสมัย (Concurrent Validity) ความเที่ยง (Reliability) และอำนาจจำแนก (Discrimination) ประกอบด้วยข้อรายการจำนวน 104 ข้อ ซึ่งวัดพฤติกรรมของครูที่ดี 6 ด้าน คือ ด้านวิชาการ (ข้อที่ 1-17) ด้านการสอน (ข้อที่ 18-42) ด้านคุณธรรม (ข้อที่ 43-57) ด้านบุคลิกภาพ (ข้อที่ 58-77) ด้านสังคม (ข้อที่ 78-92) และด้านส่วนตัว (ข้อ 93-104) ดังรายละเอียดในฉบับสมบูรณ์ที่อยู่ท้ายบทความนี้
2. แบบประเมินประสิทธิภาพในด้านความเป็นไปได้ และด้านความเหมาะสมในการนำแบบประเมินตนเองสำหรับครูมัธยมศึกษาไปใช้ในเชิงปฏิบัติเพื่อให้เกิดการพัฒนาประสิทธิภาพของครูนั้น ผู้วิจัยได้กำหนดข้อรายการไว้ 8 ข้อ คือ

- 1) ข้อคำถามชัดเจนดี
- 2) ถามได้ครอบคลุมลักษณะของครูที่ดี
- 3) ข้อรายการสามารถนำไปปฏิบัติได้
- 4) ครูได้ทราบสิ่งที่ยังบกพร่องและนำไปปรับปรุงแก้ไขในการปฏิบัติงานให้ดีขึ้น
- 5) การได้ประเมินตนเองของครู เพื่อประสิทธิภาพการทำงาน
- 6) ข้อรายการต่าง ๆ ใช้เป็นกรอบในการปฏิบัติงานได้
- 7) ผลการประเมินตนเองของครูจะเป็นประโยชน์ต่อการนิเทศงานของฝ่ายบริหาร
- 8) เป็นแนวทางในการจัดการฝึกอบรมเพื่อส่งเสริมสมรรถภาพของครู

แบบประเมินประสิทธิภาพฉบับนี้ ได้แบ่งระดับการประเมินออกเป็น 5 ระดับ โดยมีความหมายดังนี้

- |                           |             |                            |
|---------------------------|-------------|----------------------------|
| ถ้าประเมินในช่องหมายเลข 5 | หมายความว่า | ผู้ตอบเห็นด้วยอย่างยิ่ง    |
| ถ้าประเมินในช่องหมายเลข 4 | หมายความว่า | ผู้ตอบเห็นด้วย             |
| ถ้าประเมินในช่องหมายเลข 3 | หมายความว่า | ผู้ตอบไม่แน่ใจ             |
| ถ้าประเมินในช่องหมายเลข 2 | หมายความว่า | ผู้ตอบไม่เห็นด้วย          |
| ถ้าประเมินในช่องหมายเลข 1 | หมายความว่า | ผู้ตอบไม่เห็นด้วยอย่างยิ่ง |

## วิธีการวิจัย

การวิจัยครั้งนี้เป็นการติดตามผล (Follow-up Study) แบบประเมินตนเองสำหรับครูมัธยมศึกษาที่ผู้วิจัยสร้างขึ้นในปี 2522 โดยเป็นการวิเคราะห์คุณภาพของแบบประเมินตนเองสำหรับครูมัธยมศึกษาด้วยเทคนิคการวิเคราะห์แนวใหม่ คือวิธีมาตราส่วนประมาณค่าของแอนดริช และประเมินประสิทธิภาพในด้านความเป็นไปได้และด้านความเหมาะสมในการนำแบบประเมินตนเองสำหรับครูมัธยมศึกษาไปใช้ในเชิงปฏิบัติเพื่อให้เกิดการพัฒนาประสิทธิภาพของครู ดังรายละเอียดต่อไปนี้

### 1. ประชากรและกลุ่มตัวอย่าง

1.1 ประชากรในการวิจัยครั้งนี้ คือ ข้าราชการครูโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ กรุงเทพมหานครในภาคเรียนที่ 1 ปีการศึกษา 2531 จำนวนทั้งสิ้น 14,219 คน

1.2 กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ได้สุ่มมาจากประชากรในข้อ 1.1 และคำนวณหาขนาดพหุติของกลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ จำนวน 500 คน ด้วยวิธีการสุ่มแบบหลายขั้นตอน คือ สุ่มเขตบริการการศึกษา มา 4 เขต จากทั้งหมด 8 เขต และคำนวณขนาดพหุติของกลุ่มตัวอย่างในแต่ละเขตบริการการศึกษา แล้วสุ่มโรงเรียนในแต่ละเขตบริการการศึกษา โดยการสุ่มอย่างง่าย รวม 21 โรงเรียน และสุดท้ายคือคำนวณกลุ่มตัวอย่างแต่ละโรงเรียน ตามสัดส่วนที่คำนวณได้ (อ่านรายละเอียดรายงานวิจัยฉบับสมบูรณ์ 2537)

2. การเก็บรวบรวมข้อมูล การเก็บรวบรวมข้อมูลในการวิจัยครั้งนี้ มีขั้นตอนดังต่อไปนี้ คือ

2.1 ทำหนังสือขอความร่วมมือในการวิจัยจากกรมสามัญศึกษา กระทรวงศึกษาธิการ และไปติดต่อขอความร่วมมือกับทางโรงเรียน โดยทางโรงเรียนจะมีเจ้าหน้าที่ดำเนินการให้ตามรายชื่อของกลุ่มตัวอย่างที่ผู้วิจัยสุ่มได้

2.2 กลุ่มตัวอย่างแต่ละคนจะได้รับแบบประเมิน 2 ฉบับ คือ แบบประเมินตนเองสำหรับครูมัธยมศึกษาและแบบสำรวจความเป็นไปได้ และความเหมาะสมในการนำแบบประเมินตนเองสำหรับครูมัธยมศึกษาไปใช้ในเชิงปฏิบัติเพื่อให้เกิดการพัฒนาประสิทธิภาพของครู

2.3 หลังจากทีกลุ่มตัวอย่างได้รับแบบประเมินทั้ง 2 ชุด เป็นเวลาประมาณ 2 สัปดาห์ ผู้วิจัยจะไปรับคืนจากผู้ที่ทำหน้าที่รวบรวม

2.4 หลังจากนั้นตรวจสอบความถูกต้องของแบบประเมินทั้ง 2 ฉบับ คัดชุดที่ไม่สมบูรณ์ออกแล้วนำเอาคำตอบของกลุ่มตัวอย่างแต่ละคนมาลงในกระดาษลงรหัส (Coding Form) เพื่อจัดเตรียมข้อมูลลงเทปและนำไปวิเคราะห์ต่อไป

3. การวิเคราะห์ข้อมูล

3.1 การวิเคราะห์คุณภาพของแบบประเมินตนเองสำหรับครูมัธยมศึกษา มีขั้นตอนดังต่อไปนี้

1) นำข้อมูลที่ได้จากแบบประเมินตนเองมาวิเคราะห์ตัวประกอบ (Factor Analysis) โดยใช้โปรแกรม SPSS วิเคราะห์ ตัวประกอบสำคัญ (Principal Component : PC) แล้วหมุนแกนด้วยวิธีแวนิแม็กซ์ (Varimax) เพื่อตรวจสอบคุณสมบัติการวัดเพียงมิติเดียว (Unidimension Latent Space) ของแบบวัดซึ่งเป็นข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อกระทง

2) วิเคราะห์ค่าพารามิเตอร์ต่าง ๆ ของแบบประเมินตนเอง ตามวิธีการของมาตราส่วนประมาณค่าของแอนดริช โดยใช้โปรแกรม MICROSACLE

3) วิเคราะห์ความตรงเชิงทฤษฎี (Construct Validity) ของแบบประเมินตนเองสำหรับมัธยมศึกษาโดยคำนวณหาร้อยละของข้อกระทงทั้งหมดในแบบประเมินที่เหมาะสม (Fit) กับโมเดลโดยสูตรที่ใช้คำนวณดังนี้

สถิติที่ใช้วิเคราะห์ค่าความตรงเชิงทฤษฎี คือ

$$\text{Item Fit } t_i = (V_i - 1)(3/q_i) + (q_i/3)$$

เมื่อ  $V_i$  คือ Weighted Mean Square ข้อ  $i$

$q_i$  คือ Variance ของ  $v_i$

ตรวจสอบค่า  $t_i$  ที่ได้กับค่า  $t$  (t-test) จากตาราง โดยกำหนดนัยสำคัญทางสถิติที่ระดับ .01 และ  $df = n - 2$  เมื่อ  $n$  คือจำนวนข้อกระทง และคำนวณหาร้อยละของข้อกระทงทั้งหมดในแบบประเมินที่เหมาะสมกับโมเดล

4) วิเคราะห์หาความเที่ยง (Reliability) ของแบบประเมินตนเองสำหรับครูมัธยมศึกษา ตามวิธีการของมาตราส่วนประมาณค่าของแอนดริช ดังนี้

สถิติที่ใช้วิเคราะห์ค่าความเที่ยง คือ

ก. ในกรณีที่ไม่มีข้อกระทงใดในแบบวัดที่มีค่า Weighted Mean Square  $V_i$  มากกว่า 1 ประมาณค่าความเที่ยง โดยใช้สูตร

$$R = 1 - (\sigma^2_E / \sigma^2_X)$$

เมื่อ

$R$  คือ ค่าประมาณความเที่ยง

$\sigma^2_E$  คือ Mean Square Calibration Error

$\sigma^2_X$  คือ Observed Score Variance

ข. ในกรณีที่ข้อกระทงใด ๆ ในแบบวัดอย่างน้อย 1 ข้อ มีค่า  $V_i$  มากกว่า 1 ประมาณค่าความเที่ยง โดยใช้สูตร

$$R = 1 - [V(\sigma^2_E / \sigma^2_X)]$$

เมื่อ

$V$  คือ ค่าเฉลี่ย  $V_i$  ของข้อกระทงทุกข้อ ที่มีค่า  $V_i > 1$

(ดูรายละเอียดสูตรที่ใช้คำนวณ จากรายงานวิจัยฉบับสมบูรณ์ 2537)

3.2 การวิเคราะห์ประสิทธิภาพในด้านความเป็นไปได้และความเหมาะสมในการนำแบบประเมินตนเองสำหรับครูมัธยมศึกษาไปใช้ในเชิงปฏิบัติ เพื่อให้เกิดการพัฒนาประสิทธิภาพของครู กระทำโดยนำข้อมูลที่ได้จากแบบสำรวจความเป็นไปได้และความเหมาะสมในการนำแบบประเมินตนเองไปใช้มาคำนวณหาค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของคะแนนการตัดสินข้อรายการแต่ละข้อ และแปลผลการตัดสินข้อรายการแต่ละข้อจากค่าเฉลี่ยที่ได้ตามเกณฑ์ที่กำหนดขึ้น

## ผลการวิจัย\*

จากวัตถุประสงค์หลักของการวิจัยเรื่องนี้ มีข้อค้นพบที่สำคัญดังนี้

1. ผลการวิเคราะห์คุณภาพของแบบประเมินตนเองสำหรับครูมัธยมศึกษาด้วยวิธีการของมาตราส่วนประมาณค่าของแอนดริช

1.1 เนื่องจากวิธีการของมาตราส่วนประมาณค่าของแอนดริช ได้พัฒนามาจากแนวทฤษฎีการตอบสนองข้อกระทง เพราะฉะนั้นจึงมีความจำเป็นที่จะต้องตรวจสอบคุณสมบัติของการวัดเพียงมิติเดียว (Unidimension Latent Space) ของแบบประเมินฯ ซึ่งเป็นข้อตกลงเบื้องต้นที่สำคัญ ก่อนที่จะนำไปวิเคราะห์คุณภาพของแบบประเมินฯ ฉบับนี้ต่อไป

ผลจากการวิเคราะห์ตัวประกอบพบว่า ค่าไอเกนของตัวประกอบที่ 1 สูงกว่าตัวประกอบที่ 2 อย่างมาก ส่วนค่าไอเกนจากตัวประกอบที่ 2 มายังตัวประกอบที่ 3 และมายังตัวประกอบอื่น ๆ จนถึงตัวประกอบที่ 17 ตามลำดับ ค่าไอเกนสูงกว่ากันลดหลั่นลงมาไม่มากนัก จึงสรุปได้ว่าแบบประเมินตนเองสำหรับครูมัศึกษามีคุณสมบัติการวัดเพียงมิติเดียว

1.2 วิเคราะห์คุณภาพของข้อกระทงด้านความตรงและความเที่ยงของแบบประเมินฯ ด้วยวิธีการของมาตราส่วนประมาณค่าของแอนดริช ปรากฏผลดังนี้

จากผลการวิเคราะห์ข้อมูลแบบประเมินตนเองสำหรับครูมัศึกษาดังกล่าวสามารถสรุปผลการวิเคราะห์คุณภาพของแบบประเมินฯ ได้ดังนี้

1.2.1 คุณภาพของข้อกระทง ทั้ง 104 ข้อ จากการคำนวณค่าความเหมาะสม (Fit) ของข้อกระทงทั้งหมด มีค่าอยู่ระหว่าง -6.11 ถึง 12.93 สำหรับข้อกระทงที่เหมาะสม

---

\* ดูรายละเอียดได้จาก เยาวดี วิบูลย์ศรี รายงานการวิจัยเรื่อง การติดตามคุณภาพและประสิทธิภาพของแบบประเมินตนเองสำหรับครูมัธยมศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย 2537, 114 หน้า

กับโมเดลจะต้องมีค่าอยู่ระหว่าง  $-2.67$  ถึง  $2.67$  ( $-2.67 < \text{Item Fit } t_i < 2.67$ ) ดังนั้นแบบประเมินตนเองฯ ฉบับนี้จึงมีข้อกระทงที่เหมาะสมกับโมเดลถึง 73 ข้อ ตามค่าคาดหวัง (Expected Score) ของการวัดของโมเดลในการวิจัยครั้งนี้ จึงสรุปได้ว่าแบบประเมินตนเองสำหรับครูมัธยมศึกษาที่ผู้วิจัยสร้างขึ้น ประกอบด้วยข้อกระทงที่มีคุณภาพได้

1.2.2 ความตรงเชิงทฤษฎี (Construct Validity) แบบประเมินตนเองสำหรับครูมัธยมศึกษาที่ผู้วิจัยสร้างขึ้น มีข้อกระทงที่เหมาะสมกับโมเดล คิดเป็นร้อยละ 70.19 ซึ่งแสดงว่าแบบประเมินตนเองฉบับนี้ มีความตรงเชิงทฤษฎีอยู่ในเกณฑ์ที่ใช้ได้ตามโมเดลมาตรฐานส่วนประมาณค่าของแอนดริช

1.2.3 ความเที่ยง (Reliability) ของแบบประเมินตนเองสำหรับครูมัธยมศึกษาทั้งฉบับ มีค่าเท่ากับ .97 ซึ่งเป็นค่าความเที่ยงที่สูงมาก

2. ประสิทธิภาพในด้านความเป็นไปได้ และความเหมาะสมในการนำแบบประเมินตนเองสำหรับครูมัธยมศึกษาไปใช้ในเชิงปฏิบัติ พบว่ามีระดับสูง ทั้งนี้พิจารณาจากค่าเฉลี่ยของความคิดเห็นเป็นรายข้อ พบว่ามีค่าอยู่ระหว่าง 3.84 ถึง 4.07 โดยที่ข้อรายการที่ 1 คือ ข้อคำถามชัดเจนดี มีค่าเฉลี่ยสูงสุดเท่ากับ 4.07 รองลงมาได้แก่ข้อรายการที่ 8 คือเป็นแนวทางในการจัดโครงการฝึกอบรมเพื่อส่งเสริมสมรรถภาพของครู มีค่าเฉลี่ยเท่ากับ 4.01 สำหรับข้อรายการที่ 4 คือครูได้ทราบสิ่งที่ยังบกพร่องและนำไปปรับปรุงแก้ไขในการปฏิบัติงาน มีค่าเฉลี่ยต่ำสุดเท่ากับ 3.84 แต่ก็ยังอยู่ในเกณฑ์ที่ใช้ได้ดี

## อภิปรายผล

การติดตามผลแบบประเมินตนเองสำหรับครูมัธยมศึกษา ที่ผู้วิจัยสร้างขึ้นในปี 2522 ได้ชี้ให้เห็นประเด็นต่าง ๆ ซึ่งสอดคล้องกับข้อมูลที่มีผู้ศึกษาไว้ดังนี้ คือ

1) ผลการวิเคราะห์ตัวประกอบเพื่อตรวจสอบคุณสมบัติการวัดเพียงมิติเดียว พบว่าแบบประเมินตนเองฉบับนี้มีคุณสมบัติการวัดเพียงมิติเดียว ซึ่งชี้ให้เห็นว่าแบบประเมินชุดนี้มีลักษณะเป็นเอกพันธ์ (Homogeneous) ซึ่งเป็นไปตามข้อตกลงเบื้องต้นที่สำคัญของทฤษฎีการตอบสนองข้อกระทง ตามมาตรฐานส่วนประมาณค่าของแอนดริชใช้ในการวิเคราะห์ครั้งนี้ และตัวประกอบทั้ง 17 ตัวที่ได้มานั้นจะเห็นได้ว่าตัวประกอบที่ 1 สามารถอธิบายลักษณะของครูที่ดีได้ถึงร้อยละ 40.1 ส่วนตัวประกอบอื่น ๆ อีก 16 ตัว สามารถร่วมกันอธิบายคุณลักษณะของครูที่ดีได้เพิ่มขึ้นอีกเพียงร้อยละ 28.8 จึงสรุปได้ว่าแบบประเมินตนเองฉบับนี้มีตัวประกอบ

ที่เด่นชัดเพียงตัวประกอบเดียว และยังได้พบว่าองค์ประกอบของลักษณะครูที่ดีทั้ง 6 ด้านที่ผู้วิจัยคัดเลือกมานั้นมีความสามารถในการร่วมกันอธิบายลักษณะของครูที่ดีได้ถึงร้อยละ 68.9 ดังผลที่ปรากฏแล้วนั้น นอกจากนั้นยังเป็นการยืนยันว่าแบบประเมินตนเองสำหรับครูมัธยมศึกษาที่ผู้วิจัยสร้างขึ้นนี้เป็นแบบประเมินที่ดีตามแนวคิดการสังเคราะห์คุณลักษณะที่ดีของนักวิชาการที่สำคัญของไทยหลายท่าน (ก่อ สวัสดิพานิชย์ 2519, ฐะปะนีย์ นาครทรรพ 2517, บุญถิ่น อัตถากร 2514, สุรินทร์ สรสิริ 2515)

2) ผลการวิเคราะห์ความตรงเชิงทฤษฎีตามวิธีการมาตราส่วนประมาณค่าของแอนดริช พบว่าแบบประเมินตนเองฉบับนี้มีค่าความตรงในเกณฑ์ที่ใช้ได้ โดยมีข้อกระทงที่เหมาะสมกับโมเดลถึงร้อยละ 70.19 ซึ่งผลที่ได้นี้สอดคล้องกับแบบวัดความพึงพอใจในการปฏิบัติงานของครูมัธยมศึกษาที่พัฒนาโดย วิเศษ ชวระนอง (2532:110) เนื่องจากแบบวัดทั้งสองฉบับเป็นแบบวัดทางด้านจิตพิสัย (Affective Domain) ซึ่งเปิดโอกาสให้ผู้ตอบเสแสร้งได้ การได้ผลที่สอดคล้องกันเป็นการแสดงว่าแบบประเมินที่ผู้วิจัยได้สร้างมาตั้งแต่ปี พ.ศ.2522 ประกอบด้วยข้อรายการที่ทันสมัย (update) ในการใช้เพื่อประเมินคุณลักษณะของครูที่ดี

3) ผลการวิเคราะห์ความเที่ยง แบบประเมินฉบับนี้มีค่าความเที่ยงที่สูงมาก เท่ากับ .97 โดยมีข้อสรุปที่สนับสนุนดังนี้คือ

ก. ความคลาดเคลื่อนมาตรฐานของการวัดของข้อกระทงมีค่าอยู่ระหว่าง 0.07 ถึง 0.11 ซึ่งเป็นค่าความคลาดเคลื่อนที่ต่ำ ซึ่งจะทำให้ค่าความเที่ยงสูงสอดคล้องกับแนวคิดของทฤษฎีการตอบสนองข้อกระทง

ข. คะแนนโลจิสต์ของข้อกระทง มีค่าที่ใกล้เคียงกันอยู่ระหว่าง -1.05 ถึง 2.31 ดังนั้นความแปรปรวนระหว่างข้อจะต่ำทำให้สัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสูง และเป็นผลให้ค่าความเที่ยงสูงตามไปด้วย

ค. ค่าความเที่ยงที่ได้แสดงว่ามีความแปรปรวนภายในข้อกระทงสูง การตอบของครูมีการกระจายของระดับความคิดเห็นซึ่งสร้างไว้ 7 ระดับ แสดงว่าข้อคำถามมีอำนาจจำแนกความคิดเห็นของผู้ตอบได้ดี

4) ประสิทธิภาพในด้านความเป็นไปได้และความเหมาะสมในการนำแบบประเมินตนเองสำหรับครูมัธยมศึกษาไปใช้ในเชิงการปฏิบัติเพื่อให้เกิดการพัฒนาประสิทธิภาพของครู จากผลการวิจัยข้อ 2 ดังกล่าวมาแล้ว แสดงให้เห็นว่าครูมัธยมศึกษาที่เป็นกลุ่มตัวอย่างจำนวน 461 คน

พิจารณาแบบประเมินตนเองฉบับนี้แล้วเห็นว่า มีความเห็นไปได้และความเหมาะสมในการนำไปใช้ในเชิงการปฏิบัติเพื่อให้เกิดการพัฒนาประสิทธิภาพของครูในระดับสูง ดังรายละเอียดดังต่อไปนี้

ก. ขอรายการที่แสดงความเหมาะสมของแบบประเมินตนเองฉบับนี้คือ ขอรายการที่ 1 และ 2 ซึ่งแสดงถึงความชัดเจนของข้อความถามและความครอบคลุมลักษณะของครูที่ดี มีค่าเฉลี่ยของคะแนนความคิดเห็นเท่ากับ 4.07 และ 3.97 ตามลำดับ ซึ่งค่าดังกล่าวแสดงว่าขอรายการทุกข้อในแบบประเมินฉบับนี้ มีความชัดเจนและครอบคลุมลักษณะที่ดีของครูมัธยมศึกษา ตามแนวคิดของนักวิชาการด้านการศึกษาที่ได้กล่าวถึงในตอนต้น

ข. ขอรายการที่แสดงความเป็นไปได้ในทางปฏิบัติ เพื่อเพิ่มประสิทธิภาพการทำงานของตนเองคือ ขอรายการที่ 3 ถึง 6 ซึ่งแสดงถึงความเป็นไปได้ในการนำขอรายการไปใช้ในการบ่งชี้ความบกพร่องของตนเอง ในการเพิ่มประสิทธิภาพการทำงานของตนเองและเป็นการรอบในการปฏิบัติงาน มีค่าเฉลี่ยของคะแนนความคิดเห็นเท่ากับ 3.98, 3.84, 3.87 และ 3.88 ตามลำดับ ซึ่งเป็นค่าที่แสดงว่าแบบประเมินตนเองฯ ฉบับนี้มีความเป็นไปได้ที่ครูจะนำไปใช้เพื่อเพิ่มประสิทธิภาพการทำงานของตนเองอยู่ในระดับสูง และข้อค้นพบนี้สนับสนุนแนวคิดในการใช้ข้อมูลที่ได้จากแบบประเมินตนเองเพื่อพัฒนาประสิทธิภาพการทำงานของตนเอง (Steel and Ovalle 1984:667-668)

ค. ขอรายการที่แสดงความเป็นไปได้ที่ฝ่ายบริหารจะนำผลการประเมินตนเองของครูไปใช้เป็นข้อมูลเพื่อการพัฒนาศักยภาพของครู คือขอรายการที่ 7 และ 8 ซึ่งแสดงว่าผลการประเมินนี้มีประโยชน์ต่อการนิเทศงานของฝ่ายบริหารและเป็นแนวทางในการจัดโครงการฝึกอบรมเพื่อส่งเสริมสมรรถภาพของครูมีค่าเฉลี่ยเท่ากับ 3.95 และ 4.01 ตามลำดับ ซึ่งค่าดังกล่าวแสดงว่าข้อมูลที่ได้จากแบบประเมินฉบับนี้มีความเป็นไปได้ที่ฝ่ายบริหารจะนำผลไปใช้เพื่อการพัฒนาศักยภาพของครูในระดับสูง และข้อค้นพบนี้สนับสนุนแนวคิดในการที่ผู้บริหารจะนำข้อมูลจากการประเมินตนเองของผู้บังคับบัญชามาใช้ในการพัฒนาประสิทธิภาพการทำงานของผู้ใต้บังคับบัญชา (Shrauger and Osberg 1981:322-351)

## ข้อเสนอแนะ

1. เนื่องจากแบบประเมินตนเองสำหรับครูมัธยมศึกษาฉบับนี้ได้รับความสนใจและมีการนำไปใช้อย่างกว้างขวางมานานแล้ว และเมื่อได้วิเคราะห์ย้อนย้อนคุณภาพในปี 2531 จึงน่าจะ

มีการติดตามผลการปฏิบัติงานของครูโดยการนำแบบประเมินนี้ไปใช้เพื่อศึกษาประสิทธิภาพของครูในเชิงการพัฒนาคุณภาพ

2. หน่วยงานที่นำแบบประเมินตนเองฉบับนี้ไปใช้ ควรจะวางแผนและดำเนินการกระตุ้นอย่างเป็นระบบและต่อเนื่องเพื่อให้ครูเห็นความสำคัญและคุณค่าของการประเมินตนเอง

3. การแปลผลของคะแนนที่ได้จากการนำแบบประเมินตนเองสำหรับครูมัธยมศึกษาฉบับนี้ไปใช้ ผู้วิจัยได้นำเสนอในรูปเกณฑ์ปกติเปอร์เซ็นต์ไทล์ (ดูรายละเอียดจากรายงานวิจัย 2522) การแปลความหมายของคะแนน เพื่อให้ง่ายและสะดวก ขอเสนอแนะดังนี้

เกณฑ์ในการเทียบคะแนนที่ได้กับตำแหน่งเปอร์เซ็นต์ไทล์ และระดับคุณภาพในการประเมินตนเองของครูระดับมัธยมศึกษา

ช่วงคะแนน	ตำแหน่งเปอร์เซ็นต์ไทล์	ระดับคุณภาพในการประเมินตนเอง
ตั้งแต่ 567 ขึ้นไป	ตั้งแต่ PR 77 ขึ้นไป	สูง
ระหว่าง 470 ถึง 566	ระหว่าง PR 23-77	ปานกลาง
ตั้งแต่ 469 ลงมา	ตั้งแต่ PR 23 ลงไป	ต่ำ

## แบบประเมินตนเองสำหรับครูมัธยมศึกษา

**คำชี้แจง** โปรดอ่านข้อความแต่ละข้อ แล้วประเมินตัวท่านเองโดยพิจารณาว่าข้อความใดตรงกับสภาพความเป็นจริงเกี่ยวกับตัวท่าน ในระดับใด จาก 7 ระดับที่กำหนดให้ คือ

- 7 หมายถึง มากที่สุด
- 6 หมายถึง มาก
- 5 หมายถึง ค่อนข้างมาก
- 4 หมายถึง ปานกลาง
- 3 หมายถึง ค่อนข้างน้อย
- 2 หมายถึง น้อย
- 1 หมายถึง น้อยที่สุด

แล้วทำเครื่องหมาย X ลงในช่องระดับนั้น

รายการประเมินตนเอง	ตรงกับสภาพความเป็นจริง						
	7	6	5	4	3	2	1
1. มีความรู้ดีในวิชาที่สอน .....							
2. ค้นคว้าหาความรู้เพิ่มเติมเกี่ยวกับวิชาที่สอน .....							
3. เข้ารับการอบรมซึ่งให้ความรู้ใหม่ในสาขาวิชาที่สอนทุกครั้ง .....							
4. มีความรู้รอบตัวดี .....							
5. ปรึกษาและแลกเปลี่ยนความรู้กับเพื่อนครู .....							
6. สนใจวิชาการสาขาอื่นนอกเหนือจากวิชาที่สอน .....							
7. รู้แหล่งค้นคว้าหาความรู้ .....							
8. เขียนเอกสารประกอบการสอน .....							
9. แสวงหาโอกาสเพื่อการศึกษาต่อ .....							
10. ติดตามความเคลื่อนไหวทางวิชาการ .....							
11. สอนตรงตามวิชาเอก-โท ที่เคยศึกษามา .....							
12. มีความคิดริเริ่มด้านวิชาการ .....							

รายการประเมินตนเอง	ตรงกับสภาพความเป็นจริง						
	7	6	5	4	3	2	1
13. มีความรู้เรื่องหลักสูตร .....							
14. ยอมรับในวิชาการใหม่ๆ .....							
15. มีความรู้วิชาครู .....							
16. สนใจเหตุการณ์ปัจจุบัน .....							
17. เป็นสมาชิกสมาคมและ/หรือวารสารทางวิชาการ .....							
18. เตรียมการสอนล่วงหน้า .....							
19. วางแผนการสอนระยะยาว .....							
20. ใช้วิธีสอนหลาย ๆ แบบ .....							
21. กระตุ้นให้นักเรียนสนใจตลอดเวลาที่สอน .....							
22. เนื้อหาที่สอนรวบรวมจากตำราหลายเล่ม .....							
23. เปิดโอกาสให้นักเรียนแสดงความคิดเห็น .....							
24. เข้าสอนตรงตามเวลา .....							
25. สนใจนักเรียนทุกคนเท่าเทียมกันในเวลาสอน .....							
26. ใช้อุปกรณ์ประกอบการสอน .....							
27. ประเมินผลการเรียนของนักเรียนตลอดการเรียนการสอน .....							
28. ปรับปรุงการสอนของตนเองเป็นประจำ .....							
29. ให้ความเป็นกันเองกับนักเรียนเวลาสอน .....							
30. ตรวจงานของนักเรียนสม่ำเสมอ .....							
31. ส่งเสริมให้นักเรียนรู้จักค้นคว้าด้วยตนเอง .....							
32. สอนจริยธรรมควบคู่ไปกับเนื้อหาวิชาที่สอน .....							
33. พูดเสียงดัง ชัดเจน .....							
34. จัดให้มีกิจกรรมประกอบการสอน .....							
35. สอนโดยคำนึงถึงความแตกต่างระหว่างบุคคล .....							
36. เลือกตัวอย่างการสอนให้สัมพันธ์กับเนื้อหาและชีวิตประจำวัน .....							

รายการประเมินตนเอง	ตรงกับสภาพความเป็นจริง						
	7	6	5	4	3	2	1
37. ส่งเสริมให้นักเรียนมีความคิดริเริ่มสร้างสรรค์ .....							
38. มีความรู้ด้านจิตวิทยาทั่วไปและจิตวิทยาการศึกษา .....							
39. สอนเนื้อหาให้สัมพันธ์กับวิชาอื่น .....							
40. สอนตรงตามจุดมุ่งหมายที่ตั้งไว้ .....							
41. สนใจนักเรียนที่เรียนอ่อนเป็นพิเศษ .....							
42. การสอนทุกครั้งเปิดโอกาสให้นักเรียนซักถามอภิปราย .....							
43. ให้อภัยนักเรียนที่กระทำผิดแล้วยอมรับผิด .....							
44. ตัดสินข้อพิพาทของนักเรียนด้วยความยุติธรรม .....							
45. ช่วยเหลือนักเรียนที่ประสบปัญหา .....							
46. ช่วยเหลือผู้อื่นที่ประสบความเดือดร้อน .....							
47. อดกลั้นได้เมื่อถูกนักเรียนยั่วเย้า .....							
48. อดทนต่อความทุกข์ยากลำบากได้ .....							
49. ยึดมั่นในหลักของศาสนาอย่างเคร่งครัด .....							
50. ไม่ผูกพยาบาทต่อบุคคลที่ทำให้ตนเองเดือดร้อน .....							
51. มีความปรารถนาที่จะให้ผู้อื่นเป็นสุข .....							
52. เห็นใจเพื่อนครูที่มีความทุกข์ .....							
53. สนใจต่อความทุกข์ยากของนักเรียน .....							
54. รู้สึกสงสารเมื่อเห็นผู้อื่นได้รับความทุกข์ .....							
55. รู้สึกละอายใจเมื่อรู้ว่าตนได้กระทำผิด .....							
56. ระลึกถึงคุณงามความดีของผู้มีพระคุณเสมอ .....							
57. ซื่อสัตย์ต่อผู้อื่นเสมอ .....							
58. มีความมั่นใจในตนเอง .....							
59. ถือว่าการตรงต่อเวลาเป็นสิ่งสำคัญยิ่ง .....							
60. แก้ปัญหาเฉพาะหน้าได้ดี .....							

รายการประเมินตนเอง	ตรงกับสภาพความเป็นจริง						
	7	6	5	4	3	2	1
61. มีอารมณ์ขัน .....							
62. มีความรับผิดชอบสูง .....							
63. มีความสุขุม เยือกเย็น และรอบคอบ .....							
64. ยึดมั่นในเหตุผลอยู่เสมอ .....							
65. กล้าแสดงออกในทางที่ถูกต้องควร .....							
66. แต่งกายสุภาพเรียบร้อย .....							
67. เสนอแนะความคิดใหม่ๆ .....							
68. มีอารมณ์มั่นคง .....							
69. ใช้ภาษาชัดเจนและสุภาพ .....							
70. รักษาระเบียบวินัยอย่างเคร่งครัด .....							
71. เป็นผู้นำของกลุ่มอยู่เสมอ .....							
72. สามารถชักจูงให้ผู้อื่นปฏิบัติตามได้ .....							
73. ร่วมในกิจกรรมอาสาสมัคร .....							
74. ทำงานรวดเร็วและทันตามเวลาที่กำหนด .....							
75. ทำงานสำเร็จตามที่ได้รับมอบหมาย .....							
76. ระวังอารมณ์ได้เมื่อโกรธ .....							
77. ยิ้มแย้มแจ่มใสเสมอ .....							
78. ปรับตัวให้เข้ากับผู้อื่นได้ .....							
79. ช่วยเหลือกิจกรรมที่เกี่ยวข้องกับส่วนรวม หรือชุมชน .....							
80. ยอมรับฟังความคิดเห็นของผู้อื่น .....							
81. ทำงานร่วมกับหมู่คณะได้ .....							
82. ให้คำปรึกษากับบุคคลซึ่งอยู่ในที่ร่วมงาน/ท้องถิ่น เดียวกัน .....							
83. พบปะผู้ปกครองนักเรียนเมื่อมีโอกาส .....							
84. เคารพและปฏิบัติตามกฎเกณฑ์ของสังคมที่ดี .....							

รายการประเมินตนเอง	ตรงกับสภาพความเป็นจริง						
	7	6	5	4	3	2	1
85. เข้าใจสภาพแวดล้อมของสังคม .....							
86. ปฏิบัติต่อผู้อื่นอย่างเสมอต้นเสมอปลาย .....							
87. เข้าใจขนบธรรมเนียม ประเพณี และวัฒนธรรม ของสังคมที่ตนอาศัยอยู่.....							
88. คบหาสมาคมกับบุคคลหลายระดับ .....							
89. มีส่วนร่วมในการปรับปรุงสังคมให้เจริญขึ้น .....							
90. มีความจริงใจต่อบุคคลอื่น .....							
91. เข้าร่วมกิจกรรมของท้องถิ่นหรือหน่วยงานที่สังกัด .....							
92. เป็นที่เคารพนับถือของบุคคลในสังคม .....							
93. รู้จักประหยัด .....							
94. มีสุขภาพจิตดี .....							
95. ไม่สร้างปัญหายุ่งยากทางด้านการเงิน.....							
96. มีสุขภาพทางร่างกายสมบูรณ์ .....							
97. มองผู้อื่นในแง่ดี .....							
98. ไม่นำปัญหาส่วนตัวเข้ามาเกี่ยวข้องกับหน้าที่การงาน .....							
99. มีความพอใจในฐานะของตนเอง .....							
100. ไม่ได้มั่วสุมจนติดเป็นนิสัย .....							
101. มีความรักในอาชีพครู .....							
102. สมาชิกในครอบครัวอยู่ร่วมกันอย่างมีความสุข .....							
103. ไม่ชอบเล่นการพนัน .....							
104. มีประวัติส่วนตัวดี .....							

## เอกสารอ้างอิง

- วิภา เกษรศิริเจริญ. “ประสิทธิภาพของครูมัธยมศึกษาในกรุงเทพมหานคร.” วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2521.
- วิเศษ ชวระนอง. “การพัฒนาแบบวัดความพึงพอใจในการปฏิบัติงานของครูมัธยมศึกษา.” วิทยานิพนธ์ปริญญาโทมหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2532.
- เยาวดี วิบูลย์ศรี. “การสร้างแบบรายงานตนเองสำหรับครูมัธยมศึกษา.” รายงานการวิจัย ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2524.
- Andrich, D. *A Rating Formulation for Ordered Response Categories*. Measurement and Statistics Laboratory. Department of Education. University of Western Australia, 1977.
- Campbell, D.J. and Lee, C. “Self-Appraisal in Performance Evaluation : Development Versus Evaluation.” *Academy of Management* 10 (Winter 1973) : 287-295.
- Centra, John A. “Self Rating of College Teachers : A Comparison with Student Ratings.” *Journal of Educational Measurement* 10 (Winter 1973) : 287-295.
- Doyle, Kenenth O., Jr. and Webber, Patricia L. “Self Ratings of College Instruction.” *American Educational Research Journal*, 15 (Summer 1978) : 467-475.
- Fisher, C., & Russ, G. (1986) *Self and Superior Assessment*. (Jech. Rep. No. TR-ONR-4). College Station : Texas A&M University, Department of Management.
- Lord, F.M. *Applications of Item Response Theory to Practical Testing Problems*. New Jersey : Lawrence Erlbaum Associates Publishers, 1980.
- Shrauger, J., & Osberg, T. (1981) “The Relative Accuracy Self-Predictions and Judgments by Others in Psychological Assessment.” *Psychological Bulletin*. 90, 322-351.
- Steel, R., & Ovalle, N. (1984) “Self-Appraisal Based Upon Supervisory Feedback.” *Personnel Psychology*, 37, 667-68.
- Wright, B.D., and Masters, G.N. *Rating Scale Analysis*. Chicago : Mesa Press, 1982.

ขอขอบคุณ

**คุณสมชัย บุญคำศิริ**

ที่สนับสนุนค่าใช้จ่ายในการจัดพิมพ์

ขอขอบคุณ

ผู้สนับสนุนค่าใช้จ่ายในการจัดพิมพ์

**คุณศิริพรรณ เก้าเจริญ**

**คุณสมชาย เกียรติสูงเด่น**

**ศาสตราจารย์ประคอง นรรณสุด**





พิมพ์ที่โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย โทร. 2153626  
นางวัฒนา ศิวะเกื้อ ผู้พิมพ์ผู้โฆษณา มกราคม 2539  
3901-54 (1,000) 2.

ขอขอบคุณ

บริษัท ซีเทรคสากล จำกัด (ไมครือเอเชียนกรุ๊ป)

1693 ห้อง 506 อาคารศูนย์การค้าเซ็นทรัลพลาซ่า ชั้น 5

ถนนพหลโยธิน แขวงลาดยาว

เขตจตุจักร กรุงเทพฯ 10900

โทร. 541-1361-5 แฟกซ์ (662) 541-1730

ที่สนับสนุนค่าใช้จ่ายในการจัดพิมพ์



