

บทที่ 2

วรรณคดีที่เกี่ยวข้อง

การวิจัยเรื่องการประเมินอภิमानการใช้แบบจำลองชิปในการประเมินโครงการการศึกษา ผู้วิจัยได้เสนอเอกสารและงานวิจัยที่เกี่ยวข้องเป็น 4 ตอน ดังนี้

ตอนที่ 1 มโนทัศน์เกี่ยวกับการประเมิน

ตอนที่ 2 มโนทัศน์เกี่ยวกับแบบจำลองชิป

ตอนที่ 3 มโนทัศน์เกี่ยวกับการประเมินอภิमान

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 มโนทัศน์เกี่ยวกับการประเมิน

1.1 ความหมายของการประเมิน

พัฒนาการความหมายของการประเมินได้มีนักประเมินทางการศึกษาให้นิยามทางการประเมินไว้ต่างๆ กัน แล้วแต่จุดเน้นของการประเมิน ในระยะแรก Tylor ได้เป็นคนแรกที่แยกความหมายของการประเมินได้อย่างเด่นชัด โดยเน้นที่วัตถุประสงค์ของการประเมินเป็นหลัก กล่าวคือ การประเมิน หมายถึง “การวัดลักษณะที่เกิดขึ้นจริงว่าสอดคล้องกับวัตถุประสงค์เชิงพฤติกรรมที่ตั้งไว้หรือไม่” จากนั้นความหมายของการประเมินก็ได้รับการพัฒนาขึ้นเรื่อยๆ ตามแนวความคิดของนักประเมินทางการศึกษา ซึ่งได้ปรับเปลี่ยนไปตามยุคสมัย พอจะแบ่งออกได้เป็น 2 กลุ่มคือ กลุ่มนักประเมินที่เน้นการตัดสินคุณค่า (value-oriented education) ได้แก่ Scriven, Worthen and Sander, Apple, Cooiy and Lons และ Guba and Lincon ซึ่งได้ให้ความหมายของการประเมินว่า “เป็นกระบวนการในการตัดสินคุณค่าของสิ่งที่มีประเมิน” ส่วนอีกกลุ่มหนึ่งคือกลุ่มนักประเมินที่เน้นการตัดสินใจ (decision-oriented evaluation) ได้แก่ Stufflebeam et. al, Cronbach, และ Alkin ได้ให้ความหมายของการประเมินว่า “เป็นกระบวนการเสนอสารสนเทศที่เป็นประโยชน์ต่อการตัดสินใจของผู้บริหาร” (เขาวดี วิบูลย์ศรี, 2539; ศิริชัย กาญจนวาสี, 2537) จากความหมายดังกล่าว พอที่จะสรุปความหมายของการประเมินออกเป็น 2 แนวทาง คือ เป็นกระบวนการตัดสินคุณค่าของสิ่งที่มีประเมิน และเป็นกระบวนการเสนอสารสนเทศที่เป็นประโยชน์ต่อการตัดสินใจของผู้บริหาร

1.2 ประวัติและพัฒนาการของการประเมิน

Maduas et. al. 1983 อ้างถึงใน สมหวัง พิธิยานุวัฒน์ (2541) ได้แบ่งช่วงประวัติและพัฒนาการของการประเมินออกเป็นยุคต่างๆ โดยได้เพิ่มยุคที่ 1 เพิ่มเติมจึงรวมเป็น 7 ยุค ดังนี้

ยุคที่ 1 ยุคก่อนการปฏิวัติอุตสาหกรรม โดยเริ่มตั้งแต่มีการจดบันทึกจนถึง ค.ศ. 1800 ซึ่งมีเหตุการณ์สำคัญ คือ ก่อนคริสต์ศักราช 2000 ปี ประเทศจีนได้สอบคัดเลือกบุคคลเข้ารับราชการโดยใช้ข้อสอบความเรียง ก่อนคริสต์ศักราช 5 ปี Socrates และครูชาวกรีก ได้ทำการจัดสอบโดยการตั้งคำถามให้ตอบ ซึ่งก็คือการสอบปากเปล่านั่นเอง และระหว่างปี ค.ศ. 1600-1800 เป็นช่วงที่วิทยาศาสตร์เข้ามามีบทบาทต่อมนุษย์ทั้งทางด้านความคิดและการปฏิบัติ ทำให้เป็นพื้นฐานของการปฏิวัติอุตสาหกรรมในทวีปยุโรปต่อมา

ยุคที่ 2 ยุคการปฏิรูป (The Age Reform : ค.ศ. 1800-1900) เป็นช่วงของการปฏิวัติอุตสาหกรรมในทวีปยุโรปได้มีการปรับปรุงเปลี่ยนแปลงครั้งใหญ่ทั้งทางด้านวิทยาศาสตร์และเทคโนโลยี สังคม การเมือง และการศึกษา มีการปฏิรูปโครงสร้างทางสังคม การศึกษา และสาธารณสุข โดยเฉพาะอย่างยิ่งในประเทศอังกฤษและสหรัฐอเมริกา

ในด้านการวิจัยทางการศึกษา Wundt และคณะ ซึ่งเป็นนักจิตวิทยาชาวเยอรมัน ได้ทำการทดลองโดยสังเกตและบันทึกพฤติกรรมของมนุษย์อย่างเป็นระบบ นักวิชาการที่ร่วมงานกับ Wundt ในครั้งนี้ได้แก่ Cattell, Hall, Judd และ Meumann ล้วนเป็นนักวิชาการที่มีชื่อเสียงมากในศตวรรษที่ 20

ค.ศ. 1845 เริ่มมีการประเมินผลงานเพื่อประเมินประสิทธิภาพของโรงเรียนที่เมืองบอสตัน สหรัฐอเมริกา ที่เรียกว่า Boston Survey โดย Horace Mann และ Samuel Howe ได้ริเริ่มใช้คะแนนจากการสอบข้อเขียนของนักเรียน ในการประเมินประสิทธิภาพการสอนของครู และประสิทธิภาพของโรงเรียน มีการให้ข้อมูลที่เป็นระบบโดยการใช้คะแนนจากการสอบข้อเขียนของนักเรียนแทนการสอบปากเปล่า ในด้านการพิจารณาแต่งตั้งผู้บริหารนั้น Mann และ Howe ได้เป็นผู้บุกเบิกให้ใช้การประเมิน วิเคราะห์ข้อมูลจากผลสัมฤทธิ์ทางการเรียนของนักเรียน และยังได้พยายามสร้างแบบสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนที่เป็นมาตรฐานด้วย

ค.ศ. 1884 Gullton นักชีววิทยาชาวอังกฤษได้สร้างเครื่องมือวัดความสามารถทางการรับรู้ด้านเสียง จำแนกเสียง และจำแนกกระยะทาง Gullton เป็นบุคคลแรกที่ริเริ่มการวัดโดยใช้แบบสอบถาม และมาตรฐานค่า และยังเป็นบุคคลแรกที่คิดค่าสัมประสิทธิ์สหสัมพันธ์

ค.ศ. 1887-1898 Joseph M. Rice เป็นนักปฏิรูปการศึกษาชาวอเมริกัน ซึ่งเป็นบิดาทางการวิจัยทางการศึกษา และเป็นบุคคลแรกที่เป็นผู้ประเมินโครงการทางการศึกษา Rice ได้สร้างแบบสอบการสะกดคำของเด็ก

ค.ศ. 1890 Cattell ได้เป็นผู้ริเริ่มใช้คำว่า การทดสอบสมอง (mental test) ในสหรัฐอเมริกา

ยุคที่ 3 ยุคแห่งประสิทธิภาพและการทดสอบ (The Age of Efficiency and Testing) ซึ่งยุคนี้ถือเป็นยุคทองของการวัดผลทางการศึกษา เป็นการทดสอบเพื่อแสวงหาประสิทธิภาพระบบและความเป็นมาตรฐาน แบ่งออกเป็น 2 ช่วงคือ ค.ศ. 1900-1915 ได้มีการสำรวจและพัฒนาวิธีการวัดผล Thorndike เป็นนักวัดผลคนสำคัญ ได้เขียนหนังสือเรื่อง "การวัดผลทางการศึกษา" เป็นเล่มแรกของโลก ซึ่งเป็นรากฐานของการพัฒนาแบบสอบมาตรฐาน Stone เป็นบุคคลแรกที่พัฒนาแบบสอบมาตรฐาน ค.ศ. 1905 Binet นักจิตวิทยาชาวฝรั่งเศส และ Simon ร่วมกันสร้างแบบสอบวัดเชาวน์ปัญญาเป็นฉบับแรก เรียกว่า Binet Simon Scale และในช่วงที่ 2 ประมาณ ค.ศ. 1915-1930 ซึ่งเป็นระยะการขยายตัวของแบบสอบปรนัยและแบบสอบมาตรฐาน ค.ศ. 1920 McCall เป็นบุคคลแรกที่แนะนำให้ครูสร้างแบบสอบขึ้นใช้เอง (teacher-made test) และแบบสอบมาตรฐานที่มีชื่อเสียงมากในยุคนี้คือ แบบสอบ Army Alfa และ Army Beta ซึ่งสร้างขึ้นในปี ค.ศ. 1917

ในช่วงปลายยุคของประสิทธิภาพและการทดสอบนี้ มหาวิทยาลัยหลายแห่งได้ร่วมกับโรงเรียนเพื่อทดสอบประสิทธิภาพของโรงเรียน เอกลักษณะของยุคนี้คือ ยุคแห่งการทดสอบมาตรฐาน เห็นได้จากช่วงสงครามโลกครั้งที่ 1 (ค.ศ. 1914-1918) โรงเรียนได้ประเมินหลักสูตรและประสิทธิภาพของการเรียนการสอนโดยใช้แบบสอบมาตรฐาน จึงเกิดธุรกิจทางการศึกษาขึ้นในเมืองฟิลาเดลเฟีย สหรัฐอเมริกา (ค.ศ. 1916-1938) การประเมินทางการศึกษามุ่งตอบสนองความต้องการของท้องถิ่น มิได้มีแรงจูงใจในการประเมินเพื่อความเป็นสากล (generalizability)

ยุคที่ 4 ยุคทายเลอร์ (The Tylerian Age : 1930-1945) หรือยุคการประเมินตามวัตถุประสงค์ ในระยะแรกของยุคนี้ เกิดภาวะเศรษฐกิจตกต่ำมาก ทำให้นักสังคมสงเคราะห์มีบทบาทในการกำหนดนโยบายเพื่อแก้ไขปัญหาเศรษฐกิจและการศึกษา John Dewey และ Ralph Tyler เป็นบุคคลสำคัญทางการศึกษาของสหรัฐอเมริกา ในการร่วมขบวนการเคลื่อนไหวปฏิรูปการศึกษา John Dewey เป็นผู้นำทำให้เกิดการศึกษาแบบก้าวหน้า (Progressive Education) และทำให้เกิดหลักการศึกษานักการศึกษาทั่วโลก รู้จักกันดีคือการเรียนรู้จากการปฏิบัติ ส่วน Tyler ได้รับการแต่งตั้งให้เป็นผู้อำนวยการระยะยาวถึง 8 ปี ได้ศึกษาเปรียบเทียบการสอนแบบดั้งเดิมกับการสอนแบบใหม่ในระดับมัธยมศึกษาของสหรัฐอเมริกา Tyler ได้วางแนวคิดการประเมินหลักสูตรโดยการ

เปรียบเทียบผลที่เกิดขึ้นจริงกับผลสัมฤทธิ์ที่คาดหวัง ตามวัตถุประสงค์เชิงพฤติกรรม Tyler เป็นบุคคลแรกที่นิยามการประเมินผลการศึกษาว่าเป็นการวัดคุณค่าผลสำเร็จตามวัตถุประสงค์การเรียนการสอน จากแนวคิดของ Tyler ทำให้เห็นความแตกต่างของ "การวัด" และ "การประเมิน" ได้เด่นชัด ซึ่งเป็นปัจจัยสำคัญที่ทำให้ศาสตร์แห่งการประเมินได้รับการพัฒนาเป็นวิชาชีพ (professional) ในบทความของ Tyler ได้ระบุว่า ในตอนนี้ขอเพียงครูได้สามารถประเมินการสอนโดยเปรียบเทียบผลสัมฤทธิ์ทางการเรียนกับวัตถุประสงค์การสอน ซึ่งเป็นลักษณะของการประเมินที่เน้นผลลัพธ์ (outcome) มากกว่าตัวป้อน (input) และกระบวนการ (process) เหมาะกับยุคสมัย ซึ่งก่อนยุคนี้ครูยังแยกไม่ออกระหว่างการวัดกับการประเมิน ในปี ค.ศ. 1986 Tyler ได้เขียนบทความการประเมินทางการศึกษาโดยใช้แนวคิดเชิงระบบ เหมาะกับยุคสมัยที่ครูปัจจุบันจะพัฒนารูปแบบการเรียนการสอน ได้อย่างครบวงจรและเป็นระบบ

หลักการและนิยามการประเมินผลการศึกษา นับว่าเป็นรากฐานที่มีอิทธิพลต่อแนวความคิดและก่อให้เกิดการพัฒนาศาสตร์แห่งการประเมินทางการศึกษา ต่อมา Tyler ได้รับการยกย่องให้เป็นบิดาของการประเมินทางการศึกษาและเป็นผู้ก่อให้เกิดนวัตกรรมในการเขียนวัตถุประสงค์ให้ชัดเจนหรือเรียกว่าวัตถุประสงค์เชิงพฤติกรรม (behavioral objective)

ยุคที่ 5 ยุคแห่งการไร้เดียงสาทางการประเมิน (the Age of Innocence : 1946-1957) หรือยุคแห่งการเพิกเฉยทางการประเมิน (the Age of Ignorance) เนื่องจากยุคนี้เป็นยุคที่ฟื้นฟูประเทศ เกิดภัยสงครามโลกครั้งที่ 2

ในการประเมินทางการศึกษาช่วงนี้ แนวความคิดการประเมินของ Tyler ได้รับการเผยแพร่ไปอย่างกว้างขวาง แต่มีปัญหาในการเขียนวัตถุประสงค์เชิงพฤติกรรมตามลำดับขั้นของการเรียนรู้ เพื่อช่วยแก้ไขปัญหานี้ Bloom และคณะ (ค.ศ. 1956) และ Krathwohl (ค.ศ. 1964) จึงได้สร้างสารบบของวัตถุประสงค์การศึกษาริเริ่ม (taxonomies of educational objectives) ซึ่งทำให้การประเมินทางการศึกษาตามแนว Tyler มีความชัดเจนยิ่งขึ้น ยิ่งไปกว่านั้นทำให้วงการศึกษามีความชัดเจนในการกำหนดผลการศึกษาทั้งในระดับชาติ หลักสูตร และบทเรียนด้วย

ยุคที่ 6 ยุคการขยายองค์ความรู้ทางการประเมิน (The Age of Expansion : 1958-1972) เนื่องจากโซเวียตประสบความสำเร็จในการส่งดาวเทียมขึ้นสู่อวกาศเป็นประเทศแรกของโลก ทำให้สหรัฐอเมริกา รู้สึกว่าประเทศของตนล้าหลัง จึงรีบเร่งปรับปรุงการศึกษาเป็นการใหญ่โดยเน้นการศึกษาด้านวิทยาศาสตร์ คณิตศาสตร์ ภาษาต่างประเทศ และการแนะแนว โดยปรับปรุงการเรียนการสอนของรัฐบาลในแต่ละรัฐ มุ่งสร้างให้ชาวอเมริกันเป็นผู้เป็นผู้รอบรู้ทางด้านวิทยาศาสตร์และเทคโนโลยี เพื่อให้สหรัฐอเมริกาเป็นประเทศที่ยิ่งใหญ่ ส่งเสริมให้มีการพัฒนาหลักสูตรใหม่ ๆ และได้มีการลงทุน

ในโครงการพัฒนาต่างๆ มากมาย เป็นแรงผลักดันให้มีการประเมินผลอย่างเป็นระบบ ทุกโครงการต้องมีการประเมินโครงการประจำปีและมีการรายงานผล ด้วยเหตุดังกล่าว ทำให้การประเมินเจริญขึ้นเป็นศาสตร์สาขาที่เป็นความเชี่ยวชาญชำนาญการ อีกทั้งเห็นประโยชน์และคุณค่าของการประเมินโครงการอย่างกว้างขวาง ซึ่งไม่เคยมีบรรยากาศเช่นนี้ มาก่อนในประวัติศาสตร์ของการวัดและประเมินผลทางการศึกษา

ในยุคนี้ได้มีนักวิชาการเสนอแนวคิดและการประเมินมากมาย เช่น Cronbach (1963) Scriven (1967) Stake (1967) Tyler (1969) Stufflebeam et. al (1971) Provas (1971) Caro (1971) Weiss (1972) Worthen and Sanders (1973) เป็นต้น นอกจากนี้ยังมีสมาคมต่างๆ มีบทบาทในการสร้างนักประเมิน และพัฒนาทฤษฎี เช่น สมาคมเกียรตินิยมทางการศึกษา (Phi Delta Kappa) โดย Stufflebeam et. al ได้พัฒนาหลักการและโมเดล CIPP ในการประเมินโปรแกรม นอกจากนี้ทางสมาคมเกียรตินิยมทางการศึกษายังได้อบรมผู้สนใจที่จะเป็นนักประเมินด้วย

ในปี ค.ศ. 1966 Campbell and Stanley ได้เขียนหนังสือการออกแบบทดลองและกึ่งทดลองเพื่อชี้ให้เห็นว่าการวิจัยทดลองเป็นรูปแบบที่มีข้อจำกัดมากมายในการวิจัยทดลองในสนาม และเรียกร้องให้มีการนำแบบทดลองไปใช้ในการวิจัยหรือประเมินในสภาพการที่เป็นจริง

โดยสรุป ยุคนี้เป็นการประเมินยุครุ่งอรุณของการประเมินที่จะก้าวต่อไปเป็นศาสตร์แห่งวิชาชีพ เพื่อการพัฒนาคน พัฒนางาน พัฒนาองค์การและสภาพสังคมเพื่อมนุษยชาติ

ยุคที่ 7 ยุคแห่งการเป็นวิชาชีพของการประเมิน (The Age of Professionalization : 1973 - ปัจจุบัน) ตั้งแต่ปี ค.ศ. 1973 การประเมินถือว่าเป็นศาสตร์แห่งความเป็นปึกแผ่น มีศัพท์เฉพาะศาสตร์ทางการประเมิน มีองค์ความรู้ทางการประเมินระดับกระบวนการทัศน์ โมเดลและทฤษฎี มีวิธีการแสวงหาความรู้ทางการประเมิน ศาสตร์แห่งการประเมินได้ผ่านยุคแห่งการแสวงหาเอกลักษณ์มาแล้ว ทำให้การประเมินแตกต่างจากการวัดและการวิจัยอย่างเด่นชัด มีสมาคมวิชาชีพทางการประเมิน เช่น American Evaluation Association เป็นต้น มีวารสารทางการประเมินโดยเฉพาะ เช่น Educational Evaluation and Policy Analysis, Evaluation Review, New Directions for Program Evaluation, Studies in Evaluation เป็นต้น

ในด้านผู้ที่เป็นนักประเมินต้องได้รับการอบรมเล่าเรียนมาโดยเฉพาะจากสถาบันการศึกษาชั้นสูง มหาวิทยาลัยต่างๆ มหาวิทยาลัยที่บุกเบิกการสร้างการประเมินโดยตรง เช่น University of Illinois, Stanford University, Boston College, UCLA, University of Minnesota และ Western Michigan University ในปัจจุบันมหาวิทยาลัยต่างๆ ทั่วโลกโดยเฉพาะสหรัฐอเมริกา มีการสอนด้านการประเมินทางการศึกษาเป็นสาขาวิชาเฉพาะ

1.3 ความสำคัญของการประเมิน

การประเมินมีความสำคัญต่อศาสตร์ในสาขาต่างๆ มากมาย โดยผลการประเมินจะให้สารสนเทศที่เป็นประโยชน์ในการตัดสินใจด้านต่างๆ ดังที่ ศิริชัย กาญจนวาสี (2537) ได้กล่าวถึงความสำคัญของการประเมินดังนี้

ด้านการบริหาร สามารถใช้ผลการประเมินเป็นเครื่องมือของ ผู้บริหารสำหรับกำกับติดตามการดำเนินงานให้เป็นไปตามแบบแผนที่มีประสิทธิภาพและประสิทธิผล

ด้านจิตวิทยา สามารถใช้การประเมินเป็นกลยุทธ์สำหรับกระตุ้นรักษาความสนใจ และสร้างความตื่นตัวในการทำงานของผู้ปฏิบัติงาน

ด้านรัฐศาสตร์ สามารถใช้การประเมินเป็นเครื่องมือตรวจสอบความรับผิดชอบต่อการปฏิบัติ (accountability) ตลอดจนสามารถใช้เป็นกลไกในการสร้างแรงสนับสนุนจากสาธารณะ

นอกจากนี้ การประเมินยังมีประโยชน์ต่อการวางแผนและการบริหารงาน ซึ่งพอจะสรุปได้ดังนี้ (สมคิด พรหมจ้อย, 2535)

1. ช่วยให้ได้สารสนเทศต่างๆ เพื่อนำไปใช้ในการตัดสินใจเกี่ยวกับการกำหนดแผนงานและโครงการ การตรวจสอบความพร้อมของทรัพยากรต่างๆ ที่จำเป็นต้องใช้ในการดำเนินงาน ตลอดจนความเป็นไปได้ในการจัดกิจกรรม
2. ช่วยในการจัดหาสารสนเทศเชิงความก้าวหน้า ปัญหาและอุปสรรคของการดำเนินงานมาใช้ในการตัดสินใจเกี่ยวกับการปรับปรุงงาน หรือเปลี่ยนแปลงแก้ไขการดำเนินงานโครงการให้เป็นไปตามทิศทางที่ต้องการ
3. ช่วยในการจัดหาสารสนเทศเกี่ยวกับความสำเร็จและความล้มเหลวของกิจกรรมที่นำมาใช้ในการตัดสินใจ และวินิจฉัยการดำเนินการต่อไปว่าสมควรจะยกเลิกหรือขยายการดำเนินงานต่อไปอีก
4. ช่วยให้ได้สารสนเทศที่บ่งบอกประสิทธิภาพของการดำเนินงานว่าเป็นอย่างไรคุ้มค่าต่อการลงทุนหรือไม่
5. เป็นแรงจูงใจให้ผู้ปฏิบัติงานได้รับสารสนเทศเกี่ยวกับผลการปฏิบัติงานและนำสารสนเทศมาใช้ในการปรับปรุงและพัฒนางานได้อย่างมีประสิทธิภาพ

1.4 คุณลักษณะของการประเมินที่ดี

ในการทำการประเมิน ผลที่ได้จากการประเมินนั้นจะเป็นประโยชน์ต่อการตัดสินใจของผู้บริหารมากเพียงใดนั้น ขึ้นอยู่กับคุณภาพของงานประเมิน ดังที่ สมทวง พิธิยานุวัฒน์ (2525) ได้สรุปประเด็นของคุณลักษณะที่ดีในการประเมินได้ดังนี้

1. มีความตรงสูง (validity) ผลการประเมินจะมีความตรงสูงก็ต่อเมื่อเครื่องมือที่ใช้วัดมีความตรงสูง ซึ่งขึ้นอยู่กับมาตรฐานของเครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลเป็นสำคัญ
2. มีความเที่ยงสูง (reliability) การประเมินที่ดีต้องมีความคงเส้นคงวาในการประเมิน นักประเมินจึงจำเป็นต้องควบคุมการวัดทุกขั้นตอนของการประเมินให้มีความคลาดเคลื่อนน้อยที่สุด
3. ความเป็นปรนัย (objectivity) นั่นคือ เป็นการประเมินที่ปราศจากอคติ (bias) ซึ่งนักประเมินควรระมัดระวังเรื่องความผูกพันในโครงการ อาจทำให้การประเมินเอนเอียงไปในทางบวกเกินความเป็นจริงได้
4. ครอบคลุมเรื่องราวการประเมินอย่างครบถ้วน เพื่อเสนอแนะในการตัดสินใจของผู้บริหารในทุกแง่มุม
5. มีความสะดวกและประหยัด ความเป็นไปได้สูง ผลที่ได้มีความคุ้มค่ากับการลงทุน
6. ต้องให้ผลการประเมินเป็นที่ยอมรับ (credibility) ทั้งในหมู่นักประเมินอาชีพผู้บริหารโครงการ และบุคคลหรือองค์การที่เกี่ยวข้อง
7. ต้องให้ผลการประเมินแก่ผู้บริหารได้ทันท่วงที (timeliness) กล่าวคือ ผลการประเมินต้องได้ก่อนที่ผู้บริหารจะตัดสินใจในเรื่องที่เกี่ยวข้องกับโครงการที่ประเมิน
8. ในการประเมินผลโครงการประชุม ควรรบกวนผู้เข้าร่วมประชุมน้อยที่สุด เฉพาะที่จำเป็นเท่านั้น มิใช่เอากิจกรรมการประเมินผลโครงการมาเป็นภาระและอุปสรรคในการเรียนรู้ของผู้เข้าร่วมประชุม
9. ควรมีการวางแผนประชุมก่อนเข้าร่วมโครงการ ตลอดจนกำหนดแผนการประเมินให้เด่นชัดและเป็นขั้นตอนก่อนเริ่มมีโครงการ

1.5 บทบาทของการประเมิน

การประเมินผลเชิงความก้าวหน้า (formative evaluation) เน้นการประเมินเพื่อการปรับปรุงเป็นสำคัญซึ่งมักจะประเมินผลระหว่างแผนงานหรือระหว่างพัฒนาโครงการ ผลที่ได้จะช่วยตั้งวัตถุประสงค์ของโครงการให้เป็นไปตามเป้าหมายที่แท้จริง นอกจากนี้

ยังใช้ในระหว่างดำเนินโครงการจะช่วยตรวจสอบว่าโครงการดำเนินตามแผนของโครงการอย่างไร และตรวจสอบความก้าวหน้าของโครงการว่าดำเนินได้ผลอย่างไร

การประเมินผลสรุปรวม (summative evaluation) มักใช้การประเมินหลังสิ้นสุดโครงการ สำหรับโครงการที่มีระยะยาวก็อาจใช้ การประเมินผลสรุปรวมในการสรุปย่อความระยะยาวต่างๆ ข้อมูลที่ได้จะช่วยในการประเมินผลสรุปรวมนั้น ส่วนใหญ่จะเป็นการรวบรวมผลของการประเมินความก้าวหน้ามาเป็นการประเมินผลสรุปรวม ผลสรุปที่ได้จะสามารถบอกได้ว่า โครงการได้บรรลุเป้าหมายหรือไม่อย่างไร มีปัญหาอุปสรรคหรือไม่ ต้องแก้ไขอย่างไร ข้อมูลเหล่านี้จะช่วยให้ผู้บริหารโครงการตัดสินใจว่า โครงการควรจะดำเนินการต่อ ยกฐานะเป็นงานประจำหรือยกเลิกไป (เขาวดี วิบูลย์ศรี, 2538; ศิริชัย กาญจนวาสี, 2537)

ตอนที่ 2 มโนทัศน์เกี่ยวกับแบบจำลองชิป

Stufflebeam & others (1971) ได้เสนอวิธีการประเมินให้เป็นที่รู้จักโดยทั่วไปในการประเมินหลักสูตรหรือโครงการทางการศึกษา คือแบบจำลองชิป ซึ่ง Stufflebeam & others ได้ให้ความหมายของการประเมินว่า การประเมิน คือกระบวนการกำหนดข้อมูล ตลอดจนเก็บรวบรวมข้อมูล เพื่อการจัดเตรียมข้อมูลเพื่อใช้ในการตัดสินใจเลือกทางเลือกที่เหมาะสม

จากความหมายของการประเมินดังกล่าว มีประเด็นสำคัญที่น่าสนใจ คือ

1. การประเมินเป็นการกระทำเพื่อช่วยในการตัดสินใจ ดังนั้นจึงควรมีข้อสนเทศที่เป็นประโยชน์ต่อผู้ทำการตัดสินใจ

2. การประเมินเป็นกระบวนการต่อเนื่อง เพราะฉะนั้นการประเมินต้องกระทำหรือนำไปใช้ทั้งระบบของหลักสูตรหรือโครงการ

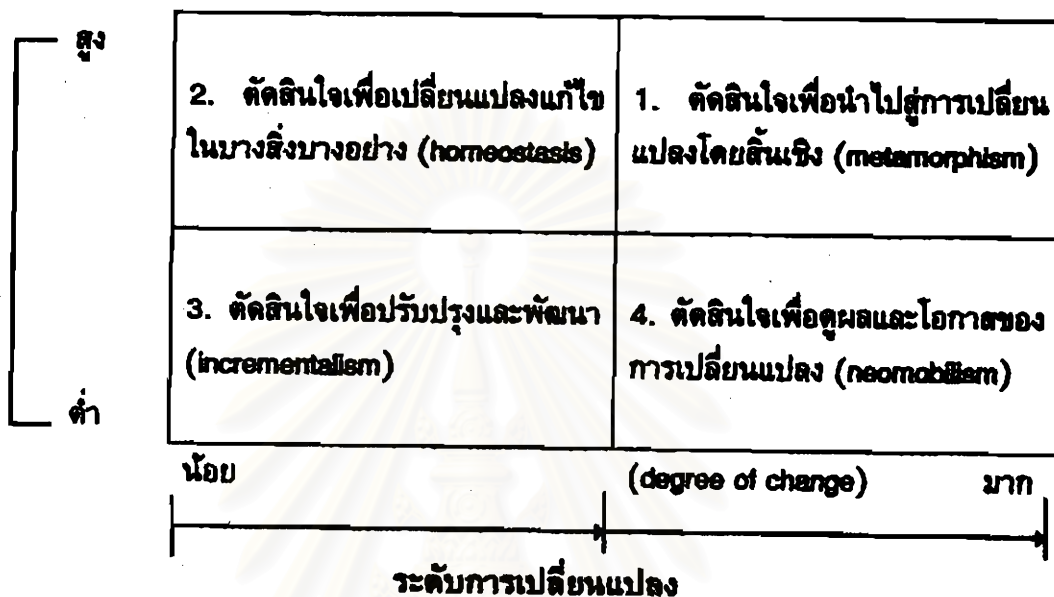
3. กระบวนการประเมินประกอบด้วยขั้นตอนที่สำคัญ 3 ขั้นตอน มีการได้มาซึ่งข้อมูล วิเคราะห์ข้อมูล และการจัดเตรียมข้อมูล ขั้นตอนเหล่านี้เป็นระเบียบพื้นฐานทางการประเมิน

4. ขั้นตอนกำหนดหรือวิเคราะห์ข้อมูล และขั้นจัดเตรียมข้อมูลในกระบวนการประเมินนั้นเป็นกิจกรรมที่ต้องทำร่วมกันระหว่างนักประเมินกับผู้ตัดสินใจ ส่วนขั้นตอนการจัดการข้อมูลเป็นกิจกรรมที่ต้องใช้เทคนิคทางการประเมินจึงเป็นหน้าที่ของนักประเมินเพียงฝ่ายเดียว (Worthen & Sander, 1973)

จากความหมายของการประเมินดังกล่าว เน้นการประเมินเพื่อนำผลการประเมินไปให้ผู้บริหารทำการตัดสินใจ ซึ่งได้แบ่งประเภทการตัดสินใจออกเป็น 4 แนวทาง (Stufflebeam & others, 1971) ดังแผนภูมิที่ 3

แผนภูมิที่ 3 การใช้แบบจำลองชิปกับแนวทางการตัดสินใจ

สารสนเทศ



แนวที่ 1 การตัดสินใจเพื่อนำไปสู่การเปลี่ยนแปลงโดยสิ้นเชิง หรือการเปลี่ยนแปลงทั้งหมด (metamorphosis) ซึ่งเป็นการเปลี่ยนแปลงของโครงสร้างทั้งระบบ การเปลี่ยนแปลงดังกล่าวนี้เป็นการเปลี่ยนแปลงซึ่งต้องมีสารสนเทศมากเพียงพอสำหรับการพิจารณาเพื่อปฏิบัติด้วยความมั่นใจ โอกาสในการที่จะตัดสินใจชนิดนี้ในการศึกษาแทบจะไม่มีเลย

แนวที่ 2 การตัดสินใจเพื่อการตัดสินใจแก้ไขบางสิ่งบางอย่าง (homeostasis) แต่ยังคงรักษาสภาพปกติได้ ดังนั้นมาตรฐานและการควบคุมคุณภาพจะเป็นสิ่งช่วยให้เกิดการเปลี่ยนแปลง ลักษณะการตัดสินใจชนิดนี้มีโอกาสเกิดขึ้นได้มากในชีวิตประจำวัน เช่น การจัดการครูอาจารย์เข้าสอน การกำหนดตารางเรียนของนักเรียน ฯลฯ

แนวที่ 3 การตัดสินใจเพื่อการปรับปรุงและพัฒนา (incrementalism) ผู้ชำนาญการจะเป็นผู้ให้ข้อมูล ความรู้ และเทคนิคต่างๆ สำหรับการทำการตัดสินใจ มุ่งหวังให้เกิดการเปลี่ยนแปลงที่ดีกว่าเดิม การตัดสินใจในแนวที่ 3 นี้ มีสารสนเทศเพียงเล็กน้อย จึงต้องอาศัยการปรึกษาจากผู้เชี่ยวชาญ

แนวที่ 4 การตัดสินใจเพื่อคุณผลและโอกาสการเปลี่ยนแปลง (neomobilsism) โดยอาศัยแนวคิดใหม่ ทฤษฎีหรือข้อค้นพบใหม่มาทดลองใช้ การตัดสินใจในแนวนี้ก่อให้เกิดการเปลี่ยนแปลงค่อนข้างสูง การพัฒนาความรู้จะเป็นตัวกระตุ้นให้เกิดการเปลี่ยนแปลงในแนวนี้

จากแนวทางการตัดสินใจทั้ง 4 ดังกล่าวแล้ว Stufflebeam & others (1971) ได้เสนอแบบจำลองซิป สามารถนำไปใช้ในการประเมินทางการศึกษาได้เป็นอย่างดี ซึ่งประกอบด้วย การประเมิน 4 ด้าน รายละเอียดในการประเมินแต่ละด้านมีดังนี้

1. การประเมินสถานะแวดล้อม (context evaluation) เป็นการประเมินที่มีจุดมุ่งหมายเพื่อให้ได้สารสนเทศมากำหนดวัตถุประสงค์ โดยระบุสภาพแวดล้อม อธิบายสภาพที่คาดหวังไว้กับสภาพความเป็นจริง ความต้องการจำเป็น (needs) และปัญหาเพื่อนำไปแก้ไข โดยวิเคราะห์ปัญหาเพื่อนำไปพัฒนาเป็นวัตถุประสงค์ การประเมินสถานะแวดล้อมนี้จะช่วยให้ได้สารสนเทศที่เป็นประโยชน์แก่ผู้บริหารในการตัดสินใจเพื่อนำไปกำหนดวัตถุประสงค์ของการประเมิน

2. การประเมินปัจจัยเบื้องต้น (input evaluation) เป็นการประเมินที่มีจุดมุ่งหมายในการนำเสนอสารสนเทศ กำหนดวิธีการในการใช้ทรัพยากรเพื่อให้ผลการประเมินบรรลุวัตถุประสงค์ที่กำหนดไว้ในทางปฏิบัติจะระบุในสิ่งต่อไปนี้ คือ ความรับผิดชอบของหน่วยงานที่รับผิดชอบ ยุทธวิธีเพื่อให้บรรลุจุดมุ่งหมายของการประเมิน การประเมินปัจจัยเบื้องต้น ช่วยให้ได้สารสนเทศที่เป็นประโยชน์แก่ผู้บริหารในการตัดสินใจด้านการวางแผนงาน

3. การประเมินกระบวนการ (process evaluation) เป็นการประเมินเพื่อกำกับติดตามการประเมิน หากจุด จุดค้อยของการประเมิน ผลการประเมินจะเป็นสารสนเทศให้ผู้บริหารใช้ในการปรับปรุงการประเมินให้สามารถดำเนินการต่อไปได้และบรรลุวัตถุประสงค์อย่างมีประสิทธิภาพ การประเมินกระบวนการจะมีบทบาทสำคัญในการให้สารสนเทศเป็นระยะๆ เพื่อเป็นการตรวจสอบการประเมินสารสนเทศที่ได้จากการประเมินกระบวนการจะนำไปวิเคราะห์อย่างถี่ถ้วน ในการประเมินกระบวนการจะทำให้ได้สารสนเทศในการตัดสินใจแก่ผู้บริหารในการปรับเปลี่ยนแผนงานดำเนินการได้อย่างเหมาะสม (implementing)

4. การประเมินผลผลิต (product evaluation) เป็นการประเมินที่มีจุดมุ่งหมายเพื่อการพิจารณาการตัดสินใจ หรือการแสดงความคิดเห็นต่อความสำเร็จของโครงการ ทั้งที่ขณะโครงการดำเนินอยู่และโครงการสิ้นสุดลง การประเมินผลผลิตจะได้ข้อมูลที่นำมาช่วยในการตัดสินใจว่ามีกิจกรรมใดบ้างที่ควรทำต่อไป และกิจกรรมใดบ้างที่ควรเลิกทำ หรือนำมาปรับปรุงใหม่ (recycling)

จะเห็นได้ว่าแบบจำลองซิปเป็นแบบจำลองที่ช่วยในการประเมินเพื่อนำสารสนเทศเสนอให้ผู้บริหารทำการตัดสินใจในทุกๆ ระยะ โดยเริ่มประเมินสถานะแวดล้อมตั้งแต่ยังไม่มีโครงการ/หลักสูตร เพื่อตัดสินใจด้านวัตถุประสงค์ ประเมินปัจจัยเบื้องต้นเพื่อตัดสินใจวางแผนงาน ประเมินกระบวนการ เพื่อตัดสินใจกำกับติดตามแผนงาน ปรับปรุงแผนงาน

เพื่อให้บรรลุผล และประเมินผลผลิตเพื่อทำการตัดสินใจ ปรับปรุง ขยายงาน หรือล้มเลิก
โครงการ/หลักสูตรต่อไป

ตอนที่ 3 มโนทัศน์เกี่ยวกับการประเมินอภิมาน (meta evaluation)

3.1 ความหมายการประเมินอภิมาน

ได้มีนักการประเมินทางการศึกษาได้ให้นิยามของการประเมินอภิมานไว้มากมาย
ซึ่งได้ให้นิยามเหมือนและแตกต่างกันไป ดังนี้มาต่อไปนี้

Cook & Gruder (1978) ได้ให้ความหมายการประเมินอภิมาน หมายถึง ศาสตร์
และศิลป์ของการประเมินงานประเมิน ทำให้ได้ข้อค้นพบที่เป็นประโยชน์ โดยมีจุดมุ่งหมาย
เพื่อทำให้การประเมินมีคุณภาพ

Schnandt & Halpern (1988) ได้ให้ความหมายการประเมินอภิมาน หมายถึง
การประเมินงานประเมินโดยถือว่าการประเมินที่ดีต้องเป็นการประเมินที่มีคุณภาพ
ปราศจากความลำเอียง มีประโยชน์ และทันต่อการนำผลการประเมินไปใช้

Scriven (1991) ได้ให้ความหมายการประเมินอภิมาน หมายถึง การประเมินงาน
ประเมินหรือเป็นงานประเมินของผู้ประเมิน โดยใช้วิธีการทางวิทยาศาสตร์หรือศีลธรรม
มาร่วมด้วย

Stufflebeam (1978 cited in Wabha, 1990) ได้ให้ความหมายการประเมิน
อภิมาน หมายถึง กระบวนการจัดหาสารสนเทศเพื่อพิจารณาการประเมินในด้าน
อรรถประโยชน์ การดำเนินการ จรรยาบรรณทางวิชาชีพ และความเพียงพอทางด้าน
การประเมิน เพื่อให้ได้สารสนเทศในการปรับปรุงการประเมินและรายงานผลการประเมิน
จุดแข็ง จุดอ่อนต่อสาธารณชน

สมหวัง พิธิยานุวัฒน์ (2541) ได้ให้ความหมายการประเมินอภิมาน หมายถึง
การตัดสินคุณค่าภายนอกและภายในของการประเมิน

อุทุมพร จามรมาน (2540) ได้ให้ความหมายการประเมินอภิมาน หมายถึง การ
ประเมินหรือการตรวจสอบการประเมินซ้ำ เพื่อให้ได้คำตอบของการประเมินคุณค่า
เหล่านั้น

นอกจากนี้ สุวิมล ร่องวาณิช (2538) ได้ให้ความหมายเพิ่มเติมว่า การประเมิน
อภิมานไม่น่าจะมีความหมายเพียงการประเมินวิธีการหรือการประเมินคุณภาพ
การประเมินอย่างเดี่ยวแต่สามารถทำได้โดยวิธีสังเคราะห์วิธีการที่ใช้ในการประเมิน
(synthesis of evaluation process) แล้วทำการประเมินคุณภาพที่ได้จากการสังเคราะห์
เหล่านั้น

จากคำนิยามดังกล่าวแนวคิดเกี่ยวกับความหมายของการประเมินนิทานพอจะสรุปได้คือเป็นการประเมินงานประเมินหรือคุณภาพของการประเมิน รวมไปถึงการประเมินการตั้งเกณฑ์วิธีการที่ใช้ในการประเมินเพื่อให้ได้สารสนเทศที่เป็นประโยชน์ในการปรับปรุงงานและ/หรือรายงานผลการประเมินในด้านจุดแข็งจุดอ่อนต่อสาธารณะ ซึ่งมีวิธีการในการประเมินนิทานแตกต่างกันไป

3.2 บทบาทของการประเมินนิทาน

บทบาทของการประเมินนิทาน สามารถแยกออกได้เป็น 2 ลักษณะ คือ การประเมินนิทานเชิงความก้าวหน้า (formative meta-evaluation) การประเมินนิทานแบบสรุปรวม (summative meta-evaluation) ซึ่งในการประเมินนิทานทั้งเชิงความก้าวหน้าและเชิงสรุปรวมมีบทบาทที่แตกต่างกันไป (Stufflebeam อ้างถึงใน Berk, 1981; Worthen & Sander, 1987; สมหวัง พิธิยานุวัฒน์, 2541) ดังนี้

1. การประเมินนิทานเชิงความก้าวหน้า (formative meta-evaluation) เป็นการประเมินงานประเมินเชิงสร้างสรรค์ กล่าวคือ เป็นการวางแผนก่อนปฏิบัติการประเมิน การประเมินนิทานเชิงความก้าวหน้าช่วยให้นักประเมินตัดสินใจว่าจะประเมินอะไร นิยามปัญหาที่ประเมิน กำหนดผู้เกี่ยวข้องในการประเมิน กำหนดวัตถุประสงค์ของการประเมินให้กระจ่าง กำหนดข้อสัญญา กำหนดสารสนเทศที่ต้องการ กำหนดวิธีการเก็บรวบรวมข้อมูลและวิธีวิเคราะห์ข้อมูลกำหนดตารางประเมิน จัดหาทีมประเมิน กำหนดงบประมาณ เตรียมข้อสรุปและข้อเสนอแนะ และเลือกรูปแบบและวิธีการรายงานผลการประเมิน โดยการประเมินนิทานในเชิงความก้าวหน้าแสดงได้ดังสมการต่อไปนี้

แผนการประเมิน และการดำเนินการ ประเมิน	=	การประเมิน ตามแนวทาง มาตรฐานที่กำหนด	=	การหลีกเลี่ยงข้อ ผิดพลาดหรือหลุมพราง ของการประเมิน
---	---	--	---	--

โดยสรุปการประเมินนิทานเชิงความก้าวหน้า เป็นการประเมินว่าการวางแผนการประเมินเป็นไปตามมาตรฐานที่กำหนดหรือไม่ ถ้าปฏิบัติตามแนวทางผลการประเมินจะถูกต้องและสมเหตุสมผลและเป็นการหลีกเลี่ยงข้อผิดพลาดโดยทั่วไปของการประเมินด้วย คุณค่าการประเมินจะมีค่าสูงสุดเมื่อดำเนินการประเมินตามมาตรฐานของการประเมินและหลีกเลี่ยงข้อผิดพลาดในการประเมิน

2. การประเมินนิทานเชิงสรุปรวม (summative meta-evaluation) เป็นบทบาทพื้นฐานของการประเมินนิทาน คือ เป็นการแสดงถึงคุณค่ารวมของงานประเมินซึ่งแสดง

ถึงความพร้อมที่จะให้ตรวจสอบได้ของนักประเมินโดยรายงานให้สาธารณชนทราบถึงความ เป็นมาตรฐานของการประเมิน ผลการประเมินอภิमानเชิงสรุปรวมจะเป็นสารสนเทศ ให้ผู้ใช้ผลการประเมินตัดสินใจว่าควรยอมรับผลการประเมินและข้อเสนอแนะจากงาน ประเมินมากน้อยเพียงใด โดยการประเมินเชิงสรุปรวม แสดงดังสมการต่อไปนี้

$\frac{\text{การประเมินภาคปฏิบัติจริงหรือส่วนประกอบ}}{\text{มาตรฐานของการประเมิน}} \leq 1.00$

คุณภาพของงานประเมินสามารถแสดงในรูปของการให้คะแนนและถ่วงน้ำหนัก การประเมินภาคปฏิบัติจริงตามมาตรฐานของงานประเมินภาคปฏิบัติจริงตามมาตรฐาน ของงานประเมิน ถ้าค่าเรโซมีค่าเข้าใกล้ 1.00 แสดงถึงความมีค่าสูงสุดของงานประเมิน ในการประเมินอภิमानเชิงความก้าวหน้าและเชิงสรุปรวม (formative and summative meta-evaluation) เป็นการประเมินอภิमानเชิงความก้าวหน้าและเชิงสรุปรวม ร่วมกัน ดังแสดงความสัมพันธ์ได้ดังสมการต่อไปนี้

$\frac{\text{การประเมินภาคปฏิบัติจริงหรือส่วนประกอบ}}{\text{มาตรฐานของการประเมิน}}$	→	1.00 เมื่อการ ดำเนินการประเมิน	$\text{การประเมินตาม} \quad \text{การหักเนื่อง}$ $\text{แนวที่กำหนด} + \text{ข้อผิดพลาดใน}$ การประเมิน
---	---	-----------------------------------	---

การประเมินอภิमानเชิงสรุปรวมจะมีค่ามากที่สุดเมื่อผู้ประเมินได้ให้ความสนใจใน ข้อติชม ปรับปรุง แก้ไขในสิ่งที่บกพร่องจากการประเมินอภิमानในเชิงความก้าวหน้า

จากการประเมินอภิमानเชิงความก้าวหน้าและการประเมินอภิमानเชิงสรุปรวม จะเห็นได้ว่า การประเมินอภิमानเชิงความก้าวหน้ามีความสำคัญต่อการประเมินอภิमान เชิงสรุปรวม เนื่องจากการประเมินอภิमानเชิงความก้าวหน้า นี้ จะได้สารสนเทศในการ แก้ไขปรับปรุงการประเมิน เพื่อหลีกเลี่ยงข้อผิดพลาดต่างๆ ในการประเมิน และเมื่อ ประเมินอภิमानเชิงสรุปรวม จะพบจุดด้อยของการประเมินน้อยลง ทำให้มีการประเมินมี คุณภาพมากขึ้นนั่นเอง

นอกจากบทบาทของการประเมินดังกล่าว ผู้ประเมินอภิमान (meta evaluator) จะเป็นผู้มีบทบาทในการประเมินอภิमान ซึ่ง Schwardt & Halpern (1988) ได้แบ่ง บทบาทของผู้ประเมินอภิमान (meta evaluator) ออกเป็น 2 แนวทางเช่นกัน คือ

1. ผู้ประเมินอภิमानเชิงความก้าวหน้า (formative meta evaluator) เป็นการ

ประเมิน งานประเมินก่อนดำเนินการโครงการ โดยจัดหาสารสนเทศเพื่อการปรับปรุงกำกับ ติดตามคุณภาพของแผนงานการประเมิน วิธีดำเนินการประเมิน ทำให้การประเมินมี คุณภาพมากขึ้น

2. ผู้ประเมินอภิमानเชิงสรุปรวม (summative meta evaluator) เป็นการประเมิน อภิमानเมื่อการประเมินเสร็จสิ้นแล้วเพื่อดูคุณภาพของการประเมินโดยทั่วไป ได้แก่ คุณค่า ภายใน ความยุติธรรม และความเชื่อถือได้ของการประเมิน

3.3 แนวคิดเกี่ยวกับการประเมินอภิमान

การประเมินอภิमान (meta evaluation) ได้รับการส่งเสริมอย่างแพร่หลายจาก หน่วยงานทางการประเมิน โดยนักประเมินที่กล่าวถึงมโนทัศน์นี้เป็นคนแรก คือ Scriven (1969)อ้างถึงใน Webha (1990) กล่าวว่า การประเมินอภิमानเป็นการประเมินชั้นการ ประเมิน เกี่ยวข้องกับการประเมินวิธีการของการประเมิน โดยผู้ประเมินอภิमानต้องเข้า ไปมีส่วนร่วมต่อการปฏิบัติการประเมิน ต่อมานักประเมินทางการศึกษาได้ให้ความหมาย ของการประเมินว่าเป็นการตรวจสอบทางการศึกษา (Schrandt & Halpern, 1988)โดย การตรวจสอบทางการศึกษามีวิธีการคือ 1) กำกับติดตามการประเมินเพื่อให้ได้คุณภาพ ของการประเมินที่เชื่อถือได้ 2) ประเมินงานประเมินอย่างถูกต้องเพื่อเสนอผลการประเมิน ต่อสาธารณะ

Stufflebeam (1974 อ้างถึงใน Webha 1990) กล่าวว่า การประเมินอภิमान (meta evaluation) สามารถที่จะช่วยให้นักประเมินทราบถึงปัญหาต่างๆ ของการประเมิน และช่วยให้สาธารณชนรับทราบผลการประเมินตามความเป็นจริง นอกจากนี้ Stufflebeam ได้สรุปมโนทัศน์ของการประเมิน (evaluation) และมโนทัศน์ของการประเมินอภิमान (meta evaluation) ไว้ได้ 8 ประเด็น ดังนี้

1. การประเมินเป็นการประเมินคุณค่าภายใน (merit) ส่วนการประเมินอภิमान เป็นการประเมินคุณค่าภายในของงานประเมิน

2. การประเมินจะสนองต่อการตัดสินใจ และแสดงถึงความรับผิดชอบในการปฏิบัติ (accountability) ส่วนการประเมินอภิमानเชิงความก้าวหน้า (formative meta-evaluation) เพื่อทำการตัดสินใจในการกำกับ ติดตามการประเมินและการประเมินเชิงอภิमानเชิง สรุปรวม (summative meta-evaluation) จะให้สารสนเทศเพื่อการแสดงความรับผิดชอบ ในการปฏิบัติ (accountability) ต่อการประเมินที่ผ่านมา

3. การประเมินควรประเมินเป้าหมาย การวางแผน การกำกับติดตาม และ ผลลัพธ์ของโครงการ ส่วนการประเมินอภิमानนั้น ควรประเมินความสำคัญของจุดประสงค์

ในการประเมิน ความพอเพียงในการวางแผนการประเมิน ความพอเพียงในการกำกับ ติดตามแผนการประเมิน ความเพียงพอด้านเทคนิคที่ใช้ในการวิเคราะห์ข้อมูล คุณภาพ และความสำคัญของผลการประเมิน

4. การประเมินจะให้สารสนเทศเชิงบรรยาย เชิงตัดสิน และข้อเสนอแนะ ส่วนการประเมินอภิमानควรจะบรรยาย ตัดสินการประเมิน และควรจะให้ข้อเสนอแนะเพื่อการปรับปรุงการประเมิน และการนำผลประเมินงานประเมินไปใช้ประโยชน์

5. การประเมินจะสนองต่อบุคคลที่เกี่ยวข้องกับการประเมินทุกฝ่าย ส่วนการประเมินอภิमानควรจะสนองต่อผู้ประเมิน และบุคคลที่สนใจในการประเมิน

6. ในการประเมินนั้น ผู้ประเมินภายในจะเป็นผู้ประเมินเชิงความก้าวหน้าเพื่อทำการตัดสินใจเกี่ยวกับโครงการ ส่วนผู้ประเมินภายนอกจะประเมินเชิงสรุปรวมเพื่อแสดงความรับผิดชอบต่อการปฏิบัติ ด้วยเหตุนี้ นักประเมินอภิमानควรจะมีการประเมินอภิमानเชิงความก้าวหน้าเพื่อทำการตัดสินใจเกี่ยวกับการประเมิน และนักประเมินอภิमानควรจะประเมินอภิमानเชิงสรุปรวมเมื่อกิจกรรมหรือการประเมินสิ้นสุดเพื่อแสดงความรับผิดชอบในการปฏิบัติต่อการประเมิน

7. การประเมินเป็นกระบวนการระบุข้อมูลที่ต้องการ และจัดเก็บวิเคราะห์ข้อมูล เพื่อใช้สารสนเทศในการตัดสินใจและใช้สารสนเทศในการแสดงความรับผิดชอบต่อโครงการ ส่วนการประเมินอภิमानจะต้องระบุคำถามที่ชัดเจน การเก็บรวบรวมข้อมูล การจัดระบบ และการวิเคราะห์สารสนเทศที่ได้รับเพื่อประยุกต์ใช้สารสนเทศในการตัดสินใจ และแสดงความรับผิดชอบต่อการประเมินอย่างเหมาะสม

8. การประเมินจะต้องมีเทคนิค มีประโยชน์ และมีประสิทธิภาพ ส่วนการประเมินอภิमानต้องมีเกณฑ์การประเมินที่เพียงพอ

นอกจากนี้ Cook & Gruder (1978) ได้จำแนกการประเมินอภิमानออกเป็น 3 แขนง คือ 1) วิจัยโดยใช้ลักษณะการประเมินที่เหมาะสม ได้แก่ การประเมินเชิงก้าวหน้า (formative evaluation) หรือ การประเมินเชิงสรุปรวม (summative evaluation) 2) วิเคราะห์จากข้อมูลทุติยภูมิ (secondary data) เป็นการประเมินซ้ำจากข้อมูลทุติยภูมิ เพื่อตอบคำถามวิจัยเดิมหรือคำถามวิจัยใหม่ โดยใช้เทคนิคสถิติที่สูงขึ้น 3) วิจัยเชิงประจักษ์ย้อนวิจัยประเมินผล นั่นคือ วิจัยย้อนวิจัย Glass (1981) ได้เพิ่มงานวิจัยออกมาอีก 1 แขนง คือ การสังเคราะห์งานวิจัย (meta analysis) ซึ่งเป็นระเบียบวิธีการศึกษาข้อเท็จจริง เพื่อตอบปัญหาใดปัญหาหนึ่ง โดยการรวบรวมงานวิจัยเกี่ยวกับปัญหานั้น หลายๆ เรื่อง มาศึกษาวิเคราะห์ และนำเสนอข้อมูลอย่างเป็นระบบเพื่อให้ได้คำตอบที่เป็นข้อยุติ การประเมินอภิमानได้แตกสาขาออกเป็น 4 สาขา Cook & Gruder (1978) ได้แสดงไว้ดังแผนภูมิที่ 4

แผนภูมิที่ 4 การแตกสาขาการประเมินอภิมาน (meta evaluation design)



Cook & Gruder (1978)

การประเมินอภิมานมีบทบาทสำคัญในการช่วยกำกับติดตามแผนการประเมิน วิธีดำเนินการประเมิน รวมทั้งประเมินคุณภาพของการประเมินอภิมานโดยรวมด้วย ศิริชัย กาญจนวาสิ (2537) ได้จำแนกรูปแบบการประเมินอภิมานได้ดังต่อไปนี้

1. การติดตามตรวจสอบการปฏิบัติการประเมิน
2. การวิพากษ์วิจารณ์
 - 2.1 การวิพากษ์วิจารณ์อย่างเป็นทางการ เช่น รูปแบบการวิจารณ์ทางปรัชญา ศิลปะ บ้านเชิง เป็นต้น
 - 2.2 การอภิปรายโดยผู้เชี่ยวชาญ

3. การเปรียบเทียบกับเกณฑ์มาตรฐาน

4. การประเมินซ้ำ

4.1 การวิเคราะห์และตีความหมายใหม่

4.2 การศึกษาเปรียบเทียบกับผลการประเมินอื่น

ส่วน Smith (1990) ได้กล่าวว่า การศึกษารายกรณี (case study) เป็นวิธีการหนึ่งที่น่ามาใช้ในการประเมินอภิमान เพื่อให้ได้ข้อมูลที่ลึกซึ้ง ทำความเข้าใจได้มากกว่าที่จะพิจารณาตัดสินถึงคุณค่าภายนอกเพียงอย่างเดียว โดยวิธีการศึกษารายกรณีนี้ จะได้สารสนเทศในการปรับปรุงการประเมิน จุดดี จุดด้อย ของการประเมินนั้นๆ

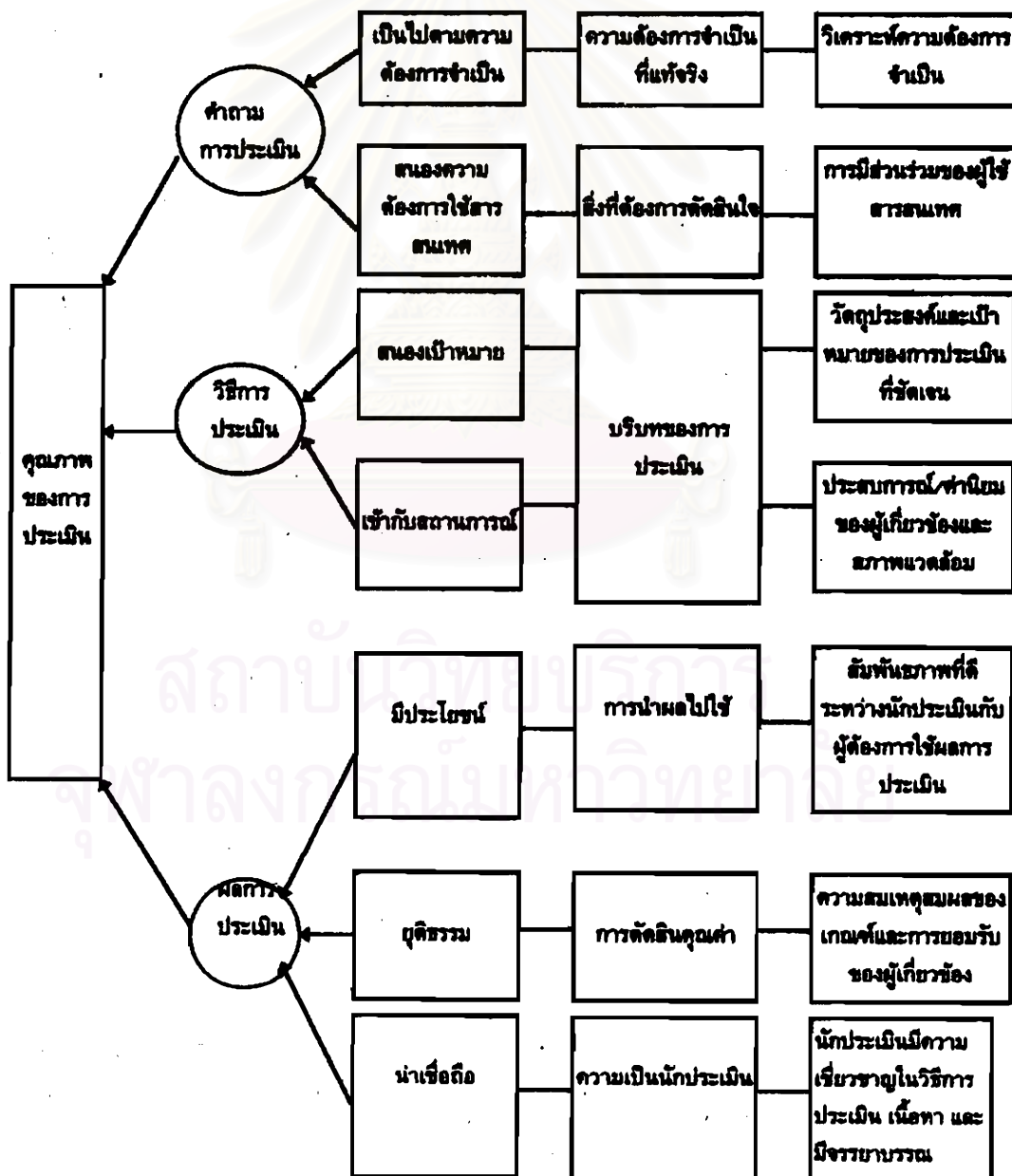
นอกจากการทำทการอภิमानโดยใช้การศึกษารายกรณีแล้ว Brinkeroff et.al(1983) อ้างถึงใน Worthen & Sander (1987) ได้เสนอวิธีดำเนินการในการประเมินอภิमान ดังตารางที่ 1

	ประเด็นของการประเมินอภิमान		
	เชิงความก้าวหน้า		เชิงสรุปรวม
	ประเมินแผนการประเมิน	ประเมินงานประเมินเชิงความก้าวหน้า	ประเมินงานประเมินหลังเสร็จสิ้น
วิธีดำเนินการในการประเมินอภิमान	จ้างที่ปรึกษาการประเมิน เช่น ผู้ประเมินผู้เชี่ยวชาญทางการวัด หรือผู้เชี่ยวชาญเฉพาะสาขาวิชา คณะผู้ทบทวนแผน เช่น กลุ่มที่ให้คำปรึกษา ทบทวนแผนในการประเมิน เช่น แผนงาน แผนการดำเนินการ	ให้ผู้สังเกตที่เป็นอิสระจากกัน เช่น ผู้ประเมินอภิमान ทีมผู้ประเมินคณะที่ปรึกษาในการประเมิน อำนวยการประเมินอภิमानเชิงความก้าวหน้า เช่น แผนการประเมินระยะต่างๆ งบประมาณ ปัจจุบัน รายละเอียดเกี่ยวกับการเก็บรวบรวมข้อมูล	ส่งรายงานการประเมินต่อผู้ประเมิน ที่ปรึกษาคณะผู้ให้คำปรึกษาผู้เชี่ยวชาญทางการประเมิน โดยให้ท่านทบทวนรายงานการประเมินเพื่อประเมินคุณภาพของการประเมิน

ตารางที่ 1 แสดงวิธีดำเนินการในการประเมินอภิमान Brinkeroff et.al (1983) อ้างถึงใน Worthen & Sander (1987)

หลักในการพิจารณาคูณภาพของการประเมินนั้น สามารถตัดสินใจจากคุณค่าของการประเมิน 3 ด้าน คือ ค่าตามการประเมิน วิธีการประเมิน และผลการประเมิน (ศิริชัย กาญจนวาสี, 2537) โดยด้านค่าตามการประเมินนั้นจะต้องเป็นไปตามความต้องการจำเป็นและสนองต่อความต้องการในการใช้สารสนเทศ ส่วนด้านวิธีการประเมินจะต้องสนองต่อเป้าหมายและเข้ากับสถานการณ์ในการประเมิน และด้านผลการประเมินจะต้องเป็นประโยชน์ ยุติธรรม และน่าเชื่อถือ ดังเสนอได้ดังแผนภูมิที่ 5

แผนภูมิที่ 5 เกณฑ์การประเมินคุณภาพของการประเมิน (ศิริชัย กาญจนวาสี, 2537)



3.4 มาตรฐานของการประเมิน

มาตรฐานของการประเมิน ตามที่คณะกรรมการพัฒนาเกณฑ์มาตรฐานสำหรับการประเมินทางการศึกษา (Finn, 1997; Mandaus Scriven & Stufflebeam, 1989; Stufflebeam & Shinkfield, 1990; ศิริชัย กาญจนวาสี, 2537; สมหวัง พิธิยานุวัฒน์, 2541) ซึ่งประกอบด้วยมาตรฐานการประเมิน 4 ด้าน คือ มาตรฐานด้านอรรถประโยชน์ (utility standards) ประกอบด้วย 8 เกณฑ์ มาตรฐานด้านความเป็นไปได้ (feasibility standards) ประกอบด้วย 3 เกณฑ์ มาตรฐานด้านความเหมาะสม (propriety standards) ประกอบด้วย 8 เกณฑ์ มาตรฐานด้านความถูกต้อง (accuracy standards) ซึ่งประกอบด้วย 11 เกณฑ์ รวมเกณฑ์ทั้งสิ้น 30 เกณฑ์ ดังรายละเอียดต่อไปนี้

1. มาตรฐานด้านอรรถประโยชน์ซึ่งเป็นการประเมินด้านกระบวนการที่ก่อให้เกิดสารสนเทศตรงกับความต้องการของผู้ใช้ผลการประเมินเป็นสำคัญ ประกอบด้วย 8 เกณฑ์ คือ

- 1.1 มีการระบุผู้เกี่ยวข้องที่ต้องการใช้สารสนเทศ
- 1.2 ความเป็นที่เชื่อถือของผู้ประเมิน ผู้ประเมินต้องรู้สิ่งที่มุ่งประเมิน
- 1.3 มีการรวบรวมข้อมูลครอบคลุมและตอบสนองความต้องการของผู้ใช้สารสนเทศของผู้เกี่ยวข้อง
- 1.4 การแปลความหมายและการตัดสินคุณค่ามีความชัดเจน
- 1.5 รายงานการประเมินมีความชัดเจนทุกชั้นตอน
- 1.6 การเผยแพร่ผลการประเมินไปยังผู้เกี่ยวข้องอย่างทั่วถึง
- 1.7 รายงานการประเมินเสร็จทันสำหรับนำไปใช้ประโยชน์
- 1.8 การประเมินส่งผลกระทบในการกระตุ้นให้มีการดำเนินการประเมินต่อไปอย่างต่อเนื่อง

2. มาตรฐานด้านความเป็นไปได้ (feasibility standards) เน้นความเป็นจริงของการประเมิน ประกอบด้วย 3 เกณฑ์ คือ

- 2.1 วิธีการประเมินสามารถนำไปใช้ปฏิบัติได้จริง
- 2.2 การเป็นที่ยอมรับของผู้ถูกประเมินและผู้ได้รับผลกระทบจากการประเมิน ตลอดจนผู้ที่เกี่ยวข้องทุกฝ่าย
- 2.3 ผลที่ได้มีความคุ้มค่ากับผลงาน สามารถนำไปปรับปรุงปฏิบัติงานขยายงาน หรือดัดแปลง

3. มาตรฐานด้านความเหมาะสม (propriety standards) มุ่งเน้นเกี่ยวกับเรื่องกฎระเบียบ ศีลธรรม จรรยาบรรณ ที่เกี่ยวข้องกับการประเมิน ประกอบด้วย 8 เกณฑ์

- 3.1 มีการกำหนดข้อตกลงอย่างเป็นทางการ
 - 3.2 มีการแก้ปัญหาของการประเมินด้วยความซื่อสัตย์ ไม่บิดเบือนความเป็นจริงอย่างสันติวิธี
 - 3.3 รายงานการประเมินอย่างตรงไปตรงมา เปิดเผย และคำนึงถึงข้อจำกัดของการประเมิน
 - 3.4 มีการให้ความสำคัญต่อสิทธิในการรับรู้ข่าวสารของบุคคลทั่วไปอย่างเปิดเผย มีการเผยแพร่ผลการประเมินและเปิดโอกาสให้ผู้สนใจได้ศึกษาผลการประเมิน
 - 3.5 มีการคำนึงถึงสิทธิส่วนตัวของกลุ่มตัวอย่าง
 - 3.6 มีการเคารพในการปฏิบัติพันธะของผู้ที่เกี่ยวข้อง รวบรวมข้อมูลจากบุคคลหลายฝ่ายที่เกี่ยวข้อง
 - 3.7 มีรายงานผลการประเมินที่สมบูรณ์ ยุติธรรม และเสนอทั้งจุดเด่น จุดด้อยของสิ่งที่ประเมิน
 - 3.8 ผู้ประเมินทำการประเมินด้วยความรับผิดชอบและมีจรรยาบรรณ
4. มาตรฐานด้านความถูกต้อง (accuracy standards) เกี่ยวข้องกับการใช้เครื่องมือที่มีคุณภาพ ตลอดจนวิธีเก็บข้อมูล จนวิเคราะห์ข้อมูลตามหลักวิชาการ ซึ่งประกอบด้วย 11 เกณฑ์
- 4.1 มีการระบุวัตถุประสงค์ของการประเมินอย่างชัดเจน
 - 4.2 มีการวิเคราะห์บริบทของการประเมินในรายงานการประเมินอย่างครอบคลุมและเพียงพอต่อการประเมินอย่างมีประสิทธิภาพ
 - 4.3 มีการบรรยายจุดประสงค์และกระบวนการประเมินอย่างชัดเจน
 - 4.4 มีการบรรยายแหล่งข้อมูลและการได้มาอย่างชัดเจน
 - 4.5 ใช้เทคนิคและเครื่องมือที่ความตรง และได้มาซึ่งสารสนเทศที่มีความถูกต้องแม่นยำ
 - 4.6 ใช้เทคนิคและเครื่องมือที่มีความเที่ยง และได้มาซึ่งสารสนเทศที่มีความคงเส้นคงวา
 - 4.7 มีการจัดระบบควบคุมสำหรับการเก็บรวบรวมข้อมูลวิเคราะห์และรายงาน
 - 4.8 มีการวิเคราะห์ข้อมูลเชิงปริมาณ
 - 4.9 มีการวิเคราะห์ข้อมูลเชิงคุณภาพ
 - 4.10 มีการลงสรุปที่มีเหตุผลและผลการประเมินสนับสนุน
 - 4.11 มีการเขียนรายงานที่มีความเป็นปรนัย

นอกจากนี้ Nyirenda (1994) ได้ร่วมกับมาตรฐานวิชาชีพการสอนแห่งชาติ (National Board for Professional Teaching Standards) หรือ NBPTS ได้ศึกษาว่า ทำอย่างไรการสอนจะบรรลุผลได้ จึงได้พัฒนาเกณฑ์การประเมินนิยามเพื่อประเมินระบบการออกไปประกอบวิชาชีพครูจีน ประกอบด้วยเกณฑ์ในการประเมินดังนี้

1. ความเป็นไปได้ทางการบริหาร (administrative feasibility)
2. ความเชื่อถือได้ต่อสาธารณชน (public credibility)
3. ความเชื่อถือได้ทางด้านวิชาชีพ (professional credibility)
4. การป้องกันทางด้านกฎหมาย (legal defeasibility)
5. ความสามารถทางการปฏิบัติทางด้านเศรษฐศาสตร์ (economic affordability)

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

Webb (1990) ได้ทำวิทยานิพนธ์ระดับปริญญาเอก ของมหาวิทยาลัยเพิทเบอร์ก (University of Pittsburgh) เรื่องการประเมินนิยามเชิงสรุปรวม (summative meta-evaluation) เกี่ยวกับการอบรมเลี้ยงดูเด็กแรกเกิดของประเทศกำลังพัฒนา 4 ประเทศ โดยได้ประยุกต์ใช้เกณฑ์มาตรฐานของคณะกรรมการการประเมินทางการศึกษา (the Joint Committee on Standards to Educational Evaluation) มีจุดประสงค์เพื่อประเมินคุณภาพของรายงานการประเมินเกี่ยวกับการอบรมเลี้ยงดูเด็กแรกเกิดของประเทศอินเดีย Sierra Leone พม่า และกานา โดยใช้มาตรฐานของคณะกรรมการพัฒนาเกณฑ์การประเมินทั้ง 30 ข้อ ดำเนินการโดย 1) เลือกรายงานการประเมินเกี่ยวกับการอบรมเลี้ยงดูเด็กแรกเกิดที่สมบูรณ์ 4 ฉบับ จากทั้งหมด 13 เรื่อง 2) นำรายงานการประเมินทั้ง 4 ฉบับไปให้ผู้เชี่ยวชาญ 4 ท่าน (ประกอบด้วย ผู้สำเร็จปริญญาเอก 3 ท่าน ปริญญาโท 1 ท่าน มีประสบการณ์ทางด้านประเมินอย่างน้อย 3 ปี) ทำการตัดสินโดยเปรียบเทียบกับเกณฑ์มาตรฐานทั้ง 30 ข้อ ประกอบด้วยมาตรฐานด้านอรรถประโยชน์ 8 ข้อ มาตรฐานด้านความเป็นไปได้ 3 ข้อ มาตรฐานด้านความเหมาะสม 8 ข้อ และมาตรฐานด้านความถูกต้อง 11 ข้อ เกณฑ์การผ่านในแต่ละมาตรฐานนั้นยึดผู้เชี่ยวชาญ 3 ใน 4 หรือร้อยละ 75 มีความเป็นสอดคล้องกันจึงถือว่ามาตรฐานนั้น ผลการประเมินปรากฏว่ารายงานการประเมินของประเทศกานา ได้รับความเห็นจากผู้เชี่ยวชาญว่ามีมาตรฐานสูงสุด 27 มาตรฐานการประเมิน คิดเป็นร้อยละ 90 รองลงมาคือ รายงานการประเมินของประเทศอินเดีย และ Sierra Leone มี 25 มาตรฐานการประเมิน คิดเป็นร้อยละ 83.30 ส่วนรายงานการประเมินของประเทศพม่า ได้รับความเห็นจากคณะผู้เชี่ยวชาญน้อยที่สุด 19 มาตรฐานการประเมิน คิดเป็นร้อยละ 63.30

Boyd & Windsor (1993) ได้ทำการประเมินอภิमानเชิงสรุปรวม (summative meta-evaluation) รายงานการประเมินโภชนาการการศึกษาของหญิงมีครรภ์ จำนวน 5 ฉบับ โดยใช้เกณฑ์การประเมินของ Haynes et. al. ซึ่งประกอบด้วยเกณฑ์ดังนี้ 1) การออกแบบการวิจัย 2) ขนาดกลุ่มตัวอย่าง ความเป็นตัวแทนของกลุ่มตัวอย่าง 3) ลักษณะประชากร 4) คุณภาพของการวัด 5) ความเหมาะสมและความเป็นตัวแทนของกลุ่มควบคุม โดยในแต่ละเกณฑ์ดังกล่าวจะมีคะแนนเต็ม 5 คะแนน รวมทั้ง 5 เกณฑ์เท่ากับ 25 คะแนน เช่น เกณฑ์ที่ 1 การออกแบบการวิจัย มีคะแนนดังนี้ จะได้ 4 คะแนน ถ้ามีกลุ่มควบคุม มีการสุ่ม ใช้แบบสอบ pretest-posttest จะได้ 3 คะแนน ถ้าออกแบบเปรียบเทียบกลุ่มควบคุมกลุ่มไม่เท่าเทียมกัน จับคู่สถานการณทางโภชนาการ อายุ สถานภาพการสูบบุหรี่ จะได้ 2 คะแนน ถ้าออกแบบเปรียบเทียบกลุ่มควบคุม กลุ่มไม่เท่าเทียมกัน จะได้ 1 คะแนน ถ้าออกแบบ pretest-posttest 1 กลุ่ม และมีคะแนนโบนัส 1 คะแนน เมื่อใช้กลุ่มเท่าเทียมกันและมีการควบคุม จากนั้นนำรายงานการประเมินมาเปรียบเทียบกับเกณฑ์ ผลการประเมินอภิमानเชิงสรุปปรากฏผลดังนี้ รายงานการประเมินของ Hunt et. al. ได้คะแนนสูงสุด 18 คะแนน, รายงานการประเมินของ Sweeney et. al., Ershoff et. al., Bowering et. al. และ Bruce and Tchado ได้ 13, 11, 10 และ 7 คะแนนตามลำดับ

จากงานวิจัยการประเมินอภิमानโภชนาการการศึกษาของหญิงมีครรภ์ดังกล่าว จะเห็นว่าเป็นการประเมินอภิमानเชิงสรุปรวม (summative meta-evaluation) โดยประเมินในลักษณะเปรียบเทียบกับเกณฑ์ที่กำหนดไว้เพื่อเสนอจุดดีจุดด้อยของการประเมินต่อสาธารณะ

Gallegos (1994) ได้ร่วมกับศูนย์เพื่อการวิจัยและการประเมินครูทางการศึกษา (the center for research on educational accountability and teacher evaluation) หรือ CREAT ได้ศึกษาการประเมินอภิमानรูปแบบการประเมินโรงเรียนในสหรัฐอเมริกา จำนวน 50 รูปแบบ (Meta Evaluation School of Evaluation Models) โดยคณะผู้ประเมินอภิमान จะทำการประเมินงานประเมินอย่างเป็นอิสระจากกันจะประเมินงานประเมิน โดยการสังเคราะห์โมเดลที่ใช้ในการประเมิน แล้วตรวจสอบกระบวนการตามเกณฑ์ดังนี้ 1) ความมุ่งหมายของการประเมิน 2) กระบวนการ 3) ขนาดของโรงเรียน 4) บทบาทในการประเมิน เป็นการประเมินเชิงความก้าวหน้า หรือเป็นการประเมินเชิงสรุปรวม 5) ลักษณะการประเมินเชิงปริมาณ เชิงคุณภาพ หรือทั้งเชิงปริมาณและเชิงคุณภาพ 6) การวางแผน/การตัดสินใจ ประเมินบริบท ปัจจัยเบื้องต้น กระบวนการ และผลผลิต (รวมทั้งการนำไปตัดสินใจ) 7) การแสดงความรับผิดชอบ (accountability) ว่ามีการพัฒนาจากความต้องการ ความเป็นธรรม ความเป็นไปได้ มีความเหมาะสมในการวางแผนทรัพยากร 8) มาตรฐานการประเมิน มาตรฐานด้านอรรถประโยชน์

มาตรฐานด้านความเป็นไปได้ มาตรฐานด้านความเหมาะสม มาตรฐานด้านความถูกต้อง
9) จุดแข็งจุดอ่อนของการประเมิน

จากการศึกษาดังกล่าวจะเห็นว่าเป็นการสังเคราะห์วิธีการประเมิน ซึ่งมีลักษณะคล้ายกับการสังเคราะห์งานวิจัย แล้วทำประเมินการประเมินที่สังเคราะห์อีกที การศึกษาลักษณะนี้จะทำให้ทราบลักษณะของการประเมินโดยทั่วไปของโมเดลการประเมินโรงเรียนดังกล่าว

Lee (1996) ได้ศึกษาการประเมินอภิมานของสถาบันเทคโนโลยีแห่งชาติ กระทรวงศึกษาธิการได้ทวน ศึกษาจากคณะผู้ประเมินและผู้มีส่วนได้ส่วนเสียจากการพิจารณาผลการประเมิน จุดหลักของการประเมินอภิมานอยู่ที่การเปรียบเทียบการประเมินว่าสัมพันธ์กับมาตรฐานด้านอรรถประโยชน์ มาตรฐานด้านความเป็นไปได้ มาตรฐานด้านความเหมาะสม มาตรฐานด้านความถูกต้อง ของสถาบันเทคโนโลยีได้ทวน โดยใช้เกณฑ์มาตรฐานความสำคัญทั้ง 30 เกณฑ์ ในการประเมินอภิมานใช้การเก็บรวบรวมข้อมูลโดยใช้แบบสอบถามและแบบสัมภาษณ์แบบเป็นทางการ จากคณะผู้ประเมินและผู้มีส่วนได้ส่วนเสีย ผลการศึกษาจะนำไปเป็นข้อมูลพื้นฐานในการปรับปรุงโมเดลการประเมินนโยบาย วิธีการและการปฏิบัติงาน

ผลการศึกษาทราบว่า 1) คณะผู้ประเมินและผู้มีส่วนได้ส่วนเสียเชื่อในเกณฑ์มาตรฐานทั้ง 30 เกณฑ์ สามารถใช้ในการกำหนดทิศทางกระบวนการและการปฏิบัติ 2) คณะผู้ประเมินและผู้มีส่วนได้ส่วนเสียให้ผลการประเมินที่เหมือนกันในมาตรฐานด้านอรรถประโยชน์ มาตรฐานด้านความเป็นไปได้ มาตรฐานด้านความเหมาะสมและ มาตรฐานด้านความถูกต้องของการประเมินสถาบันเทคโนโลยีแห่งชาติได้ทวน 1990

นอกจากนี้มหาวิทยาลัย บริกแฮม ยิง (Brigham Young University) ได้ทำการประเมินโครงการประเมินระบบการเรียนรู้อิงโรงเรียนในเมืองนิวยอร์ก (the New York city public integrated learning systems project) เพื่อดูผลการประเมินการใช้ระบบการเรียนรู้อิงบวอเทอร์ฟอร์ด (the Waterford Integrated Learning system) หรือ WILT ผู้ทำการประเมินอภิมานประกอบด้วย ดร.ฟินน์, สตีเวนส์, สตีฟเฟิลบีม และเวลเบอร์ก (Drs. Finn Stevens, Stuffiebeam and Walberg) โดยคณะผู้ประเมินจะประเมินอภิมานทั้งเชิงความก้าวหน้าและสรุปรวม (formative and summative meta-evaluation) เพื่อให้ได้ผลการประเมินในด้าน จุดแข็ง จุดอ่อน ของแผนงานและกิจกรรมการประเมิน จึงมีวิธีดำเนินการดังนี้

1. ทบทวนแผนการประเมิน วิธีดำเนินการเป็นระยะๆ และดูสารสนเทศการประเมินปัจจุบันว่าเป็นอย่างไร

2. เยี่ยมชมการใช้ระบบการเรียนรู้แบบวอเตอร์ฟอร์ด (WILS) ภายในโรงเรียนต่าง ๆ เป็นระยะ ๆ

3. มีการสนทนาแลกเปลี่ยนความคิดเห็นระหว่างผู้ประเมินอภิमानกับคณะผู้ประเมินโครงการ รวมทั้งผู้มีส่วนได้ส่วนเสียอื่น ๆ เป็นระยะ ๆ (โดยเฉพาะอย่างยิ่งผู้สนับสนุนโครงการ) เพื่อข้อมูลความก้าวหน้าในปัจจุบัน และความต้องการของผู้มีส่วนได้ส่วนเสีย

4. วิเคราะห์ระบบการประเมินของรายงานการประเมินขั้นสุดท้าย

5. เขียนร่างรายงานการประเมินอภิमानของผู้ประเมินอภิमानแต่ละคน

6. ถกปัญหาเพื่อให้ได้ฉันทามติ (consensus) เกี่ยวกับรายงานการประเมินอภิमान

7. พัฒนาเป็นรายงานการประเมินอภิमानขั้นสุดท้าย

สรุปงานวิจัยของ Finn (1997) เป็นการประเมินอภิमानเชิงความก้าวหน้าและเชิงสรุปรวม (formative and summative meta-evaluation) เพื่อกำกับติดตามการประเมินโครงการระบบการเรียนรู้ของโรงเรียนในเมืองนิวยอร์ก ให้เป็นไปตามแผนการประเมิน แล้วจากนั้นก็ทำการประเมินคุณภาพเชิงสรุปรวม เพื่อรายงานผลจุดดีจุดด้อยต่อสาธารณะ ให้เป็นที่ยอมรับได้กับบุคคลทุกฝ่าย

จากการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องเกี่ยวกับการประเมินอภิमानที่นำมาใช้ในการประเมินทางการศึกษา ได้นำมาตรฐานของคณะกรรมการพัฒนาเกณฑ์สำหรับการประเมินทางการศึกษามาใช้กันอย่างแพร่หลาย ประกอบไปด้วยมาตรฐานการประเมินทั้ง 4 ด้าน คือ มาตรฐานด้านอรรถประโยชน์ 8 เกณฑ์ มาตรฐานด้านความเป็นไปได้ 3 เกณฑ์ มาตรฐานด้านความเหมาะสม 8 เกณฑ์ มาตรฐานด้านความถูกต้อง 11 เกณฑ์ รวมทั้งสิ้น 30 เกณฑ์ ซึ่งในการประเมินอภิमानการใช้แบบจำลองชิปในการประเมินโครงการการศึกษาครั้งนี้ จะนำเกณฑ์มาตรฐานดังกล่าวมาใช้ในการประเมิน โดยประเมินอภิमानในเชิงสรุปรวม (summative meta-evaluation) เปรียบเทียบผลการประเมินกับมาตรฐานของคณะกรรมการพัฒนาเกณฑ์สำหรับการประเมินทางการศึกษาแยกตามรูปแบบการนำแบบจำลองชิปมาใช้ในการประเมินโครงการการศึกษา เพื่อให้ทราบถึงคุณภาพของงานประเมินที่ใช้แบบจำลองชิปในการประเมินโครงการการศึกษาตามเกณฑ์มาตรฐานดังกล่าว