

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยในครั้งนี้ เป็นการประเมินผลคุณลักษณะที่พึงประสงค์ของมหาบัณฑิต คณะศึกษาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ ตามความคิดเห็นของผู้บังคับบัญชา โดยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง เพื่อเป็นแนวทางการศึกษาดังนี้

1. การประเมิน

- 1.1 ความหมายการประเมินผล
- 1.2 นิยามการประเมิน
- 1.3 ปรัชญาการประเมิน
- 1.4 ประวัติและพัฒนาการของการประเมิน
- 1.5 จุดหมาย บทบาท และหน้าที่ของการประเมิน
- 1.6 ทฤษฎีการประเมิน
- 1.7 ทฤษฎีรูปแบบการประเมิน
- 1.8 ทฤษฎีการนำผลการประเมินไปใช้ประโยชน์
- 1.9 จรรยาบรรณของนักประเมิน

2. คุณลักษณะที่พึงประสงค์

3. งานวิจัยที่เกี่ยวข้อง

1. การประเมิน

1.1 ความหมายการประเมินผล

“การประเมินผล” มีนักการศึกษาได้ให้ความหมายไว้ต่าง ๆ กัน ดังนี้

กู๊ด (Good, 1973 : 220) การประเมินผล หมายถึง กระบวนการค้นหาหรือตัดสินคุณค่าหรือจำนวนของบางสิ่งบางอย่าง โดยใช้มาตรฐานของการประเมิน รวมทั้งการตัดสินใจโดยอาศัยเกณฑ์ภายในและ / หรือเกณฑ์ภายนอก

กลอนลันด์ (Gronlund, 1976 : 5 - 6) ได้อธิบายคำว่า การประเมินผลได้อย่างละเอียด ซึ่งสรุปความได้ดังนี้คือ ในเรื่องการเรียนการสอน การประเมินผล หมายถึง ขบวนการที่ใช้มาตรฐานเพื่อการตัดสินใจว่า นักเรียนนักศึกษาได้ผ่านจุดมุ่งหมายเชิงพฤติกรรมได้หรือไม่

สแตนเลย์ และฮอบกินส์ (Stanley and Hopkins, 1978 : 4 - 7) ได้ชี้ให้เห็นว่า การตัดสินใจงานทางการศึกษาเป็นการประเมินผลการศึกษาอย่างหนึ่ง

ไลด์แมนและเมอร์เอนดา (Lindeman and Merenda, 1979 : 15 - 16) ได้อธิบายความหมายคำว่า การประเมินผลนั้นมีความหมายที่ลึกซึ้งกว่าการวัดผล

เชอร์เซอร์ และลินเดน (Shertzer and Linden, 1979 : 13) การประเมินผล หมายถึง การมีความคิดเห็นเกี่ยวกับสิ่งใดสิ่งหนึ่ง การตัดสินใจความเพียงพอของสิ่งใดสิ่งหนึ่ง หรือการตัดสินใจคุณค่าของสิ่งใดสิ่งหนึ่ง

บราวน์ (Brown, 1983 : 14) การประเมินผล หมายถึง การตีความหมายในเชิงปฏิบัติการของบุคคลโดยตีค่าตีราคาในสิ่งนั้น เพื่อตอบคำถามว่าดีแค่ไหน

เมห์เรน และเลห์มาน (Mehrens and Lehman, 1984 : 5) การประเมินผล หมายถึง การตัดสินใจความสอดคล้องต้องกันระหว่างการปฏิบัติกับวัตถุประสงค์

เกย์ (Gay, 1985 : 6) ได้เขียนเป็นความหมายของคำว่า การประเมินผล เป็น 2 นัย คือ

1. การประเมินผลเป็นขบวนการที่เป็นระบบของการจัดเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูลเพื่อที่จะตัดสินใจในการที่จะสรุปผลเรื่องของจุดมุ่งหมายว่า มีปริมาณมากน้อยเท่าไรและได้มีนักเรียนนักศึกษาได้เรียนรู้แบบสัมฤทธิ์ผลหรือไม่

2. การประเมินผลเป็นขบวนการที่เป็นระบบของการเก็บรวบรวมข้อมูลและการวิเคราะห์ ข้อมูลเพื่อการตัดสินใจครั้งสุดท้าย

เชส (Chase, 1987 : 7) การประเมินผล หมายถึง การนำจำนวนที่ได้จากการวัดมาเทียบกับเกณฑ์ เพื่อตัดสินใจคุณค่าของจำนวนที่สังเกตได้

สตัฟเฟิลบีม และซิงค์ฟิลด์ (Stufflebeam and Shinkfield, 1990 : 3) ได้อธิบายว่า การประเมินผลมีความหมายที่เป็นมาตรฐานสากลว่า “การประเมินผลเป็นระบบการตัดสินใจเพื่อให้คุณค่าหรือคุณธรรมของสิ่งที่ได้รับการประเมินผล”

พอแพม (Popham, 1993 : 7) ได้เขียนเป็นความหมายในประโยคภาษาอังกฤษแปลเป็นภาษาไทยได้ว่า “การประเมินผลการศึกษาที่เป็นระบบ ข้อมประกอบด้วย การประเมินคุณค่าของเหตุการณ์ต่าง ๆ ทางด้านการศึกษา”

กังวล เทียนกัณฑ์เทศน์ (2536 : 16) การประเมินผล คือ กระบวนการประเมินค่าขั้นสุดท้ายว่าวัตถุประสงค์ของการศึกษาโดยบรรลุวัตถุประสงค์เพียงไร ทั้งนี้โดยการพิจารณาผลที่ได้จากผลของการวัดและส่วนประกอบอื่นเพื่อการประเมินค่าที่เหมาะสม

อำนาจ เลิศขยันดี (2542 : 4) การประเมินผล มีความหมายเป็นขบวนการของการค้นคว้าวิจัยเพื่อให้ได้มาซึ่งมาตรฐานที่นำมาใช้ประกอบการตัดสินใจในการที่จะลงสรุปว่า สิ่งที่ได้รับ การประเมินผลนี้มีคุณค่าสูงหรือต่ำเพียงใด ดี-เลว เหมาะ-ไม่เหมาะ ควร-ไม่ควร ผ่าน-ตก

เยาวดี วิบูลย์ศรี (2544 : 7) การประเมินผล เป็นคำที่ใช้ในการอธิบายและตัดสินคุณสมบัติบางอย่างของบุคคลหรือกลุ่มบุคคล รวมทั้งกระบวนการและโครงการต่าง ๆ

วาสนา ประवालพุกษ์ (2544 : 100) การประเมิน หมายถึง กระบวนการที่ชี้ให้เห็นว่าการจัดการศึกษาได้บรรลุจุดมุ่งหมายที่ระบุไว้ในหลักสูตรและการสอนหรือไม่ จุดมุ่งหมายทางการศึกษาที่สำคัญคือความต้องการเปลี่ยนแปลงพฤติกรรมของมนุษย์ ซึ่งหมายถึง การเปลี่ยนพฤติกรรมของนักเรียนให้เป็นไปในทางพึงปรารถนา ดังนั้น การประเมินจึงเป็นกระบวนการที่กำหนดว่ามีการเปลี่ยนแปลงพฤติกรรมเกิดขึ้นมากน้อยเพียงใด

บุญธรรม กิจปริดาบริสุทธิ (2543 : 298) การประเมินผล หมายถึง เป็นกระบวนการอย่างมีระบบและต่อเนื่องที่จะวัดความสำเร็จหรือความล้มเหลวของงาน โดยนำผลที่ประเมินได้มาปรับปรุงวิธีดำเนินงาน เพื่อให้แผนงานหรือโครงการบรรลุวัตถุประสงค์

วาสนา ประवालพุกษ์ (2544 : 16) การประเมินผล คือ การตีราคาพฤติกรรมของผู้เรียนในแง่คุณภาพ โดยการเก็บรวบรวมข้อมูลหลักฐานด้วยวิธีอื่น ๆ ที่ไม่ใช่การวัดด้วยเครื่องมือวัดเชิงปริมาณ และใช้การตัดสินใจเชิงคุณภาพ

สรุป

การประเมินผล หมายถึง กระบวนการค้นหาหรือประเมินคุณค่าที่เป็นระบบ โดยมีการเก็บรวบรวมข้อมูลและวิเคราะห์ข้อมูลแล้วนำมาเทียบกับเกณฑ์ ซึ่งเป็นการประเมินค่าขั้นสุดท้าย เพื่อตัดสินคุณค่าของสิ่งที่ประเมิน

1.2 นิยามการประเมิน (Definitions of Evaluation)

แนวคิดพื้นฐานของการประเมินมากกว่าครึ่งศตวรรษที่ผ่านมาสามารถอิงนิยามของการประเมินที่ต่างกัน (ศิริชัย กาญจนวาสี , 2545 : 13 – 15)

“การประเมิน” เป็นคำที่มีผู้กล่าวถึงกันมากที่สุดคำหนึ่ง แต่มีความเข้าใจไม่ค่อยตรงกัน พอเอ่ยถึงคำว่า “การประเมิน” บางท่านนึกถึงการกำหนดวัตถุประสงค์เชิงพฤติกรรม การติดตามผลการดำเนินงาน การสร้างเครื่องมือทดสอบ การวิเคราะห์ข้อมูล หรือการเขียนบรรยายสรุปผลการปฏิบัติงาน แท้จริงแล้วการประเมินมีความสลับซับซ้อนกว่านั้นมาก และมีความแตกต่างจากกระบวนการซึ่งเรารู้จักกันดีโดยทั่วไป เช่น “การวัดผล” หรือ “การวิจัย” เป็นต้น ความสับสนในความหมายของคำเกิดขึ้นได้เสมอ แม้แต่บรรดานักประเมินด้วยกันเองยังไม่สามารถมีความเห็นเป็นเอกฉันท์ว่าการประเมินคืออะไรกันแน่ จึงทำให้มีการให้คำนิยามของการประเมิน

อย่างหลากหลาย การทำความเข้าใจถึงความหมายของ “การประเมิน” นับได้ว่าเป็นสิ่งสำคัญต่อการทำความเข้าใจทฤษฎีและกิจกรรม การประเมิน

การประเมิน (Evaluation) กับการวัด (Measurement)

การประเมินในระยะแรกได้รับการถ่ายทอดความรู้ส่วนใหญ่มาจากการวัด การวัดผลถูกนำไปใช้ในโรงเรียนเป็นเวลานานหลายพันปีมาแล้ว เพื่อประเมินว่าผู้เรียนมีความรอบรู้ในเนื้อหาวิชามากน้อยเพียงไร ตอนแรกใช้วิธีการสังเกต สอบปากเปล่า ต่อมาเป็นการใช้แบบสอบ เพื่อให้ได้คะแนนเป็นตัวบ่งชี้ระดับผลสัมฤทธิ์และความสามารถทางสมอง การใช้แบบสอบสำหรับการวัดและประเมินผลในโรงเรียนจึงเป็นไปอย่างกว้างขวาง

ในช่วงเวลา ค.ศ. 1920–1930 นักประเมินในระยะนั้นใช้คำว่า “การวัดและการประเมิน” ในความหมายเดียวกัน เนื่องจากว่าการประเมินเป็นสิ่งที่แฝงเร้นและมีแหล่งกำเนิดมาจากศาสตร์แห่งการวัด จึงไม่น่าแปลกใจเลยเมื่อมีคนพูดถึงสองคำดังกล่าวแต่รวมความหมายถึงสิ่งเดียวกัน ซึ่งสามารถสังเกตได้จากชื่อหนังสือและงานเขียนของผู้เชี่ยวชาญทางด้านกรวัดผล เช่น ทอร์นไคและฮาเกน (Thorndike and Hagen, 1969); อีเบล (Ebel, 1965); สตัฟเฟิลบีม (Stufflebeam, 1971) ที่ใช้คำทั้งสองในความหมายเดียวกันหรือคล้ายคลึงกัน

เวลาต่อมาทั้งสองคำเริ่มแยกจากกันและมีการให้ความสำคัญต่อคำว่า “การประเมิน” มากขึ้น อย่างไรก็ตามคำว่า “การประเมิน” ยังอยู่ในลักษณะที่ใกล้ชิดมากกับคำว่า “การวัด” ตัวอย่างเช่น ทอร์นไคและฮาเกน ได้กล่าวว่า “การประเมินเป็นกระบวนการที่มีความสัมพันธ์อย่างใกล้ชิดกับการวัด เพียงแต่การประเมินมีความเฉพาะเจาะจงมากกว่าและต้องอาศัยการวินิจฉัยว่าเหมาะสมไหมดีหรือไม่ ถ้าใช้เทคนิคการวัดที่ด้อยลงจะเป็นพื้นฐานสำคัญสำหรับการประเมินที่ถูกต้อง”

การนิยามว่าการประเมิน คือ การวัดหรือในทางกลับกันมีข้อดีในแง่ที่ทำให้การประเมินฟึกคัวและพัฒนาอยู่กับการทางวิทยาศาสตร์ของการวัด ซึ่งเน้นความเป็นปรนัยและความเที่ยงตรง โดยยึดหลักการใช้เครื่องมือมาตรฐานเพื่อวัดคุณลักษณะที่แตกต่างระหว่างบุคคล การใช้สถิติวิเคราะห์เพื่อหาข้อสรุปรวมทั้งการสร้างมาตรฐานหรือบรรทัดฐานไว้สำหรับการเปรียบเทียบ อย่างไรก็ตามผลเสียที่ติดตามมาคือทำให้นักประเมินมีบทบาทเป็นนักเทคนิคการวัด มุ่งสร้างเครื่องมือวัดคุณลักษณะของบุคคลการประเมินจึงผูกติดอยู่กับการตีความหมายของคะแนนที่ได้จากเครื่องมือมาตรฐานของการวัดมากเกินไป ทำให้มองข้ามตัวแปรเชิงประเมินที่สำคัญบางตัว ซึ่งไม่สามารถวัดได้ในเชิงปริมาณ เช่น การจัดองค์กร การประสานงาน ปฏิสัมพันธ์ระหว่างองค์ประกอบต่าง ๆ ภายในและภายนอกองค์กร เป็นต้น รวมทั้งขาดพัฒนาการของการกำหนดเกณฑ์สำหรับการตัดสินใจและการวินิจฉัยคุณค่าของสิ่งที่ทำการศึกษา

การประเมิน (Evaluation) กับการวิจัยประยุกต์ (Applied Research)

เมื่อพิจารณาคุณลักษณะการประเมินแล้ว ย่อมมีลักษณะใกล้เคียงกับกิจกรรมการวิจัยเป็นอย่างมากทั้งในแง่ของความผูกพันอย่างลึกซึ้งซึ่งที่มีต่อระเบียบวิธีทางวิทยาศาสตร์ การออกแบบ (Design) เครื่องมือและการวิเคราะห์ด้วยเหตุนี้จึงมีการมองว่าการประเมินเป็นวิทยาศาสตร์ประยุกต์เท่ากับเป็นการมองว่าการประเมิน “การวิจัยประยุกต์” นั้นเองซึ่งสอดคล้องกับรอสซี (Rossi, 1972, 1982) ที่ระบุว่า การประเมินเป็นการใช้ระเบียบวิธีวิจัยทางสังคมสำหรับตัดสินใจและพัฒนาโครงการทางสังคมด้านการวางแผนการติดตามผล ประสิทธิภาพและประสิทธิผลของโครงการ ดังนั้น นักวิจัยเหล่านี้จึงเสนอแนะให้นำหลักการทางการวิจัยมาใช้เป็นหลักในงานการประเมินนั่นเอง

การมองว่าการประเมินเป็นการวิจัยประยุกต์ มีข้อได้เปรียบที่เราสามารถนำจุดเด่นของระเบียบวิธีวิจัย (ถ้าสามารถกระทำได้อย่างเหมาะสมกับสถานการณ์) มาใช้เป็นแนวทางของวิธีการประเมิน เช่น การกำหนดและทดสอบสมมุติฐาน การออกแบบการวิจัย การทดสอบทางสถิติ เป็นต้น ทำให้ผลการประเมินน่าเชื่อถือในแง่ของการปฏิบัติจุดเด่นดังกล่าวสามารถนำไปสู่ข้อจำกัดของกระบวนการตัดสินใจคุณค่าในปรัชญาของการประเมิน รวมทั้งความเป็นประโยชน์ของผลการประเมินต่อผู้ต้องการใช้สารสนเทศ

การประเมิน (Evaluation) กับการตรวจสอบความสอดคล้อง (Determining Congruence)

จากประสบการณ์ของไทเลอร์ (Tyler) ซึ่งเป็นผู้ริเริ่มการประเมิน โครงการมัธยมศึกษาของสหรัฐอเมริกาติดต่อกันเป็นเวลา 8 ปี (1932 – 1940) ไทเลอร์ได้กำหนดเกณฑ์การประเมินไว้ว่าถ้าโครงการจัดการศึกษาประสบความสำเร็จแล้วนักเรียนจะต้องเปลี่ยนพฤติกรรมไปในทิศทางที่พึงประสงค์สอดคล้องกับวัตถุประสงค์ของการจัดการศึกษานั้น

จากแนวคิดของไทเลอร์ดังกล่าว ได้สร้างมิติใหม่แก่วงการประเมินอันนำไปสู่ความหมายของการประเมินที่ว่า การประเมิน หมายถึง กระบวนการตรวจสอบความสอดคล้องระหว่าง “ผลที่ได้กับวัตถุประสงค์ที่กำหนดไว้” ซึ่งเป็นคำจำกัดความได้รับการยอมรับอย่างแพร่หลายและมีอิทธิพลอย่างยิ่งต่อการประเมินผลสืบเนื่องจนถึงปัจจุบัน

การประเมิน (Evaluation) กับการช่วยการตัดสินใจ (Assist Decision Making)

ระหว่างทศวรรษของปี 1960 การประเมินได้ขยายขอบเขตกว้างขวางขึ้นจากเดิมซึ่งจำกัดขอบเขตเพียงการตรวจสอบผลสำเร็จตามวัตถุประสงค์เป็นการแสวงหาสารสนเทศที่เป็นประโยชน์ต่อการตัดสินใจดำเนินงานต่าง ๆ ของผู้บริหาร ดังนั้นจึงมีการพัฒนาความหมายใหม่ของการประเมินเป็น “การประเมิน หมายถึง กระบวนการระบุและเสนอสารสนเทศเพื่อช่วยการ

ตัดสินใจ” ผู้ที่ให้นิยามการประเมินตามแนวทางดังกล่าวมีหลายท่านที่สำคัญ ได้แก่

ครอนบาค (Cronbach, 1963) การประเมินเป็นกระบวนการที่เป็นระบบในการเก็บรวบรวมและใช้สารสนเทศสำหรับการตัดสินใจ จุดมุ่งหมายหลักของการประเมินโครงการอยู่ที่ความต้องการทราบผลของโครงการ การเปลี่ยนแปลงที่เกิดขึ้นจากการดำเนินงานและส่วนใดของโครงการที่ควรปรับปรุงแก้ไข

อัลคิน (Alkin, 1969) การประเมินเป็นกระบวนการของการทำให้เกิดความมั่นใจในการตัดสินใจ ด้วยการคัดเลือกข้อมูลที่เหมาะสม รวบรวมและวิเคราะห์เพื่อจัดทำรายงานสรุปสารสนเทศที่เป็นประโยชน์ต่อผู้บริหาร ในการตัดสินใจเลือกทางเลือกที่เหมาะสม

สตัฟเฟิลบีม (Stufflebeam, 1971) การประเมินเป็นการระบุปัญหาการจัดการและการเสนอสารสนเทศที่เป็นประโยชน์ต่อการตัดสินใจเลือกทางเลือกวิธีการดำเนินงานที่เหมาะสม

จะเห็นได้ว่าความหมายของการประเมินดังกล่าวมีข้อดีที่สำคัญประการหนึ่ง คือ ทำให้มองเห็นหน้าที่หนึ่งของการประเมิน ได้แก่การเสนอสารสนเทศที่เป็นประโยชน์ต่อการตัดสินใจบริหารงาน ซึ่งเป็นประโยชน์ที่สำคัญยิ่งของการประเมินในการทำให้การตัดสินใจเกิดขึ้นอย่างมีเหตุมีผล การประเมินจึงขยายขอบเขตนอกเหนือไปจากการประเมินมนุษย์ เช่น ประเมินโครงการ ทรัพยากรมาตรการดำเนินงาน เป็นต้น อย่างไรก็ตามคำจำกัดความดังกล่าว ยังขาดความชัดเจนของเป้าหมายและการประเมินว่าจุดหมายที่แท้จริงของการประเมินอยู่ที่ไหน เพียงเพื่อให้ได้สารสนเทศที่เป็นประโยชน์สำหรับการตัดสินใจเท่านั้นหรือ นักประเมินหลายท่านไม่เห็นด้วยที่ว่าผู้บริหารเท่านั้นที่สำคัญ ยังมีผู้เกี่ยวข้องอื่นๆ กับการประเมินก็มีความสำคัญเช่นเดียวกัน

การประเมิน (Evaluation) กับการบรรยายอย่างลุ่มลึก (Description or Portrayal)

การประเมิน คือ การบรรยายสิ่งที่มุ่งประเมินลึกซึ่งรอบด้านนิยามนี้ได้รับอิทธิพลอย่างสูงจากแนวคิดของสเติก (Stake, 1967, 1975, 1978) สเติกและอิสลีย์ (Stake and Easley, 1978) ซึ่งเห็นว่า ผู้ประเมินจะต้องบรรยายความสัมพันธ์และความสอดคล้องของสิ่งที่คาดหวังสิ่งที่เกิดขึ้นจริงและมาตรฐาน ให้ครอบคลุมสิ่งที่มุ่งประเมินทั้งในแง่เงื่อนไขก่อนเริ่มโครงการ ปฏิสัมพันธ์ระหว่างโครงการกับผู้ร่วมโครงการ ตลอดจนผลลัพธ์ของโครงการ เพื่อสะท้อนที่ชนะและการตัดสินใจของผู้เกี่ยวข้องกับการประเมินหรือผู้มีส่วนได้ส่วนเสียจากการประเมินอย่างกว้างขวาง

การที่จะบรรยายสิ่งที่มุ่งประเมินได้อย่างลุ่มลึกนั้นสเติกเสนอแนะให้ใช้การศึกษาเฉพาะกรณี (Case Study) โดยใช้เทคนิคทางมนุษยวิทยาเพื่อตอบสนองความต้องการสารสนเทศให้ตรงกับความต้องการของผู้ใช้ผลการประเมินหลาย ๆ ฝ่าย

โดยคำจำกัดความนี้ Stake) จึงเน้นกิจกรรมการประเมินที่ “การบรรยายอย่าง ลุ่มลึก” เพื่อให้ได้สารสนเทศที่สนองความต้องการและความสนใจของผู้ใช้ ส่วนการตัดสินคุณค่า ให้ขึ้นอยู่กับดุลยพินิจของผู้ใช้สารสนเทศเป็นสำคัญ

จุดเด่นของคำจำกัดความนี้อยู่ที่การเน้นมิติสำคัญมิติหนึ่งของการประเมิน คือ การ บรรยายอย่างลึกซึ้ง ซึ่งเป็นกิจกรรมพื้นฐานที่สำคัญมากของกระบวนการประเมิน นักประเมินจึง มีบทบาทเป็นนักเก็บรวบรวมข้อมูลอย่างลุ่มลึกและเป็นนักบรรยาย (Descriptor) สารสนเทศแก่ผู้ ต้องการใช้ฝ่ายต่าง ๆ อย่างรอบด้าน แต่การที่ให้สิทธิ์แก่ผู้ใช้สารสนเทศเป็นผู้ตัดสินคุณค่าเป็น ประเด็นที่สามารถนำมาอภิปรายได้อย่างกว้างขวาง เช่น ในการประเมินใครสมควรเป็นผู้ตัดสิน คุณค่าผู้ใช้สารสนเทศสามารถตัดสินคุณค่าได้เหมาะสมกว่าผู้ประเมินหรือไม่ เป็นต้น

การประเมิน (Evaluation) กับการตัดสินคุณค่า (Determining of Worth or Value)

การประเมิน คือ การตัดสินคุณค่าของสิ่งที่มีประเมิน นิยามนี้เป็นที่ยอมรับกัน อย่างกว้างขวางในปัจจุบัน เป็นผลมาจากแนวคิดและการบุกเบิกของสคริฟเวน (Scriven, 1967, 1974, 1975) สคริฟเวนชี้ว่าที่เป้าหมายของการประเมินอยู่ที่การตัดสินคุณค่า โดยผู้ประเมินจะต้อง มีความเชี่ยวชาญในหลักการสังเกตและเหตุผล ถ้าผู้ประเมินมิได้ตัดสินคุณค่าของสิ่งที่ประเมินถือว่า ผู้ประเมินนั้นยังทำหน้าที่ไม่สมบูรณ์ แนวคิดดังกล่าวเป็นที่ยอมรับโดยนักทฤษฎีการประเมินใน เวลาต่อมา เช่น กลาส (Glass, 1969); วอร์เรนและเซนเดอร์ (Worthen and Sanders, 1973); คูลี และลอเนส (Cooley and Lohnes, 1976); เฮาส์ (House, 1980); กูบาและลินคอล์น (Guba and Lincoln, 1981) เป็นต้น

การประเมินตามคำดังกล่าวได้ถูกนำไปใช้ปฏิบัติโดยมีรูปแบบแตกต่างกัน เช่น การตัดสินคุณค่าโดยผู้เชี่ยวชาญ (Professional Judgment) ซึ่งเสนอโดยสคริฟเวน (Scriven) อีสเนอร์ (Eisner) การตัดสินคุณค่าตามมาตรฐานวิชาชีพ (Professional Review) เช่น การรับรองวิทยฐานะการพิจารณาผลงานวิชาการ การสัมภาษณ์คัดเลือกบุคลากร เป็นต้น

ข้อดีของนิยามดังกล่าว คือ การบรรลุเป้าหมายของการประเมินในแง่ของการ ตัดสินคุณค่า ภาพของการประเมินจึงมีความสมบูรณ์ขึ้นในแง่บทบาทของการวัด การบรรยาย และการตัดสินคุณค่า ถ้าผู้ประเมินมีความรอบรู้ทั้งในด้านวิธีการประเมินและเนื้อหาหรือผู้ประเมิน มีประสบการณ์ความเชี่ยวชาญและเป็นประโยชน์ แต่ถ้าผู้ประเมินขาดคุณสมบัติ อาจก่อให้เกิด ปัญหาเกี่ยวกับความเที่ยงตรงของข้อมูลที่ได้และความเหมาะสมของเกณฑ์ที่ใช้ในการตัดสิน

สรุป

ความหมายของการประเมินได้รับการพัฒนามาอย่างต่อเนื่อง เริ่มต้นจากความเข้าใจที่ว่า การประเมินเป็นสิ่งเดียวกับการวัดผลการเรียนรู้ของผู้เรียน (Measurement – Oriented) การประเมินเป็นกระบวนการศึกษาสิ่งต่าง ๆ โดยใช้ระเบียบวิธีวิจัย (Research – Oriented) การประเมินเป็นการตรวจสอบการบรรลุผลตามวัตถุประสงค์ที่กำหนดไว้ (Objectives – Oriented) การประเมินเป็นการช่วยเสนอสารสนเทศเพื่อการตัดสินใจ (Decision – Oriented) การประเมินเป็นการสนองสารสนเทศแก่ผู้เกี่ยวข้องทั้งหลายด้วยการบรรยายอย่างลุ่มลึก (Description – Oriented) และการประเมินเป็นการตัดสินใจคุณค่าของสิ่งที่มุ่งประเมิน (Judgment – Oriented) การสืบทอดและพัฒนากันมาอย่างต่อเนื่องทำให้ความหมายของการประเมินมีความชัดเจนและสมบูรณ์ยิ่งขึ้น การประเมินในอนาคตน่าจะขยับสถานภาพให้มีความหมายและบทบาทต่อสังคมมากยิ่งขึ้น โดยเฉพาะอย่างยิ่งในด้านการเป็นสารสนเทศเชิงคุณค่าที่สนับสนุนการเจรจาต่อรองเชิงสร้างสรรค์เพื่อให้บรรลุถึงเอกภาพของความคิด แนวทาง และข้อสรุปทางการเมืองและเศรษฐกิจที่ก่อให้เกิดประโยชน์สุขร่วมกันของสังคม

1.3 ปรัชญาการประเมิน

การประเมินเป็นกระบวนการให้ได้มาซึ่งความจริงในเชิงคุณค่าของสิ่งใดสิ่งหนึ่ง การเข้าถึงความจริงในเชิงคุณค่าเป็นสิ่งที่ซับซ้อนละเอียดอ่อนมาก ดังนั้นจึงจำเป็นต้องทำความเข้าใจกระบวนการที่มาของความจริงในเชิงคุณค่า นั่นคือ กระบวนการประเมินนั่นเอง การที่จะทำความเข้าใจอย่างดั่งแต่ถึงศาสตร์การประเมิน เราก็จำเป็นต้องตอบคำถามว่าทำไมต้องประเมิน จุดหมายปลายทางหรือเป้าหมายสุดท้ายของการประเมินคืออะไรและทำอย่างไรการประเมินจึงจะทำให้เข้าถึงซึ่งความจริงในคุณค่าของสิ่งใดสิ่งหนึ่งภายใต้บริบทสังคมที่กำหนดให้ เพื่อตอบคำถามสำคัญดังกล่าว เราจำเป็นต้องมาทำความเข้าใจปรัชญาที่อยู่เบื้องหลังของการประเมินนั่นเอง (สมหวัง พิธิยานุวัฒน์, 2541 : 73 – 77)

1. การประเมินตามแนวปรัชญาประโยชน์นิยม (Utilitarian Evaluation) กับ การประเมินตามแนวปรัชญาสหญาณและพหุนิยม (Intuitionist – Pluralist Evaluation)

1) การประเมินตามแนวปรัชญาประโยชน์นิยม การประเมินแนวนี้เป็นการกำหนดคุณค่าโดยประเมินจากผลกระทบ โดยรวมของโปรแกรมการศึกษาที่มีต่อผู้ที่เกี่ยวข้อง วิธีการของการประเมินตามปรัชญานี้ มีแนวโน้มสอดคล้องกับญาณปรัชญาลัทธิประโยชน์นิยม (Objectivist Epistemology) ที่เชื่อว่ามนุษย์สามารถรู้ว่าอะไรเป็นเท็จ อะไรเป็นจริง ในการตัดสินใจว่าอะไรจริง อะไรเท็จ ควรใช้มาตรฐานสากลที่แน่นอนเหมือนกันทุกคน การประเมินตามแนวประโยชน์นิยมเชื่อว่าสิ่งที่ดีที่สุดคือ สิ่งที่เกิดประโยชน์แก่บุคคลจำนวนมากที่สุด ดังนั้น

เป้าหมายที่สำคัญของการประเมินตามแนวปรัชญานี้ก็คือ การประเมินที่ก่อให้เกิดความสุขแก่สังคมให้มากที่สุด บทบาทของนักประเมิน คือการนำสารสนเทศเชิงประเมินให้ผู้รับผิชอบเป็นผู้ตัดสินใจในการประเมิน นักประเมินจะเน้นผลที่เกิดขึ้นแก่กลุ่ม ซึ่งอาจแทนด้วยค่าเฉลี่ยของคะแนนที่เพิ่มขึ้นหรือดัชนีที่แสดงถึงคุณค่าที่ดีที่สุดอันก่อให้เกิดประโยชน์กับผู้คนที่เกี่ยวข้องหรือมวลชนให้มากที่สุด โปรแกรมทางการศึกษาที่ดีที่สุดคือ โปรแกรมทางการศึกษาที่ทำให้ค่าเฉลี่ยตัวชี้วัดที่บ่งบอกคุณค่าของโปรแกรมนั้นมีค่ามากที่สุด โดยทั่วไปการประเมินโปรแกรมทางการศึกษาระดับชาติมักจะเป็นการประเมินตามแนวปรัชญาประโยชน์นิยมนักประเมินที่มีแนวทัศนะตามปรัชญาประโยชน์นิยม เช่น ไทเลอร์(Tyler, 1942); โพรวัส(Provus, 1971); สตัฟเฟิลบีมและคณะ(Stufflebeam and Others, 1971) เป็นต้น

2) การประเมินตามแนวปรัชญาสหัชญาณและพหุนิยมตามแนวปรัชญา นี้ เชื่อว่า “คุณค่า” ของโปรแกรมขึ้นอยู่กับผลกระทบของโปรแกรมที่มีต่อปัจเจกชน การประเมินตามปรัชญาสหัชญาณและพหุนิยมมีแนวโน้มสอดคล้องกับปรัชญาลัทธิอัตนัยนิยม (Subjectivist Epistemology) ที่เชื่อว่าจิตของทุกคนไม่สามารถรู้ว่าอะไรจริงอะไรเท็จได้เหมือนกัน การตัดสินใจอะไรจริงอะไรเท็จเป็นเรื่องของแต่ละบุคคล ดังนั้น สิ่งที่ดีที่สุดมีคุณค่าสูงสุดจึงเป็นสิ่งที่ประนีประนอมแต่ละบุคคล ในการประเมินจะไม่มีเกณฑ์กลางร่วมกันทั้งกลุ่ม แต่จะพิจารณาเป็นรายบุคคลและรายกลุ่มย่อยว่าได้รับประโยชน์อะไรเพียงใดการประเมินตามแนวนี้จะใช้ความแตกต่างและความหลากหลายเป็นหลักเป้าหมายของการประเมินคือการตัดสินใจคุณค่า นักประเมินตามแนวปรัชญาจะเป็นผู้พิจารณาความต้องการและคุณค่าที่หลากหลาย การตัดสินใจคุณค่าของโปรแกรมจะขึ้นอยู่กับภาระของนักประเมิน นักประเมินจึงต้องเป็นผู้รอบรู้และมีความเป็นกลาง เพื่อจะทำหน้าที่ตัดสินใจคุณค่าของโปรแกรมได้อย่างถูกต้อง นักประเมินตามแนวปรัชญาสหัชญาณและพหุนิยม เช่น สคริฟเวน (Scriven, 1967); สเต็ก (Stake, 1967); วอร์เทนและเซนเดอร์ (Worthen and Sander, 1973); อีสเนอร์ (Eisner, 1975); กูบาและลินคอล์น(Guba and Lincoln, 1961, 1989) เป็นต้น

2. การประเมินตามแนวญาณปรัชญาลัทธิปรนัยนิยม และลัทธิอัตนัยนิยม

(Objectivist and Subjectivist Epistemology) ญาณปรัชญา (Epistemology) เป็นส่วนหนึ่งของปรัชญาที่สนใจแสวงหาความรู้ว่า เราเข้าถึงความจริงได้อย่างไร โดยเฉพาะเราจะใช้มาตรการหรือเกณฑ์อะไรมาตัดสินอะไรจริง อะไรเท็จ ซึ่งญาณปรัชญาแบ่งออกเป็น 2 ลัทธิ คือ ลัทธิปรนัยนิยมกับลัทธิอัตนัยนิยม

1) การประเมินตามแนวปรัชญาปรนัยนิยม นักปรัชญาแนวนี้เชื่อว่า มนุษย์ทุกคนสามารถรู้ว่าอะไรจริงอะไรเท็จ จึงควรมีมาตรการสากลสำหรับทุกคนในการตัดสินใจว่าเป็น

จริงหรือเป็นเท็จ ถ้าบุคคลใดมีความรู้ตามมาตรฐานแล้วก็ต้องเห็นความจริงเหมือนกันหมดในการประเมินตามแนวปรัชญาปรนัยนิยมนั้น ข้อมูลที่ใช้ประเมินจะต้องมีความเป็นปรนัย กล่าวคือการรวบรวมข้อมูลและการวิเคราะห์ข้อมูล จะต้องให้ผลที่ทำซ้ำได้ตรวจสอบได้ โดยผู้อื่นที่มีความรู้ความสามารถ ในกรณีเช่นนี้ กระบวนการประเมินจะต้องมีการวางแผนเป็นขั้นตอนอย่างมีระบบ นักประเมินคนอื่นที่สามารถมาดำเนินการตามกระบวนการดังกล่าว ก็จะได้ผลประเมินทำนองเดียวกัน นักประเมินกลุ่มนี้นิยมประเมินโดยใช้วิธีการเชิงระบบ (Systematic Approach) นักประเมินที่มีความรอบรู้ เป็นสัตตบุรุษในเชิงต่อทฤษฎีวิธีประเมิน มีประสบการณ์และมีความเป็นกลางยอมเข้าถึงความจริงและคุณค่าของความจริงได้เหมือนกันทุกคน ลัทธิปรนัยนิยมเป็นสืบนี้อาจมาจากความเชื่อของนักสังคมศาสตร์กลุ่มประจักษ์นิยมนั่นเอง นักประเมินที่เชื่อในปรัชญาประเมินแนวนี้ได้แก่ ไทเลอร์ (Tyler, 1942); ครอนบาค (Cronbach, 1963); โลวิน (Lovin, 1983) เป็นต้น

2) การประเมินตามแนวปรัชญาอัตนัยนิยม นักปรัชญากลุ่มนี้เชื่อว่าจิตของมนุษย์ไม่สามารถรู้ได้ว่าอะไรจริง อะไรเท็จ ได้เหมือนกัน การตัดสินอะไรเป็นจริงเป็นเท็จ เป็นเรื่องเฉพาะบุคคลซึ่งมีลักษณะที่โดยเฉพาะ ดังนั้นจึงไม่มีเกณฑ์มาตรฐานร่วมที่แน่นอนสำหรับการประเมิน แม้ในทางปฏิบัติจะมีการกำหนดหลักเกณฑ์การประเมินก็ตามมักจะเป็นหลักเกณฑ์ที่กว้าง ๆ และมีความยืดหยุ่น เพื่อให้เป็นที่ยอมรับของทุกฝ่ายความตรงของข้ออ้างเชิงอัตนัยจะมาจากประสบการณ์ของผู้ประเมินมากกว่าจะได้จากกระบวนการทางวิทยาศาสตร์ กล่าวคือ ความตรงของผลประเมินตามแนวอัตนัยนิยมจะขึ้นอยู่กับภูมิหลัง คุณสมบัตินิสัย และความชำนาญเชิงการรับรู้ของผู้ประเมินเป็นสำคัญ วิธีการประเมินแนวนี้จะเป็นการประเมินเชิงคุณภาพหรือการประเมินด้วยวิธีธรรมชาติ ตัวอย่างนักประเมินตามแนวปรัชญาอัตนัยนิยม เช่น สเต็ก (Stake, 1976); อีสเนอร์ (Eisner, 1975); กูบาและลินคอล์น (Guba and Lincoln, 1981, 1989) เป็นต้น

3. การประเมินตามปรัชญาปฏิฐานนิยม (Positivism) กับปรัชญาต่อต้านปฏิฐานนิยม (Anti - Positivism)

1) การประเมินตามปรัชญาปฏิฐานนิยม นักปรัชญาปฏิฐานนิยมเชื่อว่ามนุษย์สามารถรับรู้โลกภายนอกได้โดยผ่านประสาทสัมผัสทั้งห้า ซึ่งทำให้เกิดประสบการณ์ ดังนั้นประสบการณ์จึงเป็นแหล่งความรู้ที่มนุษย์ได้รับจากโลกภายนอก วิธีแสวงหาความรู้คือการสังเกตและศึกษาปรากฏการณ์ต่าง ๆ โดยวิธีทางวิทยาศาสตร์ ซึ่งเป็นวิธีผสมผสานระหว่างวิธีอุปนัย (Inductive Method) กับวิธีนิรนัย (Dedutive Method) ซึ่งมีประสบการณ์และเหตุผลเป็นเครื่องมือในการแสวงหาความรู้ นักปฏิฐานนิยมเชื่อว่าความรู้ที่แน่นอนคือความรู้ที่ความสามารถอธิบายได้ พิสูจน์ได้โดยทฤษฎีทางวิทยาศาสตร์เท่านั้น การประเมินตามแนวปรัชญาปฏิฐานนิยมก็คือ การ

ประเมินโดยวิธีทดสอบหรือวิธีเชิงปริมาณ ตัวอย่างนักประเมินตามแนวปรัชญานี้ คือ ไทเลอร์ (Tyler, 1942) ; เวลช์ (Welch, 1968) เป็นต้น

2) การประเมินตามปรัชญาต่อต้านปฏิฐานนิยมหรือปรัชญาสร้างสรรค์นิยม (Constructivism) นักปรัชญาตามแนวนี้เชื่อว่า ความรู้มี 2 ประเภท คือ ความรู้ในเชิงวัตถุเกี่ยวกับโลกภายนอก สามารถศึกษาได้โดยวิธีทางวิทยาศาสตร์ ส่วนความรู้อีกประเภทหนึ่ง เป็นความรู้เกี่ยวกับมนุษย์และสังคม ซึ่งมนุษย์มีความรู้สึกนึกคิด มีจินตนาการ ความคาดหวัง ความมุ่งมั่น ศรัทธาและอื่น ๆ ที่สามารถให้ความหมายและตีความหมายของประสบการณ์ของตนเองได้ ดังนั้นแต่ละคนจึงเป็นแหล่งความรู้ โดยเฉพาะวิธีศึกษาในกรณีเช่นนี้ควรเป็นวิธีการหาความรู้จากภายใน โดยล้วงลึกให้เข้าไปถึงความรู้สึกนึกคิดถึงความจริงในเชิงคุณค่าของแต่ละบุคคล ผู้ประเมินที่จะเข้าถึงความจริงเชิงคุณค่าได้จะต้องเข้าไปคลุกคลีจนคุ้นเคย เพื่อให้เข้าถึงซึ่งความหมายในเชิงคุณค่าจะต้องเน้นความเป็นมนุษย์ของผู้ประเมินและผู้ถูกประเมิน เน้นความหมายและปรากฏการณ์โดยมุมมองจากภายในและการเน้นบริบททางสังคมและวัฒนธรรมของปรากฏการณ์ การประเมินตามแนวปรัชญานี้เรียกว่าเป็นการประเมินเชิงธรรมชาติและมีส่วนร่วม หรือการประเมินเชิงคุณภาพนั่นเอง ตัวอย่างนักประเมินแนวนี้ ได้แก่ เวลคอต (Walcott, 1976); กูบาและลินคอล์น (Guba and Lincoln, 1981, 1989); แพททอน (Patton, 1980) เป็นต้น

1.4 ประวัติและพัฒนาการของการประเมิน

การประเมินที่เราารู้จักกันทุกวันนี้เป็นผลมาจากพัฒนาการทางความคิดและประสบการณ์ของนักทฤษฎีการประเมินที่ช่วยกันสร้าง สานต่อและสะสมมาเป็นเวลาอันยาวนานจนเป็นสิ่งที่รู้จักและใช้กันอย่างกว้างขวางทุกวงการในปัจจุบัน

การประเมินเป็นกิจกรรมที่ปฏิบัติสืบต่อกันมาเป็นเวลานานตั้งแต่ครั้งโบราณกาล จากหลักฐานที่ปรากฏสามารถสืบย้อนไปได้ถึงประมาณสี่พันปีมาแล้วเมื่อ 2357 ปีก่อนคริสต์ศักราช จักรพรรดิของจีนได้ใช้การตรวจสอบรรถภาพข้าราชการทหารทุก 3 ปี เป้าหลักในการพิจารณาการเลื่อนตำแหน่งระหว่างปี 1122 B.C. – 225 B.C. (ก่อนคริสต์ศักราช) จีนได้ริเริ่มการใช้ข้อสอบคัดเลือกบุคคลเข้ารับการศึกษา นอกจากนี้เป็นที่ทราบกันมานานเช่นกันว่าปรมาจารย์ชาวกรีก เช่น ไสกราติส ได้ใช้การประเมินด้วยวาจาเป็นสื่อในกระบวนการเรียนรู้ เป็นต้น

การประเมินในลักษณะที่ปฏิบัติกันทุกวันนี้เพิ่งมีอายุได้ไม่ถึงศตวรรษ ช่วงเวลาระหว่างการประเมินครั้งโบราณกาลจนถึงยุคปัจจุบัน การประเมินได้ผ่านขั้นตอนของพัฒนาการที่มีระบบแบบแผนมากขึ้นตลอดเวลา จึงทำให้การประเมินมีความมั่นคงเป็นปึกแผ่นในลักษณะวิชาชีพแขนงใหม่ซึ่งมีความแตกต่างที่เด่นชัดยิ่งขึ้นจากศาสตร์แห่งการวัดผลและการทดสอบและมีคุณลักษณะที่แตกต่างจากการวิจัย (ศิริชัย กาญจนวาสี, 2545 : 23-24)

ลักษณะที่สำคัญอย่างหนึ่งของศาสตร์ที่ได้พัฒนาจนเป็นวิชาชีพที่มีความเป็นปึกแผ่นมั่นคง ข้อมมีการบันทึกเหตุการณ์ที่ผ่านมาอย่างเป็นระบบและมีการวิเคราะห์ประวัติศาสตร์ความเป็นมาของศาสตร์นั้น สำหรับประวัติศาสตร์แห่งการประเมินยังไม่มีผู้ใดบันทึกไว้อย่างเป็นทางการ แต่สามารถอนุมูลงจากการรวบรวมค้นคว้าและวิเคราะห์เอกสารจากแหล่งต่าง ๆ ที่เกี่ยวข้อง จะเห็นได้ว่า ประวัติศาสตร์แห่งการประเมินมีความเกี่ยวข้องอย่างใกล้ชิดกับประวัติศาสตร์แห่งการวัดผล จนยากที่จะจำแนกเป็นอิสระจากกันและกันได้ ความสัมพันธ์ดังกล่าวนี้เป็นไปในลักษณะที่การวัดเป็นกิจกรรมหนึ่งในกระบวนการประเมินและข้อมูลจากการวัดผลเป็นแหล่งสำคัญของข้อมูลสำหรับการตัดสินใจคุณค่าในการประเมิน ในการแบ่งช่วงระยะเวลาของวิวัฒนาการเกี่ยวกับความเป็นมาของการประเมิน ทั้งต่างประเทศและในประเทศมีดังนี้(สมหวัง พิธิยานุวัฒน์, 2541 : 23-34)

1. ประวัติและพัฒนาการของการประเมินในต่างประเทศ

ดังที่ทราบแล้วว่าการวัดมีความสัมพันธ์อย่างใกล้ชิดกับการประเมินและการวิจัยทางการศึกษา โดยเฉพาะอย่างยิ่งในยุคเริ่มต้นในการแบ่งประวัติและพัฒนาการของการประเมิน จึงคาบเกี่ยวกับพัฒนาการของการวัดผลการศึกษาและการวิจัยทางการศึกษาในที่นี้จะใช้การแบ่งยุคต่างๆ ตามที่เสนอโดยแมดัส(Madaus, 1983) ซึ่งเน้นประวัติและพัฒนาการของการประเมิน โครงการและการเสนอโดยกูบาและลินคอล์น (Guba and Lincoln, 1989) ดังนี้

ยุคที่ 1 ยุคก่อนการปฏิวัติอุตสาหกรรม เริ่มตั้งแต่มีการบันทึกจนถึง ค.ศ. 1800 ซึ่งมีเหตุการณ์สำคัญคือ 2000 ปีก่อนคริสต์ศักราช เจ้าหน้าที่ของจีนได้สอบคัดเลือกคนเข้ารับราชการโดยใช้ข้อสอบเรียงความ 5 ปีก่อนคริสต์ศักราช โซเครทส์ (Socrates) และครูชาวกรีกได้ใช้ประเมินโดยตั้งคำถามซึ่งรู้จักกันต่อมาคือการสอบปากเปล่านั่นเองและในช่วงปี ค.ศ. 1600 – 1800 เป็นช่วงที่วิทยาศาสตร์เข้ามามีบทบาทในความคิดและการปฏิบัติของมนุษย์มากขึ้น อันเป็นพื้นฐานของการปฏิวัติอุตสาหกรรมในทวีปยุโรปต่อมา

ยุคที่ 2 ยุคการปฏิรูป (The Age of Reform : 1800 – 1900) เป็นยุคแห่งการปฏิวัติอุตสาหกรรมในทวีปยุโรปมีการปรับปรุงเปลี่ยนแปลงครั้งใหญ่ทั้งด้านเศรษฐกิจ เทคโนโลยี สังคม การเมืองและการศึกษา มีการปฏิรูปโครงสร้างทางสังคมและโปรแกรมบริการทางการศึกษา และสุขภาพอนามัย โดยเฉพาะอย่างยิ่งในสหราชอาณาจักรอังกฤษและสหรัฐอเมริกา ในสหราชอาณาจักรอังกฤษมีความพยายามที่จะปฏิรูปสถาบันทางสังคมต่าง ๆ รวมทั้งโรงพยาบาล สถานเลี้ยงเด็กกำพร้าและโรงเรียน โดยทางราชสำนักและรัฐบาลได้แต่งตั้งคณะกรรมการขึ้นทำหน้าที่ประเมินประสิทธิภาพและการปฏิบัติของสถาบันทางสังคมเหล่านี้

ในการวิจัยทางการศึกษา โดยเฉพาะในด้านการทดลองวันด์ (Wundt) และคณะ ซึ่งเป็นนักจิตวิทยาชาวเยอรมันที่ได้ทำการสังเกตและบันทึกพฤติกรรมของมนุษย์อย่างเป็นระบบ

นักวิชาการที่ร่วมงานกับวันด์ ได้แก่ แคทเทล (Cattell), ฮอลล์ (Hall), จัดด์ (Judd) และมิวแมนน์ (Meumann) ซึ่งล้วนเป็นนักวิชาการที่มีชื่อเสียงในศตวรรษที่ 20

ค.ศ. 1845 เริ่มมีการประเมินผลงานเพื่อประเมินประสิทธิภาพของโรงเรียนต่าง ๆ ที่เมืองบอสตัน ประเทศสหรัฐอเมริกา ที่เรียกว่า Boston Survey โดย โฮราซ แมนน์ (Horace Mann) และซามูเอล ฮาว (Samuel Howe) ได้ริเริ่มการใช้คะแนนสอบข้อเขียนของนักเรียนในการประเมินประสิทธิภาพของการสอนและประสิทธิภาพของโรงเรียนมีความพยายามที่จะใช้ข้อมูลที่เป็นระบบ คือ ใช้คะแนนสอบข้อสอบข้อเขียนของนักเรียนแทนการสอบปากเปล่า เพื่อใช้ในการพิจารณาแต่งตั้งครูใหญ่แมนน์และฮาวจึงนับว่าเป็น ผู้บุกเบิกแสวงหาวิธีในการประมวลผลและวิเคราะห์ข้อมูลด้านผลสัมฤทธิ์ทางการเรียนของนักเรียน เพื่อใช้ในการพิจารณาแต่งตั้งครูใหญ่ อีกทั้งได้พยายามสร้างแบบสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนอย่างเป็นทางการเป็นมาตรฐาน

ค.ศ. 1884 นักชีววิทยาชาวอังกฤษ คือแกลตัน (Galton) ได้สร้างเครื่องมือวัดความสามารถในการได้ยินเสียง จำแนกเสียง คือแกลตันเป็นบุคคลแรกที่ริเริ่มการวัดโดยใช้แบบสอบถามและมาตรประมาณค่าและเป็นบุคคลแรกที่คิดคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์

ค.ศ. 1887 – 1898 โจเซฟ เอ็ม ไรซ์ (Joseph M. Rice) เป็นนักปฏิรูปการศึกษาชาวอเมริกัน ซึ่งถือว่าเป็นบิดาของการวิจัยทางการศึกษาและเป็นบุคคลแรกที่เป็นผู้ประเมินโครงการทางการศึกษาอย่างเป็นทางการ ไรซ์ได้สร้างแบบสอบการสะกดคำแล้วนำไปทดสอบเด็ก 33,000 คน ในโรงเรียนต่าง ๆ ทั่วประเทศสหรัฐอเมริกาตอนใต้และทางตะวันออก ผลการศึกษาพบว่า โรงเรียนที่ใช้เวลาสอนฝึกสะกดคำแตกต่างกัน แต่นักเรียนกลับมีผลสัมฤทธิ์ทางการเรียนด้านการสะกดคำไม่แตกต่างกัน ซึ่งแสดงว่าการสอนฝึกสะกดคำของโรงเรียนไม่มีผลการเรียนรู้ของนักเรียนหรือกล่าวอีกนัยหนึ่งก็คือ โรงเรียนใช้เวลาสอนอย่างไม่มีประสิทธิภาพ วิธีสอนการสะกดคำไม่มีคุณภาพ ผลการประเมินของไรซ์ในเรื่องการสะกดคำนี้ นำไปสู่การปฏิรูปวิธีการสอน ด้วยผลการศึกษาเรื่องความสามารถในการสะกดคำของเด็กอเมริกันจำนวนมากดังกล่าว ไรซ์ ซึ่งได้ชื่อว่าเป็นบุคคลแรกที่พัฒนาการประเมินผลทางการศึกษาที่มีรูปแบบและเป็นระบบในประเทศไทย โดยเฉพาะรูปแบบการทดลองเปรียบเทียบผลสัมฤทธิ์ทางการเรียน ในการวิจัยทางการศึกษาและการประเมินทางการศึกษาและรูปแบบการวัดแบบปรนัย

ค.ศ. 1890 แคทเทล (Cattell) เป็นผู้ริเริ่มใช้คำว่า การทดสอบสมอง (Mental Test) ในประเทศสหรัฐอเมริกา

ในยุคที่ 2 ของการประเมินนี้เป็นการประเมินโรงเรียน และสถาบันทางสังคมต่าง ๆ ได้เกิดประเด็นที่นักประเมินจะต้องพิจารณามากมาย เช่น ธรรมชาติและความน่าเชื่อถือของหลักฐานที่ใช้ในการประเมินความเหมาะสมของแบบประเมินที่ใช้ในการประเมิน โปรแกรมขณะ

กำลังดำเนินอยู่ในสภาพธรรมชาติ บทบาทของนักประเมินที่ค่อนข้างขัดแย้งกันคือ ประเมินเพื่อปรับปรุงพัฒนากับบทบาทของนักประเมินในการประเมินเพื่อรักษามาตรฐาน โดยการตัดสินคุณค่าของผลลัพธ์ต่าง ๆ และการแจ้งผลประเมินต่อสาธารณะชนอย่างผู้มีความรับผิดชอบ เป็นต้น คำตอบของประเด็นคำถามเหล่านี้เป็นที่มาของยุคที่ 2

ยุคที่ 3 ยุคแห่งประสิทธิภาพและการทดสอบ (The Age of Efficiency and Testing : 1900 – 1930) ในช่วงแรกของศตวรรษที่ 20 ก่อให้เกิดแนวคิดเกี่ยวกับการบริหารจัดการเชิงวิทยาศาสตร์ (Scientific Management) ขึ้นในประเทศสหรัฐอเมริกา ซึ่งต่อมาเรียกกันว่า ความเคลื่อนไหวในการสำรวจโรงเรียน (School Survey Movement) ซึ่งโรงเรียนเป็นจำนวนมากได้ทำการศึกษาศำรวจการทำงานของตนเอง เพื่อนำไปสู่การจัดระบบการดำเนินงาน (Systematization) การสร้างความเป็นมาตรฐาน (Standardization) และประสิทธิภาพของการดำเนินงานของโรงเรียนและองค์กร (School or Organizational Efficiency) ในการสำรวจโรงเรียนครั้งนี้มีการใช้ข้อสอบแบบใหม่คือ ข้อสอบแบบปรนัยเป็นจำนวนมาก แทนที่จะใช้ข้อสอบเรียงความเพียงไม่กี่ข้อ เช่น ในอดีต

ค.ศ. 1911 มีการพิมพ์ผลการวิจัยทดลองทางการศึกษาที่แสดงถึงความก้าวหน้าของการวิจัยทางการศึกษาในทวีปยุโรป โดยเฉพาะในประเทศสวีเดน

ในยุคนี้ เป็นยุคเริ่มของการประเมินผลสัมฤทธิ์ทางการเรียนแบบอิงกลุ่มเช่นที่บอสตัน มีการเปรียบเทียบประสิทธิภาพการสอนของครู โดยพิจารณาจากค่าเฉลี่ยของคำร้อยละที่นักเรียนสอบผ่านของแต่ละเมืองเป็นเกณฑ์ เป็นต้น กล่าวได้ว่ายุคนี้เป็นยุคของความเคลื่อนไหวเกี่ยวกับการทดสอบ วิทยาการทางการทดสอบก้าวหน้ามาก มีแบบสอบแบบปรนัยเกิดขึ้นในช่วงนี้ การประเมินทางการศึกษาขึ้นอยู่กับความสนใจของท้องถิ่นกับความเชี่ยวชาญของโรงเรียนรัฐบาล ส่วนกลางไม่สนใจเรื่องการประเมิน นับว่าเป็นยุคที่ท้องฟ้าทางการทดสอบ แต่ขาดการสนับสนุนและผลักดัน การประเมินจากหน่วยงานภายนอก ในยุคนี้มีนักทดสอบที่มีชื่อเสียงเกิดขึ้นในท้องถิ่นต่าง ๆ เช่น เคาทิส (Curtis), ไอริส (Ayres) และทอร์นไค (Thomdike) เป็นต้น

โดยสรุป ยุคที่ 3 นี้ ถือเป็นยุคทองของการวัดผลการศึกษา เป็นการทดสอบเพื่อแสวงหาประสิทธิภาพ ระบบ และความเป็นมาตรฐาน สามารถแบ่งเป็น 2 ช่วง คือ ค.ศ. 1900-1915 ถือเป็นระยะการบุกเบิกของการวัดผล โดยการสำรวจและพัฒนาวิธีการวัดผล ทอร์นไค (Thomdike) ถือว่าเป็นนักวัดผลสำคัญของโลก โดยได้เขียนหนังสือ “การวัดผลการศึกษา” เป็นเล่มแรกของโลกตีพิมพ์ในปี ค.ศ. 1905 ซึ่งเป็นการวางรากฐานของการพัฒนาแบบสอบมาตรฐาน สโตน (Stone) เป็นคนแรกที่พัฒนาแบบสอบมาตรฐาน ซึ่งเป็นแบบสอบวิชาเลขคณิต ค.ศ. 1905 บินเน็ต (Binet) ซึ่งเป็นนักจิตวิทยาชาวฝรั่งเศสร่วมกับไซมอน (Simon) สร้างแบบสอบเชาว์ปัญญา

เป็นฉบับแรก เรียกว่า Binet Simon Scale ในช่วงที่สองประมาณ ค.ศ. 1915 – 1930 ซึ่งเป็นระยะการขยายตัวของแบบสอบมาตรฐานและแบบสอบปรนัย เริ่มมีการใช้ทดสอบโดยชุดแบบสอบ (Battery) และเริ่มพัฒนาแบบวัดคุณลักษณะด้านอื่น ๆ มีการใช้แบบถามและแบบสำรวจต่าง ๆ มากขึ้น ค.ศ. 1920 แม็กคอล (McCall) เป็นบุคคลแรกที่แนะนำให้ครูสร้างแบบสอบขึ้นใช้เอง (Teacher Made Test) แบบสอบมาตรฐานที่มีชื่อเรียกมากในขณะนี้คือ แบบสอบ Army Alfa และ Army Beta ซึ่งสร้างขึ้นในปี ค.ศ. 1917

ในช่วงปลายของยุคแห่งประสิทธิภาพและการทดสอบนี้ มหาวิทยาลัยหลายแห่งที่มีความเชี่ยวชาญด้านการทดสอบ ได้ทำงานร่วมกับโรงเรียน เพื่อสำรวจและทดสอบประสิทธิภาพของโรงเรียน ซึ่งนับว่าเป็นจุดเริ่มต้นของศูนย์การประเมินในมหาวิทยาลัยต่าง ๆ ผู้ที่เป็นผู้นำบุกเบิกเรื่องนี้ คือ จีโอ สเตเรอร์ (George Strayer) แห่ง Teachers College, Columbia University เอกลักษณะของยุคนี้ คือ ยุคแห่งการทดสอบมาตรฐาน จะเห็นได้จากหลังสงครามโลกครั้งที่ 1 (ค.ศ. 1914 – 1918) โรงเรียนจะประเมิน ประสิทธิภาพของการเรียนการสอนและหลักสูตร โดยใช้แบบสอบมาตรฐานทำให้การสร้างแบบสอบมาตรฐานกลายเป็นธุรกิจทางการศึกษา ในเมืองพิลาเดลเฟีย ประเทศสหรัฐอเมริกา (ค.ศ. 1916 – 1938) การประเมินผลทางการศึกษาโดยแบบสอบมาตรฐาน มุ่งตอบสนองเฉพาะความต้องการของท้องถิ่นหรือโรงเรียน ไม่มีแรงจูงใจที่จะทำให้ผลประเมินมีนัยทั่วไปหรือมีความเป็นสากล (Generaliability)

ยุคที่ 4 ยุคไทเลอร์ (The Tylerian Age : 1930 – 1945) หรือยุคการประเมินตามวัตถุประสงค์ ในตอนต้นของยุคนี้เกิดเศรษฐกิจตกต่ำมากทำให้นักสังคมศาสตร์มีบทบาทในการกำหนดนโยบายต่าง ๆ เพื่อแก้ไขปัญหาตลอดจนมีความเคลื่อนไหวให้ปฏิรูประบบสังคมและการศึกษา เพื่อแก้ปัญหาดังกล่าวตลอดจนมีความเคลื่อนไหวให้ปฏิรูประบบสังคม และการศึกษา เพื่อแก้ปัญหาทางเศรษฐกิจ จอห์น ดิวอี้และราฟ ไทเลอร์ (John Dewey and Ralph Tyler) เป็นบุคคลสำคัญทางการศึกษาของสหรัฐอเมริกาในการร่วมขบวนการเคลื่อนไหวปฏิรูปการศึกษาดังกล่าว ดิวอี้ได้เป็นผู้นำให้เกิดการศึกษาแบบก้าวหน้า (Progressive Education) และเกิดหลักการศึกษานักการศึกษาทั่วโลกรู้จักกันดีคือการเรียนรู้จากการปฏิบัติ (Learning by Doing) ส่วนไทเลอร์ได้รับแต่งตั้งให้เป็นผู้อำนวยการโครงการ ปี ค.ศ. 1932 – 1940 ซึ่งเป็นโครงการวิจัยการศึกษาขนาดใหญ่ที่ต้องการศึกษาเปรียบเทียบประสิทธิภาพการสอนแบบดั้งเดิมกับการสอนแนวใหม่ในระดับมัธยมศึกษาของประเทศ สหรัฐอเมริกา ไทเลอร์ได้วางแนวคิดในการประเมินหลักสูตรและโปรแกรมการศึกษา โดยการเปรียบเทียบผลที่เกิดขึ้นจริงกับผลสัมฤทธิ์ที่คาดหวังตามวัตถุประสงค์เชิงพฤติกรรม โดยที่ไทเลอร์เป็นบุคคลแรกที่นิยามการประเมินผลการศึกษาว่าเป็นการวัดคุณค่าของผลสำเร็จตามวัตถุประสงค์ของการเรียนการสอน แนวคิดของไทเลอร์ทำให้เห็นความแตกต่าง

ระหว่าง “การวัด” และ “การประเมิน” อย่างเด่นชัด อันเป็นปัจจัยพื้นฐานที่สำคัญที่ทำให้ศาสตร์แห่งการประเมิน ได้รับการพัฒนาอย่างเป็นวิชาชีพ (Professional) และเช่นในปัจจุบัน ในบทความของ ไทเลอร์ได้ระบุอย่างเด่นชัดว่า ในขั้นตอนขอเพียงให้ครูได้สามารถประเมินได้ว่าการสอนของตน ได้ผลหรือไม่โดยการเปรียบเทียบผลสัมฤทธิ์ทางการเรียนที่เกิดขึ้นจริงกับวัตถุประสงค์ของการสอน ซึ่งมีลักษณะการสอนประเมินที่เน้นผลลัพธ์ (Outcome) มากกว่าตัวป้อน (Input) และ กระบวนการ (Process) กล่าวคือ เป็นการเน้นผลมากกว่าเหตุ ซึ่งก็เหมาะสมกับยุคสมัย ซึ่งก่อนยุคนี้นักวิชาการและครูก็ยังไม่ออกกระหว่างการวัดและการประเมิน อย่างไรก็ตามนักวิชาการบางคนที่ไม่อ่านบทความของ ไทเลอร์ก็วิพากษ์วิจารณ์ว่า ไทเลอร์ไม่สนใจประเมินเหตุทำให้การประเมินไม่มีผลต่อการปรับปรุงการเรียนการสอนเท่าที่ควรซึ่งเป็นความเข้าใจผิดของผู้วิจารณ์เอง ในปี ค.ศ. 1986 ไทเลอร์ซึ่งได้เขียนบทความต่อกับบทความ ค.ศ. 1932 โดยเสนอการประเมินทางการศึกษาโดยใช้แนวคิดเชิงระบบ ซึ่งมีการประเมินทั้ง ตัวป้อน กระบวนการ และผลลัพธ์ ครอบคลุมซึ่งก็น่าจะเหมาะสมกับยุคสมัยที่ครูในยุคปัจจุบันน่าจะพัฒนารูปแบบการประเมินการเรียนการสอน และหลักสูตรได้อย่างเป็นระบบและครบวงจร

หลักการและนิยามการประเมินผลการศึกษาที่เสนอโดย ไทเลอร์นับว่าเหมาะสมกับยุคสมัยนั้นเป็นอย่างยิ่ง เป็นที่ยอมรับกันอย่างกว้างขวางว่าหลักสูตรและการเรียนการสอนจะดีหรือไม่จะต้องประเมินที่การเรียนรู้ของศิษย์ดังคำกล่าวที่ว่า “ค่าของครูให้ดูที่ศิษย์” เป็นต้น นอกจากนี้ นิยามการประเมินทางการศึกษา หลักการและรูปแบบการประเมินที่เสนอโดย ไทเลอร์นับว่าเป็นรากฐานและมีอิทธิพลต่อแนวความคิดและยังผลให้เกิดการพัฒนาศาสตร์แห่งการประเมินทางการศึกษาต่อมา ไทเลอร์ซึ่งได้รับการยกย่องว่าเป็นบิดาของการประเมินทางการศึกษาและเป็นผู้ก่อให้เกิดนวัตกรรมของการเขียนวัตถุประสงค์ที่ชัดเจนหรือวัตถุประสงค์เชิงพฤติกรรม (Behavioral Objective)

ยุคที่ 5 ยุคแห่งความไร้เดียงสาทางการประเมิน (The Age of Innocence :
1946 – 1957) หรือยุคแห่งการเพิกเฉยทางการประเมิน (The Age of Ignorance) ทั้งนี้เป็นเพราะยุคนี้เป็นยุคฟื้นฟูประเทศ เนื่องจากภัยสงครามโลกครั้งที่ 2 ซึ่งเกิดในระหว่าง ค.ศ. 1939 – 1945 ทุกคนทุกฝ่ายก็ต้องการลืมความทุกข์ทรมานจากสงครามโลกครั้งที่ 2 ทั้งความทุกข์ไว้ในอดีต มีความพยายามแสวงหาทรัพยากร สร้างและขยายกิจกรรมและบริการต่าง ๆ แสวงหาความสุขความบันเทิงให้ชีวิต ในประเทศสหรัฐอเมริกาเร่งฟื้นฟูประเทศโดยการก่อสร้างอาคาร จัดตั้งสถาบันการศึกษา เพื่อให้มีนิสิตนักศึกษาและครูเพิ่มขึ้นมีการขยายการศึกษาอย่างกว้างขวางทั้ง โปรแกรมการศึกษาและบุคลากรทางการศึกษา การประเมินมีบ้างก็ทำไปเพื่อเอื้อประโยชน์ต่อการสานงานต่อและก่องานใหม่มากกว่าเป็นการประเมินและใช้ผลประโยชน์ตามคุณค่าที่แท้จริง ดังนั้น

การประเมินในยุคนี้จึงมิได้มุ่งไปเพื่อปรับปรุงพัฒนาตามปรัชญาของการประเมิน ข้อสังเกตก็คือ นิสิตนักศึกษาและผู้คนในยุคนี้เกือบไม่มีใครสนใจ และรับผิดชอบต่อปัญหาสังคมส่วนใหญ่มุ่งแก้ ปัญหาและสนองความต้องการของตนเองเป็นหลัก

การประเมินโดยทั่วไปในยุคนี้ เป็นความพยายามของนักการศึกษาท้องถิ่น เช่น ไทเลอร์และคนอื่น ๆ ได้ร่วมกันจัดตั้งศูนย์บริการทดสอบทางการศึกษา (Educational Testing Service) ซึ่งระยะต่อมาสถาบันแห่งนี้ได้สร้างความเจริญให้วงการทดสอบ และประเมินผลเป็นอย่างดี

ในการประเมินทางการศึกษาในช่วงนี้ แนวคิดของการประเมินไทเลอร์ได้รับการเผยแพร่ไปอย่างกว้างขวาง แต่มีปัญหาในการปฏิบัติคือการเขียนวัตถุประสงค์เชิงพฤติกรรมตาม ลำดับขั้นของการเรียนรู้ เพื่อช่วยแก้ไขปัญหาดังกล่าว บลูม (Bloom) และคณะและคราซวอล (Krahwahl) จึงได้สร้างสารบบของวัตถุประสงค์การศึกษาขึ้น (Taxonomies of Educational Objectives) ซึ่งทำให้การประเมินทางการศึกษาตามแนวของไทเลอร์ชัดเจนขึ้นยิ่งกว่านั้นทำให้ วงการศึกษาและการครูศึกษา (Teacher Education) ได้มีความชัดเจนในการกำหนดผลการศึกษา ทั้งในระดับชาติ ระดับหลักสูตร และระดับบทเรียนด้วย

โดยสรุป เอกลักษณะทางการประเมินในยุคนี้ก็คือ การประเมินการศึกษาที่ตอบสนองความต้องการของท้องถิ่น โดยประเมินตามแนวของไทเลอร์ซึ่งเน้นประเมินตามวัตถุประสงค์ ค.ศ. 1949 ไทเลอร์ที่เขียนหนังสือที่มีอิทธิพลอย่างยิ่งต่อวงการศึกษาคือการพัฒนาหลักสูตรและการประเมินผล ค.ศ. 1950 กุลิกสัน (Gulliksen) ได้เขียนหนังสือทฤษฎีการวัด ค.ศ. 1953 ลิน ไควสต์ (Lindquist) ได้เขียนหนังสือเกี่ยวกับระบบจำแนกวัตถุประสงค์ทางการศึกษาด้าน พุทธิพิสัยซึ่งผลงานเหล่านี้มีคุณประโยชน์อย่างยิ่งต่อวงการวัดและประเมินผลการศึกษาและวงการ ศึกษาโดยส่วนรวม

ยุคที่ 6 ยุคการขยายองค์ความรู้ทางการประเมิน (The Age of Expansion : 1958 – 1972) เนื่องจากสหภาพโซเวียตประสบความสำเร็จในการส่งดาวเทียมสปุตนิก 1 (Sputnik 1) ขึ้นสู่อวกาศเป็นประเทศแรกของโลกทำให้ประเทศสหรัฐอเมริกาตื่นตระหนก รู้สึกว่า ประเทศของตนด้าหลังสหภาพโซเวียตเสีย จึงรีบเร่งปรับปรุงการศึกษาเป็นการใหญ่โดยเน้นการ ศึกษาด้านวิทยาศาสตร์ คณิตศาสตร์ ภาษาดังประเทศ และการแนะแนว โดยปรับปรุงหลักสูตร และการเรียนการสอนทั้งโดยรัฐบาลและในแต่ละรัฐ โดยการลงทุนมหาศาล ริเริ่มมุ่งสร้างให้ อเมริกันชนเป็นผู้รอบรู้ทางวิทยาศาสตร์และเทคโนโลยีที่ทันสมัย เพื่อให้ประเทศสหรัฐอเมริกา เป็นประเทศที่ยิ่งใหญ่ ทันสมัยอีกครั้งหนึ่ง การส่งเสริมให้มีการพัฒนาหลักสูตรใหม่ ๆ และการ ลงทุนในโครงการพัฒนาต่าง ๆ มากมาย ย่อมเป็นแรงผลักดันให้มีการประเมินผลอย่างเป็นระบบ

ทุกโครงการต้องมีการประเมินโครงการประจำปีและต้องมีรายงาน ด้วยความต้องการดังกล่าวทำให้การประเมินเจริญขึ้นในฐานะเป็นศาสตร์สาขาที่เป็นความเชี่ยวชาญ ชำนาญการ โดยเฉพาะเกิดการยอมรับศาสตร์ว่าด้วยการประเมิน โปรแกรมและโครงการ อีกทั้งเห็นประโยชน์และคุณค่าของการประเมินโครงการและโปรแกรมอย่างกว้างขวาง ซึ่งไม่เคยมีบรรยากาศเช่นนี้มาก่อนในประวัติศาสตร์ของการวัดและประเมินทางการศึกษา

ในขณะที่มีการพัฒนากระบวนการประเมิน โปรแกรม และจากกระบวนการประเมินก็นำสู่ปฏิบัติทำให้พบว่าโมเดลและวิธีที่มีอยู่มีจุดอ่อนและข้อจำกัดมากมาย เช่น การออกแบบทดลองไม่สามารถใช้ในการประเมินในสภาพธรรมชาติหรือการประเมินในสนามได้ แบบสอบถามมาตรฐานมีความหนักแน่นคือไม่ไวพอที่จะวัดผลของหลักสูตรใหม่ในสถานการณ์เฉพาะ และที่สำคัญก็คือการประเมินที่มุ่งตอบคำถามเฉพาะ และความต้องการเฉพาะของท้องถิ่นให้ผลประโยชน์ที่มีข้อจำกัด ไม่สามารถใช้ตัดสินคุณค่าหรือประสิทธิผลของโปรแกรมระดับชาติได้

ด้วยอุปสงค์การประเมินที่มีคุณภาพและอุปทานความรู้เกี่ยวกับการประเมิน ปัจจุบันที่มีจุดอ่อนและข้อจำกัดต่าง ๆ ซึ่งไม่สามารถตอบสนองความต้องการได้ทำให้เกิดความเร่งด่วนที่ต้องขยายองค์ความรู้ด้านการประเมินอย่างรีบเร่งตามความต้องการเร่งด่วนที่จะได้กระบวนการประเมินหรือ โมเดลการประเมินที่มากกว่าการวัดด้วยข้อสอบปรนัย แล้วนำผลที่วัดได้ไปเปรียบเทียบกับวัตถุประสงค์ที่กำหนด ในการตอบสนองความต้องการทางการประเมินดังกล่าว นักวิชาการก็ได้เสนอแนวคิดและ โมเดลการประเมินมากมาย เช่น ครอนบาค (Cronbach, 1963); สคริฟเวน (Scrive, 1967); สเต็ก (Stake, 1967); ไทเลอร์ (Tyler, 1969); สตัฟเฟิลบีม (Stufflebeam, 1971); โปรวัส (Provus, 1971); คาโร (Caro, 1971); เวสส์ (Weiss, 1972); วอร์เรนและเซนเดอร์ (Worthen and Sanders, 1973) เป็นต้น นอกจากนี้สมาคมต่าง ๆ ก็มีบทบาทในการพัฒนาทฤษฎีและสร้างนักประเมิน เช่น สมาคมเกียรตินิยมทางการศึกษา (Phi Delta Kappa) โดยสตัฟเฟิลบีมได้พัฒนาหลักการและโมเดล CIPP ในการประเมินโปรแกรมแล้ว สมาคมเกียรตินิยมทางการศึกษายังจัดอบรมผู้สนใจที่จะเป็นนักประเมินด้วย

ในปี ค.ศ. 1966 แคมเบลและสแตนเลย์ (Campbell and Stanley) ได้เขียนหนังสือชื่อการออกแบบทดลองและกึ่งทดลองเพื่อการวิจัยเพื่อชี้ให้เห็นว่าการวิจัยทดลองที่เข้มงวดเป็นรูปแบบที่มีข้อจำกัดมากมายในการวิจัยทดลองในสนามและเรียกร้องให้สนใจอย่างจริงจังในการนำแบบทดลองไปใช้ในการวิจัยหรือการประเมินในสภาพการณ์จริง

โดยสรุปจึงกล่าวได้ว่ายุคนี้เป็นการประเมินยุครุ่งอรุณของการประเมินที่จะก้าวต่อไปในฐานะศาสตร์แห่งวิชาชีพเพื่อการพัฒนาคน พัฒนางาน พัฒนาองค์กรและสภาพสังคมเพื่อมนุษยชาติ