

ยุคสมัยของการประเมินผลและแนวคิดทฤษฎีของนักประเมิน



The Historical Development of Evaluation and Theorists' Concepts



ศุภามณ จันทรสกุล

บทคัดย่อ

การประเมินผล (Evaluation) เป็นการวัดผลโดยเปรียบเทียบสิ่งที่ถูกประเมินกับเกณฑ์มาตรฐานทั้งนี้เพื่อตัดสินคุณค่า ผลประโยชน์ และความคุ้มค่าที่ได้รับจากสิ่งที่ถูกประเมินนั้น ยุคสมัยของการประเมินเริ่มตั้งแต่อ่อนปี ค.ศ. 1900 และมีพัฒนาการอย่างต่อเนื่องมาจนถึงปัจจุบัน ในแต่ละยุคของการประเมินมีแนวคิดทฤษฎีของนักประเมินที่สำคัญๆ ได้แก่ แนวคิดทางการประเมินของ Michael Scriven, การประเมินซิปป์โมเดลของ Daniel L. Stufflebeam, การประเมินแบบตอบสนองของ Robert Stake, การประเมินเน้นการใช้ประโยชน์ของ Michael Patton, การประเมินโดยใช้ทฤษฎีเป็นแรงขับเคลื่อนของ Huey T. Chen, การประเมินยุคที่ 4 ของ Guba & Lincoln เป็นต้น บทความนี้นำเสนอ ยุคสมัยของการประเมินและแนวคิดทฤษฎีของนักประเมินเพื่อนำไปสู่ความรู้ ความเข้าใจ เพื่อนำแนวคิดของนักทฤษฎีไปใช้ในการประเมินผลต่างๆ ต่อไป

คำสำคัญ: การประเมินผล, การประเมินซิปป์โมเดล, การประเมินแบบตอบสนอง, การประเมินเน้นการใช้ประโยชน์, การประเมินโดยใช้ทฤษฎีเป็นแรงขับเคลื่อน, การประเมินยุคที่ 4

Abstract

The evaluation includes measurement, compares evaluation objects with standard criteria, and compares value judgment about merit, worth, significance, and benefit of evaluation objects. The period of evaluation, is the period prior to 1900 and has been developed until the recent times. Each period of evaluation has evaluation theorists' concepts such as an evaluation of Michael Scriven, The CIPP Model of Daniel L. Stufflebeam, Stake's responsive evaluation, Utilization-Focused Evaluation of Michael Patton, Theory-Driven Evaluation of Huey T. Chen, and Four Generation Evaluation Theory of Guba & Lincoln. This article presents the historical development of evaluation and theorists' concepts for knowing, understanding and applying theorists' concepts to others evaluation projects.

Keywords: Evaluation, CIPP Model, Responsive Evaluation, Utilization-Focused Evaluation (UFE), Theory - Driven Evaluation, Four Generation Evaluation Theory



ความนำ

ในชีวิตประจำวันของมนุษย์เกี่ยวข้องกับ การประเมินผลอย่างไม่อาจสามารถหลีกเลี่ยงได้ การประเมินผลเกี่ยวข้องกับเราในการเปรียบเทียบ สิ่งต่างๆ เพื่อตัดสินคุณค่าและผลประโยชน์ (benefit) ของสิ่งนั้นเมื่อเทียบกับเกณฑ์มาตรฐาน (criteria) ที่กำหนดในใจของแต่ละบุคคล ไม่ว่าจะเป็นตั้งแต่ การเลือกอาหารที่รับประทาน การใช้จ่ายใช้สอย ตลอดจนการกระทำกิจกรรมต่างๆ

การประเมินผล (Evaluation: E) เป็นการตัดสินคุณค่า (Judge: J) โดยวัดผล (Measurement: M) และตัดสินคุณค่าของสิ่งที่ถูกประเมินเทียบกับเกณฑ์ (criteria) หรือ $E = M + J$ บุคคลจะตัดสินใจประเมินผล ว่าสิ่งใด ๆ มีคุณค่าสูงจากการเปรียบเทียบกับเกณฑ์ในใจ ของแต่ละบุคคลแล้วพบว่าสิ่งนั้นมีความสำคัญ มีคุณค่า และให้ความคุ้มค่า (Scriven, 2001; Stufflebeam, & Shinkfield, 2007)

ประวัติศาสตร์ยุคสมัยของการประเมินผลเริ่ม ตั้งแต่ก่อนปี ค.ศ. 1900 โดยแบ่งเป็น 7 ยุค ได้แก่ The Age of Reform (ก่อนปี ค.ศ. 1900), The Age of Efficiency and Testing (ค.ศ. 1900-1930), The Tylerian Age (ค.ศ. 1930-1945), The Age of Innocence (ค.ศ. 1946-1957), The Age of Development or The Age of Realism (ค.ศ. 1958-1972), The Age of Professionalization (ค.ศ. 1973-1983), The Age of Expansion and Integration (ค.ศ. 1983-2001) พัฒนาการทางด้านการประเมินผลยังคงมีอย่างต่อเนื่อง มาจนถึงปัจจุบัน

ในแต่ละยุคสมัยประวัติศาสตร์การประเมินผล มีแนวคิดทฤษฎีของนักประเมินที่สำคัญ ได้แก่ แนวคิด ทางการประเมินของ Michael Scriven, การประเมิน ชิปปี้โมเดลของ Daniel L. Stufflebeam, การประเมิน แบบตอบสนองของ Robert Stake, การประเมินเน้น

การใช้ประโยชน์ของ Michael Patton, การประเมินโดยใช้ ทฤษฎีเป็นแรงขับเคลื่อนของ Huey T. Chen, การประเมินยุคที่ 4 ของ Guba & Lincoln เป็นต้น ซึ่งได้นำเสนอในบทความนี้ เนื้อหาในบทความประกอบด้วยประวัติศาสตร์ยุคสมัย ของการประเมินผล พัฒนาการของการประเมินผล ตลอดจนแนวคิดทฤษฎีของนักประเมินที่น่าสนใจในการ พัฒนางานประเมินผล

ประวัติศาสตร์และพัฒนาการของการ ประเมินผล

ยุคสมัยของการประเมินผลมีการเจริญเติบโต และพัฒนาการดังต่อไปนี้

ยุคแรก ก่อนปี ค.ศ. 1900 ซึ่งเรียกว่า “The Age of Reform” เริ่มมี Formal education evaluation เป็นครั้งแรกในอเมริกาเมื่อปี ค.ศ. 1845 มีการทดสอบ ประเมินผลสัมฤทธิ์ทางการเรียนของนักเรียนในบอสตัน ซึ่งถือเป็นจุดสำคัญของประวัติศาสตร์การประเมิน เพราะ มีการใช้คะแนนการทดสอบเพื่อประเมินประสิทธิภาพ ของโรงเรียนหรือโปรแกรมการเรียนการสอน (Stufflebeam and Shinkfield, 2007; Stufflebeam, Madaus, & Kellaghan, 2000 cited in R. Lance Hogan, 2007) ต่อมาในปี ค.ศ. 1895 Joseph Rice ได้รวบรวม คะแนนสอบในการเรียนการสอนและคณิตศาสตร์ จากนักเรียน 16,000 คนในอเมริกา และค้นพบว่า ในแต่ละวันมีการใช้เวลาในการเรียนการสอนมาก แต่ผลสัมฤทธิ์ทางการเรียนในการสอบกลับน้อย การประเมินของไรซ์ถือเป็น Formal educational program evaluation ครั้งแรกในอเมริกา (Stufflebeam and Shinkfield, 2007; Stufflebeam et al., 2000 cited in R. Lance Hogan, 2007)

ยุคที่สอง ปี ค.ศ. 1900-1930 ซึ่งเรียกว่า The Age of Efficiency and Testing Fredrick W. Taylor ได้นำแนวคิดประสิทธิภาพ (efficiency) และ

การสอบมาตรฐาน (Standardized testing) มาเพื่อค้นมาตรฐานและประสิทธิภาพในโรงเรียน (Stufflebeam and Shinkfield, 2007) มีการใช้ Objective-based tests เป็นสิ่งสำคัญในการตัดสินคุณภาพของการเรียนการสอน จะเห็นได้ว่าในช่วงแรกนี้ยังถือว่าการวัดและการประเมินเป็นสิ่งเดียวกัน (Worthen, Sanders, & Fitzpatrick, 1997 cited in R. Lance Hogan, 2007)

ยุคที่สาม ปี ค.ศ. 1930 – 1945 ซึ่งเรียกว่า “The Tylerian Age” ในช่วงต้นปี ค.ศ. 1930 มีนักทฤษฎีคนสำคัญซึ่งได้รับการกล่าวขานว่าเป็นบิดาแห่งการประเมินผลการศึกษาได้แก่ Ralph Tyler (Stufflebeam and Shinkfield, 2007; Stufflebeam et al., 2000 cited in R. Lance Hogan, 2007) ไทเลอร์ได้ทำการศึกษาเป็นระยะเวลา 8 ปี (1932-1940) โดยประเมินผลโปรแกรมการศึกษาของโรงเรียนมัธยมศึกษาตอนปลายและค้นพบว่าวัตถุประสงค์ของการเรียนการสอนเป็นสิ่งที่อธิบายพฤติกรรมในการเรียนการสอน วัตถุประสงค์จึงเป็นพื้นฐานของการประเมิน (Tyler, 1975 cited in R. Lance Hogan, 2007) การประเมินในแนวคิดของไทเลอร์จึงเป็นการประเมินอิงวัตถุประสงค์ (objective-based evaluation) โดยเทียบผลลัพธ์ที่เกิดขึ้น (outcomes) กับวัตถุประสงค์ (objectives) ว่ามีความสอดคล้องกันเพียงใด แต่ไม่ได้ให้ความสนใจมากนักกับความแตกต่างของคะแนนนักเรียนแต่ละคน การประเมินจึงเริ่มมีความชัดเจนแยกออกจากการวัดโดยมีความชัดเจนในสิ่งที่ถูกประเมินเมื่อเทียบกับเกณฑ์ซึ่งคือวัตถุประสงค์นั่นเอง

ยุคที่สี่ ปี ค.ศ. 1946 – 1957 ซึ่งเรียกว่า “The Age of Innocence” เป็นช่วงเวลาหลังสงครามโลกครั้งที่ 2 ซึ่งผู้คนต้องการลืมความเจ็บปวดซึมเศร้าจากสงครามทำให้เกิดการพัฒนาในหลายๆ อย่างเพื่อก้าวสู่ชีวิตที่ดีขึ้น ในทางการศึกษามีการขยายตัวในเรื่องของตัวบุคคล (personnel) และปัจจัยเอื้อ (facilities) (Stufflebeam, 2007; Stufflebeam and Shinkfield, 2007) ทำให้เกิดเครื่องมือและกลวิธีที่หลากหลายในการประเมิน ในปี ค.ศ. 1956 เกิด “Taxonomy of Educational Objectives” โดย Bloom, Engelhart, Furst, Hill และ Krathworth ที่แสดงให้เห็นว่าพฤติกรรมของผู้เรียนหรือผลลัพธ์ทางการเรียนที่หลากหลายถูกจำแนกตามวัตถุประสงค์

ต่าง ๆ ทางการศึกษา (Stufflebeam and Shinkfield, 2007) ในยุคนี้นักเรียนจะประเมินผลหรือไม่ประเมินผลทางการศึกษาก็ได้โดยไม่มีความกดดันจากนโยบายของรัฐบาลกลางและมลรัฐ ทุนในการประเมินมาจากองค์กรและมูลนิธิในท้องถิ่น (Stufflebeam and Shinkfield, 2007)

ยุคที่ห้า ปี ค.ศ. 1958 – 1972 ซึ่งเรียกว่า “The Age of Development or The Age of Realism” นับตั้งแต่มีการส่งยาน Sputnik 1 ในปี ค.ศ. 1957 ทำให้มีการพัฒนาของหลักสูตรต่าง ๆ โดยเฉพาะวิทยาศาสตร์และคณิตศาสตร์จึงต้องมีการประเมินผลโปรแกรมหลักสูตรการเรียนเหล่านั้น ต่อมาในปี ค.ศ. 1957 รัฐบาลสหรัฐมีนโยบายให้ทุนสนับสนุนการศึกษาสำหรับเด็กที่มีปัญหาทางการเรียน ในการประเมินผลมีการใช้การทดสอบมาตรฐานร่วมกับการประเมินความสอดคล้องของผลลัพธ์และวัตถุประสงค์ตามแนวคิดของไทเลอร์ ซึ่งผลการประเมินพบความไม่สอดคล้องเกิดขึ้นเนื่องจากนักศึกษามุ่งเทคนิคสำคัญในการตั้งวัตถุประสงค์เชิงพฤติกรรมที่ดีโดยไม่ได้เขียนวัตถุประสงค์นั้นจากข้อมูลความต้องการจำเป็น (needs) และปัญหาของนักเรียน ผลดังกล่าวทำให้เกิดความตื่นตัวทางการประเมินผล มีนักทฤษฎีทางการประเมินเกิดขึ้นอย่างมากได้แก่ Provus (ปี ค.ศ. 1969 & 1971), Hammond (ปี ค.ศ. 1967), Eisner (ปี ค.ศ. 1967) และ Metfessel & Michael (ปี ค.ศ. 1967) เสนอแนวคิดที่ปรับปรุงจากไทเลอร์ Glaser (ปี ค.ศ. 1963), Tyler (ปี ค.ศ. 1967) และ Popham (ปี ค.ศ. 1971) เสนอการทดสอบแบบอิงเกณฑ์ (criterion-referenced testing) เป็นทางเลือกจากการทดสอบแบบอิงกลุ่ม (norm-referenced testing) Cook (ปี ค.ศ. 1966) เสนอการใช้การวิเคราะห์เชิงระบบ (the systems-analysis approach) ในการประเมินโปรแกรม Scriven (ปี ค.ศ. 1967), Stufflebeam และคณะ (ปี ค.ศ. 1967 & 1971), Stake (ปี ค.ศ. 1967) นำเสนอโมเดลใหม่ของการประเมิน ซึ่งแนวคิดของเหล่านักประเมินดังกล่าวข้างต้นตระหนักถึงความจำเป็นในการประเมินเป้าหมาย (goals) ปัจจัยนำเข้า (inputs) การดำเนินงาน (implementation) การวัดผลลัพธ์ที่ตั้งใจให้เกิด (intended outcomes) และผลลัพธ์ที่ไม่ตั้งใจให้เกิด (unintended

outcomes) ของโปรแกรม รวมถึงเน้นความจำเป็นในการตัดสินเกี่ยวกับคุณค่า (merit or worth) ของสิ่งที่ถูกประเมิน จะเห็นได้ว่าพัฒนาการของการประเมินในยุคนี้มีแนวคิดการประเมินกระบวนการ การประเมินเพื่อตัดสินคุณค่า นอกเหนือไปจากการประเมินที่เน้นผลลัพธ์เพียงอย่างเดียว

ยุคที่หก ปี ค.ศ. 1973 – 1983 ซึ่งเรียกว่า “The Age of Professionalization” เป็นยุคที่มีความพยายามที่จะทำให้การประเมินมีความเป็นมาตรฐานทางวิชาชีพ โดยมีวารสารทางการประเมินเกิดขึ้นได้แก่ Educational Evaluation and Policy Analysis, Studies in Educational Evaluation, CEDR Quarterly, Evaluation Review, New Directions for Program Evaluation, Evaluation and Program Planning, Evaluation News เป็นต้น รวมถึงมหาวิทยาลัยต่าง ๆ ได้แก่ University of Illinois, Stanford University, Boston College, UCLA, University of Minnesota, Western Michigan University และ Nova University เป็นต้น โดยมหาวิทยาลัยมีการจัดอบรมและมีหลักสูตรการเรียนการสอนที่เกี่ยวกับการประเมิน นอกจากนี้ในยุคนี้ยังมีการพัฒนาเทคนิคต่าง ๆ ในการประเมิน รวมถึงการตอบสนองของความต้องการจำเป็นของผู้ใช้การประเมิน (clients of evaluation) มีการพัฒนาแนวคิดทฤษฎีทางการประเมินต่าง ๆ ได้แก่ การประเมินแบบไม่อิงวัตถุประสงค์ (goal-free evaluation) ของ Scriven, การประเมิน CIPP Model ของ Stufflebeam, การประเมิน responsive evaluation ของ Stake, การประเมินเชิงธรรมชาติ (naturalistic evaluation) โดย Guba & Lincoln, การประเมิน metaevaluation, การประเมินแบบศิลปวิทยาการของ Eisner, การประเมินความต้องการจำเป็น (needs assessment) เป็นต้น

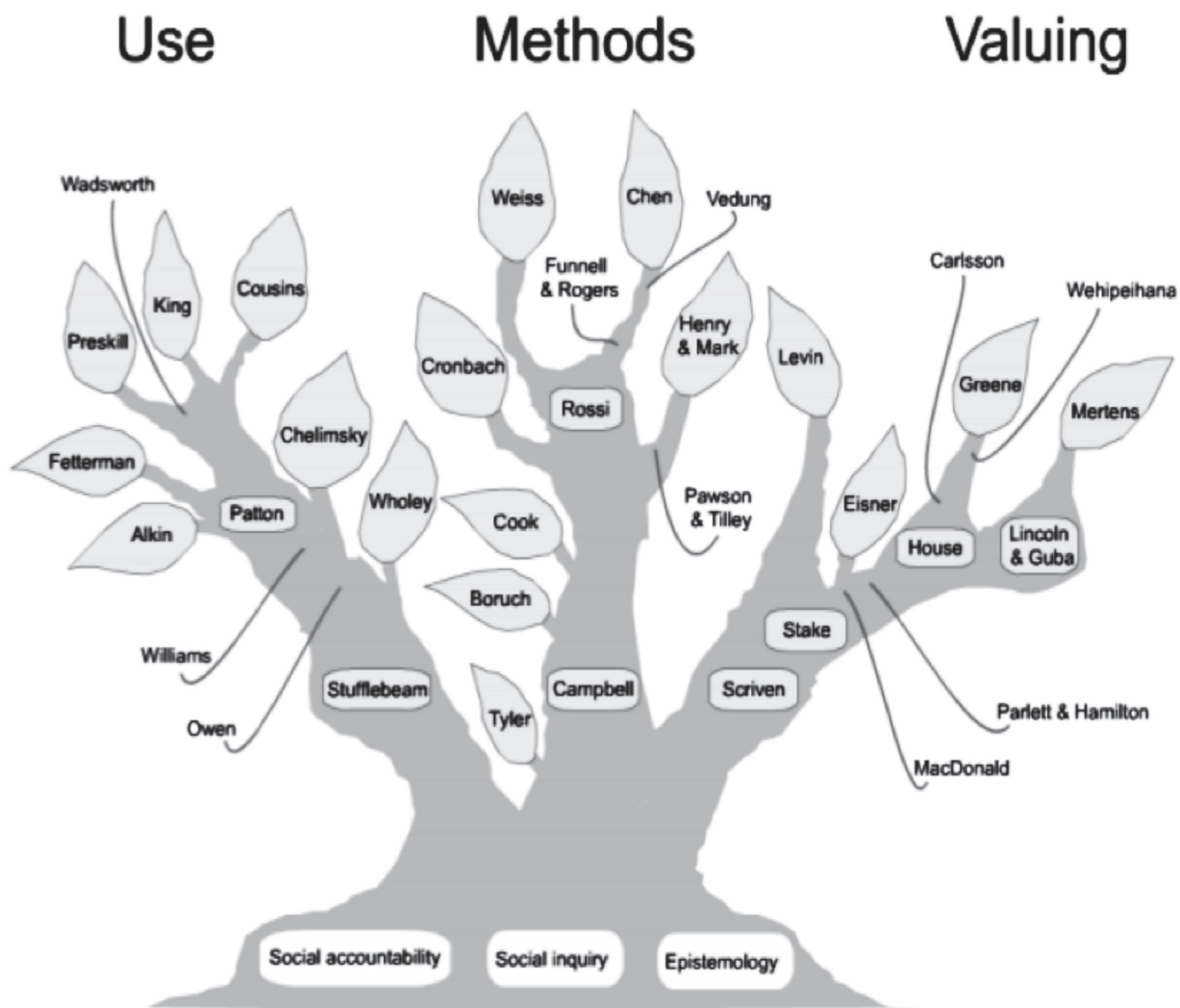
ยุคที่เจ็ด ปี ค.ศ. 1983 – 2001 ซึ่งเรียกว่า “The Age of Expansion and Integration” ในยุคนี้การประเมินมีการขยายตัวเจริญงอกงามซึ่งแสดงให้เห็นจากการมีมากกว่า 20 ประเทศที่เข้าร่วมใน professional evaluation societies มีความร่วมมือกันและการติดต่อสื่อสารกันของเหล่านักประเมิน นอกจากนี้ยังมี The Evaluation Research Society มีเครือข่ายของนักประเมิน American Evaluation Association (AEA) ผลดังกล่าว

ทำให้เกิดการบูรณาการและการแลกเปลี่ยนความคิดในการประเมิน (cross-fertilization of evaluation ideas) ตลอดจนวิธีวิทยาการในการประเมิน ในปี ค.ศ. 1995 AEA ได้จัดการประชุมและเชิญนักประเมินจากทั่วโลก ซึ่งประสบความสำเร็จและเกิดองค์กรความร่วมมือทางการประเมิน (evaluation associations) มากกว่า 20 แห่งทั่วโลก นอกจากนั้น Joint Committee on Standards for Educational Evaluation ได้กำหนดมาตรฐานต่าง ๆ ที่เกี่ยวกับการประเมิน ได้แก่ มาตรฐานการประเมินโปรแกรม (The program evaluation standards) มาตรฐานการประเมินบุคคล (personnel evaluation standards) มาตรฐานการประเมินนักเรียน (The student evaluation standards) เป็นต้น อีกทั้งมีการกำหนดมาตรฐานของนักประเมิน (AEA’s Guideline principles for evaluators)

ประวัติศาสตร์ของการประเมินตั้งแต่อดีตถึงปัจจุบันในแต่ละยุคมีนักทฤษฎีทางการประเมินที่สำคัญ ซึ่งปัจจุบันก็ยังมีกรนำแนวคิดทฤษฎีต่าง ๆ เหล่านั้นมาใช้กันอยู่ในปัจจุบันดังจะกล่าวถึงต่อไป

แนวคิดทฤษฎีของนักประเมินผลที่สำคัญ

Alkin และ Christie ได้แกะรอยแนวคิดของนักทฤษฎีทางการประเมินต่าง ๆ ในแต่ละยุคและได้เขียน Evaluation Roots ในปี ค.ศ. 2004 และปรับปรุงใหม่ในปี 2012 โดยส่วนประกอบของต้นไม้แห่งการประเมินได้แก่รากของการประเมินอยู่บนฐานของหลักความรับผิดชอบ ต่อสังคม (social accountability) หลักการสืบเสาะแสวงหาความจริงทางสังคม (social inquiry) และญาณวิทยา (epistemology) ซึ่งเป็นศาสตร์/ทฤษฎีที่เกี่ยวข้องกับความรู้ และได้จัดความรู้ทางการประเมินเป็นกิ่งก้านสาขาของต้นไม้โดยแบ่งเป็น การประเมินที่เน้นการนำไปใช้ประโยชน์ (use), การประเมินที่เน้นทางวิธีการ (Methods) และการประเมินที่เน้นด้านคุณค่า (valuing) (Alkin, M. C., & Carden, F., 2012) เมื่อนำแนวคิดของนักประเมินในแต่ละยุคมาจัดตำแหน่งบนต้นไม้และสาขากิ่งก้านของการประเมินเพื่อรวบรวมแนวคิดของนักประเมินให้เป็นหมวดหมู่จึงได้ต้นไม้แห่งการประเมินดัง ภาพ 1



ภาพ 1 ต้นไม้แห่งการประเมิน (Evaluation Theory Tree)

ที่มา. จาก “Evaluation Roots: An International Perspective,” 2012 โดย Alkin, M. C., ค้นจาก www.ipdet.org

จากภาพต้นไม้แห่งการประเมินแสดงให้เห็นถึงการเติบโตเจริญงอกงามของศาสตร์ทางการประเมินตั้งแต่อดีตถึงปัจจุบันในแต่ละยุคเกิดนักทฤษฎีทางการประเมินหลายท่าน โดยในปัจจุบันยังคงมีการนำแนวคิดทฤษฎีสำคัญๆ มาใช้อยู่ดังนี้

1. แนวคิดทางการประเมินของ Michael Scriven

แนวคิดทางการประเมินของ Scriven มุ่งในเรื่อง Consumer-Oriented Evaluation โดยเน้นความต้องการจำเป็นของ Consumer หรือเป็นการประเมินแบบ Need-Based Evaluation ตามแนวคิดของ Scriven การประเมินเป็นการตัดสินคุณค่า (merit, worth, value) ความสำคัญ

(significance) ของสิ่งที่มีงประเมินเมื่อเทียบกับเกณฑ์ (criteria) เช่นในเรื่องต้นทุน (costs) และความคุ้มค่า (benefits) โดยเกณฑ์การประเมินจำแนกได้เป็นเกณฑ์สัมบูรณ์ (absolute criteria) ซึ่งกำหนดระดับคุณภาพมาตรฐานซึ่งเป็นเกณฑ์จุดตัดที่ยอมรับได้ และเกณฑ์สัมพัทธ์ (relative criteria) ซึ่งกำหนดคุณภาพโดยการเปรียบเทียบกับสิ่งที่เคยทำในอดีต บทบาทหน้าที่ของผู้ประเมินต้องทำการตัดสินคุณค่าของสิ่งที่ประเมิน ส่วนบทบาทของการประเมินนอกจาก Goal Based Evaluation ซึ่งประเมินผลลัพธ์อิงวัตถุประสงค์ของโครงการตามแนวคิดการประเมินแบบ Tylerian แล้ว Scriven ยังได้เสนอการประเมินแบบ Goal Free Evaluation

ซึ่งประเมินผลที่เกิดขึ้นจริงทั้งหมดของโครงการทั้งผลที่คาดหวังและไม่คาดหวัง ผลทางบวกและลบ และนอกจากจะประเมินผลสรุป (Summative Evaluation) ซึ่งเป็นการประเมินผลลัพธ์ (Product Evaluation) ตัดสินคุณค่าเมื่อโครงการสิ้นสุดลงแล้ว Scriven ยังได้เสนอว่าควรมีการประเมินความก้าวหน้า (Formative Evaluation) ระหว่างการดำเนินโครงการซึ่งเป็นการประเมินกระบวนการ (Process Evaluation) ทั้งนี้เพื่อการพัฒนาปรับปรุงให้โครงการสามารถดำเนินไปได้ดี และตอบสนองความต้องการจำเป็นของ Consumer นอกจากนี้ Scriven ยังได้กล่าวถึงการประเมินแบบ Intrinsic และ Payoff โดย Intrinsic Evaluation เป็นการประเมินเกี่ยวกับเป้าหมาย, โครงสร้าง, ระเบียบวิธีการ, คุณลักษณะและทัศนคติของ staff, การอำนวยความสะดวก, ความน่าเชื่อถือ ส่วน Payoff Evaluation เป็นการประเมินเกี่ยวกับผลกระทบที่เกิดขึ้นกับ clients (Scriven, M., 1998; Scriven, M., 2001; Scriven, M., 2007; Scriven, M., 2011)

จุดเด่นในแนวคิดของ Scriven คือ Goal Free Evaluation ซึ่งประเมินผลที่เกิดขึ้นจริงทั้งหมดของโครงการทั้งผลที่คาดหวังและไม่คาดหวัง ผลทางบวกและลบ ซึ่งแตกต่างจากแนวคิดการประเมินแบบเดิมซึ่งอิงเพียงวัตถุประสงค์ของโครงการอย่างเดียว

2. แนวคิดทางการประเมิน CIPP Model ของ Daniel L. Stufflebeam

ในแนวคิดการประเมิน CIPP Model การประเมินเป็นกระบวนการที่เป็นระบบเกี่ยวข้องกับคุณภาพ คุณค่า (Merit/Worth) และความสำคัญ (Significance) ของสิ่งที่มุ่งประเมินเพื่อใช้ในการตัดสินใจ (decision making) โดยมาตรฐานของการประเมินอยู่บนหลักของประโยชน์ (Utility), ความเป็นไปได้ (Feasibility), ความชอบธรรม (Propriety), ความถูกต้อง (Accuracy) ในการประเมิน CIPP Model นักประเมิน (Evaluator) ต้องร่วมมือกับผู้ใช้การประเมิน (Client/Stakeholder) ในสิ่งที่เกี่ยวข้อง ได้แก่ ผู้ได้รับผลประโยชน์ (Beneficiaries) ความต้องการจำเป็นของผู้ใช้การประเมินในทุกขั้นตอนของ CIPP ซึ่งประกอบด้วย

Context evaluation เป็นการประเมินบริบทซึ่งเกี่ยวข้องกับความต้องการจำเป็นของผู้ได้รับผลประโยชน์ (beneficiaries' needs) สินทรัพย์ (assets) ที่มี และปัญหา (problems) ภายใต้บริบทสิ่งแวดล้อมในการประเมินซึ่งรวมถึงการเมืองในการประเมิน (political dynamics) ด้วย โดยการประเมินบริบทจะเกี่ยวข้องกับ core values ในการกำหนดเป้าหมาย (Goal-setting tasks)

Input evaluation เป็นการประเมินสิ่งนำเข้าเพื่อการวางแผน(plans) การเลือกกลยุทธ์ (strategies) ที่ตอบสนองต่อ needs และมีความเป็นไปได้ ตลอดจนงบประมาณที่เกี่ยวข้องกับแผนงาน (budgets of the selected approach)

Process เป็นการประเมินกระบวนการในการดำเนินงานของโครงการซึ่งเกี่ยวข้องกับ การติดตาม (monitor) รวมถึงเอกสาร (document) และกิจกรรมต่าง ๆ (activities)

Product เป็นการประเมินผลผลิตและผลที่ได้จากโครงการ ซึ่งเกี่ยวข้องกับการประเมินดังนี้ การประเมินผลกระทบ (impact evaluation), การประเมินประสิทธิผล คุณภาพ และความสำคัญของผลลัพธ์ที่ได้ (effectiveness evaluation), การประเมินความสำเร็จและความต่อเนื่อง (sustainability evaluation), การประเมินเพื่อขยายความสำเร็จของโครงการนำไปประยุกต์ใช้กับโครงการอื่น (transportability evaluation)

การประเมิน CIPP Model มีบทบาทของการประเมินในเชิงการประเมินความก้าวหน้า (formative evaluation) และการประเมินผลสรุป (summative evaluation) ดังนี้

- บทบาทของการประเมินความก้าวหน้า : การใช้ข้อมูลของ CIPP เป็นแนวทางชี้แนะ (guidance) ช่วยในการตัดสินใจพัฒนาปรับปรุงการดำเนินงานในขณะที่โครงการกำลังดำเนินอยู่เพื่อคุณภาพของโครงการ

- บทบาทของการประเมินผลสรุป : การใช้ข้อมูลของ CIPP ในการตัดสินคุณภาพ ความคุ้มค่า และความสำคัญของโครงการเมื่อโครงการสิ้นสุดลงแล้ว (Stufflebeam, 2007; Stufflebeam and Shinkfield, 2007)

จุดเด่นของการประเมินแบบ CIPP คือมีทั้งการประเมินแบบ Formative ซึ่งประเมินระหว่างการดำเนินโครงการและ Summative ซึ่งประเมินเมื่อโครงการสิ้นสุดลงแล้ว ดังนั้นการนำแนวคิด CIPP มาใช้ในการประเมินโครงการไม่ควรใช้แค่เพียงสรุปผลโครงการเมื่อโครงการสิ้นสุดไปแล้วเพียงอย่างเดียวเท่านั้นซึ่งเป็นการนำไปใช้ที่ผิดหลักการของ CIPP

3. แนวคิดทางการประเมิน Responsive Evaluation ของ Robert Stake สำคัญ

Stake ได้ให้ความหมายของการประเมินในแนวคิด Responsive Approach (Stake, R.E., & Abma, T. A., 2001) ดังนี้ การประเมิน (Evaluation) = คุณค่าทั้งหมดของโครงการ (Whole constellation of values held for a program) / ความคาดหวังและเกณฑ์ที่แต่ละบุคคลคิดเห็นต่อโครงการ (Complex of expectations and criteria that different people have for the program)

ลักษณะการประเมินแบบ Responsive Evaluation ในด้านจุดประสงค์ (purpose) เพื่อช่วย client ให้มองเห็นปัญหา จุดแข็งและจุดอ่อน ด้านภาพรวมของบริการ (scope of services) โดยตอบสนองความต้องการของผู้ใช้การประเมินอย่างตรงประเด็น ตอบคำถามและให้ข้อมูลผ่านการศึกษา ด้านพันธะสัญญา (contracts) โดยวัตถุประสงค์และวิธีการถูกวางโครงสร้างจากจุดเริ่มต้นและมีความยืดหยุ่นระหว่างการศึกษา ด้านความมุ่งหมายหลัก (main orientation) เกี่ยวข้องกับความคาดหวังว่าอะไรเป็นสิ่งที่สำคัญด้านการออกแบบ (designs) เป็นลักษณะ open-ended และให้รายละเอียดตอบสนองต่อคำถามสำคัญของ clients ด้านระเบียบวิธีการ (methodology) เป็นเชิงธรรมชาติ ด้านเทคนิค (preferred techniques) ใช้ case study แสดงออกถึงวัตถุประสงค์ใช้ purposive sampling และการสังเกต การรายงานเชิงบรรยายในลักษณะ storytelling ต่อ stakeholder ด้านการติดต่อสื่อสารระหว่างผู้ประเมินกับผู้รับการประเมิน (communications between evaluator and client) มีรูปแบบไม่ป็นทางการ มีลักษณะผ่อนคลายและต่อเนื่อง ด้านการแปลผล (interpretation) แปลผลคุณค่าที่แตกต่างจากมุมมอง

ของผู้คนที่เกี่ยวข้องกับการศึกษา ด้าน key trade-offs เน้นการรายงานความมีประโยชน์ของโครงการด้านการลดความอคติ (provisions for reducing bias) โดยตรวจสอบการรวบรวมข้อมูลหรือศึกษาซ้ำในทอมที่ไม่ชัดเจน การประเมิน Responsive Evaluation ทำอย่างไร ? Stake ได้ให้โครงสร้าง 12 งานของการประเมินในลักษณะเข็มนาฬิกาดังนี้

- 12 นาฬิกา : พูดคุยกับผู้เกี่ยวข้อง ได้แก่ clients, program staff เป็นต้น ซึ่งเกี่ยวข้องกับคำถามสำคัญในการประเมิน

- 1 นาฬิกา : ร่วมมือกับ client ในการกำหนดขอบเขตของโครงการ

- 2 นาฬิกา : วางภาพรวมกิจกรรมของโครงการ

- 3 นาฬิกา : กำหนดวัตถุประสงค์ของโครงการ

- 4 นาฬิกา : วิเคราะห์ประเด็นและสังเคราะห์กรอบแนวคิด

- 5 นาฬิกา : ระบ data needs และแหล่งข้อมูลบริบทของปัญหา

- 6 นาฬิกา : วางแผนกิจกรรมการเก็บรวบรวมข้อมูล ได้แก่ การสังเกต, เครื่องมือ, ตัวอย่าง เป็นต้น

- 7 นาฬิกา : สังเกตโดยใช้ Countenance model พิจารณา congruence ของ intended กับ observed พิจารณามาตรฐานและการตัดสินใจเกี่ยวกับ Antecedents, Transactions, Outcomes โดยเปรียบเทียบเชิงสัมบูรณ์และเชิงสัมพัทธ์

- 8 นาฬิกา : วิเคราะห์ข้อมูลที่เก็บรวบรวม

- 9 นาฬิกา : ตรวจสอบความตรงและวิเคราะห์คุณภาพของข้อมูลที่เก็บรวบรวมได้

- 10 นาฬิกา : ออกแบบและจัดเตรียมรูปแบบการรายงานข้อมูลที่เป็นประโยชน์สูงสุดตอบสนองต่อความต้องการข้อมูลของผู้ใช้ประโยชน์ที่แตกต่างกัน

- 11 นาฬิกา : จัดทำรายงานอย่างเป็นทางการ

จุดเด่นของการประเมินแบบ Responsive Evaluation คือการได้ข้อมูลของการประเมินที่ตอบสนองต่อความต้องการข้อมูลของผู้ใช้ประโยชน์อย่างแท้จริง

4. แนวคิดทางการประเมิน Utilization-Focused Evaluation (UFE) ของ Michael Patton

แนวคิด UFE เน้นการใช้ประโยชน์จากการประเมินของ intended user โดยค้นหามติเอกฉันท์ในการหาคุณค่า (value) และการตัดสินคุณค่า (judgments) ทั้งนี้เพื่อกระบวนการตัดสินใจ จุดประสงค์การประเมินเป็นทั้ง formative, summative และ developmental โดยใช้ข้อมูลทั้ง quantitative, qualitative และ mixed การออกแบบการประเมินเป็นได้ทั้ง naturalistic และ experimental โดยมุ่งที่ processes, outcomes, impacts, costs และ cost-benefit บทบาทของนักประเมินมีหลากหลายโดยต้องเป็นนักเจรจาต่อรอง (negotiator) เป็นผู้ฝึกอบรม (trainer) เป็นผู้อำนวยความสะดวก (group facilitator) เป็นผู้แก้ปัญหา (problem solver) สิ่งสำคัญที่ Patton เน้นคือการเจรจาต่อรองระหว่างนักประเมินกับ intended user ตั้งแต่เริ่มต้นโครงการ จนจบโครงการทั้งนี้เพราะ intended user ต้องมีส่วนเกี่ยวข้องในการประเมินตั้งแต่ต้นจนจบ ในการเจรจาต่อรองเกี่ยวข้องกับกระบวนการ Active (ระบุสิ่งที่ intended user จะถาม) – Reactive (คำตอบในสิ่งที่ intended user ถาม) – Adaptive (ความยืดหยุ่นต่อคำถามการประเมิน) Processes ขั้นตอนของ UFE (2013) (Patton, M.Q. ,2013) ดังนี้

Step 1 ประเมินและสร้างความเข้าใจ ความพร้อมในการใช้ UFE ของผู้ที่เกี่ยวข้อง

Step 2 ประเมินความรู้ ความพร้อม และสมรรถนะในการใช้ UFE ของนักประเมิน

Step 3 แนะนำให้ความช่วยเหลือ สร้างความรู้สึกเป็นเจ้าของการประเมินของ primary intended users

Step 4 ทำงานร่วมกับ primary intended users ในสถานการณ์ต่างๆ ของการประเมิน

Step 5 ระบุและจัดเรียงความสำคัญของวัตถุประสงค์การประเมินร่วมกับ primary intended users

Step 6 พิจารณาและสร้างกระบวนการใช้การประเมินที่เหมาะสมร่วมกับ primary intended users

Step 7 ระบุและช่วย primary intended users ตั้งคำถามและจัดเรียงความสำคัญของคำถามการประเมิน

Step 8 ตรวจสอบว่า primary intended users ได้พิจารณาประเด็นเกี่ยวกับกระบวนการใช้และผลลัพธ์ของโปรแกรมการประเมิน

Step 9 พิจารณากรอบแนวคิดทฤษฎีในการประเมินที่เหมาะสม

Step 10 เจรจาต่อรองวิธีที่เหมาะสมเพื่อสร้างความน่าเชื่อถือของข้อค้นพบ

Step 11 อภิปรายและเลือกระเบียบวิธีที่เหมาะสม

Step 12 ก่อนเก็บรวบรวมข้อมูล จำลองการใช้ข้อค้นพบ

Step 13 เก็บรวบรวมข้อมูล

Step 14 วิเคราะห์แปลผลข้อค้นพบ

Step 15 รายงานผลการประเมิน

Step 16 ติดตามการใช้ข้อค้นพบจากการประเมิน

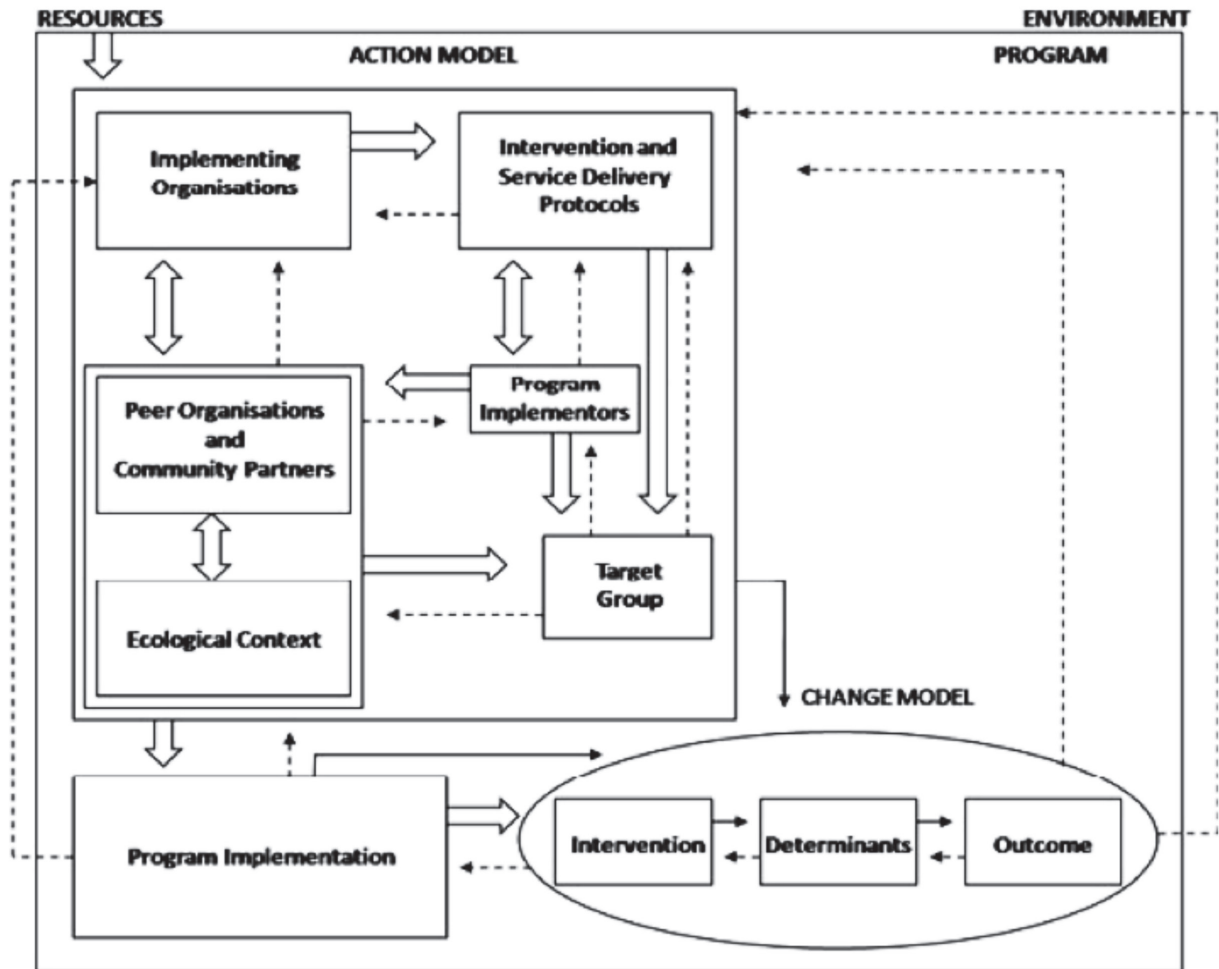
Step 17 การประเมินของการประเมิน (Meta-evaluation)

จุดเด่นของการประเมิน UFE คือ เน้นการใช้ประโยชน์จากการประเมินของ intended user โดยค้นหามติเอกฉันท์ในการหาคุณค่า (value) และการตัดสินคุณค่า (Judgments) ของสิ่งที่ประเมิน แต่อย่างไรก็ตาม UFE มีข้อจำกัดในเรื่องความลำเอียงและคอร์รัปชันในกลุ่ม intended users โดยถ้า intended users ไม่ได้นำเสนอประเด็นตามความสนใจของ stakeholder และถ้านักประเมินไม่มีความสามารถสูงพอในทักษะเจรจาต่อรองและมีทักษะในการประเมิน (evaluation competence) ที่เพียงพอไม่ควรใช้การประเมิน UFE

5. แนวคิดทางการประเมิน Theory - Driven Evaluation ของ Huey T. Chen

การประเมินโดยใช้แรงขับเคลื่อนทางทฤษฎีเป็นการออกแบบและประเมินโดยใช้ทฤษฎีเป็นกรอบแนวคิดในการประเมินเรียกว่า “Program theory” ซึ่งเป็นเซตข้อตกลงของ stakeholders เกี่ยวกับการกระทำอะไรที่จำเป็นในการแก้ปัญหาและทำไมปัญหาถึงตอบสนองต่อการกระทำนั้น จุดมุ่งหมายของการประเมินโดยใช้แรงขับเคลื่อนทางทฤษฎีไม่เพียงแต่ประเมินว่า intervention ใดเหมาะสมหรือไม่เหมาะสมแต่คำนึงถึงว่าทำไมจึงใช้ intervention และจะใช้อย่างไรที่ทำให้ได้สารสนเทศที่มีความจำเป็นสำหรับ stakeholders การประเมินโดยใช้แรงขับเคลื่อนทางทฤษฎีแตกต่างจาก Black-box Evaluation ซึ่งเป็นการประเมิน intervention ที่มีผลกระทบต่อ outcomes โดยไม่สนใจกระบวนการเปลี่ยนแปลง (transformation) ระหว่าง intervention และ outcomes หรือ Minimalist Theory นอกจากนี้การประเมินโดยใช้แรงขับเคลื่อนทางทฤษฎียังมีความแตกต่างจาก Method-driven Evaluation ซึ่งใช้วิธีการวิจัยเป็นฐานสำหรับการทำการประเมินโดยใช้วิธีเชิงปริมาณ (Quantitative), วิธีเชิงคุณภาพ (Qualitative), วิธีผสมผสาน (Mixed) นอกจากนี้ Method-driven Evaluation ไม่สนใจความคิดเห็นของ stakeholders ซึ่งเกี่ยวข้องในการประเมิน การประเมินโดยใช้แรงขับเคลื่อนทางทฤษฎีอยู่บนพื้นฐานรายละเอียดข้อตกลง

(Descriptive assumptions) ซึ่งเรียกว่า change model เกี่ยวข้องกับอะไรเป็นกระบวนการเชิงสาเหตุที่คาดหวังให้เกิดในการบรรลุเป้าหมายโปรแกรมประกอบไปด้วย 3 องค์ประกอบคือ 1) เป้าหมายและผลลัพธ์ (Goals and Outcomes) - เป้าหมายสะท้อนความต้องการเพื่อเติมเต็มความต้องการจำเป็นที่ไม่บรรลุได้แก่ สุขภาพเจ็บป่วย ความยากจน เป็นต้น ส่วนผลลัพธ์ต้องเป็นรูปธรรมสามารถวัดได้ตามเป้าหมาย, 2) Determinants – กลไกชั้นกลางซึ่งช่วย treatment/intervention ให้บรรลุความต้องการจำเป็น/ผลลัพธ์, 3) Intervention or Treatment – กิจกรรมในโปรแกรมมีผลต่อการเปลี่ยนแปลง Determinants นอกจาก change model แล้วการประเมินโดยใช้แรงขับเคลื่อนทางทฤษฎียังอยู่บนพื้นฐาน Prescriptive assumptions ซึ่งเรียกว่า action model เกี่ยวข้องกับอะไรเป็นกิจกรรมที่ต้องทำในโปรแกรมเพื่อสร้างความเปลี่ยนแปลงที่ต้องการโดยมีการวางแผนอย่างเป็นระบบดังนี้ Implementing Organization – การดำเนินการขององค์กร, Program Implementers – สมรรถนะของบุคลากรในองค์กร, Peer Organizations/Community Partners - ความร่วมมือระหว่างองค์กร, Intervention and Service Delivery Protocols - เนื้อหาและกิจกรรมรายละเอียดขั้นตอนปฏิบัติ, Ecological Context – บริบททางนิเวศวิทยา, Target Population – ประชากรเป้าหมาย ความสัมพันธ์เชื่อมโยงดังภาพ 2



ภาพ 2 กรอบแนวคิดของ Program Theory ทาง การประเมินผล

ที่มา. จาก “Theory-driven evaluation: Conceptual framework, application and advancement,” 2012 โดย Huey T. Chen, ค้นจาก www.springer.com

จุดเด่นของการประเมิน Theory - Driven Evaluation คือ การใช้ทฤษฎีเป็นแรงขับเคลื่อนในการออกแบบและเป็นกลไกในการประเมิน

6. แนวคิดทางการประเมิน Four Generation Evaluation Theory ของ Guba & Lincoln

Guba และ Lincoln ได้แบ่งยุคของการประเมินออกเป็น 4 ยุคดังนี้ ยุคแรก First Generation Evaluation: ยุควัดผล (Measurement) ยุคนี้เริ่มประมาณต้นปี ค.ศ. 1900 การประเมินมุ่งในเรื่องของการวัดผลเป็นหลัก ยุคที่ 2 Second Generation Evaluation: ยุคบรรยาย (Description) มุ่งการบรรยายจุดแข็ง จุดอ่อนตามวัตถุประสงค์เป็นหลัก นักทฤษฎีการประเมินคนสำคัญ

ในยุคนี้ได้แก่ Tyler ยุคที่ 3 Third Generation Evaluation: ยุคตัดสิน (judgment) มุ่งการตัดสินคุณค่าของสิ่งที่ประเมิน นักทฤษฎีการประเมินคนสำคัญในยุคนี้ได้แก่ Scriven, Eisner จาก 3 ยุคที่ผ่านมาของการประเมินพบว่าการประเมินยังไม่สามารถรวบรวมคุณค่าในเชิงพหุมิติ จึงเกิดยุคที่ 4 คือ Fourth Generation Evaluation: ยุคการตอบสนองและสร้างสรรค์ (responsive constructivist) โดยมุ่งการประเมินอย่างสร้างสรรค์ และตอบสนองต่อความต้องการของ stakeholders โดยการประเมินต้องอาศัยความร่วมมือจากทุกฝ่ายเพื่อตอบสนองต่อ stakeholders การประเมินประกอบด้วย Claims, Concerns, Issues – CCI การประเมินมี 4 ระยะ (Moreau, A. K., & Clarkin, L. C., 2012) ดังนี้

ระยะที่ 1 การระบุ stakeholders และดึงข้อมูล CCI โดย

- C : Claims – ให้ stakeholder เสนอและยืนยันสิ่งที่พึงประสงค์ของโครงการ
- C : Concerns – ให้ stakeholder เสนอสิ่งไม่พึงประสงค์ของโครงการ
- I : Issues - ค้นหาประเด็นเหตุผลของความไม่สอดคล้อง

ระยะที่ 2 การเปิดโอกาสให้ stakeholders แลกเปลี่ยนกันและได้ข้อมูลใหม่ ๆ

ระยะที่ 3 นักประเมินรวบรวมข้อมูลเพื่อเตรียมการสนทนาระหว่างกลุ่ม stakeholder เกี่ยวกับ CCI

ระยะที่ 4 นักประเมินปล่อยให้กลุ่ม stakeholder อภิปรายตามข้อมูลในระยะ 3 เพื่อหามติเอกฉันท์

ในการประเมินแบบ Four Generation Evaluation Theory มีจุดเด่นคือ (Claims, Concerns, Issues--CCI) ผลลัพธ์ที่ได้จะเป็นข้อยุติสมบูรณ์และเป็นมติเอกฉันท์ อันนำไปสู่การปฏิบัติต่อไป

สรุป

ยุคสมัยของการประเมินผลในระยะแรก (ก่อนปี ค.ศ. 1900) การประเมินผลเน้นการวัดผลเป็นหลัก จนปี ค.ศ. 1930 Tyler แสดงแนวคิดทางการประเมินผลแบบอิงวัตถุประสงค์โดยการประเมินผลลัพธ์ของโครงการต้องพิจารณาตามวัตถุประสงค์ของโครงการ (Goal-based evaluation) Tyler จึงเป็นบิดาของการ

ประเมินผล ในยุคที่ห้าของการประเมินผลมีนักทฤษฎีทางการประเมินผลเกิดขึ้นหลายท่านซึ่งมีแนวคิดทฤษฎีทางการประเมินผลต่าง ๆ ได้แก่ Scriven แสดงแนวคิดทางการประเมินมุ่งการตัดสินคุณค่าของสิ่งที่ประเมิน และการประเมินผลควรประเมินตามผลลัพธ์ที่เกิดขึ้นจริง ไม่จำเป็นต้องอิงวัตถุประสงค์ของโครงการเพียงอย่างเดียว (Goal-free evaluation) อีกทั้งการประเมินผลควรประเมินทั้งในขณะที่โครงการกำลังดำเนินอยู่ (Formative evaluation) และเมื่อโครงการสิ้นสุดลงแล้ว (Summative evaluation), Stufflebeam มีแนวคิดทางการประเมินผลโดยใช้ CIPP Model, Robert Stake แสดงแนวคิดทางการประเมินแบบตอบสนองต่อ stakeholders, Michael Patton แสดงแนวคิดทางการประเมินโดยเน้นการใช้ประโยชน์ของ stakeholders, Huey T. Chen แสดงแนวคิดทางการประเมินโดยใช้ทฤษฎีเป็นแรงขับเคลื่อน และ Guba & Lincoln แสดงแนวคิดทางการประเมินแบบ Four Generation Evaluation Theory ดังได้กล่าวถึงในบทความ ในการนำแนวคิดของนักทฤษฎีทางการประเมินผลไปใช้ในโครงการควรพิจารณาถึงความเหมาะสมของบริบทและจุดเน้นในการประเมินผลของแต่ละโครงการเพื่อให้การประเมินผลมีความเป็นไปได้และเกิดประสิทธิภาพ ประสิทธิผลสูงสุด ตลอดจนเกิดผลสำเร็จของโครงการ อย่างไรก็ตาม พัฒนาการทางด้านการประเมินผลยังคงมีอย่างต่อเนื่อง มีแนวคิดทฤษฎีใหม่ ๆ ทางทางการประเมินผลเกิดขึ้น รวมถึงวิธีการใหม่ ๆ ทางทางการประเมินผลที่น่าสนใจติดตาม ทั้งนี้เพื่อให้การประเมินผลสามารถทำให้องค์กรมีการพัฒนาที่ยั่งยืนต่อไป

เอกสารอ้างอิง

- Alkin, M. C., & Carden, F. (2012). Evaluation Roots: An International Perspective. *Journal of MultiDisciplinary Evaluation*, 8(17), 102-118.
- Chen, H. T. (2012). *Theory-driven evaluation: Conceptual framework, application and advancement*. Retrieved from <http://www.springer.com>.

- Hogan, L. R. (2007). The Historical Development of Program Evaluation: Exploring The Past and Present. *Online Journal of Workforce Education and Development*, 2(4), (Fall). Retrieved from <http://opensiuc.lib.siu.edu>
- Madaus, G. F., & Stufflebeam, D. L. (2000). *Program Evaluation: A Historical Overview*. Retrieved from <http://link.springer.com>.
- Moreau, A. K., & Clarkin, L. C. (2012). Thinking beyond measurement, description and judgment: Fourth generation evaluation in family-centered pediatric healthcare organizations. *Journal of MultiDisciplinary Evaluation*, 8(17), 33-45.
- Patton, M.Q. (2003). *Qualitative evaluation checklist*. Retrieved from <http://www.wmich.edu>
- Patton, M.Q. (2013). *Utilization-focused evaluation (U-FE) checklist*. Retrieved from <http://www.wmich.edu>
- Scriven, M. (2011). *Evaluating evaluations: A meta-evaluation checklist*. Retrieved from <http://michaelscriven.info>
- Scriven, M. (2001). Evaluation: Future tense. *American Journal of Evaluation*, 22(3), 301-307.
- Scriven, M. (1998). Minimalist theory: The least theory that practice requires. *American Journal of Evaluation*, 19(1), 57-70.
- Scriven, M. (2007). *The logic of evaluation*. Retrieved from <http://www.coris.uniroma1.it>
- Stake, R. E., & Abma, T. A. (2001). Stake's responsive evaluation: Core ideas and evolution. *New Directions for Evaluation*, 92, (Winter), 7-21.
- Stufflebeam, D. L. (2007). *CIPP evaluation model checklist: A tool for applying the CIPP Model to assess long-term enterprises*. Retrieved from <http://www.wmich.edu>.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, & applications*. San Francisco, CA., John Wiley & Sons.
- Youker, B. W. (2010). *Goal-free evaluation and goal-based evaluation two program evaluation approaches one synthesized report*. Retrieved from <http://www.wmich.edu>

