

## บทที่ 1

### ประวัติและพัฒนาการของการประเมิน

#### ประวัติและพัฒนาการของการประเมินในต่างประเทศ

ประวัติและพัฒนาการของการประเมินในต่างประเทศ จะใช้การแบ่งยุคตามที่เสนอโดย Madaus และคณะ และการเสนอโดย Guba และ Lincoln Madaus และคณะ (Madaus et al. 1983) ได้แบ่งยุคการประเมินออกเป็น 7 ยุค ดังนี้

ยุคที่ 1 ยุคก่อนการปฏิวัติอุตสาหกรรม (เริ่มตั้งแต่มีการบันทึกจนถึง ค.ศ. 1800) ซึ่งมีเหตุการณ์สำคัญ คือ 2000 ปีก่อนคริสต์ศักราช เจ้าหน้าที่ของจีนได้สอบคัดเลือกคนเข้ารับราชการ โดยใช้ข้อสอบเรียงความ 5 ปีก่อนคริสต์ศักราช Socrates และครูชาวกรีก ได้ใช้การประเมิน โดยตั้งคำถามซึ่งรู้จักกันต่อมา คือ การสอบปากเปล่านั้นเอง และในช่วงปี ค.ศ. 1600-1800 เป็นช่วงที่วิทยาศาสตร์เข้ามามีบทบาทในความคิดและปฏิบัติของมนุษย์มากขึ้น อันเป็นพื้นฐานของการปฏิวัติอุตสาหกรรมในทวีปยุโรปต่อมา

ยุคที่ 2 ยุคการปฏิรูป (The Age of Reform: 1800-1900) เป็นยุคแห่งการปฏิวัติอุตสาหกรรมในทวีปยุโรป มีการปรับปรุงเปลี่ยนแปลงครั้งใหญ่ทั้งด้านเศรษฐกิจ เทคโนโลยี สังคม การเมือง และการศึกษา มีการปฏิรูปโครงสร้างทางสังคม และโปรแกรมบริการทางการศึกษาและสุขภาพอนามัย โดยเฉพาะอย่างยิ่งในสหราชอาณาจักรอังกฤษและสหรัฐอเมริกา ในสหราชอาณาจักรอังกฤษ มีความพยายามที่จะปฏิรูปสถาบันทางสังคมต่าง ๆ รวมทั้งโรงพยาบาล สถานเลี้ยงเด็กกำพร้าและโรงเรียน โดยทางราชสำนักและรัฐบาลกลาง ได้แต่งตั้งคณะกรรมการขึ้นทำหน้าที่ประเมินประสิทธิภาพและการปฏิบัติของสถาบันทางสังคมเหล่านี้

ในการวิจัยทางการศึกษา โดยเฉพาะในด้านการทดลองของ Wundt นักจิตวิทยาชาวเยอรมัน ได้ทำการสังเกตและบันทึกพฤติกรรมของมนุษย์อย่างเป็นระบบ นักวิชาการที่ร่วมงานกับ Wundt ได้แก่ Cattell, Hall, Judd และ Meumann ซึ่งล้วนเป็นนักวิชาการที่มีชื่อเสียงในศตวรรษที่ 20

ค.ศ. 1845 เริ่มมีการประเมินผลงานเพื่อประเมินประสิทธิภาพของโรงเรียนต่าง ๆ ที่เมืองบอสตัน ประเทศสหรัฐอเมริกา ที่เรียกว่า Boston Survey โดย Horace Mann และ Samuel Howe ได้ริเริ่มการใช้คะแนนสอบข้อเขียนของนักเรียนในการประเมินประสิทธิภาพ

ของการสอนและประสิทธิภาพของโรงเรียน มีความพยายามที่จะใช้ข้อมูลที่เป็นระบบ คือ ใช้คะแนนสอบข้อเขียนของนักเรียนแทนการสอบปากเปล่า เพื่อใช้ในการพิจารณาแต่งตั้งครูใหญ่ Mann และ Howe จึงนับว่าเป็นผู้บุกเบิกแสวงหาวิธีในการประมวลผลและวิเคราะห์ข้อมูลด้านผลสัมฤทธิ์ทางการเรียนของนักเรียน เพื่อใช้ในการพิจารณาแต่งตั้งครูใหญ่ อีกทั้งได้พยายามสร้างแบบสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนอย่างเป็นมาตรฐาน

ค.ศ. 1884 นักชีววิทยาชาวอังกฤษ คือ Galton ได้สร้างเครื่องมือวัดความสามารถในการได้ยินเสียง จำแนกเสียง และจำแนกระยะทาง Galton เป็นบุคคลแรกที่ริเริ่มการวัดโดยใช้แบบสอบถาม และมาตราส่วนประมาณค่า และเป็นบุคคลแรกที่คิดคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์

ค.ศ. 1887-1898 Joseph M. Rice เป็นนักปฏิรูปการศึกษาชาวอเมริกัน ซึ่งถือว่าเป็นบิดาของการวิจัยทางการศึกษา และเป็นบุคคลแรกที่เป็นผู้ประเมินโครงการทางการศึกษาอย่างเป็นทางการ Rice ได้สร้างแบบสอบการสะกดคำแล้วนำไปทดสอบเด็ก 33,000 คน ในโรงเรียนต่าง ๆ ทั่วประเทศสหรัฐอเมริกาตอนใต้และทางตะวันออก ผลการศึกษาพบว่าโรงเรียนที่ใช้เวลาสอนฝึกสะกดคำแตกต่างกัน แต่นักเรียนกลับมีผลสัมฤทธิ์ทางการเรียนด้านการสะกดคำไม่แตกต่างกัน ซึ่งแสดงว่าการสอนฝึกสะกดคำของโรงเรียนไม่มีผลต่อการเรียนรู้ของนักเรียน หรือกล่าวอีกนัยหนึ่งก็คือ โรงเรียนใช้เวลาสอนอย่างไม่มีประสิทธิภาพ วิธีสอนการสะกดคำไม่มีคุณภาพ ผลการประเมินของ Rice ในเรื่องการสะกดคำนี้ นำไปสู่การปฏิรูปวิธีการสอน จากผลการศึกษาเรื่องความสามารถในการสะกดคำของเด็กอเมริกันจำนวนมากดังกล่าว Rice จึงได้ชื่อว่าเป็นบุคคลแรกที่พัฒนาการประเมินผลทางการศึกษาที่มีรูปแบบและเป็นระบบในประเทศสหรัฐอเมริกา โดยเฉพาะรูปแบบการทดลองเปรียบเทียบผลสัมฤทธิ์ทางการเรียนในการวิจัยทางการเรียนในการวิจัยทางการศึกษา และรูปแบบการวัดแบบปรนัยในการประเมินทางการศึกษา

ค.ศ. 1890 Cattell เป็นผู้ริเริ่มใช้คำว่า การทดสอบสมอง (mental test) ในประเทศสหรัฐอเมริกา

ในยุคที่ 2 นี้การประเมินโรงเรียน และสถาบันทางสังคมต่าง ๆ ได้เกิดประเด็นที่นักประเมินจะต้องพิจารณามากมาย เช่น ธรรมชาติและความน่าเชื่อถือของหลักฐานที่ใช้ในการประเมินความเหมาะสมของแบบประเมินที่ใช้ในการประเมินโปรแกรมขณะกำลังดำเนินอยู่ในสภาพธรรมชาติ บทบาทของนักประเมินที่ค่อนข้างขัดแย้งกัน คือ ประเมินเพื่อปรับปรุงพัฒนากับบทบาทของนักประเมินในการประเมินเพื่อรักษามาตรฐานโดยการตัดสินคุณค่าของผลลัพธ์ต่าง ๆ และการแจ้งผลประเมินต่อสาธารณะอย่างผู้มีความรับผิดชอบ เป็นต้น

ยุคที่ 3 ยุคแห่งประสิทธิภาพและการทดสอบ (The Age of Efficiency and Testing: 1900-1930) ในช่วงแรกของศตวรรษที่ 20 ก่อให้เกิดแนวคิดเกี่ยวกับการบริหารจัดการเชิงวิทยาศาสตร์ (scientific management) ขึ้นในประเทศสหรัฐอเมริกา ซึ่งต่อมาเรียกกันว่า ความเคลื่อนไหวในการสำรวจโรงเรียน (school survey movement) ซึ่งโรงเรียนเป็นจำนวนมากได้ทำการศึกษาสำรวจการทำงานของตนเอง เพื่อนำไปสู่การจัดระบบการดำเนินงาน (systematization) การสร้างความเป็นมาตรฐาน (standardization) และประสิทธิภาพของการดำเนินงานของโรงเรียนและองค์กร (school or organizational efficiency) ในการสำรวจโรงเรียนครั้งนี้มีการใช้ข้อสอบแบบใหม่คือ ข้อสอบแบบปรนัยเป็นจำนวนมาก แทนที่จะใช้ข้อสอบเรียงความเพียงไม่กี่ข้อเช่นในอดีต

ค.ศ. 1911 มีการพิมพ์ผลการวิจัยทดลองทางการศึกษาที่แสดงถึงความก้าวหน้าของการวิจัยทางการศึกษาในทวีปยุโรป โดยเฉพาะในประเทศสวีเดน

ในยุคนี้เป็นยุคเริ่มการประเมินผลสัมฤทธิ์ทางการเรียนแบบอิงกลุ่ม เช่น ที่บอสตัน มีการเปรียบเทียบประสิทธิภาพการสอนของครู โดยพิจารณาจากค่าเฉลี่ยของคำร้อยละที่นักเรียนสอบผ่าน ของแต่ละเมืองเป็นเกณฑ์ เป็นต้น กล่าวได้ว่า ยุคนี้เป็นยุคของความเคลื่อนไหวเกี่ยวกับการทดสอบ วิทยาการทางการทดสอบก้าวหน้ามาก มีแบบสอบแบบปรนัยเกิดขึ้น ในช่วงนี้การประเมินทางการศึกษาขึ้นอยู่กับความสนใจของท้องถิ่นกับความเชี่ยวชาญของโรงเรียน รัฐบาลส่วนกลางไม่สนใจเรื่องการประเมิน นับว่าเป็นยุคที่เฟื่องฟูทางการทดสอบแต่ขาดการสนับสนุนและผลักดันจากหน่วยงานภายนอก ในยุคนี้นักทดสอบที่มีชื่อเสียงเกิดขึ้นในท้องถิ่นต่าง ๆ เช่น Curtis, Ayres, Thorndike เป็นต้น

โดยสรุป ยุคที่ 3 นี้ ถือเป็นยุคทองของการวัดผลการศึกษา เป็นการทดสอบเพื่อแสวงหาประสิทธิภาพ ระบบ และความเป็นมาตรฐาน สามารถแบ่งเป็น 2 ช่วง คือ ค.ศ. 1900-1915 ถือเป็นระยะการบุกเบิกของการวัดผลโดยการสำรวจและพัฒนาวิธีการวัดผล Thorndike เป็นนักวัดผลที่สำคัญของโลก โดยได้เขียนหนังสือ “การวัดผลการศึกษา” เป็นเล่มแรกของโลก ตีพิมพ์ในปี ค.ศ. 1905 ซึ่งเป็นการวางรากฐานของการพัฒนาแบบสอบมาตรฐาน Stone เป็นคนแรกที่พัฒนาแบบสอบมาตรฐาน ซึ่งเป็นแบบสอบวิชาเลขคณิต ใน ค.ศ. 1905 Binet ซึ่งเป็นนักจิตวิทยาชาวฝรั่งเศส ร่วมกับ Simon สร้างแบบสอบเชาวน์ปัญญาเป็นฉบับแรก เรียกว่า Binet Simon Scale ในช่วงที่สองประมาณ ค.ศ. 1915-1930 ซึ่งเป็นระยะการขยายตัวของแบบสอบมาตรฐานและแบบสอบแบบปรนัย เริ่มมีการใช้ทดสอบโดยชุดแบบสอบ (battery) และเริ่มพัฒนาแบบวัดคุณลักษณะด้านอื่น ๆ มีการใช้แบบสอบตามและแบบสำรวจต่าง ๆ

มากขึ้น ค.ศ. 1920 McCall เป็นบุคคลแรกที่แนะนำให้ครูสร้างแบบสอบขึ้นใช้เอง (teacher-made test) แบบสอบมาตรฐานที่มีชื่อเสียงมากในระยะนี้ คือ แบบสอบ Army Alfa และ Army Beta ซึ่งสร้างขึ้นในปี ค.ศ. 1917

ในช่วงปลายของยุคแห่งประสิทธิภาพและการทดสอบนี้ มหาวิทยาลัยหลายแห่งที่มีความเชี่ยวชาญด้านการทดสอบได้ทำงานร่วมกับโรงเรียนเพื่อสำรวจและทดสอบประสิทธิภาพของโรงเรียน ซึ่งนับว่าเป็นจุดเริ่มต้นของศูนย์การประเมินในมหาวิทยาลัยต่าง ๆ ผู้ที่เป็นผู้นำบุกเบิกเรื่องนี้ คือ George Strayer แห่ง Teachers College, Columbia University เอกลักษณ์ของยุคนี้ คือ ยุคแห่งการทดสอบมาตรฐาน จะเห็นได้จากหลังสงครามโลกครั้งที่ 1 (ค.ศ. 1914-1918) โรงเรียนจะประเมินประสิทธิภาพของการเรียนการสอนและหลักสูตรโดยใช้แบบสอบมาตรฐานทำให้การสร้างแบบสอบมาตรฐานกลายเป็นธุรกิจทางการศึกษาในเมืองฟิลาเดลเฟีย ประเทศสหรัฐอเมริกา (ค.ศ. 1916-1938) การประเมินผลทางการศึกษาโดยแบบสอบมาตรฐานมุ่งตอบสนองเฉพาะความต้องการของท้องถิ่นหรือโรงเรียน ไม่มีแรงจูงใจที่จะทำให้ผลการประเมินมีความเป็นสากล

ยุคที่ 4 ยุคไทเลอร์ (The Tylerian Age: 1930-1945) หรือยุคการประเมินตามวัตถุประสงค์ ในตอนต้นของยุคนี้เกิดเศรษฐกิจตกต่ำมาก ทำให้นักสังคมศาสตร์มีบทบาทในการกำหนดนโยบายต่าง ๆ เพื่อแก้ไขปัญหา ตลอดจนมีความเคลื่อนไหวให้ปฏิรูประบบสังคมและการศึกษา เพื่อแก้ปัญหาทางเศรษฐกิจ John Dewey และ Ralph Tyler เป็นบุคคลสำคัญทางการศึกษาของสหรัฐอเมริกาในการร่วมขบวนการเคลื่อนไหวปฏิรูปการศึกษาดังกล่าว John Dewey ได้เป็นผู้นำทำให้เกิดการศึกษาแบบก้าวหน้า (Progressive Education) และเกิดหลักการศึกษานักการศึกษาทั่วโลกรู้จักกันดี คือ การเรียนรู้จากการปฏิบัติ (learning by doing) ส่วน Tyler ได้รับแต่งตั้งให้เป็นผู้อำนวยการโครงการ 8 ปี (ค.ศ. 1932-1940) ซึ่งเป็นโครงการวิจัยการศึกษาขนาดใหญ่ ที่ศึกษาเปรียบเทียบประสิทธิภาพการสอนแบบดั้งเดิมกับการสอนแนวใหม่ ในระดับมัธยมศึกษาของประเทศสหรัฐอเมริกา Tyler ได้วางแนวคิดในการประเมินหลักสูตรและโปรแกรมการศึกษา โดยการเปรียบเทียบผลที่เกิดจริงกับผลสัมฤทธิ์ที่คาดหวังตามวัตถุประสงค์เชิงพฤติกรรม โดยที่ Tyler เป็นบุคคลแรกที่นิยามการประเมินผลการศึกษาว่าเป็นการวัดคุณค่าของผลสำเร็จตามวัตถุประสงค์ของการเรียนการสอน แนวคิดของ Tyler ทำให้เห็นความแตกต่างระหว่าง “การวัด” และ “การประเมิน” อย่างเด่นชัด อันเป็นปัจจัยพื้นฐานที่สำคัญที่ทำให้ศาสตร์แห่งการประเมินได้รับการพัฒนาอย่างเป็นวิชาชีพ (professional) เช่นในปัจจุบัน ในบทความของ Tyler ได้ระบุอย่างเด่นชัดว่า ในขั้นตอน

ขอเพียงให้ครูได้สามารถประเมินได้ว่าการสอนของตนบรรลุผลหรือไม่โดยการเปรียบเทียบผลสัมฤทธิ์ทางการเรียนกับวัตถุประสงค์ของการสอน ซึ่งมีลักษณะการประเมินที่เน้นผลลัพธ์ (outcome) มากกว่าตัวป้อน (input) และกระบวนการ (process) กล่าวคือ เป็นการเน้นผลมากกว่าเหตุ ซึ่งก็เหมาะสมกับยุคสมัย ซึ่งก่อนยุคนี้นักวิชาการและครูก็ยังแยกไม่ออกระหว่างการวัดและการประเมิน อย่างไรก็ตาม มีนักวิชาการบางคนที่ไม่ได้อ่านบทความของ Tyler ก็วิพากษ์วิจารณ์ว่า Tyler ไม่สนใจประเมินเหตุทำให้การประเมินไม่มีผลต่อการปรับปรุงการเรียนการสอนเท่าที่ควรซึ่งเป็นความเข้าใจผิดของผู้วิจารณ์เอง ในปี ค.ศ. 1986 Tyler ได้เขียนบทความต่อจากบทความปี ค.ศ. 1932 โดยเสนอการประเมินทางการศึกษาโดยใช้แนวคิดเชิงระบบ ซึ่งมีการประเมินทั้งตัวป้อน กระบวนการ และผลลัพธ์ครบถ้วน

หลักการและนियามการประเมินผลการศึกษาที่เสนอโดย Tyler นับว่าเหมาะสมกับยุคสมัยนั้นเป็นอย่างยิ่ง เป็นที่ยอมรับกันอย่างกว้างขวางว่า หลักสูตรและการเรียนการสอนจะดีหรือไม่จะต้องประเมินที่การเรียนรู้ของศิษย์ ดังคำกล่าวที่ว่า “ค่าของครูให้ดูที่ศิษย์” เป็นต้น นอกจากนี้นियามการประเมินทางการศึกษา หลักการและรูปแบบการประเมินที่เสนอโดย Tyler นับว่าเป็นรากฐานและอิทธิพลต่อแนวความคิด และยังผลให้เกิดการพัฒนาศาสตร์แห่งการประเมินทางการศึกษาต่อมา Tyler จึงได้รับการยกย่องว่าเป็นบิดาของการประเมินทางการศึกษา และเป็นผู้ก่อให้เกิดนวัตกรรมของการเขียนวัตถุประสงค์ที่ชัดเจน หรือวัตถุประสงค์เชิงพฤติกรรม (Behavioral objective)

ยุคที่ 5 ยุคแห่งความไร้เดียงสาทางการประเมิน (The Age of Ignorance: 1946-1957) หรือยุคแห่งการเพิกเฉยทางการประเมิน (The Age of Ignorance) ทั้งนี้เป็นเพราะยุคนี้เป็นยุคฟื้นฟูประเทศ เนื่องจากภัยสงครามโลกครั้งที่ 2 ซึ่งเกิดในระหว่าง ค.ศ. 1939-1945 ทุกคนทุกฝ่ายก็ต้องการลืมความทุกข์ทรมานจากสงครามโลกครั้งที่ 2 ทั้งความทุกข์ไว้ในอดีต มีความพยายามแสวงหาทรัพยากร สร้างและขยายกิจกรรมและบริการต่าง ๆ แสวงหาความสุขความบันเทิงในชีวิต ในประเทศสหรัฐอเมริกาเร่งฟื้นฟูประเทศโดยการก่อสร้างอาคาร จัดตั้งสถาบันการศึกษาเพื่อให้มีนิสิตนักศึกษาและครูเพิ่มขึ้น มีการขยายการศึกษาอย่างกว้างขวาง ทั้งโปรแกรมการศึกษาและบุคลากรทางการศึกษา การประเมินมีบ้างก็ทำไป เพื่อเอื้อประโยชน์ต่อการสานงานต่อและก่องานใหม่มากกว่าเป็นการประเมินและใช้ผลประโยชน์ตามคุณค่าที่แท้จริง ดังนั้น การประเมินในยุคนี้จึงมิได้มุ่งเพื่อปรับปรุงพัฒนาตามปรัชญาของการประเมิน ข้อสังเกตก็คือ นิสิตนักศึกษาและผู้คนในยุคนี้เกือบไม่มีใครสนใจ และรับผิดชอบต่อปัญหาสังคม ส่วนใหญ่มุ่งแก้ปัญหาและสนองความต้องการของตนเองเป็นหลัก

การประเมินโดยทั่วไปในยุคนี้เป็นความพยายามของนักการศึกษาท้องถิ่น เช่น Tyler, Lindquist และคนอื่น ๆ ได้ร่วมกันจัดตั้งศูนย์บริการทดสอบทางการศึกษา (Educational Testing Service) ซึ่งระยะต่อมาสถาบันแห่งนี้ได้สร้างความเจริญให้วงการทดสอบและประเมินผลเป็นอย่างยิ่ง

การประเมินทางการศึกษาในช่วงนี้ แนวคิดการประเมินของ Tyler ได้รับการเผยแพร่ไปอย่างกว้างขวาง แต่มีปัญหาในการปฏิบัติ คือ การเขียนวัตถุประสงค์เชิงพฤติกรรมตามลำดับขั้นของการเรียนรู้ เพื่อช่วยแก้ไขปัญหาดังกล่าว Bloom และคณะ (ค.ศ. 1956) และ Krathwohl (ค.ศ. 1964) จึงได้สร้างสารบบของวัตถุประสงค์การศึกษาขึ้น (taxonomies of educational objectives) ซึ่งทำให้การประเมินทางการศึกษาตามแนวของ Tyler ชัดเจนขึ้น ยิ่งไปกว่านั้น ทำให้วงการศึกษาและการฝึกหัดครูได้มีความชัดเจนในการกำหนดผลการศึกษาทั้งในระดับชาติ ระดับหลักสูตร และระดับบทเรียนด้วย

โดยสรุป เอกลักษณะทางการประเมินในยุคนี้ก็คือ การประเมินทางการศึกษาที่ตอบสนองความต้องการของท้องถิ่น โดยประเมินตามแนวของ Tyler ซึ่งเน้นประเมินตามวัตถุประสงค์ ค.ศ. 1949 Tyler เขียนหนังสือที่มีอิทธิพลอย่างยิ่งต่อวงการการศึกษาชื่อ การพัฒนาหลักสูตรและการประเมินผล ค.ศ. 1950 Gulliksen ได้เขียนหนังสือทฤษฎีการวัด ค.ศ. 1953 Lindquist ได้เขียนหนังสือหลักการออกแบบทดลอง และใน ค.ศ. 1956 Bloom และคณะ ได้เขียนหนังสือเกี่ยวกับระบบจำแนกวัตถุประสงค์ทางการศึกษาด้านพุทธิพิสัย ซึ่งผลงานเหล่านี้ล้วนมีคุณประโยชน์อย่างยิ่งต่อวงการวัดและประเมินผลการศึกษา และวงการศึกษาโดยรวม

ยุคที่ 6 ยุคการขยายองค์ความรู้ทางการประเมิน (The Age of Expansion: 1958-1972) เนื่องจากสหภาพโซเวียตรัสเซียประสบความสำเร็จในการส่งดาวเทียมสปุตนิก 1 (Sputnik 1) ขึ้นสู่อวกาศเป็นประเทศแรกของโลก ทำให้ประเทศสหรัฐอเมริกาตื่นตระหนกรู้สึกว่าประเทศของตนล้าหลัง สหภาพโซเวียตรัสเซียจึงริบเร่งปรับปรุงการศึกษาเป็นการใหญ่ โดยเน้นการศึกษาด้านวิทยาศาสตร์ คณิตศาสตร์ ภาษาต่างประเทศ และการแนะแนว โดยปรับปรุงหลักสูตรและการเรียนการสอนทั้งโดยรัฐบาลและในแต่ละรัฐ โดยการลงทุนมหาศาลริเริ่มมุ่งสร้างให้อเมริกันชนเป็นผู้รอบรู้ทางวิทยาศาสตร์และเทคโนโลยีที่ทันสมัย เพื่อให้ประเทศสหรัฐอเมริกาเป็นประเทศที่ยิ่งใหญ่ทันสมัยอีกครั้งหนึ่ง การส่งเสริมให้มีการพัฒนาหลักสูตรใหม่ ๆ และการลงทุนโครงการพัฒนาต่าง ๆ มากมาย ย่อมเป็นแรงผลักดันให้มีการประเมินผลอย่างเป็นระบบ ทุกโครงการต้องมีการประเมินโครงการประจำปีและต้องมีรายงาน ด้วยความต้องการดังกล่าวทำให้การประเมินเจริญขึ้นในฐานะเป็นศาสตร์สาขาที่เป็น

ความเชี่ยวชาญชำนาญการ โดยเฉพาะเกิดการยอมรับศาสตร์ว่าด้วยการประเมินโปรแกรมและโครงการ อีกทั้งเห็นประโยชน์และคุณค่าของการประเมินโครงการและโปรแกรมอย่างกว้างขวาง ซึ่งไม่เคยมีบรรยากาศเช่นนี้มาก่อนในประวัติศาสตร์ของการวัดและประเมินทางการศึกษา

ในขณะที่มีการพัฒนากระบวนการประเมินโปรแกรม และจากกระบวนการพัฒนากระบวนการประเมินก็นำสู่ปฏิบัติ ทำให้พบว่าโมเดลและวิธีการที่มีอยู่มีจุดอ่อนและข้อจำกัดมากมาย เช่น การออกแบบทดลองไม่สามารถใช้ในการประเมินในสภาพธรรมชาติหรือการประเมินในสนามได้ แบบสอบถามมาตรฐานมีความหนัก กล่าวนคือ โมเดลที่จะวัดผลของหลักสูตรใหม่ในสถานการณ์เฉพาะ และที่สำคัญก็คือ การประเมินที่มุ่งตอบคำถามเฉพาะ และความต้องการเฉพาะของท้องถิ่นให้ผลประโยชน์ที่มีข้อจำกัด ไม่สามารถใช้ตัดสินคุณค่าหรือประสิทธิผลของโปรแกรมระดับชาติได้

ด้วยอุปสงค์การประเมินที่มีคุณภาพ และอุปทานความรู้เกี่ยวกับการประเมินที่มีจุดอ่อนและข้อจำกัดต่าง ๆ ซึ่งไม่สามารถตอบสนองความต้องการได้ ทำให้เกิดความเร่งด่วนที่ต้องขยายองค์ความรู้ด้านการประเมินอย่างรีบเร่งตามความต้องการเร่งด่วน ที่จะได้กระบวนการประเมิน หรือโมเดลการประเมินที่มากกว่าการวัดด้วยข้อสอบปรนัย แล้วนำผลที่วัดได้ไปเปรียบเทียบกับวัตถุประสงค์ที่กำหนด ในการตอบสนองความต้องการทางการประเมินดังกล่าว นักวิชาการก็ได้เสนอแนวคิดและโมเดลการประเมินมากมาย เช่น Cronbach (1963), Scriven (1967), Stake (1967), Tyler (1969), Stufflebeam et al (1971), Provus (1971), Caro (1971), Weiss (1972), Worthen and Sanders (1973) เป็นต้น นอกจากนี้สมาคมต่าง ๆ ก็มีบทบาทในการพัฒนาทฤษฎี และสร้างนักประเมิน เช่น สมาคมเกียรตินิยมทางการศึกษา (Phi Delta Kappa) โดย Stufflebeam et al ได้พัฒนาหลักการและโมเดล CIPP ในการประเมินโปรแกรมและสมาคมเกียรตินิยมทางการศึกษายังจัดอบรมผู้สนใจที่จะเป็นนักประเมินด้วย

ในปี ค.ศ. 1966 Campbell and Stanley ได้เขียนหนังสือชื่อการออกแบบทดลองและกึ่งทดลองเพื่อการวิจัย เพื่อชี้ให้เห็นว่าการวิจัยทดลองที่เข้มงวดเป็นรูปแบบที่มีข้อจำกัดมากมาย และเรียกร้องให้สนใจอย่างจริงจังในการนำแบบทดลองไปใช้ในการวิจัยหรือการประเมินในสภาพการณ์จริง

โดยสรุป จึงกล่าวได้ว่า ยุคนี้เป็นยุครุ่งอรุณของการประเมินที่จะก้าวต่อไปในฐานะศาสตร์แห่งวิชาชีพ เพื่อการพัฒนาคน พัฒนางาน พัฒนาองค์การและสภาพสังคมเพื่อมนุษยชาติ

ยุคที่ 7 ยุคแห่งการเป็นวิชาชีพของการประเมิน (The Age of Professionalization: 1973-ปัจจุบัน) ตั้งแต่ปี ค.ศ. 1973 เป็นต้นมา การประเมินถือว่าเป็นศาสตร์ที่มีความเป็น

ปีกแผ่น มีศัพท์เฉพาะศาสตร์ทางการประเมินโดยเฉพาะ มีองค์ความรู้ทางการประเมินในระดับ กระบวนทัศน์ โมเดลและทฤษฎี และมีวิธีการแสวงหาความรู้ด้านการประเมิน ศาสตร์แห่งการ ประเมินได้ผ่านยุคแห่งการแสวงหาเอกลักษณ์มาแล้ว ทำให้การประเมินแตกต่างจากการวัด และการประเมินก็แตกต่างจากการวิจัยอย่างเด่นชัด มีสมาคมวิชาชีพทางการประเมิน เช่น American Evaluation Association เป็นต้น มีวารสารทางการประเมินโดยเฉพาะ เช่น Educational Evaluation and Policy Analysis, Evaluation Review, New Directions for Program Evaluation, Studies in Evaluation เป็นต้น

ในด้านผู้ที่เป็นนักประเมินก็ต้องได้รับการอบรมเล่าเรียนมาโดยเฉพาะจากสถาบันการศึกษา ชั้นสูงหรือมหาวิทยาลัยต่าง ๆ มหาวิทยาลัยที่บุกเบิกการสร้างบัณฑิตทางการประเมินโดยตรง เช่น University of Illinois, Stanford University, Boston College, UCLA, University of Minnesota และ Western Michigan University ในปัจจุบันมหาวิทยาลัยต่าง ๆ ทั่วโลก โดยเฉพาะในประเทศสหรัฐอเมริกา มีการสอนทางด้านประเมินทางการศึกษาเป็นสาขาวิชาเฉพาะ

สำหรับประเทศไทยสถาบันที่บุกเบิกให้มีการผลิตบัณฑิตระดับปริญญาโททางด้าน การวัด และประเมินผลการศึกษา คือ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และคณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ส่วนในระดับปริญญาเอกทางด้าน การวัดและประเมินผล การศึกษา มีการเปิดหลักสูตรครั้งแรกในประเทศไทยและในกลุ่มประเทศอาเซียน ที่คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปี พ.ศ. 2524

นอกจากนี้ Guba และ Lincoln (1989) ได้แบ่งยุคของการประเมินผลออกเป็น 4 ยุค ดังนี้

**ยุคที่ 1 ยุคการวัดผล (Measurement: 1904-1942)** ในยุคนี้การวัดผลและการ ประเมินผลมีความหมายเหมือนกัน โดยถือว่าเป็นสิ่งเดียวกัน มีอาจแยกจากกันได้ ยุคนี้เป็นยุคของ การทดสอบ เป็นยุคการวัดคุณลักษณะต่าง ๆ ของนักเรียน เช่น ผลสัมฤทธิ์ เชาวน์ปัญญา บุคลิกภาพ ความถนัด เป็นต้น การทดสอบทางอ้อมเป็นเอกลักษณ์ของการประเมินผลใน ยุคแรกนี้ ซึ่งอาจมีเหตุจากปัจจัยที่เรียกร้องความถูกต้องเหมาะสม (legitimation) ที่จะศึกษา พัฒนาการที่สำคัญของมนุษย์ โดยวัดความแตกต่างของมนุษย์ และอีกปัจจัยหนึ่งที่ผลักดันให้ เกิดการวัดทางอ้อม คือ ความเคลื่อนไหวด้านการจัดการเชิงวิทยาศาสตร์ที่เกิดขึ้นใน การธุรกิจและอุตสาหกรรม แรงผลักดันเหล่านี้ก่อให้เกิดโปรแกรมการทดสอบในโรงเรียน ทำให้มีแรงผลักดันการสร้างแบบสอบและแบบวัดต่าง ๆ ดังนั้น ยุคแรกของการประเมิน ถือเป็นยุคของการวัดผล ผู้ประเมินมีบทบาทในเชิงเทคนิค มีความสามารถใช้เครื่องมือหรือสร้าง เครื่องมือในการสอบวัด สำนึกทางเทคนิคการวัดและการประเมินทางการศึกษายังดำรง



อยู่ถึงปัจจุบัน จะเห็นได้จากจะเข้าเรียนก็ต้องมีการสอบวัด จะสำเร็จการศึกษาได้ก็ต้องผ่านการสอบวัด การคัดเลือกคนเข้าทำงานก็ต้องผ่านการสอบวัด แม้ตำราก็ยังคงใช้คำคู่ว่าการวัดและการประเมิน ทางการศึกษาอยู่นั่นเอง กล่าวโดยสรุป “นักประเมิน” ในยุคแรกนี้ก็คือ “นักวัดผล” นั่นเอง

ยุคที่ 2 การบรรยายลักษณะ (Description: 1942-1967) ในยุคนี้การประเมินมุ่งแก้ไขจุดอ่อนของยุคแรก ที่เน้นเป้าหมายของการประเมิน (การวัด) อยู่ที่นักเรียน ข้อมูลเกี่ยวกับนักเรียนไม่อาจช่วยในการปรับปรุงหลักสูตรหรือโปรแกรมการศึกษาได้ การประเมินในยุคนี้เป็นการอธิบายลักษณะโดยพรรณนาถึงจุดแข็งและจุดอ่อนของวัตถุประสงค์ที่กำหนดขึ้น บทบาทของนักประเมิน คือ ผู้บรรยาย หรือพรรณนาลักษณะตามขอบเขตของวัตถุประสงค์ นักประเมินจะต้องมีความรู้เทคนิควิธี เช่นเดียวกับยุคแรกแต่การวัดผลจะไม่ได้มาแทนที่บทบาทของการประเมินผลอีกต่อไป การวัดผลเป็นเครื่องมือหรือเป็นส่วนหนึ่งของการประเมิน กล่าวคือ การประเมินในยุคนี้ก็เป็นผลมาจากผลงานในโครงการ 8 ปีของไทเลอร์ ที่นิยามว่า การประเมินเป็นการพิจารณาว่าได้ผลบรรลุตามวัตถุประสงค์เพียงใดนั่นเอง

ยุคที่ 3 การตัดสิน (Judgement: 1967-1989) เป็นยุคของการเรียกร้องให้นักประเมินทำหน้าที่ผู้พิพากษาดัดสินคุณค่าของสิ่งใดสิ่งหนึ่ง นักประเมินยุคนี้ต้องรักษาเอกลักษณ์ของนักเทคนิควิธี นักบรรยายหรือพรรณนาลักษณะตามวัตถุประสงค์ และจะต้องทำหน้าที่ตัดสินด้วย นักประเมินจะต้องใช้ดุลยพินิจพิจารณาอย่างรอบคอบ โดยใช้ข้อมูลจากการวัดและใช้เกณฑ์ภายนอกที่เชื่อถือได้ ประกอบการใช้วิจารณ์คุณภาพประเมินคุณค่าของสิ่งต่าง ๆ ด้วยความเป็นกลาง ในยุคนี้มีการเสนอรูปแบบการประเมินมากมายเพื่อช่วยให้การตัดสินของนักประเมินเป็นไปอย่างมีประสิทธิภาพ เช่น Stake (1967), Provus (1971), Stufflebeam et al (1971), Scriven (1973), Eisner (1979) เป็นต้น รูปแบบการประเมินเหล่านี้เห็นพ้องต้องกันว่า การตัดสินเป็นส่วนหนึ่งของการประเมิน ดังนั้น การตัดสินหรือการใช้ดุลยพินิจจึงเป็นสัญลักษณ์ของนักประเมินในยุคที่ 3

ยุคที่ 4 การตอบสนองและสร้างสรรค์ (Responsive Constructivist: 1989-ปัจจุบัน) การประเมินในยุคที่ 4 สนองเงื่อนไขหลัก 2 ประการ คือ การประเมินที่สนองตอบข้อเรียกร้องความกังวลใจ และปัญหาของผู้มีส่วนได้ส่วนเสีย (stakeholders) และเป็นการประเมินตามกระบวนการที่สนับสนุนการสร้างสรรค

การประเมินในยุคที่ 4 เป็นผลลัพธ์จากกระบวนการประเมินและบทบาทของผู้ประเมิน (Evaluator) หลักการใหม่ของการประเมินมีความจำเป็นเพื่อที่จะช่วยชี้นำการปฏิบัติ จุดยืนของการประเมินในยุคที่ 4 จึงมีลักษณะเด่นที่แตกต่างไปจากการประเมินใน 3 ยุคที่ผ่านมา รวมถึงชี้

ให้เห็นบทบาทของผู้ประเมินที่เปลี่ยนแปลงไป และมีบทบาทเพิ่มขึ้นดังนี้

1. การประเมินในยุคที่ 4 มองว่า การประเมินเป็นกระบวนการทาง “สังคมการเมือง” (Sociopolitical) ปัจจัยทางสังคม การเมือง และวัฒนธรรม ควรเป็นสิ่งสำคัญที่สามารถส่งผลกระทบต่อความถูกต้องของการประเมิน เป็นองค์ประกอบที่มีความหมายต่อกระบวนการประเมิน ถ้าปราศจากสิ่งเหล่านี้การประเมินจะไร้ประโยชน์และไร้ความหมาย การประเมินในยุคที่ 4 ยอมรับว่ากิจกรรมทุกอย่างของมนุษย์อยู่ภายใต้กรอบของการเมืองและวัฒนธรรม การตระหนักถึงปัจจัยเหล่านี้ไม่ได้หมายความว่า จะมีการประเมินที่ประนีประนอมเกิดขึ้นในการประเมิน แต่จะยังยกย่องระดับของการประเมิน เพราะสิ่งเหล่านี้ได้กลายเป็นส่วนหนึ่งของการสืบสวนและการตั้งคำถาม

2. การประเมินเป็นกระบวนการของความร่วมมือ เป็นกระบวนการที่มีเป้าหมายไปสู่การพัฒนาให้เกิดความเห็นพ้องต้องกันของผู้ที่ถูกประเมิน (evaluator) เป้าหมายนี้จะไม่สามารถสำเร็จลงได้ถ้าผู้มีส่วนได้ส่วนเสียไม่ให้ความร่วมมือ แต่การที่กล่าวเช่นนี้ไม่ได้หมายความว่า การประเมินทุกครั้งจะต้องจบลงด้วยความเห็นพ้องต้องกัน (consensus) ของทุกฝ่าย แต่ตรงกันข้ามการประเมินบางครั้งอาจจะจบลงด้วยการกระทำที่จะนำไปสู่การปรึกษาหารือ ซึ่งผู้ที่มีส่วนร่วมอาจจะตกลงที่จะไม่เห็นด้วย แต่ความพยายามและใคร่ครวญอย่างจริงจังจะช่วยให้ผู้มีส่วนได้ส่วนเสียมาถึงจุดที่อย่างน้อยทุก ๆ ฝ่ายยอมรับถึงประเด็นที่ตกลงกันไม่ได้ และพยายามที่จะแก้ไขให้ประเด็นนั้นได้รับความเห็นพ้องต้องกันจากทุก ๆ ฝ่าย

**หลักการสำคัญของการประเมินในยุคที่ 4**

การประเมินในยุคที่ 4 นี้ มีหลักการและแนวคิดเกี่ยวกับการประเมินที่ควรทำความเข้าใจ และระลึกถึงเสมอ เมื่อจะทำการประเมินโครงการหรือหลักสูตร คือ (Guba, 1989: 263-264, ส. วาสนา ประมวลพจนานุกรม 2534: 31-33)

1. การประเมินเป็นกระบวนการที่ผู้ประเมินและกลุ่มบุคคลที่เกี่ยวข้องกับโครงการร่วมมือ และตกลงใจที่จะวางแนวทางการประเมินและศึกษาคูณค่าของสิ่งที่ประเมิน

บทแทรก 1 ก. การประเมินจะช่วยให้เห็นความจริง

บทแทรก 1 ข. การประเมินคุณค่าของโครงการจะช่วยให้เกิดผลการประเมินที่ต่อเนื่องและเกิดการเปลี่ยนแปลงในโครงการ โดยการช่วยให้เกิดการปรับปรุงให้ดีขึ้น เพิ่มเติมหรือเปลี่ยนแปลงอันเป็นการสร้างเสริมให้โครงการนั้นมีคุณค่ายิ่งขึ้น

2. การประเมินเป็นกระบวนการในการเพิ่มเติมสะสม และแปลความหมายข้อมูลอย่างต่อเนื่องไม่สามารถตัดตอนจากกันได้ตลอดกระบวนการ

บทแทรก 2 ก. ส่วนต่าง ๆ ในกระบวนการประเมินไม่สามารถจะแยกมาพิจารณาอย่างอิสระได้

3. การประเมินเป็นกระบวนการเฉพาะสำหรับสภาพการณ์นั้น ๆ เท่านั้น ซึ่งขึ้นอยู่กับบริบทของโครงการ กลุ่มบุคคล และการแปลความหมายของสภาพการณ์นั้นไม่สามารถจะลงสรุปอ้างอิงไปในสภาพการณ์อื่นได้

บทแทรก 3 ก. ข้อมูลจากการประเมินโครงการหรือหลักสูตรในสภาพหนึ่งไม่สามารถจะนำมาใช้ในสภาพอื่นได้ แต่อาจปรับหรือใช้เป็นแนวทางในการศึกษาได้ในลักษณะการนำประสบการณ์จากโครงการอื่นมาประยุกต์ใช้

4. การประเมินเป็นกระบวนการทางสังคมที่เกี่ยวกับวัฒนธรรม และแนวคิดทางการเมือง ซึ่งจะต้องมีอิทธิพลจากแนวคิดของสังคมและการเมืองมาเกี่ยวข้องด้วยเสมอ

5. การประเมินเป็นกระบวนการเรียนการสอน ผู้ประเมิน ผู้ถูกประเมิน นายทุนและทุกคนที่เกี่ยวข้อง ต่างก็เรียนรู้จากกันและกัน และช่วยกันเสริมสร้างภูมิความรู้จากโครงการนั้น

6. การประเมินเป็นกิจกรรมที่ต่อเนื่อง สะท้อนและขยายแนวความคิดให้กว้างออกไป เพราะข้อค้นพบแต่ละประเด็นสามารถสร้างองค์ความรู้ และนำไปปรับปรุงแก้ไขโครงการอยู่ตลอดเวลา ดังนั้น การประเมินจะต้องทำอย่างต่อเนื่อง เพิ่มพูนและปรับปรุงอยู่เสมอ

บทแทรก 6 ก. การประเมินที่ดีจะทำให้มีปัญหาคือจะต้องตอบมากขึ้นกว่าการตอบปัญหา

บทแทรก 6 ข. การประเมินที่ดีไม่มีจุดสิ้นสุด

7. การประเมินเป็นกระบวนการที่ต้องทำทันทีทันใด ไม่สามารถที่จะวางรูปแบบอย่างชัดเจนล่วงหน้าได้ว่าจะเน้นที่จุดไหน นั่นคือ สามารถจะปรับปรุงแบบได้ตลอดเวลาขึ้นอยู่กับผู้เกี่ยวข้องและกิจกรรมที่กำลังดำเนินอยู่

บทแทรก 7 ก. การประเมินเป็นกระบวนการที่ไม่สามารถจะทำนายผลได้

8. การประเมินเป็นกระบวนการที่ทุกคนที่เกี่ยวข้องจะต้องร่วมกันรับผิดชอบ และดำเนินการมากกว่าจะมีใคร กลุ่มใดเป็นผู้กำหนดให้ทำ

9. การประเมินเป็นกระบวนการที่ผู้ประเมินและผู้เกี่ยวข้องจะต้องดำเนินการศึกษาผลที่ได้ในลักษณะของความสัมพันธ์ระหว่างสิ่งที่ค้นพบ และการพิจารณาสร้างแนวคิดในลักษณะของการเสริมสร้างองค์ความรู้จากข้อสรุปที่ได้เป็นพื้นฐานของการศึกษาต่อไป

บทแทรก 9 ก. การประเมินเป็นกระบวนการร่วมที่รวบรวมการเสริมสร้างความรู้จากกลุ่มผู้มีผลประโยชน์เกี่ยวข้องกับกลุ่มผู้ประเมิน ให้เข้ามาประสานสัมพันธ์กันในลักษณะของผลลัพธ์ที่ได้ตามที่ทุกคนตกลงเห็นพ้องต้องกัน โดยทุกฝ่ายมีความกระจำใจในเป้าหมาย ดังนั้น การประเมินจะต้องก่อให้เกิดกิจกรรมการเรียนรู้แก่ทุกฝ่าย

บทแทรก 9 ข. การประเมินเป็นกระบวนการช่วยเหลือซึ่งกันและกัน ในระหว่างกลุ่มบุคคลในการควบคุมเกี่ยวกับวิธีการและการตัดสินใจ ดังนั้น การประเมินจึงเป็นกิจกรรมที่ก่อให้เกิดการรวมพลังสำหรับทุกฝ่าย

บทแทรก 9 ค. การประเมินเป็นกระบวนการที่จะช่วยกำจัดความแตกต่างระหว่างหลักการใช้ความคิดแบบพื้นฐานหรือการประยุกต์แนวคิด ดังนั้น การประเมินจึงก่อให้เกิดความเข้าใจและมองปัญหาได้ชัดเจนว่ามีความต้องการที่จะแก้ปัญหาอย่างไร

บทแทรก 9 ง. การประเมินเป็นกระบวนการที่ต้องมีการกระทำที่เหมาะสม ที่กำหนดให้ผู้ประเมินต้องเข้าไปเกี่ยวข้อง มีปฏิสัมพันธ์ชนิดตัวต่อตัวกับบุคคลในกลุ่มต่าง ๆ ที่เกี่ยวข้อง เพราะการประเมินอย่างมีประสิทธิภาพไม่สามารถจะเกิดขึ้นได้ หากบุคคลในกลุ่มต่าง ๆ ที่เกี่ยวข้องไม่ได้เข้ามาปะทะสังสรรค์กันโดยตรง

10. ผู้ประเมินจะต้องมีบทบาททั้งทางตรงและทางอ้อมที่จะต้องรับผิดชอบต่อการประเมินในยุคที่ 4 นี้

บทแทรก 10 ก. ผู้ประเมินจะต้องเป็นผู้ชำนาญไม่ใช่เป็นเพียงผู้ใช้เครื่องมือเก็บรวบรวมและวิเคราะห์ข้อมูลเท่านั้น เป็นผู้อธิบายไม่ใช่เป็นเพียงผู้ทราบประวัติความเป็นมา และเป็นผู้ตัดสินไม่ใช่เป็นเพียงผู้สนับสนุนในกระบวนการตัดสินใจ

บทแทรก 10 ข. ผู้ประเมินเป็นผู้ให้ความร่วมมือสนับสนุนทางการเมือง เป็นผู้กำกับเวที เป็นผู้ปรับสภาพตามความเป็นจริง เป็นคนกลางแลกเปลี่ยนและประสานความรับผิดชอบในผลที่เกิดขึ้นทุก ๆ วิธีทาง

11. ผู้ประเมินไม่เพียงแต่จะเป็นผู้มีความรู้ความชำนาญเป็นพิเศษเท่านั้น หากยังต้องมีคุณสมบัติส่วนตัวที่เหมาะสมอื่น ๆ ที่สำคัญคือ ความอดทน ความมีมนุษยสัมพันธ์ เปิดเผยใจกว้าง ปรับตัวได้ดี และอารมณ์ดีอยู่เสมอ

การประเมินโครงการหรือหลักสูตรมีการพัฒนามาในรูปแบบของการประเมินอย่างต่อเนื่อง ไม่มีจุดจบซึ่งจะมีลักษณะเป็นกิจกรรมหนึ่งที่สอดแทรกอยู่ เพื่อที่จะเป็นผู้คอยดูแลคุ้มครองโครงการหรือหลักสูตรให้ดำเนินไปตามทิศทางและแนวทางที่พึงประสงค์ สามารถที่จะปรับแก้กิจกรรมให้เหมาะสมกับสภาพปัจจุบันอยู่เสมอ ในส่วนของข้อมูลก็ไม่ได้เน้นเฉพาะข้อมูลที่เป็นปริมาณเท่านั้น หากยังเน้นข้อมูลในลักษณะรายละเอียด เจาะลึก อันมีความลุ่มลึกมากยิ่งขึ้นหากผู้ประเมินมีอคติต่อโครงการ ดังนั้น ผู้ประเมินในยุคนี้จึงจำเป็นต้องเป็นบุคคลที่มีคุณธรรมสูง นอกเหนือไปจากความรู้ความสามารถในหลักการประเมินผลอีกด้วย

อย่างไรก็ตาม ไม่มีสูตรสำเร็จสำหรับการประเมินโครงการที่จะให้ผลสมบูรณ์ โครงการหรือหลักสูตรต่าง ๆ ย่อมมีแนวทาง จุดเน้น และส่วนประกอบต่าง ๆ แตกต่างกันไป ดังนั้น ผู้ประเมินจะต้องปรับแนวความคิดเกี่ยวกับการประเมินให้เหมาะสมกับแต่ละโครงการหรือหลักสูตร ในลักษณะของการกำหนดรูปแบบเฉพาะตัว (Tailor made) ไม่ใช่เป็นแบบการประเมินสำเร็จรูป (Ready made)

## ประวัติและพัฒนาการของการประเมินในประเทศไทย

การจัดการศึกษาของไทย ตามประวัติการจัดการศึกษา อาจแบ่งเป็น 3 ช่วง คือ (โชติ เพชรชื่น ม.ป.ป., บุญชม ศรีสะอาด 2520, คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย และคณะกรรมการประถมศึกษาแห่งชาติ 2527)

ช่วงที่ 1 การศึกษาตั้งแต่สมัยสุโขทัยถึงกรุงรัตนโกสินทร์ตอนต้น (พ.ศ. 1286-2414) การจัดการศึกษาในช่วงนี้ไม่มีหลักสูตร การศึกษาจัดกันที่วัด พระสงฆ์เป็นครู ไม่จัดชั้นเรียน ไม่มีการสอบไล่ ต่างสถานศึกษาต่างก็สอนกันไปตามความถนัดและความสนใจของตน ดังนั้น จึงไม่มีการประเมินหลักสูตร

ช่วงที่ 2 การศึกษาก่อนการเปลี่ยนแปลง (พ.ศ. 2414-2475) ช่วงนี้มีหลักสูตร และหลักสูตรเน้นวิชาอ่าน เขียน และคิดเลข นโยบายการศึกษาในช่วงนี้ มุ่งผลิตคนให้เข้ารับราชการ ในตอนหลังคนก็มุ่งเรียนเพื่อเข้ารับราชการมากขึ้นจนไม่มีตำแหน่งที่จะบรรจุให้ทำงาน จึงได้ปรับปรุงหลักสูตรเพื่อเน้นให้ผู้เรียนสามารถนำความรู้ไปปรับปรุงอาชีพการงาน ในช่วงนี้จึงถือว่ามี การประเมินหลักสูตร แต่ไม่ปรากฏหลักฐานว่าได้มีการประเมินกันเป็นระบบหรือไม่ ในช่วงนี้มีเหตุการณ์สำคัญดังนี้

2423 สมเด็จพระยาดำรงราชานุภาพ ผู้บังคับการทหารมหาดเล็กคิด (วางโครงการ) วิธีสอบไล่หนังสือไทยขึ้นให้ปรากฏว่าใครได้เรียนรู้ตลอดหลักสูตร โดยมีประกาศนียบัตรเป็นสำคัญ

2427 จัดการสอบไล่หนังสือเป็นครั้งแรกในประเทศไทย ชั้นประโยค 1 (เรียน 4 ปี ควรเทียบได้ชั้นประถมศึกษาปีที่ 4) ของโรงเรียนพระตำหนักสวนกุหลาบ สันนิษฐานว่าข้อสอบเป็นข้อสอบแบบเรียงความและเป็นข้อสอบที่ครูออกเอง (Teacher made tests) วันศุกร์ เดือน 5 ขึ้น 12 ค่ำ พระบาทสมเด็จพระจุลจอมเกล้าเจ้าอยู่หัวพระราชทานรางวัล (ประกาศนียบัตร) แก่ผู้สอบไล่ได้ประโยค 1 รุ่นแรก

2433 ออกพระราชบัญญัติการสอบวิชาการ

2451 ตั้งกองสอบไล่วิชาการขึ้นที่กระทรวงธรรมการ

2454 หลักสูตรฉบับ พ.ศ. 2454 มีการสอบปากเปล่า และการอ่านฟังเสียงสำเนียง เสนาะด้วย ในชั้นมัธยมบริบูรณ์ชาย (ชั้นมัธยมปีที่ 8 หรือชั้น 12 แบบสากล)

2470 พระยาเมธาธิบดีได้สร้างแบบสอบเขาวนขึ้นเป็นครั้งแรกเรียกว่า “แบบทดสอบเขาวนของพระยาเมธาธิบดี” ซึ่งมีลักษณะคล้ายกับแบบทดสอบ อาร์มี อัลฟา ของกองทัพบก สหรัฐอเมริกา แสดงว่า ความรู้เรื่องการวัดผลการศึกษาของต่างประเทศเริ่มเข้ามามีบทบาทในการวัดผลการศึกษาของไทย โดยมีช่วงระยะเวลาห่างกันประมาณ 10 ปี

ช่วงที่ 3 การศึกษาหลังการเปลี่ยนแปลงการปกครอง (พ.ศ. 2475-2521) ช่วงนี้มีหลักสูตรชัดเจน นโยบายการศึกษาช่วงนี้เน้นฝึกคนให้สามารถปรับตัวเข้ากับระบอบการปกครอง ในตอนต้น ๆ หลักสูตรได้เน้นความสำคัญด้านวิชาชีพและพัฒนาการด้านสติปัญญา แต่ในตอนหลังมุ่งพัฒนาเด็กทุกด้าน การประเมินหลักสูตรได้ทำกันเป็นระบบมากขึ้น กล่าวคือ มีการตั้งคณะกรรมการพิจารณานโยบาย มีการตั้งคณะทำงานประเมินหลักสูตร มีการศึกษาผลการวิจัยเกี่ยวกับหลักสูตรของหน่วยงานและสถาบันการศึกษาต่าง ๆ ประกอบมีการระดมความคิดจากผู้รู้หรือผู้ชำนาญการทุกสาขา และวิธีการประเมินก็เป็นวิธีการวิทยาศาสตร์มากยิ่งขึ้นกว่าแต่ก่อน

ก่อนการปรับปรุงเปลี่ยนแปลงหลักสูตรประถมศึกษาและมัธยมศึกษา พ.ศ. 2503 ได้มีการประเมินหลักสูตรดังกล่าวอย่างกว้างขวาง คือ ประเมินตั้งแต่จุดมุ่งหมาย โครงสร้าง เนื้อหา รายวิชา เวลาเรียน ความเหมาะสมกับระดับอายุของนักเรียน การบริหารและการนิเทศ วิธีสอนวิธีเรียน วัสดุหลักสูตร ฯลฯ ในช่วงนี้มีเหตุการณ์สำคัญคือ

2478 หม่อมหลวงดุษฎี ชุมสาย ได้ดัดแปลงแบบสอบสแตนฟอร์ดบินเน็ต (Stanford Binet) ให้เป็นภาษาไทย หลังจากที่แบบสอบและแบบวัดบินเน็ตเริ่มใช้ได้ประมาณ 30 ปี และในปี พ.ศ. 2488 ก็ได้สร้างแบบสอบรูปทรงเรขาคณิตขึ้น ทดสอบกับบัณฑิตจุฬาลงกรณ์มหาวิทยาลัยจำนวน 1,500 คน หม่อมหลวงดุษฎี ชุมสาย ใช้ข้อสอบแบบใหม่คือ ข้อสอบแบบปรนัยในการสอบไล่สิ้นปีการศึกษาที่โรงเรียนกุเก็ตวิทยาลัย (มีนาคม 2478)

2479 ห้างนิพนธ์บรรณาการตีพิมพ์เผยแพร่หนังสือเรื่อง “วิธีวัดผลทางการศึกษาและจิตวิทยา” เรียบเรียงโดย หม่อมหลวงดุษฎี ชุมสาย

2493 แผนกวิชาครุศาสตร์ คณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ได้เปิดสอนแบบสอบและมาตรฐานในโรงเรียน (Test and Measurement in School) เป็นรายวิชา 2 หน่วยกิต ในหลักสูตรอนุปริญญา ครุศาสตร์ ในปีเดียวกันคือ พ.ศ. 2493 รัฐบาลได้จัดตั้งหน่วยงานกลางขึ้นมาทำหน้าที่ประสานการรับความช่วยเหลือจากสหรัฐอเมริกา ซึ่งทำให้เกิดการประเมินผลโครงการตามความช่วยเหลือที่ได้รับจากต่างประเทศ ซึ่งต่อมา พ.ศ. 2497 ให้ชื่อ “คณะกรรมการดำเนินงานตามโครงการความช่วยเหลือทางเศรษฐกิจและวิชาการ” ของสหรัฐอเมริกา (กศว.) และคณะกรรมการดังกล่าวได้รับการยกฐานะเป็นกรมวิเทศสหการ ในปี พ.ศ. 2506 โดยที่หน้าที่ประการหนึ่งของกรมวิเทศสหการคือ การประเมินผลโครงการความช่วยเหลือที่ได้รับจากต่างประเทศเพื่อเสนอหน่วยงานต่างประเทศซึ่งเป็นเจ้าของทุนได้พิจารณาให้ความช่วยเหลือต่อไป

2495 หลวงสวัสดิสารศาสตร์พุทธิ อธิบดีกรมวิสามัญศึกษา ตั้งกรรมการบัญญัติศัพท์ทางครุศาสตร์ มี ดร. กมล เกฬพิจิตร เป็นประธาน ได้บัญญัติศัพท์ด้านการสอน ที่สำคัญคือ

ทดสอบ ปรนัย อัตนัย ความถนัดเชาวน์ ผลสัมฤทธิ์

2496 ได้เริ่มทดลองใช้ข้อสอบปรนัยขึ้นตลอดระบบของโรงเรียนที่โรงเรียนเบญจม-ราชรังสฤษฎ์ จังหวัดฉะเชิงเทรา อาจารย์ไพบูลย์ รัตนมังคละ เป็นผู้ริเริ่มและอบรมครูประจำการให้รู้จักออกข้อสอบปรนัยเมื่อวันที่ 6 กรกฎาคม พ.ศ.2496 ดร. แนบ บุญสิทธิสร้างแบบทดสอบความถนัดวิชาช่างให้กรมอาชีวศึกษา กระทรวงศึกษาธิการ เพื่อการแนะแนวอาชีพ

2497 หน่วยแนะแนว โรงเรียนเบญจมราชรังสฤษฎ์ จังหวัดฉะเชิงเทรา ได้จัดการวัดเชาวน์นักเรียนโดยดัดแปลงแบบสอบ Kullmann-Anderson เป็นภาษาไทย กระทรวงศึกษาธิการรับข้อเสนอแนะของรายงานการศึกษาและค้นคว้าเรื่องการวัดผลการศึกษาจากการประชุมอบรมครูใหญ่สังกัดกองโรงเรียนรัฐบาลและกองโรงเรียนราษฎร์ที่วิทยาลัยวิชาการศึกษาบางแสน ให้ปรับปรุงแก้ไขวิธีสอบและให้คะแนนของกระทรวงศึกษาธิการตามประมวลศึกษาพิเศษ ภาค 2 เพื่อใช้วิธีสอบแบบปรนัยบางวิชา

2498 ประชุมครูใหญ่ทั่วประเทศไทย ณ โรงเรียนเตรียมอุดมศึกษา เรื่องการพัฒนาการศึกษาแผนใหม่ (ทฤษฎีการออกข้อสอบแบบปรนัย) อาจารย์ไพบูลย์ รัตนมังคละ เป็นวิทยากร นำทั้งภาคทฤษฎีและปฏิบัติ มีรายงานการประชุมและตัวอย่างแบบสอบปรนัยพิมพ์ออกเผยแพร่อย่างละเอียด เป็นหนังสือชุด 4 เล่ม อาจารย์ ประชุมสุข อาชาวำรุง นิเทศครูวิทยาศาสตร์ทั่วประเทศไทย ให้ใช้เทคนิคการสอบวัดทัศนคติทางวิทยาศาสตร์ของนักเรียนกรมวิสามัญ กระทรวงศึกษาธิการ ได้พิมพ์คู่มือครูเรื่องการสอบปรนัยเล่มแรกชื่อ “หลักการทดสอบเบื้องต้น” โดย อาจารย์ไพบูลย์ รัตนมังคละ เป็นผู้เรียบเรียง หนา 165 หน้า

2499 พิมพ์เผยแพร่หนังสือตำราเรียนเรื่องการวัดผลการศึกษาลเล่มแรกสำหรับนิสิตนักศึกษาครูชื่อ “หลักและกลวิธีการวัดผลการศึกษาแผนใหม่” เรียบเรียงโดย อาจารย์ไพบูลย์ รัตนมังคละ ในปีเดียวกัน กระทรวงศึกษาได้ตั้งกรรมการสร้างแบบทดสอบเชาวน์ขึ้นโดยมี หม่อมหลวงต๋อย ชุมสาย เป็นประธานกรรมการ และยึดทฤษฎีของ Thurstone เป็นหลัก

2502 กระทรวงศึกษาธิการ จัดตั้งสำนักงานทดสอบขึ้นที่วิทยาลัยวิชาการศึกษาประสานมิตร มี หม่อมหลวงต๋อย ชุมสาย เป็นผู้อำนวยการสำนักงาน ในปีเดียวกันคือ พ.ศ. 2502 มีการจัดตั้งสำนักงานพัฒนาเศรษฐกิจและสังคมแห่งชาติ จากชื่อเดิมคือ “สำนักคณะกรรมการเศรษฐกิจและสังคมแห่งชาติ” มีการเริ่มจัดทำแผนพัฒนาเศรษฐกิจฉบับที่ 1 ในปี พ.ศ. 2504 เป็นต้นมา ทำให้เกิดแรงปฏิบัติในการประเมินแผนงานและโครงการต่าง ๆ ของส่วนราชการ เพื่อเสนอต่อสำนักงานสภาพัฒนาเศรษฐกิจและสังคมแห่งชาติ

2508 วิทยาลัยวิชาการศึกษาประสานมิตร เปิดหลักสูตรการศึกษามหาบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา

2511 แผนกวิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จัดห้องสมุดแบบสอบ เริ่มสะสมแบบสอบมาตรฐานที่สำคัญ ๆ ทั้งของไทยและต่างประเทศ ในปีเดียวกัน กรมการฝึกหัดครูได้ยกฐานะงานสร้างแบบสอบมาตรฐานของสำนักทดสอบขึ้น เป็นโครงการหนึ่งในแผนพัฒนาการศึกษาของชาติ

2513 ดร. ชวาล แพร์ตันกุล ได้สร้างแบบสอบมาตรฐานชุดแรก พร้อมกับเกณฑ์ปกติระดับชาติ (national norm) สำหรับนักเรียนชั้นประถมศึกษาปีที่ 7 ได้สำเร็จโดยเป็นแบบสอบสำหรับวัดผลสัมฤทธิ์ในการเรียนจำนวน 23 ฉบับ และสำหรับวัดความถนัดทางการเรียนอีก 7 ฉบับ นับว่าเป็นครั้งแรกในประวัติการศึกษาของชาติที่สร้างแบบสอบมาตรฐานระดับชาติโดยคณาจารย์ไทย ในปีเดียวกัน พวงทอง ไสยวรรณ สร้างเกณฑ์ประเมินผลโรงเรียนระดับมัธยมศึกษาเพื่อรับรองวิทยฐานะและการเปรียบเทียบคุณภาพ

2515 สภาการศึกษาแห่งชาติ ตั้งกรรมการสร้างแบบทดสอบคัดเลือกนิสิตนักศึกษาเข้ามหาวิทยาลัยร่วมกันมี ศาสตราจารย์ ประชุมสุข อาชวอำรุง เป็นประธานได้วางแผนเต็มรูปทดลองสร้างแบบสอบมาตรฐาน 4 ฉบับ และปรับปรุงการออกข้อสอบสัมฤทธิ์ผลวิชาต่าง ๆ คือ

1. แบบสอบความถนัดทางวิชาการทั่วไป (Scholastic Aptitude Tests) มี ดร. วิลาส สิงหวีสัย เป็นประธาน
2. แบบสอบความถนัดพิเศษ (Special Aptitude Test) ศาสตราจารย์ ดร. ประชุมสุข อาชวอำรุง เป็นประธาน
3. แบบความสนใจในการเลือกวิชาชีพ มี รองศาสตราจารย์ ดร. สวัสดิ์ ปทุมราช เป็นประธาน
4. แบบสอบบุคลิกภาพและลักษณะทางสังคม มีศาสตราจารย์ ดร. ประชุมสุข อาชวอำรุง เป็นประธาน และโครงการปรับปรุงแบบสอบสัมฤทธิ์ผลวิชาต่าง ๆ ซึ่งคณะกรรมการสอบคัดเลือกเป็นผู้ออก โดยมี ศาสตราจารย์ ดร. อุบล เรียงสุวรรณ เป็นประธาน

ในปีเดียวกัน (4-5 กุมภาพันธ์ 2515) สำนักทดสอบทางการศึกษาวิทยาลัยวิชาการศึกษาประสานมิตร (ปัจจุบันคือ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร) โดยมี ดร. ชวาล แพร์ตันกุล เป็นหัวหน้าสำนัก ได้จัดการประชุมทางวิชาการเรื่องการวัดผลการศึกษาเป็นครั้งแรก โดยมีครูอาจารย์ทุกระดับทั่วประเทศเข้าร่วมประชุมมากกว่า 2,400 คน

2516 สภาการศึกษาแห่งชาติจัดประชุมคณะกรรมการออกข้อสอบคัดเลือกเข้ามหาวิทยาลัยทุกสาขาวิชาเป็นครั้งแรกที่คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เรื่องเทคนิคการออกข้อสอบวิชาต่าง ๆ



2517 สำนักทดสอบ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ร่วมกับกรมแรงงาน กระทรวงมหาดไทย และ ยูซอม สร้างแบบสอบชุดสำหรับคัดสรรจำแนกประเภทอาชีพ ประกอบด้วยแบบสอบ 3 ประเภท คือ แบบสำรวจความสนใจเกี่ยวกับอาชีพ แบบสอบความรู้เกี่ยวกับอาชีพและแบบสอบความถนัด

2519 คณาจารย์มหาวิทยาลัยสงขลานครินทร์ได้ทดลองสร้างเกณฑ์มาตรฐานประเมินผลตนเอง เกี่ยวกับการสอบและสถานภาพของอาจารย์มหาวิทยาลัย

2522 บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย เปิดหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการศึกษาและประเมินผลการศึกษา

2524 กระทรวงศึกษาธิการได้กระจายอำนาจการประเมินผลการเรียนให้โรงเรียนและกลุ่มโรงเรียน ซึ่งเป็นไปตามเจตนารมณ์ของหลักสูตรฉบับปรับปรุง พุทธศักราช 2521

2526 สำนักงานทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ ได้เริ่มประเมินคุณภาพทางการศึกษา (national assessment) โดยประเมินทุก 2 ปี เริ่มประเมินคุณภาพการศึกษาระดับประถมศึกษา (ป.6) ในปีการศึกษา 2526 และได้เริ่มประเมินคุณภาพการศึกษาระดับมัธยมศึกษา (ม.3 และ ม.6) ในปีการศึกษา 2527

2527 เมื่อวันที่ 3-5 เมษายน 2527 กองวิชาการ สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ร่วมกับ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ได้จัดการสัมมนาและนิทรรศการเรื่อง 100 ปีของการทดสอบ เพื่อเทิดพระเกียรติพระบาทสมเด็จพระจุลจอมเกล้าเจ้าอยู่หัว ที่ทรงมีพระราชกรณียกิจในการทดสอบการวัดและประเมินผลตลอดจนเพื่อประมวลสาระสำคัญของการวัดเชาวน์ปัญญาและความถนัดและการวัดผลสัมฤทธิ์และการวัดด้านจิต อารมณ์

## แบบฝึกหัด

1. Madaus และคณะแบ่งยุคการประเมินออกเป็นกี่ยุค อะไรบ้าง อธิบายแต่ละยุค โดยสรุป
2. Guba และ Lincoln แบ่งยุคการประเมินออกเป็นกี่ยุค อะไรบ้าง อธิบายโดยสรุป
3. การประเมินในประเทศไทยมีประวัติและพัฒนาการมาอย่างไร
4. หลักการสำคัญของการประเมินในยุคที่ 4 การตอบสนองและสร้างสรรค์ ตามการแบ่งยุคของ Guba และ Lincoln มีอะไรบ้าง

## บรรณานุกรม

- คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยและกองวิชาการ สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ (2527). รายงานการสัมมนาเรื่อง 100 ปีของการทดสอบ. 3-5 เมษายน.
- โชติ เพชรชื่น (ม.ป.ป.) เทคนิคการประเมินหลักสูตร. สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- บุญชม ศรีสะอาด (2520). การวัดและประเมินผลการศึกษา. คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม.
- ส.วาสนา ประวาสพฤษ (2534). “พัฒนาการของการประเมินโครงการ” วารสารการวัดผลการศึกษา ปีที่ 13 ฉบับที่ 38 กันยายน-ธันวาคม, 23-34.
- สมหวัง พิธิยานุวัฒน์ (2541). วิธีวิทยาการประเมิน. กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- Berk, R.A. and P.H. Rossi (19 90). **Thinking about Program Evaluation**. Newbury Park: Sage Publications.
- Guba, E.G. (1987). **Naturalistic Evaluation**. Jossey-Boss, California, U.S.A.
- Guba, E.G. and Y.S. Lincoln (1989). **Fourth Generation Evaluation**. Newbury Park: Sage Publications.
- Madaus, G.F., M.S. Scriven and D.L. Stufflebeam (1983). **Evaluation Models: Viewpoints on Educational and Human Services Evaluation**. Boston: Kluwer-Nijhoff Publishing.