



LAS CONSECUENCIAS NO DESEADAS DE LAS POLÍTICAS DE EVALUACIÓN DEL APRENDIZAJE

Leonel Pérez Expósito
Universidad Autónoma Metropolitana Unidad Xochimilco

Área temática: 12. Evaluación educativa

Línea temática: Desarrollo teórico y conceptual de la evaluación educativa

Tipo de ponencia: Aportaciones teóricas

Resumen:

La ponencia desarrolla un enfoque teórico de los sistemas de evaluación que busca ampliar nuestra comprensión sobre la distancia entre las consecuencias previstas y no deseadas de las políticas de evaluación. Desarrolla un enfoque de las políticas de evaluación basado en un análisis crítico de su “objeto”: los sistemas de evaluación. Muestra cómo diferentes trabajos utilizan la noción de “sistema” para ensamblar un proyecto de regulación y control de las prácticas de evaluación. Este enfoque cibernético prevalece en los documentos de política y caracteriza el futuro como un escenario prístino de coherencia y orden. Basándose en la concepción de los sistemas sociales de Giddens, y mediante el análisis de las políticas de evaluación de los Estados Unidos y de Finlandia, el documento argumenta que los sistemas de evaluación deben considerarse sistemas de práctica, a fin de comprender mejor las consecuencias imprevistas de estas políticas, su inevitabilidad y la distancia entre los documentos de política pública y las prácticas de evaluación reales en los sistemas educativos.

Palabras clave: sistemas de evaluación; sistemas cibernéticos; sistemas de práctica; políticas de evaluación

Introducción

En diferentes países, las políticas de evaluación están en el centro de intensos debates en el campo de la educación; en parte, porque algunas de sus consecuencias en las escuelas con frecuencia contravienen sus propósitos originales. Una política de evaluación con el objetivo de mejorar el aprendizaje y la equidad a través de un fuerte sistema de rendición de cuentas, termina promoviendo prácticas para aumentar las puntuaciones de las pruebas, pero sin ganancias similares en el aprendizaje. Otra, cuyo propósito es motivar a los estudiantes a aprender, resulta en una disminución en el gusto de los estudiantes y el compromiso con la escuela. ¿Por qué las políticas de evaluación tienden a producir resultados no intencionales? ¿Y en qué medida es probable que se eviten esas consecuencias imprevistas? La ponencia tiene como objetivo responder a estas preguntas. Desarrolla un argumento que parte del enfoque sociológico de Ball sobre la ontología de la política educativa (Ball 1993, 2015; Ball, Maguire y Braun 2012), con el fin de sostener que las políticas de evaluación son discursos y textos sobre, y prácticas de, *sistemas de evaluación*. De un examen crítico sobre esta noción en el campo de la evaluación educativa, el artículo muestra cómo, si bien hay una tendencia hacia un uso no reflexivo del término, diferentes obras encuentran en el concepto de ‘sistema’ el recurso ideal para avanzar una visión sobre cómo debe/debe llevarse a cabo la evaluación del aprendizaje en las instituciones educativas; un artefacto conceptual para armar un proyecto de regulación y control de las prácticas de evaluación. A esto lo llamo el enfoque cibernético de los sistemas de evaluación, que revela el estatus ontológico de la “política como discurso”, ya que se centra en la producción de prácticas que se controlan artificialmente. En esta lógica, las políticas de evaluación representan la situación actual como mejorable y caracterizan el futuro deseable como un escenario prístino de coherencia y orden. A pesar de la prevalencia de la visión cibernética entre académicos y formuladores de políticas, los sistemas de evaluación normalmente muestran resultados involuntarios. La frustración a menudo sigue la ambición de control. Sobre la base de la concepción de Giddens (1984) acerca de las prácticas recurrentes como sistemas sociales, sostengo la necesidad de mirar los sistemas de evaluación como sistemas de práctica. Este enfoque sociológico explica por qué las formas en que la evaluación se lleva a cabo en las escuelas son mucho menos controlables que lo que los sistemas abstractos avanzados en los documentos de política esperan, y revela el carácter ontológico de las políticas como “textos” y “puesta en práctica”. Las políticas de evaluación están lejos de ser solamente *implementadas* en las escuelas. Más bien, los *agentes cognoscentes* involucrados en tales esfuerzos interpretan, traducen y recontextualizan su significado. Esto abre la puerta a una variedad de “prácticas de política” con consecuencias tanto previstas como no deseadas que se retroalimentan al sistema de evaluación, provocando ya sea equilibrio o desequilibrio. Para mostrar la aplicabilidad empírica del enfoque de los sistemas de evaluación como sistemas de práctica, la ponencia examina algunos de los resultados no intencionales de las políticas de evaluación de Estados Unidos y Finlandia.

Desarrollo

En la vasta bibliografía sobre evaluación educativa, la noción de *sistemas de evaluación* se ha utilizado ampliamente en los últimos años. Sin embargo, el término raramente es problematizado o conceptualizado en detalle, y su significado es bastante difuso, parcial u oblicuo. En general, en los trabajos de investigación y en los documentos de política, el término puede entenderse en tres formas no mutuamente excluyentes :1) un sistema de evaluación significa múltiples evaluaciones que funcionan coherentemente como un todo, y el análisis se centra en los criterios normativos para lograr la coherencia (Frederiksen y Collins 1989, Roberts, Wilson y Draney 1997, Wilson y Sloane 2000, Consejo Nacional de investigación 2001, Coladarci 2002, Consejo Nacional de investigación 2006, Black and Wiliam 2007, Consejo Nacional de investigación 2010, Darling-Hammond et al. 2013, Consejo Nacional de investigación 2014, Chattergoon y Marion 2016); 2) un sistema de evaluación comprende diferentes componentes que son necesarios en el proceso de desarrollar e implementar una estrategia de evaluación determinada o un conjunto de estrategias (por ejemplo, harlen 2007, Departamento de educación primaria y secundaria de Massachusetts 2015); o 3) el carácter sistémico se entiende bajo ciertas principios que se asocian comúnmente a las teorías de sistemas abiertos y complejos, como las nociones de interdependencia y retroalimentación (centro de BEAR 2016, Kennedy 2005, consejo nacional de investigación 2006).

El enfoque cibernético de los sistemas de evaluación

Incluso cuando algunos trabajos avanzan una visión de los sistemas de evaluación basados en conceptos que pertenecen a las teorías de sistemas abiertos y complejos, la gran mayoría visualiza sistemas cerrados de evaluación. La falta de apertura puede explicarse por la aspiración de control. La bibliografía que utiliza la noción de sistemas de evaluación está plagada de “*deberes*”. Con pocas excepciones (por ejemplo, harlen 2007), hay una carencia de trabajos descriptivo-analíticos. Lo que prevalece es un enfoque normativo, en el que la idea de sistema aparece como un recurso para el diseño de mecanismos de regulación y control de las prácticas de evaluación en diferentes niveles del sistema educativo. A esto lo llamo el enfoque cibernético de los sistemas de evaluación.

En el siglo XX el término *cibernética* fue definido por Norbert Wiener como ‘ el estudio científico del control y la comunicación en el animal y la máquina ’ (Wiener 1961). Se refería al análisis de los mecanismos de retroalimentación que regulan los sistemas complejos, con la esperanza de descubrir cómo controlarlos y “desarrollar la capacidad de sintetizar sistemas artificiales con capacidades similares” (Asaro 2010, 20). Como mencioné, gran parte del trabajo en los sistemas de evaluación se ocupa de los sistemas cerrados, mientras que la cibernética se desarrolló bajo un modelo de INPUT/OUTPUT de sistemas abiertos. Sin embargo, la idea de mirar los sistemas con la intención de controlarlos es algo que ambos comparten.

El enfoque cibernético de los sistemas de evaluación es lo que vemos en las diferentes políticas de evaluación. Los documentos de política (leyes y reglamentos, programas, textos curriculares, etc.) establecen normas, criterios, componentes y relaciones para desarrollar un sistema de evaluación que tiene como objetivo

“arreglar” las prácticas actuales de evaluación en las escuelas y especificar niveles del sistema educativo. Las políticas de evaluación pretenden producir un “nuevo” orden a través de las prácticas “corregidas” que proponen. Al hacerlo, el sistema debe diseñarse para ser controlado intencionalmente (artificialmente).

Los sistemas de evaluación como sistemas de práctica

Incluso cuando las políticas de evaluación se desarrollan bajo un enfoque cibernético, los sistemas de evaluación normalmente muestran resultados no intencionales. La frustración a menudo sigue la ambición de control, y las prácticas de evaluación que estas políticas pretenden regular difícilmente pueden controlarse intencionalmente; en parte, porque son prácticas sociales institucionalizadas, estructuradas a través del tiempo y en condiciones espaciales concretas (Giddens 1984). Desde una perspectiva sociológica, la idea de la práctica puede definirse en términos simples como “formas socialmente reguladas de hacer las cosas” (van Leeuwen 2008, 5). Pero dicha regulación dista mucho de la noción de regulación por diseño; más bien, se refiere a las “propiedades estructurales” que permiten la recursividad de las prácticas sociales; sin las cuales no se pudieron recrear.

Giddens (1984) entiende estas prácticas recursivas como “sistemas sociales”, en parte, porque identifica dos mecanismos de retroalimentación en ellos. La primera, la “retroalimentación homeostática del sistema”, se refiere a “bucles causales” a través de los cuales “las consecuencias imprevistas de la acción se alimentan de nuevo para reconstituir las circunstancias de iniciación” (27). El segundo se llama “autorregulación reflexiva”, a través del cual “los actores estratégicamente localizados buscan de forma reflexiva regular las condiciones generales de la reproducción del sistema, ya sea para mantener las cosas como son o cambiarlas.” (27-28)

Ambos tipos de retroalimentación sólo son posibles si asumimos que, en el enfoque de Giddens, los actores humanos son “conocedores” de sus actos en prácticas sociales recursivas. Si bien hay “condiciones no reconocidas” en el flujo de nuestras acciones sociales, todos desarrollamos una conciencia “práctica” y otra “discursiva” que nos permite ser monitores reflexivos de tales acciones y su contexto. Este “seguimiento reflexivo” es la fuente de las consecuencias tanto imprevistas como previstas de los actos de los agentes en los sistemas sociales. La conciencia práctica en el enfoque de Giddens se refiere a ese conocimiento tácito que nos permite saber cómo actuar en nuestras prácticas sociales cotidianas. La conciencia discursiva, por su parte, no significa que los agentes conozcan las causas objetivas y las consecuencias de sus actos, o incluso sus motivos, pero si se le solicita, ella o él es capaz de proporcionar una explicación de sus acciones.

Es difícil clasificar estrictamente las prácticas de evaluación en las escuelas como “sistemas sociales” en el sentido de las prácticas recursivas sociales. Como señala Giddens (1984), la interdependencia de los sistemas sociales hace sumamente difícil delinear claramente uno de otro. Sin embargo, su enfoque es útil para entender por qué los sistemas de evaluación son, en la práctica, raramente controlables en la lógica cibernética.

En primer lugar, las prácticas de evaluación en las escuelas se han estructurado a través de un período de tiempo significativo y como resultado de acciones interconectadas de maestros, estudiantes, padres, autoridades educativas, legisladores, responsables políticos, entre otros agentes. Las prácticas de evaluación, como un sistema, no son el resultado de un solo orquestador.

En segundo lugar, su carácter sistémico implica que esta compleja red de acciones interdependientes produce resultados involuntarios que se alimentan de nuevo al sistema. Los educadores, los padres, los estudiantes, los encargados de formular políticas, todos son actores “conocedores” de las prácticas en las que están involucrados. Como resultado del seguimiento reflexivo de sus actos en los sistemas de evaluación, son capaces de organizar acciones colectivas “ya sea para mantener las cosas como son o cambiarlas”. Si bien algunas de estas acciones logran producir sus resultados previstos, que se alimentan de las prácticas de evaluación en forma de “autorregulación reflexiva”, todas ellas también tendrán consecuencias imprevistas para el sistema de evaluación. Esta es la retroalimentación como “retroalimentación homeostática”.

Metodología y fuentes de información

A continuación, ilustraré cómo el enfoque de los sistemas de evaluación como sistemas de práctica nos ayuda a comprender mejor la distancia entre las consecuencias previstas y no deseadas de las políticas de evaluación. Abordo los casos de No Child Left Behind Act y la política finlandesa de evaluación formulada en 2004. El análisis se apoya en la evidencia producto de una revisión crítica de la literatura (Jesson, Matheson, y Lacey 2011), y un análisis de contenido (De la Garza 1988, Corbin y Strauss 2007, Scott 1990) de los documentos expuestos en el cuadro 1. Este análisis integra herramientas, estrategias y recursos de la teoría fundamentada (Corbin y Strauss 2007), y el análisis crítico del discurso (Fairclough 2003, van Dijk 2009, Wodak y Meyer 2009).

Cuadro 1: Documentos de política incluidos en el análisis

FINLANDIA DOCUMENTOS	FECHAS	ESTADOS UNIDOS DOCUMENTOS	FECHAS
LEY DE EDUCACIÓN BÁSICA BEA	1996 Y ACTUALMENTE EN VIGOR	NO CHILD LEFT BEHIND ACT 2001 NCLB	PROMULGADA EN 2002, TOTALMENTE EN VIGOR HASTA 2016-17 AÑO ESCOLAR, Y EN PARTE EN VIGOR HASTA EL AÑO ESCOLAR 2017-2018
DECRETO DE EDUCACIÓN BÁSICA CAMA	1996 Y ACTUALMENTE EN VIGOR		
CURRÍCULO NACIONAL FINLANDÉS PARA EDUCACIÓN BÁSICA 2004 (FNC)	PUBLICADO EN 2004 Y EN VIGOR HASTA EL AÑO ESCOLAR 2016-17		
EL SISTEMA EDUCATIVO FINLANDÉS Y PISA	PUBLICADO EN 2009 (DOCUMENTO NO REGLAMENTARIO)		

El NCLB Act en los Estados Unidos establece como uno de sus principales propósitos “asegurar que los estudiantes cumplan desafiantes logros académicos estatales y los estándares de contenido y aumenten su logro educativo en general, pero especialmente los desfavorecidos “ (U.S. Departamento de educación

2002, 1440). Para lograrlo despliega una política de evaluación sobre la base de los principios de la Rendición de Cuentas basada en Pruebas estandarizadas (RCP).¹

La política de NCLB fue avanzando a través de las acciones colectivas y a menudo coordinadas de los encargados de formular políticas públicas, legisladores, empresarios y emprendedores de derechos civiles, grupos de expertos, padres, algunos educadores, entre otros agentes (Rhodes 2012). Vieron en RCP una solución a los problemas en las prácticas actuales de evaluación y/o un camino hacia el avance de sus propios intereses en relación con el sistema educativo. En otras palabras, el apoyo a la RCP fue el resultado de su “vigilancia reflexiva” sobre el sistema educativo en general y sobre el sistema de evaluación en particular. Estos grupos han sido lo suficientemente poderosos como para sacar adelante las políticas de RCP, pero otros actores no se limitan a cumplir pasivamente estos lineamientos. Algunos han actuado de acuerdo con lo que creen que estas políticas pretenden lograr, mientras que otros han organizado la resistencia activa y la oposición (Rhodes 2012). Sin embargo, las diversas acciones dentro de este rango han mostrado consecuencias previstas y no deseadas que impactan en el sistema a través de una retroalimentación negativa o positiva.

Uno de los ejemplos bien documentados de tal funcionamiento sistémico es la inflación de los *puntajes*: “las ganancias en las puntuaciones de las pruebas utilizadas para la rendición de cuentas que son notablemente mayores que las ganancias reales en el aprendizaje de los estudiantes” (Holcombe, Jennings, y koretz 2013). Si vemos a la RCP desde un enfoque cibernético como un proyecto de sistemas de evaluación controlados, la inflación de la puntuación es claramente un resultado involuntario. Sin embargo, las fuentes de inflación de la puntuación son las prácticas que comprenden acciones de maestros, desarrolladores de pruebas, autoridades educativas, directores, padres, estudiantes, entre otros, que tienen consecuencias deliberadas y no intencionales. Entre estas prácticas, la preparación de la prueba como *enseñanza a la prueba* ha ganado considerable atención (pedulla et al. 2003, Hamilton et al. 2007). Ya sea como iniciativa individual, o como parte de las estrategias de escuelas o distritos, los educadores centran su enseñanza en el contenido y el formato de la prueba estandarizada. El objetivo es preparar a los estudiantes para responder con éxito los elementos de prueba, en lugar de guiar un proceso de aprendizaje sobre los conocimientos y habilidades en ella evaluados. En algunos casos, la inflación de la puntuación es, en parte, un resultado intencional de maestros, estudiantes, directores o las autoridades educativas. En otros, la distinción entre la preparación adecuada para la prueba y la enseñanza para la prueba es ambigua (Firestone, Monfils y Schorr 2004, Holcombe, Jennings y koretz 2013), y los educadores no son conscientes de los efectos inflacionarios de sus actos. Al final, ambas, las consecuencias previstas y no deseadas de las acciones de

¹ Estas políticas se basan en la lógica básica de exigir a los estudiantes que tomen pruebas estandarizadas y les atribuya altas apuestas a maestros, escuelas y distritos, ofreciendo recompensas cuando los resultados de las pruebas mejoren y se sancionen cuando no (Hamilton et al. 2002). Suponiendo que un aumento en las puntuaciones de las pruebas es igual a una mejora en el aprendizaje de los estudiantes, los defensores de TBA creen que la educación puede mejorarse bajo esta lógica.

los agentes, contribuyen a un resultado (inflación de la puntuación) que contradice los propósitos y la justificación de la RCP, y se *retroalimenta de nuevo* al sistema que busca regular.

En el caso finlandés, diferentes trabajos han encontrado una baja satisfacción escolar y en el gusto por la escuela entre los estudiantes finlandeses (Linnakylä y Malin 2008, Haapasalo, Välimaa, y Kannas 2010, Rönkä et al. 2016), así como una disminución en el compromiso con la escuela (OCDE 2013, Linnakylä y malin 2008). Estos hallazgos están en tensión con uno de los propósitos del sistema de evaluación finlandés: la motivación de los estudiantes para aprender. Como se afirma en la *ley de educación básica y el currículo nacional finlandés* (FNC), el primer objetivo de la evaluación es “guiar y alentar el aprendizaje y estudiando” (nbe 2004, 260, parlamento finlandés 1998, 10).

Problemas como la baja satisfacción con la escuela y el poco gusto por estudiar no han sido provocados exclusivamente por el sistema de evaluación. Sin embargo, según Kupiainen, Hautamäki, y karjalainen (2009) “una de las razones que los estudiantes finlandeses dan de por qué no les gusta la escuela “es [...] la multitud de exámenes que sienten que tienen” (21). Ciertamente, la política de evaluación de Finlandia rechaza claramente el uso de pruebas estandarizadas de alto riesgo y fomenta la recopilación de “una diversidad de pruebas” para evaluar el aprendizaje de los alumnos (nbe 2004, 260). Sin embargo, no promueve formatos de evaluación alternativos, porque el principal recurso utilizado por los maestros es los “exámenes recurrentes que miden el logro curricular de los estudiantes”, ya sea compuestos por ellos o por desarrolladores externos (Kupiainen, Hautamäki, y karjalainen 2009). Como argumenta Simola (2015, 2005), hay una fuerte tradición conservadora entre los maestros finlandeses con respecto a sus estilos de enseñanza, lo que podría explicar la preferencia por los exámenes sobre otras formas de recolectar evidencia.

Por lo tanto, los problemas como la disminución del compromiso estudiantil con la escuela son en parte consecuencias imprevistas del sistema de evaluación finlandés; el resultado del choque entre el enfoque cibernético a los sistemas de evaluación mostrados en la política formal y las prácticas de evaluación de los docentes estructuradas a lo largo de la historia.

Conclusiones

El documento ha ofrecido un enfoque teórico distintivo a la noción de sistemas de evaluación; un concepto que se utiliza cada vez más en la literatura relacionada con evaluación educativa y en diferentes programas y estrategias de evaluación. Este enfoque explica la distancia entre las políticas de evaluación previstas y no intencionadas y muestra la aplicabilidad para el análisis académico de dichas políticas en diferentes contextos.

Referencias

- Asaro, Peter. 2010. "Whatever Happened to Cybernetics?" In *Geist in der Maschine. Medien, Prozesse und Räume der Kybernetik*, edited by Günther Friesinger, Johannes Grenzfurthner, Thomas Ballhausen and Verena Bauer, 39-49. Vienna: Verlag Turia & Kant.
- Ball, Stephen J. 1993. "What Is Policy? Texts, Trajectories and Toolboxes." *Discourse: Studies in the Cultural Politics of Education* 13 (2):10-17. doi: 10.1080/0159630930130203.
- Ball, Stephen J. 2015. "What is policy? 21 years later: reflections on the possibilities of policy research." *Discourse: Studies in the Cultural Politics of Education* 36 (3):306-313. doi: 10.1080/01596306.2015.1015279.
- Ball, Stephen J., Meg Maguire, and Annette Braun. 2012. *How schools do policy : policy enactments in secondary schools*. London ; New York: Routledge.
- BEAR Center. 2016. "The BEAR Assessment System." <http://bearcenter.berkeley.edu/page/bear-assessment-system>.
- Black, P., and D. Wiliam. 2007. "Large-scale assessment systems: Design principles drawn from international comparisons." *Measurement: Interdisciplinary Research and Perspectives* 5 (1):1-53. doi: 10.1080/15366360701293386.
- Chattergoon, Rajendra, and Scott Marion. 2016. "Not as Easy as It Sounds: Designing a Balanced Assessment System." *State Education Standard* 16 (1):6-9.
- Coladarci, Theodore. 2002. "Is it a house... or a pile of bricks?" *Phi Delta Kappan* 83 (10):772.
- Corbin, Juliet, and Anselm Strauss. 2007. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. California: Sage Publications, Incorporated.
- Darling-Hammond, Linda, Joan Herman, James Pellegrino, Jamal Abedi, J Lawrence Aber, Eva Baker, Randy Bennett, Edmund Gordon, Edward Haertel, and Kenji Hakuta. 2013. "Criteria for high-quality assessment." Stanford, CA: Stanford Center for Opportunity. Retrieved March 21:2014.
- De la Garza, Enrique. 1988. *Hacia una Metodología de la Reconstrucción: Fundamentos, Crítica y Alternativas a la Metodología y Técnicas de Investigacion Social*. México DF: Porrúa, UNAM.
- Fairclough, Norman. 2003. *Analysing discourse : textual analysis for social research*. London: Routledge.
- Finnish Parliament. 1998. *Basic Education Act*.
- Firestone, William A., Lora Frances Monfils, and Roberta Y. Schorr. 2004. *The ambiguity of teaching to the test : standards, assessment, and educational reform*. Mahwah, NJ: L. Erlbaum Associates Publishers.
- Frederiksen, John R., and Allan Collins. 1989. "A Systems Approach to Educational Testing." *Educational Researcher* 18 (9):27-32.
- Giddens, Anthony. 1984. *The constitution of society : outline of the theory of structuration*. Cambridge: Polity Press.
- Haapasalo, Ilona, Raili Välimaa, and Lasse Kannas. 2010. "How comprehensive school students perceive their psychosocial school environment." *Scandinavian Journal of Educational Research* 54 (2):133-150.
- Hamilton, Laura S, Brian M Stecher, Julie A Marsh, Jennifer Sloan McCombs, Abby Robyn, Jennifer Russell, Scott Naftel, and Heather Barney. 2007. "How Educators in Three States Are Responding to Standards-Based Accountability under No Child Left Behind. Research Brief." RAND Corporation.
- Hamilton, Laura S., Brian M. Stecher, Stephen P. Klein, and National Science Foundation (U.S.). 2002. *Making sense of test-based accountability in education*. Santa Monica, CA: Rand.
- Harlen, Wynne. 2007. *Assessment of learning*. London: Sage Publications Ltd.

- Holcombe, Rebecca, Jennifer L Jennings, and Daniel Koretz. 2013. "The roots of score inflation: An examination of opportunities in two states' tests." In *Charting reform, achieving equity in a diverse nation*, edited by Gail L. Sunderman. Charlotte, North Carolina: Information Age Publishing.
- Jesson, Jill, Lydia Matheson, and Fiona M. Lacey. 2011. *Doing your literature review : traditional and systematic techniques*. London: SAGE.
- Kennedy, Cathleen A. 2005. "The BEAR assessment system: A brief summary for the classroom context." Berkeley, CA: Berkeley Evaluation & Assessment Research Center.
- Kupiainen, Sirkku, Jarkko Hautamäki, and Tommi Karjalainen. 2009. *The Finnish education system and PISA*. Vol. 46: Ministry of Education Helsinki.
- Linnakylä, Pirjo, and Antero Malin. 2008. "Finnish students' school engagement profiles in the light of PISA 2003." *Scandinavian Journal of Educational Research* 52 (6):583-602.
- Massachusetts Department of Elementary and Secondary Education. 2015. *2014 MCAS and MCAS-Alt Technical Report*. Malden: Massachusetts Department of Elementary and Secondary Education.
- National Research Council. 2001. *Knowing what students know : the science and design of educational assessment*. Washington, DC: National Academy Press.
- National Research Council. 2006. *Systems for state science assessment*. Washington, DC: National Academies Press.
- National Research Council. 2010. *State assessment systems: exploring best practices and innovations : summary of two workshops*. Washington, D.C.: National Academies Press.
- National Research Council. 2014. *Developing assessments for the Next Generation Science Standards*. Washington, D.C.: The National Academies Press.
- NBE. 2004. *National Core Curriculum for Basic Education*. edited by Finnish National Board of Education.
- OECD. 2013. *PISA 2012 Results: Ready to Learn (Volume III): Students' Engagement, Drive and Self-Beliefs*. OECD Publishing.
- Pedulla, Joseph J, Lisa M Abrams, George F Madaus, Michael K Russell, Miguel A Ramos, and Jing Miao. 2003. *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston: National Board on Educational Testing and Public Policy.
- Rhodes, Jesse H. 2012. *An education in politics : the origins and evolution of No Child Left Behind, American institutions and society*. Ithaca, N.Y.: Cornell University Press.
- Roberts, Lily, Mark Wilson, and Karen Draney. 1997. "The SEPUP assessment system: An overview." University of California, Berkeley: BEAR Report Series, SA-97-1.
- Rönkä, Anna Reetta, Vappu Sunnari, Arja Rautio, Markku Koiranen, and Anja Taanila. 2016. "Associations between school liking, loneliness and social relations among adolescents: Northern Finland Birth Cohort 1986 study." *International Journal of Adolescence and Youth*:1-14. doi: 10.1080/02673843.2015.1136659.
- Scott, John. 1990. *A matter of record: documentary sources in social research*. Cambridge: Polity.
- Simola, Hannu. 2005. "The Finnish miracle of PISA: Historical and sociological remarks on teaching and teacher education." *Comparative education* 41 (4):455-470.
- Simola, Hannu. 2015. *The Finnish education mystery : historical and sociological essays on schooling in Finland*. London ; New York: Routledge.
- U.S. Department of Education. 2002. *The No Child Left Behind Act of 2001*. Public Law 107-110. Washington DC.

Van Dijk, Teun A. 2009. "Critical discourse studies: a sociocognitive approach." In *Methods of critical discourse analysis*, edited by Ruth Wodak and Michael Meyer, ix, 204 p. London; Thousand Oaks Calif.: SAGE.

Van Leeuwen, Theo. 2008. *Discourse and practice: new tools for critical discourse analysis*. New York; Oxford: Oxford University Press.

Wiener, Norbert. 1961. *Cybernetics; or, Control and communication in the animal and the machine*. 2d ed. New York.: M.I.T. Press.

Wilson, Mark, and Kathryn Sloane. 2000. "From Principles to Practice: An Embedded Assessment System." *Applied Measurement in Education* 13 (2):181-208. doi: 10.1207/S15324818AME1302_4.

Wodak, Ruth, and Michael Meyer. 2009. "Critical discourse analysis: History, agenda and methodology." In *Methods of critical discourse analysis*, edited by Ruth Wodak and Michael Meyer, ix, 204 p. London; Thousand Oaks Calif.: SAGE.