



VOL. 13, Nº 2 (2009)

ISSN 1138-414X

Fecha de recepción 12/04/2009

Fecha de aceptación 05/08/2009

TIMSS, PISA, IGLU Y DEMÁS: RAZÓN Y SINRAZÓN DE LOS ESTUDIOS INTERNACIONALES DE RENDIMIENTO ESCOLAR

*TIMSS, TREADS, IGLOO and others: reason and injustice of the
international school achievement studies*



Wilfried Bos

Universidad Técnica de Dortmund

Institut für Schulentwicklungsforschung

E-mail: officebos@ifs.uni-dortmund.de

Knut Schwippert

Universidad de Hamburgo

Institut für Erziehungswissenschaften. Abt. Allg.,

Interkulturelle und International vergleichende EW

E-mail: knut.schwippert@uni-hamburg.de

*La versión original de este estudio apareció con el título "TIMSS, PISA, IGLU & Co. Vom Sinn und Unsinn internationaler Schulleistungsuntersuchungen: Schulleistungen im Internationalen Vergleich", en *Bildung und Erziehung*, 2002, vol. 55, no1, pp. 5-23.

Traducción de Leandro Reyno

Resumen:

Treinta años después de que los Estados Unidos tuvieran que vérselas con la llamada “crisis del Sputnik”, Alemania ha recibido su contrapartida bajo la forma de la llamada “crisis del TIMSS”. A partir de la publicación, a mediados de los años 90, de los primeros resultados de este estudio internacional en las áreas de matemática y ciencias naturales, los alemanes se han visto obligados a reconocer que su nivel de educación es mucho más bajo de lo esperado. La sorpresa fue grande, debido a que hacía años que Alemania no participaba regularmente en estudios internacionales de evaluación a gran escala. Como resultado de la “crisis”, se puso un énfasis mucho mayor en la investigación educativa de carácter empírico. Una consecuencia positiva de ello fue que pudo realizarse una investigación más amplia en este campo; otra consecuencia, esta vez negativa, fue que muchas personas comenzaron a hablar acerca de estos estudios sin tener una comprensión adecuada de los mismos. En el presente artículo se pasa brevemente revista a los estudios realizados en Alemania, se reseñan algunas falsas interpretaciones que han cobrado notoriedad y, por último, se señala las perspectivas que pueden esperarse de la investigación futura y la utilización de sus resultados tanto en la investigación como en la práctica.

Palabras clave: Estudios internacionales (TIMSS-PIRLS-PISA) - IEA (Association for the Evaluation of Educational Achievement) - Análisis comparado - Matemáticas escolares - Evaluación - Evaluación institucional - Sistema educativo alemán.

Abstract:

Thirty years after the US-Americans had to deal with the so called Sputnik shock Germany found its counterpart in the TIMSS shock. Since in the middle of the nineties the first results of the international mathematics and science study have been published, the Germans had to realize that they are far below the level of education they expected. It was a big surprise because Germany was not participating regularly in international large scale assessments since years. After the shock much more emphasis was put on empirical educational research. The advantage is, that much more research was undertaken in this field again - the disadvantage is that many people started to talk about studies without any appropriate understanding. The article reviews briefly studies undertaken in Germany, some famous miss-interpretations, and gives at the end a prospect about what can be expected from future research and how results can be used both in research and practice.

Key words: International studies (TIMSS-PIRLS-PISA) - IEA (Association for the Evaluation of Educational Achievement) - comparative Analysis - Mathematics - Evaluation - institutional evaluation - German Educational System.

0. Introducción

“La catástrofe educativa alemana”: tal el título que, ya en 1964, dio PICTH a sus documentos y análisis acerca de la educación en Alemania, no sin hacer referencia al atraso que podía comprobarse respecto del extranjero. Testimonio empírico de ello lo brindó el *First International Science Study* (FISS, cf. SCHULTZE 1974), en el que los alumnos alemanes, a comienzos de los años 70, se ubicaron en la mitad inferior de la clasificación. Y ya en la misma época MATHIESEN (1974) titulaba un artículo suyo en el periódico *Die Zeit* de la siguiente manera: “La educación alemana, con resultados desastrosos en la comparación internacional. Los alumnos alemanes en los últimos puestos”. Como consecuencia inmediata de ello, se impulsó, es cierto, una serie de reformas educativas, pero los responsables se abstuvieron durante veinte años de comprobar su efectividad por medio de la comparación internacional.

“Alemania, casi el ‘farolillo rojo’ en la comparación internacional de rendimiento escolar”: titulares como éste podían leerse en diversos periódicos alemanes a finales de 2001

en su cobertura del estudio PISA (*Programme for International Student Assessment*). Aun antes de que el extenso informe sobre PISA (cf. BAUMERT et al. 2001) llegara a las librerías, ya toda una serie de asesores en política educativa, de variado color partidario, y no pocos expertos en la materia, se lanzaron a ofrecer propuestas de reforma —algunas nuevas y otras en parte viejas— para sacar la educación alemana de su triste situación. Y al menos en esta oportunidad podemos estar seguros de que no se habrá de esperar décadas para comprobar empíricamente su efectividad. Si bien Alemania no cuenta con una tradición arraigada en la realización de estudios de rendimiento escolar, la Conferencia de Ministros de Cultura y Educación ha establecido, por medio de una serie de normativas para el control de la calidad educativa, que se realice una comprobación periódica internacional del rendimiento escolar en distintos niveles y asignaturas. En las siguientes páginas abordaremos la manera en que estos estudios están conformados y los resultados que se pueden esperar, o no, de ellos.

1. Estudios de rendimiento escolar a nivel nacional e internacional

En esta sección hemos de exponer los adelantos técnicos y metodológicos que han hecho posibles los recientes estudios internacionales de rendimiento escolar. Nos centraremos en las tecnologías de información, ya que fue la creciente capacidad de procesamiento de datos de los sistemas informáticos lo que posibilitó la realización de estudios de rendimiento escolar a gran escala. Esbozaremos, primero, la historia de los estudios de rendimiento escolar tomando como ejemplo una de las instituciones más destacadas en la materia: la *International Association for the Evaluation of Educational Achievement* (IEA), una asociación sin fines de lucro con sede en los Países Bajos, conformada, por regla general, por instituciones universitarias y expertos gubernamentales. Luego hemos de repasar algunos de los más importantes estudios de rendimiento escolar realizados en Alemania. Debe señalarse que no se trata de un inventario exhaustivo, sino que se pretende, por medio de algunos ejemplos, solamente dar una idea del cambio que la opinión política (y educativa) ha experimentado en las últimas décadas en lo relativo a los estudios de rendimiento escolar.

Aunque los procedimientos estadísticos a considerarse en los estudios escolares comparados a gran escala fueron, en gran medida, desarrollados hace ya décadas, sólo en los últimos años han llegado a ser aplicables en la práctica. Hace no mucho tiempo, el procesamiento de un programa con ordenadores de gran tamaño podía llevar una noche entera. Hoy en día, en cambio, si bien los estudios son cada vez más amplios y las rutinas de cálculo cada vez más exigentes, el procesamiento de datos, con un ordenador personal corriente, es cuestión sólo de minutos o segundos. La aplicabilidad práctica de los ordenadores personales se debe no sólo al aumento de su capacidad de procesamiento de datos, sino también a un modo de empleo más accesible a los usuarios. No obstante, la investigación educativa empírica requiere de una rigurosa aplicación de procedimientos de análisis establecidos, y la interpretación de sus datos exige un alto grado de pericia, la que, a su vez, supone una profunda capacitación metodológica.

Estos avances de la informática han supuesto un beneficio para la investigación pedagógica empírica, entre otras muchas áreas de aplicación. Mientras que, hasta mediados de los años 1980, no había alternativa posible al uso de grandes ordenadores, en 2001 el escalamiento de la prueba piloto del *Progress in International Reading Literacy Study* (CAMPBELL et al. 2001) pudo ya calcularse con ordenadores personales comunes. Además de la impresionante aceleración en el procesamiento de datos, los avances informáticos han supuesto también un beneficio en lo que respecta a la transmisión de datos e informaciones. HUSÉN y POSTLETHWAITE (1996, p. 131), relatan, por ejemplo, que en 1967, en el marco del *First International Mathematics Study* (FIMS), las respuestas de los examinados se

almacenaban en tarjetas perforadas y se enviaban luego por correo bajo esa misma forma al centro internacional de coordinación para su posterior procesamiento. A comienzos del siglo XXI, en cambio, la información de los cuestionarios se volcó a datos electrónicos (el reemplazo de estos cuestionarios en “formato papel” es sólo cuestión de tiempo) y se envió en segundos de un confín al otro de la Tierra para su procesamiento.

La facilidad de la comunicaciones y una transferencia de datos sin mayores complicaciones han hecho posible también la consolidación de instituciones que cuentan con la experiencia necesaria para extraer muestras aleatorias o establecer escalamientos. La participación de tales instituciones internacionales de investigación y la simplificación del intercambio de informaciones han logrado que se amplíe no sólo el espectro de los estudios, sino también su calidad. El hecho de que quienes organizan, realizan y controlan los estudios sean especialistas de los respectivos países participantes (*National Research Coordinator*: NRC), se ha visto reflejado de manera positiva, en virtud de su proximidad lingüística y cultural con el sistema educativo en cuestión. Y así como se ha revelado ventajosa la experiencia práctica de los NRC provenientes de instituciones nacionales, puede decirse lo mismo de la consolidación de instituciones especializadas en la realización de estudios a nivel internacional.

Aparte de esta transformación en la estructura organizativa de los tests de la IEA, se produjo también una evolución en los tests mismos, tal como se ha de detallar más adelante. A comienzos de los años 90, en los tests elaborados por la IEA se procesaba la totalidad de las preguntas dadas al conjunto de los examinandos en un cuadernillo. Así, en el *Reading Literacy Study* de 1991 (ELLEY 1994) se buscó registrar la comprensión en lectura de los alumnos sobre la base de unos cuadernillos de tareas que debían resolverse en más o menos una hora. En estudios internacionales de rendimiento escolar más recientes, se procede sistemáticamente a presentar a los alumnos sólo una parte de la totalidad del test. Este procedimiento, denominado “multimatriz”, se describirá más adelante en el presente trabajo, tomando como base un estudio de lectura realizado en 2001, el *Progress in Reading Literacy Study*, PIRLS. Se trata del mismo estudio que en Alemania se conoce por el acrónimo IGLU (*Internationale Grundschul-Lese-Untersuchung*; Estudio internacional de lectura en la escuela primaria). Cabe hacer notar que IGLU contiene una ampliación del estudio internacional PIRLS para dar cabida a problemas de interés puramente nacional.

En la etapa de elaboración de PIRLS, un consejo asesor internacional recomendó extender el test en cada área de lectura –textos literarios y textos informativos– por un espacio poco superior a dos horas. En la oportunidad se argumentó, sobre la base de la experiencia recogida en tests anteriores, que sólo con una duración tal podían obtenerse resultados confiables en ambos dominios. Sin embargo, dado que, por razones tanto de organización como de motivación, resulta poco recomendable someter a alumnos de una edad promedio de diez años a un test de cuatro horas (dos horas por cada tipo de texto), se decidió por último que los alumnos hicieran sistemáticamente sólo una parte del test en un lapso total de una hora. Luego se determinaron los parámetros que debían servir de foco al estudio (capacidades personales y dificultad del problema), tomando en consideración los datos sistemáticamente faltantes por medio de un análisis de ítems dentro del marco de la *Item Response Theory* (IRT). De tal modo resultó posible presentar a los alumnos un ítem de cada dominio, equivalente a una duración de test de dos horas, y al mismo tiempo, obtener valores confiables para cada examinando en ambos dominios, sin exigir por ello sobremanera a los alumnos participantes.

Tabla 1: Estudios de la IEA: Tipo y extensión de la participación alemana*

Estudio	Año	Áreas evaluadas	Grupos de edad (población)	Núm. de países participantes	Participación de Alemania (país/estado)
FIMS: First International Mathematics Study	1964	Matemática	13 años Preuniversitario	12	sólo Hessen y Schleswig-Holstein
“Six Subject Study”					
•Inglés •Francés •Literatura	1970/71	Inglés Francés Literatura	14 años Preuniversitario	21 en total	sólo componentes de lectura en inglés
•FISS: First International Science Study	1970/71	Ciencias Naturales	10 años 14 años Preuniversitario	19 en total	Representativa
•Study of Reading Comprehension	1970	Comprensión en lectura		15 en total	sin part. alemana
•Civic Education	1971	Educación cívica		10 en total	10 estados federados
SIMS: Second International Mathematics Study	1980/81	Matemática	13 años Preuniversitario	20/15	sin part. alemana
SISS: Second International Science Study	1983/84	Ciencias Naturales	10 años 14 años Preuniversitario	23 en total	sin part. alemana
Study of Writing Composition	1985/87	Composición de textos	11° año escolar (escuelas de jornada completa)	14	sólo Hamburgo (Hamburger Aufsatzstudie)
Classroom Environment Study	1985/86	Matemática	5° y 6° año escolar	10	sólo Múnich (Münchner Hauptschulstudie)
Pre-Primary-Education Study	1986/89	Calidad de vida de niños en edad preescolar	alrededor de 4 años	11	representativa (estados de la RFA)
RL: Reading Literacy	1990/91	Comprensión en lectura	9 años 14 años	26/31	representativa (estados de la RFA y la ex-RDA)
ComPed: Computers in Education Study	1989 y 1992	Utilización de ordenadores	10 años 13 años Preuniversitario	21	9 estados Pobl. II y III
TIMSS: Third International Mathematics and Science Study (= TIMSS 1995)	I.1994/95 II. 1995 III. 1995/96	Matemática Ciencias Naturales	9 años 13 años Preuniversitario	40	representativa (exc. Baden-Württemberg) Pobl. II y III
CIVICS: Civic Education Study	1999 2000	Educación cívica	14 años 16/19 años	28	representativa (sólo 14 años)
SITES: Second Information Technology in Education Study	I.1998/99 II.2000/01 III.2001/05	Utilización de ordenadores	10 años 14 años Preuniversitario	26	representativa (módulos II y III)
TIMSS Repeat (= TIMSS 1999)	1999	Matemática Ciencias Naturales	13 años	38	sin part. alemana
PIRLS: Progress in International Reading Literacy (en Alemania: IGLU)	2001	Comprensión en lectura	10 años	38	Representativa
TIMSS 2003	2001/03	Matemática Ciencias Naturales	9 años 14 años	s/d	sin part. alemana

* Respecto de los datos contenidos en la Tabla, véase la nota al final del artículo

La mayoría de los estudios de investigación educativa comparada de los que Alemania ha tomado parte en los últimos años han sido estudios organizados y realizados por la IEA. Esta institución viene llevando a cabo estudios de rendimiento escolar a gran escala (los llamados *system monitoring studies*) en distintos campos desde la década de 1960. Además de la comprensión en lectura, en las últimas décadas se realizaron estudios centrados en los conocimientos de los alumnos en las áreas de matemática y ciencias naturales, así como

también en lenguas extranjeras y educación cívica. Ciertas áreas han sido sometidas a prueba en más de una ocasión. En la Tabla 1 se ofrece una sinopsis de los estudios de la IEA. Si bien los tests de rendimiento escolar registran, por lo general, las competencias de los alumnos, el primer objetivo de este tipo de estudio no es evaluar el rendimiento a nivel individual ni a nivel de clases o escuelas. Su objetivo es, antes bien, obtener un conocimiento sistemático acerca de los sistemas educativos, los cuales se evalúan en forma paralela. En los *system monitoring studies* no es el alumno como individuo lo que es objeto de evaluación, sino la totalidad del sistema educativo (TUIJMAN y POSTLETHWAITE 1994).

Como lo había hecho ya en el Tercer Estudio Internacional de Matemática y Ciencias Naturales (TIMSS), la IEA encomendó la planificación central y la organización de PIRLS al experimentado centro de estudios del *Boston College* (BC, EE. UU.). La elaboración de los cuestionarios se llevó a cabo conjuntamente con la *National Foundation for Educational Research* (NERF, Inglaterra y Gales). Mediante las aportaciones de instituciones tales como el *Educational Testing Service* (ETS, EE. UU.) en el caso de PIRLS y el *Australian Council for Educational Research* (ACER, Australia) en el de TIMSS, fue posible aplicar los métodos de escalamiento hasta el momento más elaborados para la determinación de las capacidades personales y el grado de dificultad de los problemas. *Statistics Canada* (StatCan, Canadá) tuvo a su cargo la selección de las muestras aleatorias pertinentes, de modo que todos los países participantes en cada estudio quedaran debidamente representados. El *Data Processing Center* de la IEA (DPC, Hamburgo, Alemania) asumió la tarea de elaborar los resultados, encargándose además de establecer el necesario flujo de datos entre los centros nacionales y las instituciones participantes en el estudio.

Entre los estudios de la IEA con participación alemana, el que mayor estado público ha tomado es TIMSS, realizado a mediados de los años 90. Sin embargo, Alemania ha tomado parte también en otros estudios de la IEA. En la Tabla 1 se ofrece un cuadro sinóptico que refleja la medida de la participación alemana. Como puede observarse claramente, hasta no hace mucho tiempo la participación alemana en estudios internacionales de rendimiento escolar no había sido muy amplia. Ello se debió, sin duda, a cierta ignorancia de parte de la política, pero también a la tendencia humanística de nuestra disciplina, que estaba más interesada en debates normativos propensos a la ideologización que en una comprobación empírica de los hechos relevantes a las ciencias de la educación.

Como consecuencia de los resultados obtenidos en los estudios internacionales de rendimiento escolar, se dio comienzo –como puede apreciarse en la Tabla 2– a una serie de estudios nacionales en diversos estados federados.

Tras la conmoción causada por TIMSS (el llamado *TIMSS-Schock*) hacia finales de los años 1990, pudo observarse que distintas instituciones abrigaban concepciones diferentes acerca de la forma de tales estudios y de la representatividad de sus resultados. Como hemos visto, se planteó toda una serie de expectativas no realizadas (e irrealizables) respecto de TIMSS. Sólo con el paso del tiempo fue posible, por medio de la publicación de algunos trabajos, poner en conocimiento de las personas interesadas las potencialidades, pero también las limitaciones de este tipo de estudios.

La publicación de los primeros resultados de TIMSS a mediados de los años 1990 fue objeto de una crítica que provenía, precisamente, de las escuelas. El *Boston College*, responsable de la dirección internacional del estudio, dio a conocer a una opinión pública no preparada para ello una serie de datos descriptivos que no contenían ninguna ayuda ni aclaraciones para su lectura. Ello fue causa de que se extrajeran conclusiones equivocadas y se interpretaran los datos de manera errónea y populista.

Tabla 2: Participación de los estados federados alemanes en estudios nacionales de rendimiento escolar y en el estudio internacional de la OCDE

Estudio	Año	Áreas evaluadas	Grupo de edad	Estados federados
LAU (Aspekte der Lernausgangslage und Lernentwicklung)	1996	Alemán	5° año escolar	
	1998	Matemática	7° año escolar	Hamburgo
	2000	Solución de problemas	9° año escolar	(Relevamiento total)
	2002	1ª Lengua Extranjera	11° año escolar	
MARKUS (Mathematik Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontexte)	2000	Matemática	8° año escolar	Renania-Palatinado (Relevamiento total)
QUASUM (Qualitätsuntersuchung an Schulen zum Unterricht in Mathematik)	1999	Matemática	5° año escolar 9 año escolar	Brandemburgo
PISA (Programme for International Student Assessment)	2000	Comprensión en lectura, Matemática y Ciencias Naturales	15 años	representativo (toda Alemania)

Para hacer frente a estas críticas, los científicos alemanes que habían tomado parte en TIMSS procedieron a publicar los datos previamente corroborados, tomando en cuenta además las condiciones contextuales de las que se disponía. Así, siguiendo las pautas trazadas en ARNOLD (1999), en dos libros dedicados a TIMSS (BAUMERT, BOS y LEHMANN 2000a y 2000b) se llegó a definir con claridad las limitaciones y potencialidades del estudio. En el primero de estos volúmenes se hizo una neta distinción entre los interrogantes que pueden resolverse y aquellos cuya solución queda pendiente, haciendo evidente así —como también lo señala DITTON (1999, p. 89)— la discrepancia existente entre las expectativas que generan los *system monitoring studies* y sus potencialidades.

2. Investigación de base y planificación educativa

La comparación internacional de sistemas educativos —incluidos los estudios internacionales de rendimiento escolar— debería constituir la base de toda planificación educativa a nivel nacional. ¿Qué tipo de contenidos se enseña? ¿Bajo qué forma de escuela? ¿Con qué nivel y con cuál éxito? ¿Por qué en ciertos sistemas escolares se enseñan los números fraccionarios negativos sólo en el séptimo año, mientras que en otros ello ocurre ya en el quinto año? ¿Qué conclusiones pueden extraerse de este último hecho? ¿Cuáles contenidos se consideran importantes en otros países? ¿Cuáles condiciones de enseñanza/aprendizaje ejercen una influencia positiva en el rendimiento escolar? ¿Cuáles de ellas son efectivas tanto interdisciplinaria como interculturalmente? ¿Cuáles otras están ligadas a una disciplina o cultura? Este tipo de información debe seguramente concitar el interés de los planificadores en materia educativa.

Por otra parte, ciertas cuestiones educativas alemanas –sobre las cuales, por diversas consideraciones de tipo político, no se realizan estudios comparados entre los estados federados– sólo pueden aclararse recurriendo a los estudios comparados internacionales. ¿Cuál debe ser la duración del período escolar? ¿Son suficientes doce años, o deberían ser trece, con el fin de preparar a los jóvenes para la educación terciaria? Ya en el estudio SISS había podido constatarse que los alumnos de Hong Kong, tras doce años de escolaridad en un sistema poco selectivo, eran capaces de obtener en Física resultados similares a los que los alumnos del sistema inglés –altamente selectivo– lograban sólo en el décimotercer año. En lo que respecta a la duración del período escolar, la selectividad y la tasa de expansión, estos resultados se vieron confirmados recientemente en TIMSS. ¿Cuál es la edad más conveniente para el ingreso al sistema escolar? En algunos países la escolarización comienza ya a los cuatro años de edad, mientras que en otros sólo a los siete. ¿Qué consecuencias tiene esto al final del ciclo secundario? Fueron precisamente los estudios comparados internacionales los que lograron poner en vista de los planificadores el hecho de que las mayor parte de las discusiones sobre educación, fuertemente influidas desde la política, en realidad no se dirigía las áreas más problemáticas del sistema educativo alemán. La comparación internacional –TIMSS, PISA– logró así demostrar, por ejemplo, la obsolescencia del debate acerca de la organización escolar. Tanto los sistemas escolares de estructura segmentada –por caso, el de Suiza– como los países con escuelas integradas o comprensivas –como Suecia– están en condiciones de brindar un rendimiento óptimo. La organización centralizada (Francia) o descentralizada (Suiza) de los exámenes finales tampoco desempeña ya un papel de la importancia que las distintas “fracciones educativas” alemanas tendían a adjudicarles en vista del rendimiento escolar. Otro aspecto que se ha sobreestimado es el del tamaño de los cursos y el del tiempo nominal de clase. Las magnitudes que resultan realmente relevantes para el buen rendimiento de los alumnos son, antes bien, la manera en la que se enseña, el grado de efectividad en el empleo del tiempo de clase y la reacción de los docentes ante el tamaño, homogeneidad o heterogeneidad de un curso dado. Es sólo mediante los estudios internacionales de rendimiento escolar que es posible obtener datos de este tipo, ya que los estudios puramente nacionales no producen por sí mismo la varianza del sistema que sería necesaria para la investigación de tales problemas.

La práctica de elaboración de tests y cuestionarios de rendimiento aplicada en los diversos estudios de la IEA refleja, en última instancia, lo que RAWLS (1977) denomina “justicia en el test” y ARNOLD (1999) “equidad (*fairness*) en las reglas de los tests”. En los estudios internacionales de rendimiento escolar, ello se manifiesta en la exclusión de diferenciaciones arbitrarias y en el establecimiento de un equilibrio apropiado entre exigencias encontradas. Ahora bien, no siempre es posible contemplar los intereses de todos los países participantes. La realización de encuentros de expertos en el marco de los estudios de la IEA es prueba de la voluntad de aunar en los tests tanto tales exigencias de los países como la “justicia” y la “equidad” que deben imperar en ellos. A final de cuentas, tests como TIMSS o PISA sólo pueden llevarse a cabo cuando los principios que guían el estudio cuentan con la aceptación de todos los países participantes.

Como se ha mencionado ya, estos instrumentos internacionales son susceptibles de ampliación por medio de estudios complementarios a nivel nacional. El Instituto Max Planck de Desarrollo Humano de Berlín –responsable del TIMSS en Alemania– hizo uso de tal alternativa. Basándose en las buenas experiencias recogidas en dicha ampliación del marco internacional, también la dirección nacional del estudio IGLU, con sede en la Universidad de Hamburgo, decidió la realización de estudios complementarios. Por ello, en el estudio IGLU no sólo se han registrado la competencia en lectura, sino que se han añadido las competencias en matemática y escritura, así como también los conocimientos básicos en materia de ciencias naturales. Se ampliaron, además, los cuestionarios de contexto, de modo

de dar cabida a aspectos complementarios de importancia. De tal manera, este *system monitoring study*, de base internacional, consiguió reunir la investigación del rendimiento escolar con la del desarrollo de la escuela en sí, utilizando para ello preguntas que tienen como destinatarios a los responsables de la educación en general.

Se ampliaron a nivel nacional los cuestionarios dirigidos a los padres, de modo tal de poder suplementar el análisis del entorno familiar con informaciones más concretas que las que puede brindar un niño de diez años. Otros cuestionarios tenían como destinatarios a los docentes y su objeto era el de registrar informaciones más detalladas acerca de los conceptos de enseñanza aplicados. A esta ampliación nacional de los cuestionarios de IGLU siguieron varios estudios realizados por distintos estados federados. Por ejemplo, se investigó en Brandemburgo la relación entre rendimiento escolar y capacidad motora. En Baviera y Baden-Württemberg se estudió la relación entre la recomendación de un alumno, expedida en la escuela primaria, para su paso a cualquiera de los niveles escolares superiores (*Gymnasium*, *Hauptschule*, *Realschule*) y las notas y evaluaciones personales del alumno. En Turingia, por otra parte, se incorporó un estudio sobre el desarrollo de la escuela. Sobre la base de una muestra ampliada correspondiente al estado de Bremen, se realizó una comparación entre las distintas formas escolares. Por último, con el fin de reunir informaciones acerca del progreso en el aprendizaje, un equipo de investigación de la Fundación Alemana de Investigación (DFG: *Deutsche Forschungsgemeinschaft*) centrado en la calidad educativa (*BiQua*), incluyó partes de IGLU en los cursos de algunas clases del tercer y cuarto año primario, estableciéndose una medición periódica de los resultados.

Pero, en el marco de IGLU, se recaban informaciones relevantes no sólo dentro del sistema escolar (primario). A efectos de establecer el punto de contacto con la enseñanza secundaria, se convoca a expertos en este nivel educativo para evaluar las preguntas de rendimiento dirigidas a los alumnos. En este módulo (rotulado como “Encuesta de destinatarios”) se requiere de expertos de los distintos estados federados (por ejemplo, a directores de institutos universitarios) una evaluación de los tests de rendimiento dados a los alumnos. Las informaciones así obtenidas pueden utilizarse para comprobar si las competencias pronosticadas por los expertos se verifican en el rendimiento posterior de los alumnos.

Sin embargo, el propósito de obtener un cuadro total del aprendizaje y la enseñanza se ve sujeto a límites. Para no exigir desmedidamente a los alumnos participantes, los cuestionarios fueron elaborados de manera tal de que pudieran resolverse en un tiempo aceptable. Esta limitación temporal tuvo como consecuencia que se reuniera, en última medida, una multiplicidad de datos y pudiera darse sólo cuenta parcial de la complejidad escolar. La intención de los participantes en el estudio es la de brindar una imagen lo más completa posible del “modelo” de la realidad de la enseñanza, que debería tener por fin una formación de docentes profesional, actualizada y más cercana a la práctica. Pero, del mismo modo, debe darse cuenta de las informaciones requeridas a los participantes administrativos. Ahora bien, todos los participantes en el estudio concuerdan en que el modelo aplicado en la ocasión sirve para describir sólo una sección de la realidad. BOSKER y SCHEERENS (1994) advierten, no sin razón, que las suposiciones acerca de los modelos de efectividad escolar determinan de antemano los cuestionarios. Por otra parte, los estudios más recientes se orientan a resultados y áreas no estudiadas antes (cf. AITKIN y LONGFORD 1986). Si bien ello contribuye a un mayor grado de teorización y al avance científico, se corre el riesgo de limitar seriamente los cuestionarios y de caer en una tendencia demasiado unilateral hacia la “corriente mayoritaria” en el área de la ciencia de la educación o la psicología.

3. Una aportación a la evaluación externa y el desarrollo de la escuela

Los estudios de evaluación constituyen un aspecto que ya no puede excluirse del reciente debate acerca del sistema educativo alemán; es por ello que se les ha asignado un lugar en el estudio PIRLS.

En el caso de la evaluación interna, realizada por miembros de una escuela dada, los resultados, su evaluación y las conclusiones que pueden extraerse de ellos, se discuten, por lo general, entre los docentes del establecimiento. En determinados casos, pueden también incorporarse a este proceso alumnos y otros colegas (BECKER, ILSEMANN y SCHRANZ 2001). En el caso de una evaluación externa, en cambio, se prevé que la evaluación de los datos esté a cargo de personas no involucradas en el proceso escolar, siguiendo criterios preestablecidos. Tampoco es necesario que estos expertos externos estén presentes en el recabamiento de los datos, ya que para ello pueden emplearse los cuestionarios usuales en los estudios de comparación escolar a gran escala. En la evaluación externa, los expertos contribuyen sólo a la explicación de los datos tras su recabamiento.

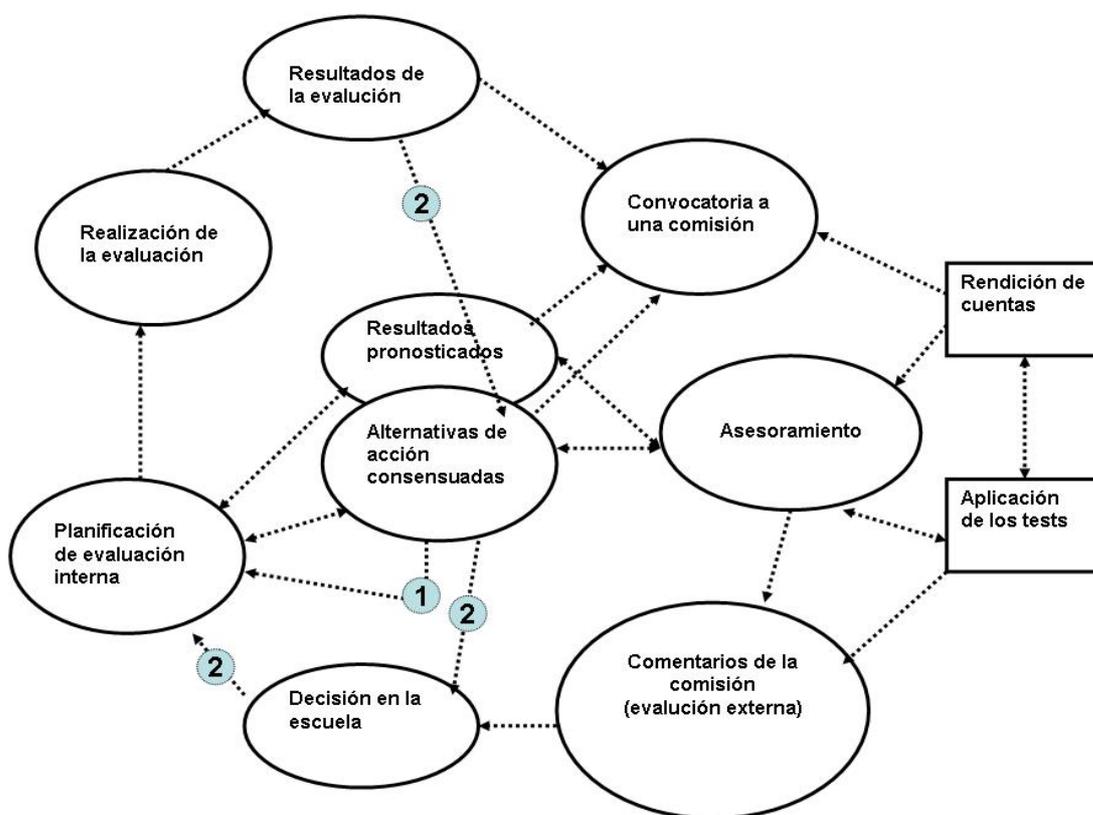
En los modelos de evaluación planteados por KEMPFERT y ROLFF (1999) y ampliados por BOS y POSTLETHWAITE (2000), la evaluación interna y externa se presentan no como modelos en competencia mutua, sino como modelos complementarios. ¿Qué es lo que distingue a un estudio de evaluación de otro tipo de estudios? WOTTAWA y THIERAU (1998) mencionan el criterio de utilidad como una de sus características más salientes. Señalan, además, que todos los participantes en el proceso de evaluación deberían acordar previamente su proceder frente a las posibles consecuencias que surjan de los datos. KLIEME, BAUMERT y SCHWIPPERT (2000), basándose en datos obtenidos en evaluaciones externas, apuntan, en cambio, al papel determinante que desempeñan las alternativas de acción consensuadas previamente a la ejecución de los tests. Por ejemplo, si alguno de los participantes desconoce las responsabilidades personales que surgen de los tests, es probable que se generen conflictos en el momento de la evaluación de los datos, poniendo así en riesgo el proceso que se espera desarrollar. A efectos de dar solidez al proceso de evaluación, es necesario que los participantes establezcan previamente alternativas de acción consensuadas. En este punto, el carácter interno o externo de la evaluación carece de importancia. En el Diagrama 1, la interdependencia entre el pronóstico de los resultados y las alternativas de acción se ve representada por flechas dobles que apuntan tanto en una como en otra dirección. Sólo en caso de preverse determinadas consecuencias posibles en el comienzo del proceso de evaluación se acordarán alternativas de acción, de modo que los participantes que resulten personalmente afectados no puedan oponerse a ellas, siendo objetivas y oportunas. Si los datos indican cambios en el proceder de una persona determinada, puede producirse espontáneamente una negativa emocional a adoptar las acciones necesarias tendientes a un mejoramiento.

Si se opta por una evaluación externa, debe convocarse a una comisión consultiva externa compuesta por expertos en la materia. Sólo estos expertos estarán en condiciones –con previo conocimiento de los tests aplicados– de pronosticar los resultados posibles.

En el Diagrama 1, las etapas denominadas “Resultados pronosticados” y “Alternativas de acción consensuadas” no resultan intercambiables, en la conexión entre evaluación interna y externa, por la etapa “Planificación de evaluación interna”, dado que allí se incorpora la comisión de expertos directamente en la fase de planificación. En el caso de una evaluación interna, las dos etapas primeramente mencionadas pueden conducir directamente a la planificación. En tal situación, es importante tener en cuenta, sin embargo, que si se tienen resultados concretos, las decisiones a tomarse dentro de la escuela pueden y deben sólo llevarse a cabo por medio de alternativas de acción previamente consensuadas. En los

estudios internacionales de rendimiento escolar, los resultados pueden, en principio, darse a conocer incluso a nivel de escuela y curso. En el marco del estudio IGLU se ponen en conocimiento de los docentes que así lo soliciten los resultados de obtenidos por el curso en el que enseñan. Esta información incluye no sólo la documentación relativa al nivel de rendimiento de los alumnos del curso, sino que también aporta una serie de condicionamientos y principios de interpretación que facilita al docente la comprensión y clasificación de los resultados. Sin embargo, todavía no se ha realizado una investigación sistemática acerca del manejo que las escuelas hacen de estos datos.

El nivel de aceptación de cada estudio depende estrechamente de las informaciones de las que se disponga antes, durante y después del estudio. Aparte de una adecuada preparación, en los estudios de rendimiento escolar cobra un papel de especial importancia la presentación de los resultados. Es por esta razón que, ya desde la puesta en marcha de un estudio escolar comparado, debe considerarse y hacerse público el modo en que dicha presentación ha de efectuarse.



En los siguientes párrafos se expondrán los factores que han de tenerse en cuenta en la presentación de la información.

La presentación de los resultados contribuye de manera determinante al éxito de una evaluación, tanto interna como externa. En este punto, se da por supuesta la necesidad de una presentación correcta y adecuada de los resultados. Especial atención debe dedicarse a la presentación de valores estadísticos. Sobre todo en vista de los participantes no del todo entendidos, debe realizarse en la presentación una distinción clara entre los datos relevantes y los datos marginales (KLIEME, BAUMERT y SCHWIPPERT 2000). Ya en la etapa de planificación de la evaluación debe dejarse en claro si se han de aportar informes y, en caso afirmativo, ante quién hacerlo. Resulta entonces necesario que se consensúen tanto la forma

que han de adoptar tales informes como las posibles guías para la interpretación de los resultados y el desarrollo de medidas de intervención. En la tarea de formación de consenso respecto de las alternativas de acción debe repararse también en si ciertos resultados pueden comunicarse a un nivel tan personal que sólo las personas interesadas tengan conocimiento de ellos, o si deben ponerse, dentro del proceso acordado, a disposición del conjunto de los participantes. Esta última opción brindaría al proceso un grado máximo de transparencia. La discusión abierta de todos los resultados supone, a su vez, cierto grado de profesionalismo de parte de los participantes para reflexionar acerca de su propio proceder. Además de su escaso grado de transparencia, los informes personalizados haría más difícil la realización de los controles de calidad necesarios. En tal caso, debería dejarse en claro, aun antes de tomar la decisión de realizar una evaluación, si los procesos que la evaluación genera han de constatarse por medio de otras evaluaciones posteriores o si se trata de una evaluación única. Si dentro del marco de una evaluación externa (auxiliar) se aplican tests (estandarizados), se dispondrá, por regla general, de resultados que asumen la forma de índices estadísticos y que pueden requerir, bajo determinadas circunstancias, de consultas ulteriores.

4. Consideraciones finales y perspectivas

Los *system monitoring studies* proporcionan un conocimiento acerca de los sistemas que puede contribuir a un mejoramiento sostenido de la calidad de los sistemas educativos. Los estudios realizados por la IEA reúnen, además de datos que sirven a docentes y directores de escuela como ayuda para el mejoramiento escolar, informaciones que resultan relevantes también a nivel administrativo. En la interpretación de los resultados debe distinguirse entre datos de alcance nacional y datos de alcance internacional.

Es indiscutible que la enseñanza en clases con muchos alumnos resulta pedagógicamente más exigente que en clases con pocos alumnos. Cuando los pedagogos reclaman clases más reducidas, suelen aducir, aparte del argumento anterior, que una frecuencia de clases creciente afecta de modo negativo la calidad de la enseñanza y, por ello, también el rendimiento de los alumnos (RAMSEGER 2000, *GEW* 4.12.2001). Ahora bien, sobre la base de distintos reanálisis tanto del *Reading Literacy Study* (SCHWIPPERT 2002) como de TIMSS (BAUMERT, BOS y LEHMANN 2000a), ha podido constatarse, sin embargo, la similitud en los resultados obtenidos por docentes con distintas frecuencias de clase. Sólo se han observado efectos de escasa magnitud en cursos excepcionalmente reducidos o excepcionalmente grandes (MOSER, RAMSEIER, KELLER y HUBER 1997). No ha podido encontrarse indicios de que los alumnos de clases particularmente grandes se vean sistemáticamente perjudicados. Este dato aporta cierta tranquilidad, pero es, a la vez, impopular, ya que —como se ha mencionado— la enseñanza en cursos grandes plantea a los docentes un nivel de exigencia mayor que en cursos reducidos.

En una serie de publicaciones se ha llegado a acusar a los organizadores de TIMSS de crear un clima adverso a la *Gesamtschule* [escuela integrada o comprensiva] (PREUSS-LAUSITZ 2000). Se les reprocha haber hecho una interpretación errónea de los datos y no estar en condiciones de juzgar acerca de la situación de la *Gesamtschule*.

En BAUMERT, BOS y LEHMANN (2000a y 2000b) pueden encontrarse los datos que acreditan plenamente la aptitud de TIMSS-III para establecer tal género de comparaciones, a la vez que se afirma que no existe indicio alguno de una “campaña” en contra de la *Gesamtschule* [escuela integrada o comprensiva]. Por el contrario, el estudio demuestra que los alumnos de las *Gesamtschulen* —si se tiene en cuenta la composición característica de su alumnado— no obtuvieron resultados sistemáticamente malos ni en comparación internacional

ni dentro de Alemania. Los sistemas escolares integrados y segmentados están representados por igual tanto en los países con altos rendimientos como en aquellos con rendimientos bajos. Lo que sí debe ser objeto de una indagación legítima es el coste considerablemente alto por alumno que esta forma escolar tiene en Alemania, si se lo compara con sus “rendimientos” (entendido esto solamente como ‘rendimiento en el test’).

El trabajo de ROSS (1997) pone de manifiesto la manera en la que un análisis poco adecuado del tamaño de los cursos puede dar lugar a interpretaciones erróneas en los estudios internacionales de rendimiento escolar. Basándose en los datos aportados por TIMSS, Ross muestra que, si se considera la totalidad de los países participantes, el rendimiento de los alumnos se incrementa cuanto más grandes son las clases y menor el salario de los docentes. Ross se funda en ello para exponer el reclamo –formulado como broma provocadora– de agrandar las clases y reducir el salario de los docentes. Se trata, claro está, de una interpretación escandalosamente errónea de los datos, ya que se realiza un análisis que abarca el conjunto de los sistemas educativos, dejando de lado toda diferencia cultural. Trasladado esto a la publicación de los primeros resultados del estudio bajo la forma de una “lista jerárquica” de países, ello sólo puede significar que un control insuficiente de los contextos escolares y sistémicos ha de dar lugar a interpretaciones erróneas. Sin embargo, con un análisis apropiado y una corroboración cuidadosa de los datos, es seguro que estudios tales como TIMSS e IGLU han de efectuar una contribución valiosa al desarrollo de las enseñanzas y los sistemas escolares.

En los últimos años, la investigación internacional en materia educativa ha hecho varios avances. Como ya se ha señalado, el desarrollo en el área informática ha posibilitado la realización de estudios de rendimiento escolar a gran escala. El avance del instrumental y de la organización de los tests ha llevado asimismo a la conformación de un modelo cada vez más detallado de la compleja estructura que componen la “enseñanza” y la “transmisión de conocimientos”, facilitando así su descripción y comprensión. También Alemania –en la medida en que ha tomado parte en estos estudios– ha sabido sacar provecho de este progreso, sin duda positivo. Ello se hace evidente, por ejemplo, en la desmitificación de los estereotipos relacionados con las “clases grandes” y el reciente “debate sobre la reforma escolar”.

Los últimos –y tal vez aterradores– datos de este tipo de investigación los ha proporcionado la publicación, en diciembre de 2001, de los resultados del estudio realizado por la Organización para la Cooperación y el Desarrollo Económico (OCDE) acerca del nivel de instrucción de los jóvenes de quince años. PISA ha venido a certificar el insuficiente nivel de instrucción de alumnos y alumnas alemanes. Sobre un total de 31 países participantes, Alemania ocupó en lectura un lugar muy alejado de las primeras posiciones, quedando en el puesto 21°. Un resultado que prácticamente se repite también en las áreas de matemática y ciencias naturales (puesto 20° en ambas). Estos escalofriantes datos reclaman una explicación. Y dado que, por la estructura misma de la investigación, no puede culparse de ello solamente, de manera arbitraria, a alumnos perezosos o poco inclinados al estudio, las causas de la catástrofe deben buscarse en el sistema educativo en su totalidad. El espanto suscitado por estos resultados tiene ahora por consecuencia que incluso viejos adversarios de tales estudios reclamen *stante pede* que se realicen nuevas investigaciones y que lleguen a hacer, al mismo tiempo, una revisión de las investigaciones anteriores. Así, TIMSS e IGLU aparecen bajo una luz distinta. IGLU cobra especial interés por el hecho de centrarse en la escuela primaria y porque se ha comprendido, tras la comprobación de la catástrofe educativa de finales de 2001, que no puede seguir analizándose el problema sólo de manera parcial, sino que se requiere de un análisis de conjunto. Es gracias al compromiso de los investigadores alemanes que se han opuesto en los últimos años a la corriente general en

materia de educación, adhiriendo a los estudios internacionales escolares comparados, que hoy podemos contar con métodos y estudios adecuados que proporcionen indicios tendientes a la recuperación del nivel de rendimiento del sistema educativo, de modo tal que las alumnas y alumnos alemanes no vuelvan a ocupar en lo futuro las últimas posiciones.

Notas:

Los datos reproducidos en la Tabla 1 se han tomado de los estudios que se detallan en la siguiente publicación:

DEGENHART, E. R. (1990). *Thirty Years of International Research. An annotated bibliography of IEA publications (1960-1990)*. Amsterdam: IEA.

Ciertos datos relacionados con la realización de tests, población y países participantes resultan contradictorios a la luz de distintas publicaciones. Los datos incluidos en la Tabla se basan, por ello, en la confrontación de fuentes diversas.

Otros datos se han tomado de:

ELLEY (1994); HUSÉN y POSTLETHWAITE (1996); BAUMERT, BOS y LEHMANN (2000a); SCHWIPPERT, ITZLINGER y BOS (1999).

Los datos correspondientes a la Tabla 2 pueden encontrarse en las siguientes direcciones electrónicas:

- LAU: <http://www.hamburger-bildungserver.de/lau/> (consultado el 3.1.2002)
 - MARKUS: <http://www.rhrk.uni-kl.de/~zentrum/markus/markus.html> (consultado el 3.1.2002)
 - QUASUM: http://www2.rz.hu-berlin.de/empir_bf/QuaSUm.htm (consultado el 3.1.2002)
 - PISA: <http://www.mpib-berlin.mpg.de/pisa/dt/Pisastart.htm> (consultado el 3.1.2002)
-

Referencias bibliográficas

- Aitkin, M. y Longford, N. (1986). Statistical Modelling Issues in School Effectiveness Studies. *Journal of Royal Statistical Society*, 149 (serie A), pp. 1-49.
- Arnold, K.-H. (1999). *Fairneß bei Schulsystemvergleichen*. Münster: Waxmann.
- Baumert, J., Bos, W. y Lehmann, R. (2000a). TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn (vol. 1). Opladen: Leske & Budrich.
- Baumert, J., Bos, W. y Lehmann, R. (2000b). TIMSS/III: Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn (vol. 2). Opladen: Leske & Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. y Weiß, M. (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske & Budrich.

TIMSS, PISA, IGLU y demás:

- Becker, G., Ilsemann, C. v. y Schratz, M. (2001). Qualität entwickeln: Evaluieren. En: *Friedrich Jahresheft*.
- Bos, W., Lankes, E.-M., Prenzel, M., Valtin, R. y Walther, G. (2001). Lesen, Mathematik und Naturwissenschaft in der Grundschule -welche Leistungen erbringen deutsche Grundschul Kinder? En: Institut für International und Interkulturell Vergleichende Erziehungswissenschaft der Universität Hamburg (coord.). *Internationale Grundschul-Leseuntersuchung IGLU/PIRLS Progress in International Reading Literacy Study* (pp. 3-17). Opladen: Leske & Budrich.
- Bos, W. y Postlethwaite, T. N. (2000). Möglichkeiten, Grenzen und Perspektiven internationaler Schulleistungsforschung. En Rolff, H.-G., Bos, W., Klemm, K., Pfeiffer, H. y Schulz-Zander, R. (coords.). *Jahrbuch der Schulentwicklung*. Vol. 11 (pp. 365-386). Weinheim: Juventa.
- Bosker, R. y Scheerens, J. (1994). Alternative Models of School Effectiveness Put to the Test. *International Journal of Educational Research*, 21, pp. 159-180.
- Campbell, J.R., Kelly, D.L., Mullis, I.V.S., Martin, M.O. y Sainsbury, M. (2001). *PIRLS Progress in International Reading Literacy Study*. 2ª edición. Boston: Boston College.
- Ditton, H. (1999). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. *Zeitschrift für Pädagogik*, 41, pp. 73-92.
- Elley, W. B. (1994). *The IEA Study of Reading Literacy: Achievement and Instruction in Thirty-Two School Systems*. Exeter: Pergamon.
- GEW (Gewerkschaft Erziehung und Wissenschaft) 4.12.2001. disponible en: http://www.gew.de/standpunkt/aschlagzeilen/schule/pisa/presseerkaerungen/texte/d_prekrlaerung4.htm (consultado el 3.1.2002)
- Husén, T. y Postlethwaite, T.N. (1996). A Brief History of the International Association for the Evaluation of Educational Achievement (IEA). *Assessment in Education*, 3 (2), pp. 129-141.
- Kempfert, G. y Rolff, H.-G. (1999). *Pädagogische Qualitätsentwicklung*. Weinheim: Beltz.
- Klieme, E., Baumert, J. y Schwippert, K. (2000). Schulbezogene Evaluation und Schulleistungsvergleiche -Eine Studie im Anschluß an TIMSS. En: Rolff, H.-G., Bos, W., Klemm, K., Pfeiffer, H. y Schulz-Zander, R. (coords.). *Jahrbuch der Schulentwicklung*. Vol. 11 (pp. 387-419). Weinheim: Juventa.
- Matthiesen, H. (1974, 20.9). Die deutschen Schüler auf dem letzten Platz. *Die Zeit*.
- Moser, U., Ramseier, E., Keller, C. y Huber, M. (1997). *Schule auf dem Prüfstand*. Zürich: Rüegger.
- Picht, G. (1964). *Die deutsche Bildungskatastrophe*. Analyse und Dokumentation. Olten/Friburgo: Walter.
- Preuss-Lausitz, U. (2000). Die -fast autonome -Missbrauchsmöglichkeiten der Black Box Forschung und die Publikationsformen der Forscher. En: Gewerkschaft Erziehung und Wissenschaft (coord.). *Was leisten Leistungsvergleiche (nicht)?* (pp. 9-15). Pfungstadt: Spritzer.
- Ramseger, J. (2000). disponible en: http://212.83.35.53/wissen/zeitschriften/e-w/2000/2000-3/texte/d_s-12.htm (consultado el 3.1.2002).
- Rawls, J. (1977). *Gerechtigkeit als Fairness*. Friburgo/München: Karl Alber [Justicia como equidad: materiales para una teoría de la justicia. Madrid: Tecnos, 1986].
- Ross, K.N. (1997). Research and Policy: a complex mix. IIEP Newsletter, 15 (1), pp. 1-4.

TIMSS, PISA, IGLU y demás:

Schultze, W. (1974). Die Leistungen im naturwissenschaftlichen Unterricht in der Bundesrepublik im internationalen Vergleich. En: Deutsches Institut für Internationale Pädagogische Forschung (coord.). *Mitteilungen und Nachrichten* (número especial). Fráncfort: DIPF.

Schwippert, K. (2002). *Optimalklassen: Mehrebenenanalytische Untersuchungen*. Münster: Waxmann.

Schwippert, K., Itzlinger, U. y Bos, K. (1999). Database Construction. En: Pelgrum, W.J. y Anderson, R.E. (coords.). *ICT and the Emerging Paradigm for Life Long Learning* (pp. 345-361). Enschede: Universidad de Twente.

Tuijnman, A.C. y Postlethwaite, T.N. (1994). *Monitoring the Standards of Education*. Oxford: Pergamon.

Wottawa, H. y Thierau, H. (1996). *Lehrbuch Evaluation*, 2ª edición. Berna: Hans Huber.