



Programa de Promoción de la Reforma  
Educativa en América Latina y el Caribe

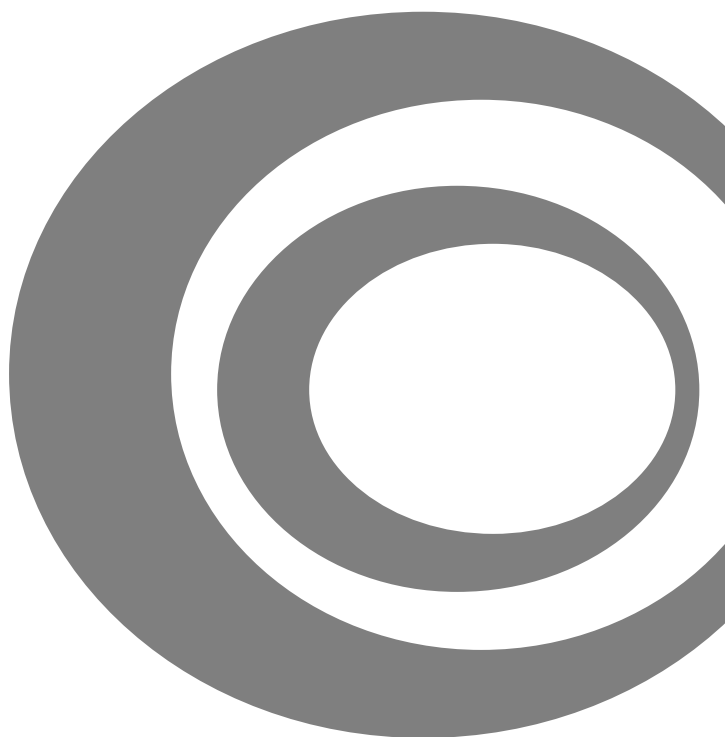
## **Grupo de Trabajo sobre Estándares y Evaluación**

**Los estándares de  
desempeño  
educacional y la  
economía**

**Henry M. Levin**



Grupo de Análisis para el Desarrollo



## LOS ESTÁNDARES DE DESEMPEÑO EDUCACIONAL Y LA ECONOMÍA<sup>1</sup>

Henry M. Levin

*Tomado y traducido de **Educational Researcher**, vol. 27, no. 4, mayo de 1998, pp. 4 - 10 con autorización de los editores. El autor ocupa la cátedra David Jacks de Educación Superior y Economía de la Universidad de Stanford. Actualmente es profesor visitante de Teachers College en la Universidad de Columbia., BOX 181, 525 W. 120 Street, Nueva York, NY 10027. Se especializa en economía de la educación y en reforma educacional.*

*Uno de los argumentos más esgrimidos a favor de la reforma educacional de los E.E.U.U. en general y a favor del establecimiento de estándares de desempeño escolar en particular, ha sido la afirmación de que crearán una fuerza laboral de primer nivel mundial. Detrás de esta expectativa está el supuesto de que hay evidencia fuerte y comprobada de que estándares y desempeños escolares más elevados y mejores son las claves para una mayor productividad laboral. Pero una revisión de la evidencia disponible sugiere sólo una débil relación entre los puntajes en las pruebas y la productividad económica y virtualmente ninguna evidencia sobre la validez predictiva de los nuevos estándares de desempeño. Este artículo sugiere que el movimiento a favor de los estándares educacionales se ha apoyado en la argumentación económica principalmente por su capacidad persuasiva para estimular la reforma educacional, antes que por tener alguna evidencia sólida sobre los vínculos entre estándares educacionales específicos y el logro económico.*

Uno de los principales argumentos a favor del establecimiento de estándares educacionales es que preparan a los estudiantes para cumplir roles productivos en la fuerza laboral. De hecho, la conversión del movimiento a favor de los estándares en una manía de la política nacional data mayormente de la publicación en 1983 del documento *Una nación en riesgo*, que daba a entender que los estándares poco exigentes de las escuelas estadounidenses habían conducido a un desarme unilateral de la nación en la competencia económica mundial (Comisión Nacional para la Excelencia en la Educación 1983). Supuestamente, necesitábamos formar una fuerza laboral de primer nivel mundial para las batallas siguientes y el establecimiento y logro de los estándares educacionales para todos los estudiantes representaban el manifiesto para una movilización económica. Poco se dijo acerca de estándares educativos para cultivar la ciudadanía en una democracia o para promover un desarrollo personal y social saludables. Y, con la notable excepción de Hirsch (1987), se dedicó poca atención a la preocupación sobre estándares para el conocimiento y la cohesión cultural. Eran los argumentos económicos los que parecían incitar a los educadores y políticos a la acción. Es notable, en efecto, que el más importante esfuerzo nacional por establecer estándares educacionales sea patrocinado por una organización denominada Centro Nacional para la Educación y la Economía.<sup>2</sup>

Menciono esto al principio porque creo que cualquier lectura imparcial de la reciente historia de los estándares educacionales verá que el argumento más explícito es la fundamentación económica. Sin embargo, las conexiones específicas entre

---

<sup>1</sup> Este artículo es una revisión del trabajo presentado en el simposium, "Perspectivas respecto al establecimiento de estándares", en la reunión anual de la Asociación Americana para la Investigación Educativa (*American Educational Research Association*) del 26 de marzo de 1997 en Chicago. El autor desea expresar su aprecio a la Fundación Russell Sage por financiar su estadía como académico visitante cuando este artículo fue escrito.

<sup>2</sup> Aparentemente, el centro emplea el término *economía* en su nombre para enfatizar la centralidad de la economía para justificar la reforma educativa.

estándares educacionales y resultados económicos han sido el aspecto menos analizado de los estándares. No hay duda de que la educación en general es un importante determinante de los ingresos<sup>3</sup>. Pero existe una enorme distancia entre este hecho y la aseveración de que nuevos estándares de desempeño educacional para los estudiantes conducirán a una mayor productividad económica<sup>4</sup>. Estudios que han intentado establecer categorías de rasgos deseables del trabajador para el mercado laboral del futuro se han limitado a esbozar trazos gruesos en vez de estándares específicos y mensurables (Rumberger & Levin, 1989; Comisión de la Secretaría para el Logro de Destrezas Necesarias, 1991, 1992). De hecho, el Panel sobre Economía de la Reforma Educacional, un grupo de economistas especializados en la economía de la educación, concluyó que aunque los retornos a la inversión en capital humano son importantes, no les era posible identificar resultados educacionales específicos que pudieran explicar la productividad laboral (Hanushek, 1994)<sup>5</sup>. Como sugeriré en este artículo, hay buenas razones para ello, pues no se sabe casi nada respecto a la relación entre estándares específicos y la productividad económica, aun en una economía estática con tecnología, mercados, organización e incentivos invariables. La capacidad predictiva de estándares educacionales específicos para un alto desempeño económico en una economía futura es aun más oscura<sup>6</sup>.

Aquí es importante tener en claro a qué me estoy refiriendo con *estándares de desempeño*. En este artículo, llamaré estándares de desempeño a los requerimientos de conocimientos específicos y de proficiencia analítica (y en algunos casos, hábitos de trabajo y destrezas interpersonales) esperados de los estudiantes en niveles específicos del sistema educativo, medidos apropiadamente con instrumentos de evaluación con el objetivo de establecer el nivel de suficiencia de la preparación educativa de un individuo<sup>7</sup>. Es decir, me estaré refiriendo a estándares de evaluación que se espera logren los estudiantes. Por supuesto está implícito en esta definición que la suficiencia será calificada de acuerdo a algunos criterios externos -- en este caso, la validez predictiva de las evaluaciones educacionales sobre la productividad laboral de individuos que logran o no logran los estándares<sup>8</sup>.

En este contexto, los estándares de desempeño escolar son vistos como algo más que sólo amplias categorías de resultados educacionales deseables para los estudiantes. Por ejemplo, la Comisión de la Secretaría para el Logro de Destrezas

---

<sup>3</sup> Para revisar la evidencia, véase Cohn y Geske (1990, pp.34-69).

<sup>4</sup> Los tratamientos teóricos por parte de los economistas de la relación entre estándares de un lado y productividad e ingresos de otro, no abordan la magnitud de los lazos empíricos sino simplemente asumen que no son triviales. Véase, por ejemplo, Betts (1996) y Costrell (1994).

<sup>5</sup> Hay una larga historia de uso de modelos de ecuaciones estructurales para estimar los coeficientes de la educación escolar, la experiencia laboral, la raza, y el género, respecto a los ingresos. En general, los estudios estadísticos encuentran que los retornos económicos a años adicionales de escolaridad son sustanciales y exceden los retornos al capital físico. Véase, por ejemplo, Mincer (1974). Este antiguo trabajo fue criticado por considerarse que exageraba los ingresos asociados con la escolaridad, debido a que se esperaba que variables omitidas respecto a la habilidad, la calidad de la escuela y características socioeconómicas estén correlacionadas tanto con la productividad y con los ingresos como con la cantidad de años de escolaridad de los individuos. Trabajos más recientes han encontrado que los coeficientes estimados de la escolaridad están subvalorados, mayormente por errores de medición y la endogeneidad de la escolaridad. Véase, por ejemplo, Blackburn y Neumark (1995).

<sup>6</sup> Bailey y Merritt (1997) han demostrado que cuando determinadas industrias han intentado establecer estándares de destrezas específicas que esperan sean traducidas a pautas educacionales para las escuelas y los estudiantes, el ímpetu de sus esfuerzos parece resultar incompatible con las orientaciones más amplias de las reformas educativas.

<sup>7</sup> Algunos de los temas aquí presentados han sido abordados conceptualmente en Wiley y Haertel (1996).

<sup>8</sup> Una discusión útil sobre los temas de validez respecto a los estándares de desempeño se encuentra en Linn (1996).

Necesarias (1991), bajo el auspicio del Departamento de Trabajo de los EEUU, ha establecido una serie de metas generales que considera deseables en la fuerza de trabajo futura. Ellas incluyen la identificación, organización, planificación y asignación de recursos; destrezas interpersonales que incluyen la participación en equipos; la adquisición y uso de información; la comprensión de interrelaciones complejas desde una perspectiva de sistemas y el trabajo con una variedad de tecnologías. Las dimensiones generales y sus detalles proceden de entrevistas y discusiones con empleadores y trabajadores, así como de análisis de centros de trabajo de alta productividad. Esos criterios son sin duda razonables y tienen validez nominal, en el sentido de que parecen ajustarse en términos generales a las demandas laborales en ciertos tipos de centros de trabajo. Pero son tan amplios que invitan a muchas interpretaciones y aplicaciones diferentes respecto a sus dimensiones precisas -- es decir, cuáles son los componentes específicos de conocimientos y comportamientos y cómo pueden ser éstos medidos. En buena parte, son guías exhortatorias para preparar a los futuros trabajadores, antes que pautas prescriptivas.

En cambio, voy a considerar los estándares de desempeño estudiantil como prescriptivos en el establecimiento de criterios y medidas de evaluación escolar más precisos sobre lo que se espera que logren los estudiantes. Es decir, me voy a concentrar en las implicancias económicas de los estándares tal como están incorporados tanto en los actuales sistemas de pruebas como en los sistemas de evaluación de desempeño que están recién surgiendo (Kane & Mitchell 1996). Aunque algunos puedan decir que el actual sistema de pruebas estandarizadas no debería ser visto como parte del movimiento a favor de los estándares, ese sistema ha sido y sigue siendo utilizado de esa manera. Los estudiantes son seleccionados para oportunidades educativas adicionales y los sistemas educativos nacionales son evaluados en comparaciones internacionales de logros. La mayoría de estas pruebas son tradicionales, aun cuando algunas de las versiones más recientes plantean problemas y reflexiones más complejos y son calificados en base a criterios en vez de ser calificados en base a normas. En realidad, gran parte del impacto del informe *Una nación en riesgo* se debió a la presuposición de que la pérdida de la competitividad económica nacional a principios de los años 80 estaba directamente relacionada con los bajos puntajes de los EEUU en las comparaciones internacionales de logros, sobre la base de criterios tradicionales. En cambio, voy a sostener que el argumento de que los resultados de las pruebas tradicionales han estado estrechamente vinculados al desempeño laboral de los individuos tiene poco sustento y que no hay ninguna evidencia de que los nuevos estándares tengan validez predictiva más alta.

Esta discusión es particularmente importante para los investigadores educacionales porque ilustra cómo creencias fuertemente afianzadas, que movilizan tanto a la opinión pública en general como a los diseñadores de políticas, pueden tener poco sustento en resultados de investigación. Como preludeo a esta discusión, llama la atención que hay pocas pruebas de que la recesión económica y el descenso de la competitividad de la economía estadounidense a fines de los años 70 y a principios de los 80 fueron causados por una fuerza laboral deficientemente educada, pese a que el informe *Una nación en riesgo* sustentó sus reclamos en dicha aseveración. Esto no niega que haya habido amplios márgenes para mejorar las escuelas en los años 80 y 90 por muchas otras razones. Y cabe resaltar que ningún economista haya atribuido el formidable desempeño de la economía de los EEUU desde 1992 a reformas educacionales exitosas en reacción al informe *Una nación en riesgo*. Las reformas han tenido solo resultados modestos hasta ahora y sólo una pequeñísima parte de la fuerza laboral podría posiblemente haberse beneficiado con estas reformas antes de entrar al mercado laboral de principios de los años 90 (probablemente 5% o menos). El eminente economista laboral Clark Kerr (1991, p.30) examinó una serie de evidencias sobre el argumento de que la educación es la clave para la competitividad de la nación

y concluyó: “Pocas veces en el curso del diseño de políticas en los EEUU, tantas convicciones firmes han sido mantenidas por tanta gente sobre la base de tan pocas pruebas”<sup>9</sup>.

Incluso la explicación más extendida sobre la emergencia de nuestros competidores europeos y asiáticos daba gran significancia a la “alta calidad” de sus sistemas educativos. Sin embargo, estudios econométricos rigurosos sobre el crecimiento económico de las potencias económicas asiáticas como Singapur, Hong Kong, Corea, y Taiwan no brindan sustento a esa interpretación<sup>10</sup>. La evidencia sugiere que su crecimiento económico y competitividad extraordinarios se debieron fundamentalmente a la inversión económica masiva, y en segundo término, al crecimiento de sus fuerzas de trabajo, no a la calidad de sus sistemas educacionales (Kim & Lau 1996; Krugman 1994)<sup>11</sup>. Y debe señalarse que en esta década, Europa Occidental ha tenido el doble de tasa de desempleo y tasas anémicas de crecimiento económico en comparación a los EEUU, mientras que las tasas de desempleo en Japón se han duplicado y el crecimiento económico es apenas perceptible.

En resumen, sostengo que la inversión educativa es importante para los resultados económicos, pero aún no se ha demostrado qué aspectos precisos de la educación hacen la diferencia. Cuando nos colocamos en el contexto de los estándares de desempeño escolar que se establecen en cierta medida para crear una fuerza laboral “de primer nivel mundial”, es apropiado un escepticismo saludable, pues hay poca evidencia de que, tanto los instrumentos tradicionales de medición de logros, como las medidas empleadas por la Evaluación Nacional del Progreso Educativo (NAEP), o las nuevas medidas propuestas para establecer estándares nacionales, tengan alta validez predictiva sobre la productividad en los centros laborales<sup>12</sup>. Estoy abierto a la posibilidad de que esa vinculación pueda establecerse con el tiempo mediante estudios de validez y nuevos refinamientos de medidas, pero es engañoso sugerir que actualmente tenemos evidencias que sustenten esas afirmaciones.

### **Validez predictiva**

La importancia económica de la evaluación de los estándares de desempeño se reduce a su capacidad predictiva para explicar la productividad económica en el centro de trabajo, tanto en el ámbito individual como en el social. La diferencia entre estos es instructiva. En el plano individual, es posible que la persona que logra los estándares pueda aprender nuevas habilidades con más rapidez y con mayor proficiencia, y pueda ser capaz de hacer un trabajo superior y realizar una mejor asignación de recursos del centro laboral. En el plano social es posible que el logro generalizado de los estándares pueda dar lugar a mejores arreglos e interacciones organizacionales entre individuos y organizaciones, de manera que la productividad económica se eleve más allá del individuo. La noción de que resultados educacionales específicos, en

---

<sup>9</sup> Algunos de los capítulos en Resnick y Wirt (1996) sugieren una visión más optimista, pero sin ninguna evidencia respecto a la validez predictiva.

<sup>10</sup> Estos países han estado entre aquéllos que obtienen más altos puntajes en las comparaciones nacionales de logros. Aunque recientemente han sufrido reveses económicos, se espera que se recuperen rápida y fuertemente.

<sup>11</sup> En 1960, el ingreso per cápita en los EEUU era aproximadamente siete veces mayor que el de Hong Kong. Hacia el principio de los años 90 eran casi equivalentes a las tasas de cambio vigentes.

<sup>12</sup> Parece que esta interpretación tiene un paralelo en el debate entre Hanushek (1989) y Hedges, Laine y Greenwald (1994). Ambos concuerdan en que el gasto escolar puede hacer una diferencia en el logro estudiantil cuando los fondos son asignados adecuadamente, pero los detalles precisos o los estándares para aquello que es “apropiado” son desconocidos mayormente como un fenómeno genérico. Es posible que sean mejor conocidos en los contextos específicos de los centros educativos.

contraposición con un mayor número de años de escolaridad o más grados académicos, incrementarán la productividad, no es nueva. Aunque no fue abordada o analizada directamente en los primeros estudios sobre capital humano, se asume implícitamente que una mayor productividad está ligada al incremento de destrezas y conocimientos que resultan de una mayor escolaridad (Becker, 1964).

Debido a que la manera más frecuente de medir lo aprendido ha sido el desempeño escolar en pruebas estandarizadas de logros, el enfoque se estrechó rápidamente al poder predictivo de los resultados de esas pruebas para explicar diferencias en la productividad de los trabajadores que se ven reflejadas en sus ingresos y en las calificaciones que les dan sus supervisores. Hace unos 20 años, una revisión de la evidencia disponible sobre tanto la validez concurrente como la predictiva de los puntajes de las pruebas cognitivas respecto a la productividad laboral, demostró que había muy poca relación (Levin, 1978). Si bien la discusión sobre los estándares, su formulación y su medición se han vuelto más ricos y sofisticados, no hay evidencia de que su validez predictiva respecto a la productividad laboral sea muy diferente. En lo que sigue, argumentaré que no hay más evidencia de que los actuales enfoques sobre la evaluación de desempeño hayan demostrado validez predictiva alta, pese a las sugerencias de que su uso creará una fuerza laboral más capaz<sup>13</sup>.

### **Prediciendo ingresos**

La bibliografía relacionada a este tema ha producido resultados decepcionantes, para decir lo menos, cuando se le compara con las expectativas y la sabiduría convencional. Históricamente se ha asociado una diferencia de una desviación estándar en los puntajes de las pruebas -- controlando por raza, género, nivel educativo y experiencia -- con un 3 o 4% de diferencia en los ingresos (Bishop, 1989). Debemos tener en cuenta que ningún proyecto educativo ha demostrado la capacidad de producir consistentemente mejorías de una desviación estándar en los puntajes de las pruebas, de manera que este criterio de ganancias educacionales debería ser visto como una exageración de lo que las políticas educativas pueden rendir. Por ejemplo, el ampliamente publicitado estudio de Coleman y Hoffer (1987) que pregonó el mayor éxito de las escuelas secundarias privadas sobre las públicas, encontró una diferencia típica de 0.06 en el logro a favor de los estudiantes de las escuelas privadas en las siete áreas evaluadas, aproximadamente equivalente a 12 puntos de la Prueba de Aptitud Escolástica (*Scholastic Aptitude Test - SAT*) combinando la parte verbal con la matemática.

Un estudio reciente ha suscitado la atención con su evidencia de que la relación entre logro estudiantil e ingresos parece estar creciendo. Murnane, Willett y Levy (1995), empleando la prueba con más larga relación estadística con los ingresos, encontraron que entre 1978 y 1986, el impacto estadístico de los puntajes en matemáticas sobre los ingresos se había elevado. Entre los hombres, el efecto aparente de un incremento de una desviación estándar en los puntajes de las pruebas había aumentado de una diferencia de 3% en las ganancias, a una de 7.4%; entre las

---

<sup>13</sup> Por ejemplo, Khattri y Sweet (1996,p.4) afirman que la presión por las evaluaciones de desempeño provino de los educadores para reemplazar las pruebas referidas a normas y basarlas en enfoques constructivistas del aprendizaje, como también “la preocupación por parte de la comunidad empresarial de que los estudiantes que ingresan a la fuerza laboral no eran lo suficientemente capaces como para competir en una economía crecientemente globalizada”. Sin embargo, aunque su ensayo está incluido en un volumen sobre implementación de evaluaciones de desempeño compilado por Kane y Mitchell (1996), el texto no contiene ninguna discusión sobre las especificidades de la validez predictiva de los estándares de evaluación de desempeño para el desempeño laboral.

mujeres el incremento fue de 8.5% a 15.5%. Pero adviértase que aun usando el más “poderoso” vaticinador de ingresos, el impacto aparente del incremento de los puntajes de las pruebas es minúsculo, especialmente cuando se considera lo que la política educativa puede ofrecer respecto a la elevación de los puntajes para la población. Por ejemplo, ¿qué pasaría si aplicamos los resultados de Coleman y Hoffer (1987) respecto a la superioridad de desempeño de los que asisten a escuelas secundarias privadas, para predecir la superioridad de ingresos de los graduados de las escuelas privadas sobre los de las escuelas públicas? Utilizando los estimados de Murnane, Willett, y Levy de 1986 (1995), los varones graduados de escuelas privadas recibirían 4 centavos más por hora que aquéllos de la escuela pública, debido a sus mejores puntajes y las mujeres procedentes de escuelas privadas ganarían 6 centavos más la hora. También es importante señalar que los hombres con puntajes equivalentes a una desviación estándar por encima del promedio de 1986, tenían ingresos que eran alrededor de \$1 menos por hora que el de los hombres con puntajes promedio 1978. Los salarios reales habían caído considerablemente a lo largo de este periodo y ni siquiera grandes mejorías en puntajes hubieran permitido recuperar las remuneraciones.

La débil relación entre los puntajes de las pruebas y los ingresos de los adultos reaparece virtualmente en todos los estudios sobre el tema y ya figuraba en trabajos tempranos como aquellos resumidos en Levin (1978). Este resultado se mantiene aun entre aquéllos que se encuentran en los últimos lugares de la distribución educacional. Por ejemplo, Berlin y Sum (1988) encontraron que cada grado escolar adicional culminado estaba asociado con un incremento salarial cuatro veces más grande que el equivalente a un año de destrezas básicas medidas por los puntajes de las pruebas. Y la culminación del último año de educación secundaria estaba asociada con un incremento diez veces mayor que el incremento de ingresos anuales que generaba una mejoría en los resultados de las pruebas equivalente a un grado escolar.

### **Prediciendo las calificaciones de los supervisores**

Dado el interés por hacer que los puntajes de las pruebas de los Estados Unidos sean comparables o superiores a los de aquéllos de nuestros principales competidores económicos, la evidencia que relaciona los puntajes de las pruebas con indicadores de productividad económica tales como las remuneraciones rinde resultados endebles. Pero en la medida en que los observadores discrepan de la teoría económica neoclásica según la cual en un mercado competitivo los ingresos reflejarán la productividad, resulta útil acudir a otras medidas de productividad. La principal medida alternativa son las calificaciones sobre el desempeño de un trabajador que hacen sus supervisores.

El esfuerzo más conocido por usar puntajes de pruebas para predecir el desempeño laboral se basó en la Batería de Pruebas de Destreza General (*General Ability Test Battery - GATB*), usada por el Departamento de Empleos de los Estados Unidos para clasificar a los trabajadores para posibles empleadores. La GATB incluye subpruebas de inteligencia, aptitud verbal y aptitud numérica, como también una serie de otras mediciones. Pero la validez predictiva de la GATB respecto a las calificaciones de los supervisores de los empleados en diferentes trabajos es del orden de 0.25, aun ajustando por errores muestrales, según un panel del Consejo Nacional de Investigación (*National Research Council -Hartigan & Wigdor, 1989*). Es decir, cerca del 6% de la varianza en las calificaciones de los supervisores puede ser explicada por diferencias en los puntajes GATB, pese a que el Departamento de Empleo de los EEUU ha usado los resultados de la GATB para clasificar a los candidatos a empleos sobre la base de su productividad predecible ante potenciales empleadores. Otras pruebas también parecen manejarse con este grado de exactitud en la predicción de las calificaciones de los supervisores (Schmitt, Gooding, Noe & Kirsch 1984).

Incluso estas simples correlaciones exageran el caso porque no toman en cuenta que los niveles educacionales (años de escolaridad) no están siendo controlados en los estudios de validez que constituyen la base de la investigación. Es decir, parte de la correlación observada entre los puntajes de las pruebas y la medida de productividad se debe, en realidad, a su covarianza con el impacto no medido del nivel educacional sobre ambas medidas.

Llama mucho la atención cómo estas correlaciones tan modestas son usadas por algunos psicólogos industriales para sostener que si se coloca en los puestos de trabajo primero a los candidatos con los más altos puntajes se elevará el producto en cientos de billones dólares (Levin, 1989). Primero, usan valores predictivos de 0.6 o más, a pesar de los hallazgos del Consejo Nacional para la Investigación respecto a que un valor más apropiado está cerca de 0.25 (Hartigan & Wigdor 1989). Segundo, asumen un inmenso número de candidatos para cada puesto de trabajo, a todo lo ancho del rango de habilidades, sin darse cuenta de que tales proporciones de selección requerirían una tasa de desempleo del rango del 90% para cumplir con esa condición. Tercero, calculan el valor económico de las diferencias en las calificaciones de los supervisores asumiendo que los supervisores laborales son capaces de proporcionar estimados precisos del valor económico del desempeño de los trabajadores con diferentes calificaciones, pese al hecho de que estudios econométricos y contables muy complejos no logran calcular el impacto económico de las diferencias de actividad del trabajador y que pocos supervisores están enterados de cualquiera de las variables financieras subyacentes. Cuarto, parten de un pequeño número de ocupaciones que han sido sometidas a estudios piloto y generalizan a cientos de ocupaciones que ellos dicen que combinan tareas que fueron analizadas en los estudios originales.

### **Los nuevos estándares**

He puesto de relieve la débil validez predictiva de los puntajes de las pruebas tradicionales respecto a la productividad económica porque existe un importante cuerpo de investigación que procura medir esas relaciones. Además, los puntajes de las pruebas estandarizadas han sido el “patrón de oro” para medir la calidad educativa en el pasado, y no deberíamos olvidar que aun constituyen la forma predominante de evaluación de la escuela y del estudiante. Más aun, se encuentran en el centro de la atención pública y tienen defensores importantes en las páginas editoriales de los medios de prensa, en las legislaturas del estado, en los consejos escolares y entre los administradores y diseñadores de políticas educacionales. Muchos defensores de los estándares utilizan los resultados de la Evaluación Nacional del Progreso Educativo (NAEP) o los diversos estudios internacionales de logros educacionales como principal evidencia para juzgar el nivel de preparación económica de los estudiantes de los EEUU (v.gr., Bishop, 1989). Aun cuando la NAEP y algunos de los estudios internacionales más recientes sobre logros educativos se basan en esfuerzos de evaluación más auténtica y se califican de diferente manera, hay poca evidencia de su validez predictiva respecto a los resultados económicos. El movimiento a favor de los nuevos estándares va más allá de estos proyectos, al intentar integrar los marcos de contenido curricular con evaluaciones auténticas de resultados medidos por productos de su trabajo, que son reflejos más realistas de los contextos y las situaciones que los escolares y adultos enfrentan, en vez de meros puntajes estandarizados de logro.

Antes de comentar sobre los criterios de evaluación como predictores del éxito económico, es importante señalar que estoy impresionado con la calidad del trabajo sobre los nuevos estándares publicado por el Centro Nacional para la Educación y la

Economía y la Universidad de Pittsburgh<sup>14</sup>. Su informe "Estándares de Desempeño" (***New Standards***, 1997<sup>a</sup>, 1997b) es rico en ideas, aplicaciones, ilustraciones, ejemplos de trabajo escolar y comentarios. Encuentro que son una fuente muy fértil para discusiones sobre curriculum, estándares y demostraciones de la proficiencia estudiantil. En este sentido, pueden constituir un gran activo para mi Proyecto de Escuelas Aceleradas (***Accelerated Schools Project***) (Hopfenberg, Levin, et al. 1993; Levin 1997) y otros movimientos reestructuradores de la escuela, al suministrar un foro de reflexión y una fuente de contribución para la construcción de estrategias de aprendizaje poderosas. Este es también el enfoque para el uso de los nuevos estándares sugerido por Darling Hammond (1994), quien los ve en el contexto de "informar" la definición de los estándares locales antes que forjar la adopción de unos externos. Es decir, ellos suministran la base para una discusión muy informada en las comunidades escolares "empoderadas" que deben establecer e implementar sus propios objetivos. Aun si una escuela decide embarcarse en la adopción de estos estándares y sus actividades y evaluaciones por entero, el desafío de la implementación sería abrumador. Los estándares son muy detallados y matizados, como deberían serlo, sirviendo -- en mi opinión -- de caudal fecundo de ideas útiles para la planificación, implementación y evaluación del curriculum y de la enseñanza. Desde mi punto de vista, éste es un papel más apropiado que su adopción en bloque, en su forma original, por todas las escuelas y profesores, una acción que es más probable que conduzca a una obediencia mecánica que a una aplicación reflexiva.

El atractivo de los nuevos estándares es que ellos parecen tener alta validez nominal respecto a muchas de las metas que tenemos para nuestros muchachos. Por ejemplo, sus unidades de aprendizaje aplicado abordan las metas establecidas por los informes SCANS (***Secretary's Commission on Achieving Necessary Skills*** 1991, 1992). En el trabajo etnográfico de nuestro proyecto sobre los "Requerimientos de trabajadores para organizaciones y tecnologías cambiantes" (***Worker Requirements for Changing Organizations and Technology***), observamos dimensiones que se superponen con aquellas del informe SCANS e intentamos traducirlas en actividades y metas educacionales (Rumberger & Levin, 1989). Creemos que hemos establecido un valor nominal para estos criterios, pero no una validez predictiva en el sentido tradicional.

Una razón por la cual se puede preferir este enfoque más amplio es que los requisitos laborales no son rígidos, cambian de un trabajo a otro y a lo largo del tiempo. Es más fácil establecer criterios generales para el éxito que atributos educacionales específicos en forma mensurable que vaticinarán el éxito en una amplia variedad de situaciones diferentes. Por ello, sabemos que el número de años de educación predice mucho mejor los ingresos y el éxito ocupacional que cualquier atributo educativo específico tal como los puntajes en las pruebas o la escuela de donde alguien procede. Esto es cierto aún después de controlar por raza, género, ubicación geográfica y status socioeconómico de los padres<sup>15</sup>. Hay algo respecto al nivel educativo (años de escolaridad) como fenómeno genérico en sí mismo, que parece aplastar lo específico. En mi opinión, esto explica en gran medida el fracaso en general de la educación vocacional, con su énfasis en destrezas específicas, en lo que respecta a lograr mucho

---

<sup>14</sup> De hecho, creo que deberían ser ampliamente accesibles a bajo costo y me pregunto porqué son vendidas comercialmente a precios altos, dados los subsidios y donaciones sustanciales que suministran las agencias gubernamentales y las principales fundaciones, presumiblemente para reducir su costos y promover su uso.

<sup>15</sup> Véase por ejemplo, las revisiones de los determinantes de los ingresos en Levy y Murnane (1992), de los premios a los niveles educacionales altos en Kevin Murphy y Finis Welch (1989), así como también las referencias en las notas 3 y 5.

impacto en los centros de trabajo<sup>16</sup>. Es probable que sean las destrezas genéricas con valor nominal las más valiosas y las que puedan ser moldeadas a través del entrenamiento y la experiencia laboral para adaptarse a las necesidades específicas de diferentes puestos de trabajos, diferentes centros laborales y diferentes tiempos, a medida que la tecnología, los productos y los servicios cambian a lo largo de la vida del trabajador.

Sin embargo, los sistemas de evaluación de desempeño y los estándares de desempeño son mucho más específicos que esto. Ellos están compuestos por tareas específicas en cada dominio supuestamente diseñadas para medir el grado de preparación de los escolares como trabajadores que ingresarán a la fuerza laboral. No deberíamos dejarnos convencer fácilmente de que esas evaluaciones, con sus propiedades psicométricas y mediciones aparentemente precisas, tendrán alta validez predictiva respecto a la productividad en el centro de trabajo. Sin duda se formularán esas alegaciones, pero hasta que no sean validadas, no deberíamos tener ninguna razón para creer que tienen una mayor capacidad de predicción que las anteriores mediciones o que se ha demostrado que están crucialmente vinculadas al futuro de la fuerza laboral y de la economía. Se requiere un escepticismo saludable hasta que se suministre la evidencia.

Una interesante ilustración de por qué es apropiado ser escépticos respecto a la idea de que nuestra fuerza laboral es seriamente deficiente y que los nuevos estándares vendrán a su rescate, es el enorme éxito que han tenido los fabricantes extranjeros que trasplantan sus operaciones a los EEUU. En *Una nación en riesgo* y otros informes se pensaba que tanto las firmas japonesas como las alemanas eran altamente exitosas debido a la calidad de sus sistemas educativos y la consecuente calidad de sus fuerzas de trabajo. Sin embargo cuando firmas japonesas como Honda, Toyota y Nissan establecieron sus operaciones en los EEUU empleando trabajadores locales de áreas difícilmente conocidas por la calidad de su educación, encontraron que podían producir automóviles con la misma eficiencia y con la misma alta calidad que en Japón (e.g., véase referencias en Levin & Kelley 1994; Levine 1996). Tal vez aun más cuestionador de la visión ingenua de que los puntajes altos en las pruebas y los estándares educacionales son los principales determinantes de la productividad de la fuerza laboral, es que la BMW y la Mercedes han invertido cientos de millones de dólares en el establecimiento de plantas de ensamblaje automotriz en áreas relativamente rurales de Carolina del Sur y Alabama, respectivamente, dos estados que ocupan los últimos lugares de la jerarquía educacional de los EEUU respecto a puntajes en pruebas e inversión educacional. Sin embargo, se espera que ambos esfuerzos sean altamente lucrativos, tanto en el mercado doméstico como el exportador, incluyendo automóviles para el mercado europeo. Tal vez una ventaja de la fuerza laboral estadounidense es su maleabilidad para satisfacer requerimientos de distintos sistemas de producción, antes que las destrezas específicas que posee en un momento determinado.

Probablemente, lo más importante es la manera en que esas firmas están organizadas y administradas, como también una alta inversión de capital, un buen producto, un proceso eficiente de producción, el énfasis en el trabajo en equipo, una

---

<sup>16</sup> Por esta razón, el presente enfoque de la educación vocacional es integrar más la preparación académica y la ocupacional. Esto está más explicado en las publicaciones del Centro Nacional de Investigación sobre Educación Vocacional de la Universidad de California, Berkeley. Por ejemplo, véase Grubb (1955<sup>a</sup>, 1955<sup>b</sup>), Benson (1997) y Steinberg (1997). Los magros o incluso negativos rendimientos de la educación vocacional se encuentran en toda la bibliografía pertinente. Para un buen estudio reciente que documenta este resultado, véase Crawford, Johnson, y Summers (1997).

estabilidad laboral razonable y un entrenamiento suficiente de los trabajadores para sus puestos de trabajo (Brown, Reich, & Stern 1993; Ichniowski, Shaw, & Prennushi, en prensa; Levin & Kelley 1994; Levine 1996). La organización y la administración dan especial énfasis a los incentivos para trabajar productivamente en equipo y para recompensar la calidad en la producción. El entrenamiento es intensivo y continuo, actualizando constantemente las destrezas de los trabajadores, incrementando no sólo el valor del trabajador para la firma, sino incrementando el vínculo del trabajador a través de destrezas específicas para el puesto de trabajo que pueden no ser directamente transferibles a otras firmas (Becker 1964; Mincer 1974; Ziderman 1978, cap. 2).

## Resumen

Por el momento, no hay estándares específicos de evaluación de desempeño que hayan sido validados como sólidos predictores de la productividad económica o la calidad de la fuerza laboral, a pesar de constituir ambos argumentos la justificación fundamental de los estándares. En el mejor de los casos, se podrían identificar estándares generales que tienen validez nominal respecto a las necesidades de los trabajadores futuros, como aquellos establecidos por la Comisión de la Secretaría para el Logro de Destrezas Necesarias (1991). Estos están propuestos también con un grado alto de generalidad, de manera que los que toman las decisiones locales en los distritos y centros escolares puedan desarrollar sus propios enfoques para lograr los estándares. Las empresas pueden crear programas de entrenamiento que provean las destrezas específicas y que pueden aprovechar los fundamentos más generales.

Estos hallazgos conducen a varias observaciones más generales. Primero, la percepción de una debilidad económica nacional es un móvil mucho más poderoso para poner en marcha una reforma educativa y nuevos estándares que las amenazas percibidas a la cultura, la ciudadanía, la equidad o el deterioro educacional. Los numerosos puntos de vista recalcitrantes y contendientes respecto a lo último parecen aglutinarse fácilmente en torno a lo primero. Segundo, la creencia de que el progreso económico depende crucialmente del establecimiento de estándares de desempeño educacionales para la futura “economía de información” es tan fuerte que no necesita ninguna evidencia para ganar adherentes. Tercero, aunque hay claramente alguna relación entre el desempeño escolar y la productividad del centro de trabajo, parece ser mucho más modesta que lo usualmente asumido en nuestro sistema de creencias. Cuarto, un desafío muy importante para los promotores de los estándares educacionales y la productividad laboral será el generar estudios de buena calidad para validar los supuestos que encierran las pruebas.

Finalmente, es peligroso reforzar la creencia de que los estándares de desempeño educacional existentes y los que están emergiendo tienen altos poderes predictivos con relación a la productividad laboral. Esa creencia podría llevar a establecer un sistema en el ámbito local, estatal o federal que certifique postulantes a empleos que cumplen con unos “nuevos estándares” arbitrariamente adoptados y es posible que los empleadores creen que esos estándares han sido validados respecto a la productividad laboral. Por cierto, esto es lo que el Departamento de Empleo de los EEUU intentó hacer cuando usó el orden de méritos de GATB como base para recomendar postulantes a los empleadores potenciales<sup>17</sup>. Si los empleadores creen que estos sistemas de evaluación y medición son válidos, es posible que ellos menosprecien otras medidas que usarían normalmente, como los *curriculum vitae*, las

---

<sup>17</sup> Véase Hartigan y Wigdor (1989). He estado presente en discusiones entre empleadores y educadores insistiendo en que las pruebas Regents del estado de Nueva York sean usadas de esta manera.

entrevistas y las recomendaciones – criterios que son comparables a los puntajes en pruebas en cuanto a su habilidad predictiva (Schmitt, Gooding, Noe, & Kirsch 1984) y probablemente muy superiores cuando se usan conjuntamente. Dada la falta de evidencia respecto a la validez predictiva de las nuevas mediciones y la baja validez predictiva de las mediciones tradicionales, venerar a los estándares como base para la selección para el empleo está preñado de gran peligro. Esa política, en efecto, tiene un fuerte potencial para reducir de un solo golpe tanto la eficiencia económica nacional como la equidad en la selección para el empleo, precisamente lo opuesto a lo que se reclama y se intenta hacer.

## Referencias

- Bailey, T., & Merritt, D. (1997). Industry skill standards and education reform. *American Journal of Education*, 105, 401-436.
- Becker, G. (1964). Human capital: *A theoretical and empirical analysis*. Nueva York: Columbia University Press.
- Benson, C. S. (1997). New vocationalism in the United States: Potential problems and outlook. *Economics of Education Review*, 16, 201-212.
- Berlin, G., & Sum, A. (1988) *Toward a more perfect union: Basic skills, poor families and our economic future*. (Occasional Paper 3). Nueva York: Ford Foundation Project on Social Welfare and the American future.
- Betts, J.R. (1996). *The impact of educational standards on the level and distribution of earnings* (Discussion Paper 96-29). La Jolla, California: Departamento de Economía, Universidad de California, San Diego.
- Bishop, J.(1989) Incentives for learning: Why American high School students compare poorly to their counterparts overseas. En Commission on Workforce Quality and labor Market Efficiency, Departamento de Trabajo de los E.E.U.U. (ed.), *Investing in people: Background papers* (Vol. 1, pp. 1-84). Washington, DC: U.S. Government Printing Office.
- Blackburn, M.,& Neumark, D. (1995) Are OLS estimates of the return to schooling biased of Economics and downward? *Another look. Review Statistics*. 77, 217-229.
- Brown, C., Reich, M., & Stern, D. (1993). Becoming a high-performance work organization: The role of security, employee involvement, and training. *The International Journal of Human Resource Management* 4, 247-275.
- Cohn, E. & Geske, T.G. (1990). *The economics of education* (3ra. Ed.) Nueva York: Pergamon Press.
- Coleman, J., & Hoffer, T. (1987) *Public and private high schools*. Nueva York: Basic Books.
- Costrell, R. (1994) A simple model of educational standards. *American Economic Review*. 84, 956-971.
- Crawford, D.L., Johnson, A.W., & Summers, A.A. (1997). Schools and labor market outcomes. *Economics of Education Review*. 16, 255-269.
- Darling-Hammond, L. (1994). National standards and assessment: Will they improve education? *American Journal of Education*. 102, 478-510.
- Grubb, W.N. (ed.). (1995<sup>a</sup>) *Education through occupations in American high Schools: Approaches to integrating academic and vocational education*. (Vol. 1). Nueva York: Teachers College Press.
- Grubb, W.N. (ed.). (1995<sup>b</sup>) *Education through occupations in American High schools: The challenges of implementing curriculum integration*. (Vol. 2). Nueva York: Teachers College Press.
- Hanushek, E. (1989) The impact of differential expenditures on school performance. *Educational Researcher*. 18(4), 45-62.
- Hanushek, E. (et al.). (1994). *Making schools work*. Washington, DC: The Brookings Institution.
- Hartigan, J.A., & Wigdor, A.K. (eds.). (1989). *Fairness in employment testing: Validity generalization, minority issues, and the general aptitude test battery*. Washington, DC: National Academy Press.
- Hedges, L., R., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*. 23(3), 5-14.
- Hirshc, E. D., Jr. (1987). Cultural literacy: *What every American needs to know*. Boston: Houghton Mifflin.

Hopfenberg, W., Levin, H. M., Chase, C., Christensen, G., Moore, M., Soler, P., Brunner, I., Keller, B., & Rodriguez, G. (1993). ***The accelerated schools resource guide***. San Francisco: Jossey-Bass.

Ichniowski, C., Shaw, K., & Prennushi, G. (en prensa). The impact of human resource management practices on productivity: A study of steel finishing lines. ***American Economic Review***. (También disponible como National Bureau of Economic Research (NBER) trabajo No.5333. (1995). Cambridge, MA: Author)

Kane, M., & Mitchell, R. (eds). (1996). ***Implementing performance assessment***. Mahwah, NJ: Lawrence Erlbaum Associates.

Kerr, C. (1991, febrero 7). Is education really all that guilty? ***Education Week***. (pp. 30).

Khatti, N., & Sweet, D. (1996). Assessment reform: Promises and challenges. En M.B. Kane & R. Mitchell (eds.), ***Implementing performance assessment*** (pp. 1-22). Mahwah, NJ: Lawrence Erlbaum Associates.

Kim, J. I., & Lau, L. (1996). The sources of economic growth in the East Asian newly industrialized countries [número especial]. ***Canadian Journal of Economics***. 29, S448-S454.

Krugman, P. (1994) The myth of Asia's miracle. ***Foreign Affairs***. 73, 62-78.

Levin, H. M. (1978). Educational performance standards: Image or substance? ***Journal of Educational Measurement***. 15, 309-319.

Levin, H. M. (1989). Ability testing for job selection: Are the economic claims justified? En B. Giffor (ed.), ***Test policy and the politics of opportunity allocation: The workplace and the law***. (pp. 211-232). Boston: Kluwer Academic Publishers.

Levin, H. M. (1997). Raising school productivity: An efficiency approach. ***Economics of Education Review***. 16, 303-311.

Levin, H. M., & Kelley, C. (1994). Can education do it alone? ***Economics of Education Review***. 13, 97-108.

Levine, D. I. (1996). ***Reinventing the workplace: How business and employees can both win***. Washington, DC: The Brookings Institution.

Levy, F., & Murnane, R. (1992). U. S. earnings levels and earnings inequality: A review of recent trends and proposed explanations. ***Journal of Economic Literature***. 30, 1333-1381.

Linn, R. L. (1996). Work readiness assessment: Questions of validity. En L. B. Resnick & J. G. Wirt (eds), ***Linking school and work: Roles for standards and assessment***. (pp. 245-266). San Francisco: Jossey-Bass Publishers.

Mincer, J. (1974). ***Schooling, experience and earnings***. Nueva York: Columbia University Press.

Murnane, R. J., Willett, J.B., & Levy, F. (1995). The growing importance of cognitive skills in wage determination. ***Review of Economics and Statistics***. 77, 251-266.

Murphy, K., & Welch, F. (1989). Wage premia for college graduates: Recent growth and possible explanations. ***Educational Researcher***. 18(4), 17-26.

National Commission on Excellence in Education. (1983). ***A nation at risk***. Washington, DC: U.S. government Printing Office.

"New standards". (1997<sup>a</sup>). ***Performance standards: English language arts, mathematics, science, and applied learning***. Vol. 1. ***Elementary schools***. Vol. 2. ***Middle school***. Vol. 3. ***High school***. Washington, DC: National Center on Education and the Economy.

Resnick, L. B., & Wirt, J. G. (eds.). (1996). ***Linking school and work: Roles for standards and assessment***. San Francisco: Jossey-Bass.

Rumberger, R., & Levin, H. M. (1989). Schooling for the modern workplace. En Commission on Workforce Quality and Labor Market Efficiency, Departamento de Trabajo de los E.E.U.U. (eds.), ***Investing in people: Background papers***. (Vol. 1, pp. 85-144). Washington, DC: U.S. Government Printing Office.

Schmitt, N., Gooding, R.Z., Noe, R.A., & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology*. 37, 407-422.

Secretary's Commission on Achieving Necessary Skills, U.S. Department of Labor. (1991). *What work requires of schools*. Washington, DC: U.S. Department of Labor.

Secretary's Commission on Achieving necessary Skills, U.S. Department of Labor. (1992). *Learning a living: A blueprint for high performance - A SCANS report for America 2000*. Washington, DC: U.S. Department of Labor.

Steinberg, A. (1997). *Real learning, real work*. Nueva York: Routledge.

Wiley, D. E., & Haertel, E. H. (1996). Extended assessment tasks: Purposes, definitions, scoring, and accuracy. En M.B. Kane & R. Mitchell (eds.), *Implementing performance assessment*. (pp. 61-90). Mahwah, NJ: Lawrence Erlbaum Associates.

Ziderman, A. (1978). *Maspower training: Theory and policy* (Macmillan Studies in Economics). Londres: Macmillan Press.