

Translated by Camilo Hurtado-Parrado with permission from SEAB and the authors.

Original article: Schwartz, I. S., & Baer, D. M. (1991). SOCIAL VALIDITY ASSESSMENTS: IS CURRENT PRACTICE STATE OF THE ART? *Journal of applied behavior analysis*, 24, 189-204. <https://doi.org/10.1901/jaba.1991.24-189>.

Journal of Applied Behavior Analysis

1991, 24, 189-204

NÚMERO 2 (VERANO 1991)

## Evaluaciones de Validez Social: ¿Qué tan actualizada está su implementación?<sup>1</sup>

ELENE S. SCHWARTZ AND DONALD M. BAER

UNIVERSITY OF KANSAS

El uso de los comentarios evaluativos de los consumidores para orientar la planificación y evaluación de programas suele denominarse valoración de la validez social. Han surgido diferentes puntos de vista sobre su papel y valor en el análisis aplicado de la conducta, y cada vez son más comunes las evaluaciones estereotipadas de la validez social. Este artículo argumenta que las aplicaciones actuales de las evaluaciones de validez social se están desviando del punto originalmente propuesto para ellas. Por lo tanto, se proponen varias sugerencias para mejorar la actual evaluación de la validez social, incluyendo (a) la ampliación de la definición de los consumidores para el conocimiento de la variedad de miembros de la comunidad capaces y susceptibles de afectar a la supervivencia de un programa, (b) aumentar el rigor psicométrico de las evaluaciones de validez social, (c) ampliar la evaluación a poblaciones hasta ahora subrepresentadas, (d) generalizar la aplicación de evaluaciones de validez social bien diseñadas, (e) aumentar la participación significativa de los consumidores en la planificación y evaluación de programas conductuales, y (f) educar a los consumidores para que tomen decisiones mejor informadas sobre los programas.

*Palabras clave:* validez social, medición, evaluación, metodología, diseminación.

En 1978, Wolf introdujo formalmente la cuestión de la validez social en el campo del análisis aplicado de conducta. Esencialmente, observó cómo raramente se había preguntado a los consumidores de programas de análisis de conducta sobre su aceptación de los procedimientos, objetivos y personal de un programa; advirtió con base en su experiencia que la no aceptación podría preceder a un desastroso rechazo de los programas por parte de los

consumidores; y recomendó una cuidadosa evaluación futura de la satisfacción del consumidor desde ese punto de vista (Wolf, 1978). El análisis aplicado de conducta ha progresado ampliamente desde 1978 y, de hecho, la validez social ha sido a menudo de interés primordial para los estudiosos que trazan la historia del campo y predicen su futuro.

Sin embargo, la opinión sobre la contribución de la validez social a la disciplina

---

<sup>1</sup> La preparación de este artículo ha sido financiada en parte por becas del Departamento de Educación de EE.UU. (HO24J80003) y del Instituto Nacional de Salud Infantil y Desarrollo Humano (HO03144).

Expresamos nuestro agradecimiento a Julie Cross Hoko y Mary Todd por su ayuda en la preparación de este manuscrito, y a Susan A. Fowler, Kathryn K. Ramp, Montrose M. Wolf, Edward K. Morris y Deborah Altus por sus útiles comentarios sobre versiones anteriores de este trabajo. Las solicitudes de copias deben dirigirse a Ilene S. Schwartz, Juniper Gardens Children's Project, 1614 Washington Blvd., Kansas City, Kansas 66102, o a Donald M. Baer, Department of Human Development, University of Kansas, Lawrence, Kansas 66045.

dista mucho de ser unánime. La validez social ha sido considerada por algunos como una guía importante para el futuro del campo (e.g., Baer, Wolf & Risley, 1987) y denunciada por otros como una desviación de la naturaleza científica de nuestra práctica investigadora (e.g., Barrett, 1987). A pesar de estas diferencias, las medidas de validez social se están convirtiendo en algo casi habitual en la literatura sobre conducta.

En este artículo se analiza la práctica actual de la evaluación de la validez social, se proponen elementos que deberían incluirse en cualquier procedimiento de evaluación de la validez social, se ofrece una taxonomía propedéutica de los consumidores para orientar la elaboración y difusión de estas evaluaciones, y se sugieren estrategias que pueden ayudar a configurar la práctica actual para producir evaluaciones más precisas de las conductas asociadas a la validez social.

### **¿Qué es la Validez Social?**

El propósito de las valoraciones de validez social es evaluar la aceptabilidad o viabilidad de una intervención programada. La mayoría de las veces, la evaluación de la validez social se realiza pidiendo a personas distintas de los planificadores o experimentadores del programa que rellenen algún tipo de cuestionario (Kazdin, 1977; Wolf, 1978). El objetivo de estas evaluaciones es anticipar el rechazo de un programa antes de que se produzca; por lo tanto, en las evaluaciones deben participar todos los consumidores relevantes del programa. Se debe preguntar a los consumidores sobre la aceptabilidad de los objetivos del programa, los métodos, el personal, los resultados y la facilidad de integración de los componentes del programa en el estilo de vida actual de los consumidores. A continuación, esta

información debe utilizarse inmediatamente, así como en la futura planificación, implementación y evaluación del programa. Así pues, las evaluaciones de la validez social no pretenden ser, y de hecho no son, medidas dependientes primarias; lógicamente, sólo pueden complementar las medidas objetivas de la conducta que son las medidas dependientes primarias (cf. Barrett, 1987; Dietz, 1978; Michael, 1980; Pierce & Epling, 1980; Scheirer, 1978).

Idealmente, la evaluación de la validez social es un proceso que consta de dos partes: en primer lugar, se recoge una muestra precisa y representativa de las opiniones de los consumidores; a continuación, esa información se utiliza para mantener prácticas satisfactorias o efectuar cambios en el programa para mejorar su viabilidad en la comunidad. El fracaso en la aplicación de cualquiera de las partes de este proceso socava el concepto de validez social y, a menudo, también desacredita la evaluación de la validez social. Una evaluación de la validez social útil no debería (a) evaluar las opiniones de una comunidad limitada (o equivocada), (b) evaluar incorrectamente las opiniones de la comunidad relevante, o (c) evaluar correctamente las opiniones de la comunidad pertinente, pero no utilizar esa información para cambiar el programa y/o la opinión de sus consumidores sobre él. En resumen, es igual de inadecuado muestras las críticas y los elogios y no responder a ellos como responder a ellos sin investigar sus fuentes y sus correlatos (i.e., su exactitud).

Recabar la opinión de los consumidores sobre los programas propuestos y en curso puede establecer lazos dosificadores entre la comunidad y el programa; esto puede dar al público de un programa una sensación de control compartido, lo que a

su vez puede evitar las quejas y el abandono del programa (Baer et al., 1987; Giordano, 1977; Lebow, 1982). Sin embargo, es fundamental hacer que esta información sea funcional, incorporándola a la programación futura y explicando a los consumidores cómo funciona el proceso de retroalimentación (Warfel, Maloney & Blase, 1981). Solicitar repetidamente opiniones sin explicar cómo se utilizarán y sin poner en práctica ninguna de las sugerencias puede enseñar a los consumidores que sus opiniones no son útiles y que los solicitantes son engañosos. Esto puede invalidar futuros intentos de evaluar la validez social y, a su vez, amenazar la viabilidad del programa (i.e., la participación y adopción del programa por parte de los consumidores).

### **¿Qué es la Invalidez Social?**

La invalidez social no es simplemente la ausencia de evaluaciones positivas por parte de los consumidores y la comunidad relevante cuando se les pregunta; no es simplemente no preguntar; no es simplemente preguntar y luego ignorar lo que se oye; y no es simplemente lo inverso de la validez social. Baer (1987) describió la invalidez social como la conducta de los consumidores que no sólo desaprueban algún componente del programa en curso, sino que van a hacer algo sobre su desaprobación. Ese algo puede consistir en retirarse del programa, alentar a otros a hacer lo mismo, quejarse a los de la comunidad y los medios de comunicación o, más sutilmente, no aplicar algunos o todos los procedimientos del programa, a pesar de las respuestas positivas en los cuestionarios.

Los consumidores descontentos no son miembros de un grupo homogéneo; existe una gran variación en las causas y manifestaciones del descontento (Aaker & Day, 1971). Es decir, algunos harán algo más

respecto a su descontento, otros no; y, de los que hacen algo más, sus manifestaciones de desaprobación pueden variar mucho. Por lo tanto, los casos de invalidez social deben evaluarse individualmente para rastrear el origen a variables específicas del programa o del entorno y explorar qué ocurrirá, si se modifica el programa en respuesta a en respuesta a esa retroalimentación y también si no se hace.

La invalidez social es difícil de predecir a partir de los resultados de las evaluaciones típicas de validez social. El hallazgo más consistente en las revisiones de la literatura sobre satisfacción del consumidor en servicios de salud mental, servicios médicos y terapia de conducta es que los consumidores generalmente valoran los programas de tratamiento de forma positiva (Bornstein & Rychtarik, 1983; Fuqua & Schwade, 1986; Lebow, 1982; McMahon & Forehand, 1983; Ware, Davies-Avery, & Stewart, 1978). Los consumidores, incluso cuando se les da la oportunidad de quejarse, parecen conformes, especialmente con los servicios médicos y de salud mental (Lebow, 1982; Ware, 1978). Sin embargo, la experiencia común sugiere que, a pesar de las valoraciones positivas, algunos de esos programas son rechazados. Algunos se rechazan crudamente, y muchos más sutilmente, sobre todo en forma de interrupción de los procedimientos del programa cuando personal del programa lo abandona. Así pues, esa aprobación verbal fue menos que precisa y, por tanto, menos que válida; llamarla validez social es frívolo, y basarse en ella dificulta la predicción de la invalidez social.

La predicción de la validez e invalidez social requiere evaluaciones que evocan la información veraz de lo que a los consumidores les gusta y no les gusta de un programa. El primer objetivo de las

evaluaciones de validez social debe ser reunir información precisa y útil sobre posibles problemas, no fomentar falsos elogios de los consumidores. Los consumidores descontentos deben quejarse, y que lo más pronto posible, al personal del programa. De hecho, puede que los consumidores que hacen públicas sus quejas suelen estar más satisfechos con la resolución del problema que los consumidores que se quejan en privado o no lo hacen en absoluto (Bearden & Oliver, 1985). Las respuestas positivas a las evaluaciones de validez social refuerzan al personal del programa y, a veces, complacen o apaciguan organismos reguladores o de financiación, pero si lo hacen a costa de dejar en silencio a un consumidor descontento, el reforzamiento inicial implícito en los consumidores que no se quejan puede sustituirse por el castigo del fracaso del programa (Bornstein & Rychtarik, 1983).

### **Qué no es la Validez e Invalidez Social**

La variedad de significados atribuidos a la validez social ha crecido más allá de su utilidad y se ha convertido en confusión. Su significado original era, y sigue siendo sencillo: Al aplicar programas en entornos reales, evalúe con antelación hasta qué punto esos programas serán para el público pertinente. La evaluación de la validez social es una técnica defensiva. Está orientada hacia la detección de inacceptabilidad en cualquiera de las tres fuentes principales: los objetivos del programa, sus métodos y su personal.

Pero casi inmediatamente después de la introducción de este concepto, surgieron aplicaciones bastante diferentes de las evaluaciones de la validez social, y en los últimos años su variedad ha florecido. El temor a que los analistas de la conducta sustituyeran las medidas de validez social, obviamente subjetivas, por las medidas de

resultados de los programas, obviamente objetivas, es una de las críticas más antiguas a esta metodología. Sin embargo, nunca se pretendió que éste fuera el objetivo de las evaluaciones de la validez social. De hecho, las evaluaciones de validez social siempre se propusieron en el supuesto de que un programa conductual tenía conductas objetivo distintas a las declaraciones de agrado o desagrado por los objetivos, los procedimientos y el personal del programa. Siempre se supuso que esas conductas objetivo se medirían de la forma más directa, serían conductuales, objetivas y fiables como fuera posible, y que la validez social se evaluaría como una segunda cuestión importante, relevante no para la eficacia del programa, sino para su viabilidad (cf. Kazdin, 1977; Wolf, 1978).

Sin embargo, irónicamente, parece que surgió casi de inmediato una contingencia implícita para producir medidas positivas de validez social, como si fueran conductas objetivo. Los desarrolladores del programa parecían asumir que un criterio importante para la publicación de su programa era una evaluación positiva de la validez social. No se sabe lo que esta contingencia implícita puede haber hecho a la exactitud de esos intentos de evaluación de la validez social. Por supuesto, un programa que logre sus objetivos de forma exhaustiva y efectiva en cuanto a costes, y que además guste a sus consumidores, es sin duda el objetivo último del análisis aplicado de la conducta. Sin embargo, el objetivo de la evaluación de la validez social es identificar, de entre todos los programas que logran sus conductas objetivo de forma completa y efectiva en cuanto a costes, y también de entre todos los programas que no tienen tanto éxito, cuáles gustan menos a los consumidores que

otros. Los dos puntos de la evaluación de la validez social son:

1. Es importante para el avance y la supervivencia del análisis aplicado de la conducta saber de antemano qué programas gustan y cuáles no, por lo que la publicación de evaluaciones negativas de la validez social es tan valiosa e importante como la publicación de las positivas.

2. Es importante iniciar el análisis de por qué algunos programas gustan y otros no, para que, con el tiempo, la evaluación de la validez social pueda convertirse en una predicción calculada en lugar de una alerta temprana o un aval evaluado empíricamente. Sin la publicación de una amplia muestra de los casos negativos para compararlos con los positivos, no es probable que se alcance este segundo objetivo.

Si la experiencia del análisis de la conducta es, en efecto, que todos sus programas parecen socialmente válidos, que es lo que podría deducirse de su literatura hasta ahora, entonces es casi seguro que la validez social se está evaluando a menudo de manera espuria. El análisis aplicado de la conducta simplemente no puede ser tan bueno. Es más probable que las técnicas actuales de evaluación de la validez social sean a menudo social y psicométricamente inválidas. Están exigiendo implícitamente a sus consumidores que respondan positivamente a sus breves, sencillas, anodinas y poco exigentes escalas de 7 puntos. Estamos animando a los consumidores a "fingir ser buenos", en la terminología de los investigadores del desarrollo de pruebas que llevan al menos cinco décadas sufriendo problemas similares.

En un reciente panel de discusión de la *Association for Behavior Analysis* (ABAi, Asociación para el Análisis de Conducta) sobre los usos actuales de las evaluaciones de validez social, Cataldo (comunicación

personal, mayo de 1990) señaló una contingencia bastante similar:

La necesidad de refutar por parte de los patrocinadores del programa, los financiadores y cualquier otra audiencia atenta a la crítica frecuente de que los consumidores modernos no gustan de los programas de análisis de conducta. De hecho, tanto los medios de comunicación como nuestros competidores conceptuales sugieren a menudo que los programas de conducta no son congruentes con el *Zeitgeist* contemporáneo del mentalismo, autonomía personal y la necesidad de analizar cada problema social para encontrar su origen en los fallos de sus clientes. Enfrentados a esta necesidad, las evaluaciones de validez social pueden parecer una defensa útil. Pero, como en el caso anterior, la naturaleza de esa contingencia es reforzar las valoraciones positivas y castigar las negativas, degradando así la precisión de esas técnicas de validez social. Por lo tanto, es una contingencia destructiva para la función original de proporcionar alertas tempranas del rechazo del programa por parte de los consumidores del programa.

En la misma mesa redonda de ABA, Geller (comunicación personal, mayo de 1990) sugirió que las evaluaciones de validez social a veces se ven ahora como técnicas para descubrir los nuevos e importantes objetivos sociales a los que deberían dirigirse los programas de análisis de conducta (y otros) y para reafirmar la importancia social de ciertos objetivos pasados aún no alcanzados a escala social. Si las evaluaciones de la validez social van a utilizarse realmente con este fin, surgirán inmediatamente dos problemas:

1. Si el objetivo de las evaluaciones de validez social es detectar el rechazo del programa (por parte de quienes pueden y es probable que lo hagan), el muestreo debe tener como objetivo representar sólo

a esas poblaciones. Hasta ahora eso es relativamente fácil y barato en nuestra pequeña escala de investigación y aplicación. Pero, si de lo que se trata es de la importancia social, entonces cualquier concepto de importancia social derivado de los deseos de la propia sociedad requerirá una muestra aleatoria de la sociedad en su conjunto, y esto es algo extraordinariamente difícil y caro.

2. Sin embargo, muchos conceptos de importancia social son ortogonales, y a veces antitéticos, a los deseos momentáneos de la sociedad en su conjunto. La verdadera liberación de la mujer y la igualdad funcional de las minorías raciales, étnicas y religiosas son ejemplos de objetivos programáticos que habrían mostrado poca validez social cuando empezaron y siguen mostrando un apoyo poco unánime. Sin embargo, ¿quién podría declarar ahora que son objetivos sin importancia social? Así pues, la importancia social tiende a resolverse, y quizá siempre deba resolverse, mediante el proceso político habitual de una sociedad democrática o incluso parcialmente democrática, en lugar de mediante algo parecido a evaluaciones de validez social.

### **Evaluando la Validez Social**

#### *Identificando los consumidores del programa*

Qué preguntar a su público, quién constituye su audiencia y cómo evaluar a su público de forma fiable son cuestiones clave en las evaluaciones de validez social. Las revisiones de esta literatura muestran un acuerdo general sobre las preguntas que hay que hacer; sin embargo, las cuestiones sobre a quién hay que preguntar y cómo hacerlo no son tan apreciadas (e.g., Fuqua & Schwade, 1986; Mc-Mahon & Forehand, 1983). Wolf (1978) resumió sucintamente tres tipos de preguntas a formular:

1. ¿Son los objetivos de los procedimientos importantes y relevantes para los cambios deseados en el estilo de vida?

2. ¿Las técnicas utilizadas son aceptables para los consumidores y la comunidad, o cuestan demasiado (e.g., en términos de esfuerzo, tiempo, incomodidad, ética o similares)?

3. ¿Están satisfechos los consumidores con el resultado, tanto con el cambio de conducta previsto como con cualquier efecto secundario imprevisto?

Una vez elegidas las preguntas para evaluar la validez social, surge la cuestión más difícil: ¿a quién preguntar? Existe un consenso general en que se debe preguntar a los "consumidores del programa" (Bornstein y Rychtarik, 1983; Kazdin, 1977; Larsen, Attkisson, Hargreaves y Nguyen, 1979; Lebow, 1982; Wolf, 1978). Sin embargo, el problema reside en identificar a estos consumidores, especialmente a los que controlan la viabilidad, directa o indirecta, del programa. En la actualidad, los consumidores suelen ser simplemente los destinatarios del programa. Pero bien podría ser cierto que las personas o agencias que contrataron el programa y ciertos miembros de la comunidad en general también son consumidores, especialmente si el criterio es el control de la viabilidad del programa. Por ejemplo, los vecinos de al lado de un programa de hogar para adultos con discapacidades de desarrollo se convierten en su público, y pueden quejarse a la comunidad en general si no les gusta. Los cónyuges de los trabajadores de una fábrica pueden apoyar un programa de seguridad si les gusta, o pueden criticar destructivamente el procedimiento si no les gusta. Los compañeros de un niño agresivo o socialmente retraído pueden formar parte de un programa para cambiar las conductas del niño o pueden sabotearlo en cuanto el personal del programa se

ausenta. ¿No son los contribuyentes consumidores de todos los programas financiados con fondos públicos?

Las características clave de estos ejemplos son que muchas personas distintas de los destinatarios del programa son consumidores pasivos de los programas de tratamiento, y si deciden convertirse en consumidores activos, pueden apoyar o criticar el programa. En la actualidad casi no disponemos de análisis sobre lo que convierte a los no receptores en consumidores activos y lo que hace que estos consumidores apoyen o critiquen los programas.

En el mundo empresarial, un consumidor es cualquier persona que compra bienes o servicios o hace que se compren (Engel, Kollat & Blackwell, 1973; Troelstrup, 1974). La amplitud de esta definición es apropiada tanto para los negocios como para la aplicación de programas conductuales, pero los términos consumidor y compra necesitan un mayor desarrollo antes de que puedan contribuir a aumentar la precisión de las evaluaciones de validez social. Para ello, puede ser útil clasificar a los consumidores en consumidores directos, consumidores indirectos, miembros de la comunidad inmediata y miembros de la comunidad ampliada. La pertenencia a estas categorías podría determinarse en función de los siguientes criterios:

*Consumidores directos.* Los consumidores directos son los destinatarios primarios de la intervención del programa. Pueden haber comprado o "contratado" el programa o pueden haber sido remitidos por otra persona. (Así, un niño con discapacidades de desarrollo es a menudo el consumidor directo, aunque rara vez "contrata" el programa). Estos consumidores pueden afectar a la viabilidad del programa directamente y en cualquier momento, participando o mediante negativas

selectivas o generalizadas a participar. Ejemplos de consumidores directos son los padres que participan en un programa de formación de padres, los monitores y aprendices de un programa de monitorización por pares, los conductores que son remitidos a un programa de seguridad vial debido a su historial de infracciones de tráfico y los clientes de un restaurante que participan en un programa para aumentar la selección de platos bajos en grasa.

*Consumidores indirectos.* Los consumidores indirectos compran o contratan el programa para otra persona o se ven muy afectados por el cambio de conducta al que se dirige la intervención, pero no son sus destinatarios. Los consumidores indirectos pueden afectar directamente a la viabilidad del programa al seguir comprándolo o al negarse a hacerlo, e indirectamente al comportarse como clientes satisfechos, cónyuges o amigos de los consumidores directos y, por tanto, como publicidad del programa. Ejemplos de consumidores indirectos incluyen los padres de un niño con discapacidades de desarrollo que aprende a vestirse de forma independiente, los administradores de una empresa que encarga un programa para mejorar las condiciones de seguridad en una fábrica y los familiares de los participantes en un programa de climatización de viviendas.

Miembros de la comunidad inmediata. La comunidad inmediata son aquellas personas que interactúan con los consumidores directos e indirectos de forma habitual, normalmente a través de la proximidad durante situaciones laborales, escolares o sociales. Estos consumidores pueden afectar a la viabilidad del programa indirectamente, a través de la interacción (o falta de interacción) con los consumidores directos e indirectos. Algunos ejemplos de miembros de la comunidad inmediata son

los vecinos de un hogar de grupo para delincuentes juveniles, el conductor habitual del autobús en la ruta que utilizan los adultos con discapacidades del desarrollo y los niños que ocasionalmente juegan en el mismo patio que un niño cuya conducta agresiva grave fue objeto de tratamiento.

*Miembros de la comunidad ampliada.* La comunidad ampliada incluye a aquellas personas que probablemente no conocen a los consumidores directos e indirectos ni interactúan con ellos, pero que viven en la misma comunidad. Algunos ejemplos son los directores de periódicos, que pueden considerar que los programas de tratamiento son noticia por su interés humano o por su mala política social, y sus lectores, influenciados o enfadados; los cajeros de supermercado, que pueden considerar que los adultos con discapacidades del desarrollo que utilizan las habilidades de compra recién adquiridas son conmovedores o problemáticos; los contribuyentes, que pueden protestar por el pago de un programa o exigir más de él y menos de otro; las Madres contra la Conducción en Estado de Ebriedad y organizaciones similares; y los legisladores y burócratas, que pueden regular la existencia o inexistencia del programa.

El siguiente ejemplo se diseñó para ilustrar estas clases abstractas de consumidores, sus funciones como compradores y las interrelaciones entre las distintas funciones:

Una compañía de seguros se pone en contacto con uno de sus clientes, los propietarios de una empresa minera a cielo abierto, y le comunica que, a menos que ponga en marcha un programa de seguridad para sus empleados en la mina y documente una disminución del índice de accidentes laborales, sus tarifas de seguro se cuadruplicarán y su póliza podría ser cancelada. Los propietarios de las minas

responden contratando a consultores de seguridad para que apliquen y documenten un programa de seguridad permanente para los mineros (véase Fox, Hopkins & Anger, 1987). Los consumidores directos son los empleados que trabajan en la mina, incluidos los mineros, los conserjes y el personal administrativo. Sin embargo, estos consumidores no compraron el programa. Los compradores del programa son los propietarios de la mina, por lo que éstos constituyen un grupo de consumidores indirectos. Otro grupo de consumidores indirectos está formado por las familias y amigos de los consumidores directos, que se benefician indirectamente de la mayor seguridad de sus seres queridos y sostén de la familia y también pueden beneficiarse de cualquier sistema de incentivos que utilice el programa de seguridad. La compañía de seguros, aunque ha impulsado el programa, no es un consumidor directo ni indirecto. Es un miembro de la comunidad inmediata; sólo se ve afectada monetariamente por el resultado del programa y puede no tener ninguna otra relación con el programa o sus participantes. La comunidad ampliada incluye a los contribuyentes, que pueden beneficiarse de que menos trabajadores cobren indemnizaciones por incapacidad, y a los comerciantes, que se benefician de que los trabajadores gasten tanto su paga extra como sus incentivos en metálico por seguir las directrices de seguridad.

Este ejemplo de clases de consumidores, a pesar de su pertenencia en cierto modo indefinida, proporciona una lógica para determinar a quién se debe inducir a la hora de evaluar la validez social. El simple hecho de reconocer la existencia de diferentes grupos de consumidores y sus posibles intereses en los resultados del programa demuestra que la adopción o el rechazo del programa no pueden

predecirse con seguridad a partir de una muestra restringida de sólo consumidores directos (cf. Mathews & Fawcett, 1979). En este caso, la comunidad inmediata y la comunidad ampliada representan a los consumidores que realmente pueden controlar la supervivencia del programa, y este caso no carece de representatividad.

### *Ejecución de evaluaciones de la validez social*

Una vez decididas las cuestiones de a quién y qué preguntar en una evaluación de la validez social, la siguiente cuestión difícil es cómo recopilar información de forma válida, fiable y rentable. Esta cuestión plantea problemas especiales a los analistas de conducta. La mayoría de las evaluaciones de validez social se basan en el uso de entrevistas, cuestionarios o encuestas administradas por el experimentador (e.g., Fuqua & Schwade, 1986; McMahon & Forehand, 1983; Schwartz, 1991). La naturaleza subjetiva de este tipo de información, unida a las posibles variables de confusión de las contingencias sociales proporcionadas por el experimentador (a menudo denominadas en otras disciplinas "características de la demanda"), a menudo dificultan la interpretación de estos datos (véase Azer & Schwade, 1986; McMahon & Forehand, 1983; Schwartz, 1991). que estos datos sean difíciles de interpretar (véase Azrin, Holz y Goldiamond, 1961, para un caso empírico al respecto).

Los datos subjetivos pueden ser valiosos para el análisis de la conducta. Skinner (1953, 1963) reconoció la importancia de la conducta privada en el estudio de la conducta humana; de hecho, diferenció las conductas públicas y privadas por su accesibilidad a la observación, no por el grado en que son fenómenos lícitos. Aun así, es poco probable que se pueda etiquetar de

forma predecible o uniforme a los sentimientos. Por ejemplo, en muchas evaluaciones de validez social se pide a los encuestados que respondan a preguntas describiendo (por ejemplo, en una escala de 7 puntos) lo satisfechos que están. "Satisfacción" describe un estado controlado por una amplia gama de condiciones de estímulo diferentes para personas diferentes. Para algunos individuos, "satisfecho" puede ser una calificación media, para otros puede ser extraordinaria. El reto, como señaló Skinner hace mucho tiempo, es buscar los aspectos públicos de la conducta privada.

Las respuestas de los cuestionarios de validez social son ciertamente públicas; la cuestión pasa a ser determinar cuándo son acompañamientos tempranos de las conductas que realmente determinan la viabilidad del programa y cuándo son los falsos positivos que pueden ponerla en peligro. Han surgido algunos métodos para evaluar la validez social que citan una gama más amplia de conductas pero formulan los mismos tipos de preguntas que las medidas subjetivas más utilizadas. Estas técnicas incluyen (a) el uso de escalas de valoración del afecto definidas operacionalmente para evaluar la conducta "emocional" de un niño mientras participa en la intervención objetivo (e.g., Dunlap, 1984; Dunlap & Koegel, 1980), (b) permitir a los consumidores elegir la intervención después de haber sido expuestos a dos o más intervenciones diseñadas para abordar la misma conducta objetivo (e.g., Harris, 1986), (c) evaluar experimentalmente diferentes tasas de la conducta objetivo en el entorno natural para determinar la tasa óptima y utilizar esta información para determinar el objetivo de la intervención (e.g., Jones & Azrin, 1969; Warren, Rogers-Warren, & Baer, 1976), (d) pedir a ejecutantes competentes de la conducta

objetivo que juzguen la adecuación de ejemplos de producto permanente producidos por los consumidores directos de la intervención (e.g., Schepis, Reid, & Fitzgerald, 1987), y (e) pedir a jueces pares y expertos que comparen muestras de conducta fotografiadas o grabadas en vídeo antes y después de la intervención (e.g., Friman & Hove, 1987).

Aunque estos procedimientos animan a los analistas de conducta aplicada a utilizar medidas de validez social que se basan en un muestreo más amplio de conductas observables y probablemente relevantes, representan la minoría de los procedimientos utilizados. Una buena evaluación de la validez social consiste en hacer las preguntas adecuadas, a las personas adecuadas, de una manera apropiada. Antes de intentar trazar un rumbo para el futuro en el ámbito de la evaluación de la validez social, debemos evaluar las prácticas actuales en este campo. Deberíamos saber (a) quién evalúa la validez social; (b) a quién, qué y cómo preguntan; y (c) qué hacen con la información. Schwartz (1991) intentó responder a estas preguntas en una revisión de la frecuencia y variedad de las evaluaciones de la validez social de las que se informó en la producción de un año de siete revistas de conducta. Encontró que el 29% de los 139 artides revisados informaban de alguna forma de evaluación de la validez social (pero nótese que en JABA el 41% de los 34 artides informaban de alguna medida de validez social). Sin embargo, estos artículos utilizaron métodos incoherentes para recopilar sus datos de validez social y un lenguaje conceptual incoherente para describirlos. Aunque se informó de un número alentador de métodos interesantes e innovadores para evaluar la validez social, la mayoría de los estudios cuestionaban sólo a los consumidores directos o indirectos y se basaban

únicamente en el uso de datos de cuestionarios para evaluar la validez social.

### **Direcciones Futuras**

Estos resultados sugieren tres direcciones futuras para mejorar las evaluaciones de validez social: (a) Más investigadores y profesionales deberían realizar y responder a las evaluaciones de validez social y comenzar a informar los resultados aparentes de hacerlo, tanto negativos como positivos, aunque solo sean anécdotas; (b) se debe muestrear una mayor variedad de consumidores y miembros de la comunidad, y nuevamente se deben informar los resultados aparentes de hacerlo, aunque solo sea como anécdotas; y (c) la investigación debe apuntar al descubrimiento de precursores de invalidez social más objetivos y más confiables que los que están disponibles en la actualidad, tal vez al principio examinando esas anécdotas emergentes en busca de puntos en común interesantes.

Cada una de estas áreas se discutirá individualmente, aunque las mejoras en cada área pueden afectar a las otras, y aunque las mejoras en cualquiera de las dos primeras probablemente no reduzcan sustancialmente el riesgo actual de proceder de manera inapropiada. Por ejemplo, si los investigadores que actualmente informan evaluaciones de validez social encontraran formas más confiables de predecir la validez social y la invalidez social, el análisis de conducta aplicado sería una mejor disciplina para el conocimiento, pero solo una pequeña proporción de los programas actuales se beneficiaría. Por otro lado, si todos los investigadores de la conducta comenzaran a usar las evaluaciones actuales de validez desconocida, el campo no estaría más cerca de su objetivo de resolver

problemas importantes con métodos efectivos y socialmente aceptables.

### *Incrementando los reportes de la Validez Social*

El primer problema es la tasa actual de uso de evaluaciones de validez social. ¿Hay una razón suficiente para hacer que las evaluaciones de validez social sean tan rutinarias en la aplicación y evaluación de programas como lo es la evaluación de la confiabilidad de la medición en nuestras prácticas de investigación actuales? Creemos que si la hay. Los resultados de los programas de investigación no se consideran creíbles ni replicables sin evaluar la confiabilidad y validez de sus procedimientos de medición. En el caso de los programas de aplicación, no es la credibilidad de sus resultados, sino los propios programas los que están en riesgo cuando proceden sin una evaluación fiable de la validez social. Tal vez sea solo ese cambio en dónde buscar la validez y la confiabilidad lo que explica la actual atención constante a la confiabilidad de la medición y la mínima atención a la validez social. Hasta ahora, la mayoría de los profesionales del campo que informan en las revistas se formaron como investigadores más que como profesionales; la mayoría de ellos aprendieron la aplicación en el trabajo. Si la mayoría de sus aplicaciones han sobrevivido hasta ahora, al menos mientras estuvieron directamente involucrados, es posible que ni siquiera cuestionen la viabilidad de sus programas. Así, ni su socialización profesional ni su experiencia actual han castigado consistentemente la ausencia de valoraciones de validez social, al menos no con el mismo vigor que se castigaba cualquier ausencia de confiabilidad de las medidas.

Podemos resolver esta disparidad esperando las inevitabilidades del futuro o,

tal vez, podemos usar la conducta verbal ahora para cambiar la conducta profesional lo suficiente como para evitar algunos de ellos. Una técnica que puede facilitar este cambio en nuestra conducta profesional es establecer criterios para realizar y reportar medidas de validez social que se deben cumplir para su publicación en revistas de conducta. Estos criterios existen para otras cuestiones metodológicas; tal vez sea hora de extenderlos a la evaluación de la validez social.

El campo del análisis aplicado de la conducta ha hecho mucho para demostrar la como la conducta humana tiene regularidades, es decir, tiene leyes o principios, y desarrollar una tecnología para el cambio de conducta. Ha tenido mucho menos éxito en la comercialización y difusión de sí mismo y de su tecnología (e.g., Geller, 1989, 1990; Morris, 1985; Pennypacker, 1986; Seekins & Fawcett, 1984). Es al menos posible que el uso generalizado de evaluaciones precisas de validez social mejoraría algo de esa condición, aunque solo sea mostrando a los consumidores de programas que a los conductistas les importa lo que piensan en lugar de controlar sus pensamientos (como sugiere la oposición).

Tal vez las evaluaciones de validez social se conviertan en un procedimiento estándar cuando los profesionales e investigadores dejen de usar las evaluaciones de validez social como nada más que el cumplimiento de la ley (Enmienda de los Centros Comunitarios de Salud Mental, PL 94-63) y como un método económico para calmar las preocupaciones de los participantes vocales del programa y miembros de la comunidad. Sin embargo, el problema puede resultar notablemente similar al de muchos programas de educación para la salud: los estudiantes aprenden lo que se les enseña sobre cómo preservar o

mantener su salud, pero no comienzan a practicar lo que han aprendido hasta que han visto a suficientes compañeros sufrir las consecuencias de no hacerlo. Desafortunadamente para ese análisis, la invalidez del programa y sus desastres posteriores rara vez se informan en revistas, convenciones o informes de los medios, por lo que los desarrolladores de programas aprenden sobre ellos y sus tasas solo a través de anécdotas casuales. ¿Sería saludable una columna de obituario del programa en cada número de revistas relevantes?

Incluso las evaluaciones precisas de la validez social serán de poca utilidad para la viabilidad del programa, a menos que se realicen de manera prescriptiva y no correctiva. La validez social de los objetivos, métodos y resultados anticipados del programa debe conocerse antes del comienzo del programa y debe evaluarse periódicamente durante la implementación. Sin embargo, si se les pide a los consumidores que completen cuestionarios de satisfacción frecuentes, es mejor que lo hagan para beneficiar al consumidor de manera tangible: o se deben mejorar los componentes que no le gustan, o se debe ofrecer una razón excelente y creíble para no seguir las recomendaciones del consumidor. De lo contrario, el consumidor se entera de que las valoraciones de validez social no sólo son inútiles sino que en cierto modo son fraudulentas, y por tanto el procedimiento de valoración se convierte en un elemento más de la invalidez social del programa. El componente de evaluación del Modelo de Enseñanza-Familia ofrece un excelente ejemplo del lado positivo de esa relación: incluye un sistema continuo de evaluación del consumidor que responde constantemente a los

consumidores, ya sea con cambios o explicaciones (Blase, Fixsen & Phillips, 1984; Phillips, Phillips, Fixsen & Wolf, 1972; Warfel et al., 1981).

#### *Ampliar la participación de los consumidores*

Del mismo modo que puede resultar difícil animar a los profesionales a realizar evaluaciones de la validez social con más frecuencia, también puede resultar difícil animar a los creadores de programas a realizar un muestreo más amplio de sus consumidores. La comunidad ampliada rara vez ha estado representada en las evaluaciones de validez social, y cuando se les ha incluido, lo más probable es que se les pregunte sobre la aceptabilidad de los objetivos del programa, aunque es probable que tengan intereses igualmente fuertes en los métodos y resultados del programa (Schwartz, 1991). Los analistas de conducta aplicada, que normalmente no controlan directamente la viabilidad de sus programas, difícilmente pueden considerarse aplicados sin solicitar y utilizar la opinión de la comunidad (Baer, Wolf & Risley, 1968).

Algunos investigadores ya están afrontando este reto. El Modelo de Enseñanza-Familia<sup>2</sup> ofrece un ejemplo para realizar entrevistas de satisfacción del consumidor con personas con discapacidades del desarrollo (Strouse, 1988). Fawcett y sus colegas (e.g., Seekins, Fawcett, & Mathews, 1987; Seekins, Mathews, & Fawcett, 1984; Whang, Fawcett, & Mathews, 1984) validaron intervenciones basadas en la comunidad con paneles de expertos de la comunidad. Por ejemplo, los jueces de un programa de formación en liderazgo fueron dos politólogos, un consultor en

<sup>2</sup> Teaching-Family Model en el original.

desarrollo organizativo, un urbanista y un consultor en comunicaciones (Seekins et al., 1984); los jueces de un programa de formación en habilidades sociales relacionadas con el trabajo fueron empresarios locales (Whang et al., 1984); y los jueces de un programa de enseñanza de habilidades de defensa del consumidor fueron un político local y un administrador público (Seekins et al., 1987).

A medida que los analistas de conducta comienzan a incluir un espectro más amplio de consumidores en las evaluaciones de validez social, debe abordarse la cuestión de la ponderación diferenciada de las opiniones de los distintos grupos de consumidores, especialmente cuando esas opiniones son contradictorias. El objetivo de las evaluaciones de validez social es proporcionar información que ayude a garantizar la supervivencia del programa. Por lo tanto, la información de los consumidores más directamente relacionada con la viabilidad del programa es la que debe tener más peso. La cuestión de qué grupo de consumidores afecta más directamente a la viabilidad del programa es empírica; sin embargo, aún no se dispone de los datos necesarios para responder a estas preguntas. Por lo tanto, los analistas de conducta deberían empezar a recopilar estos datos incluyendo a miembros de las cuatro clases de consumidores (i.e., directos, indirectos, miembros de la comunidad inmediata y miembros de la comunidad ampliada) en las evaluaciones de validez social y relacionando la satisfacción y la insatisfacción de los consumidores de cada grupo con la longevidad del programa.

#### *La precisión de la Evaluación de la Validez Social*

La precisión de las evaluaciones de la validez social es la cuestión central de los trabajos futuros. La validez social es

intrínsecamente una medida complementaria; su función no es evaluar la eficacia del programa, sino su aceptabilidad y viabilidad. Del mismo modo, su propósito no es comparar programas, sino protegerlos contra el rechazo o el sabotaje. En lugar de intentar comparar la aceptabilidad de programas o componentes de programas, quizá deberíamos revisar los componentes evaluados de los programas y recopilar un menú de elementos de diferentes programas calificados favorablemente por los consumidores y un menú paralelo de los calificados desfavorablemente. A continuación, podemos examinar estos componentes del programa para determinar si tienen algo en común que explique sus calificaciones similares, a pesar de la disimilitud de los programas en los que operaban. Son las similitudes en los componentes de los programas mejor valorados (y en los peor valorados) las que pueden proporcionar respuestas sobre cómo ofrecer intervenciones de cambio de conducta aceptables y eficaces.

La mejora de la construcción de evaluaciones de validez social es un objetivo importante del análisis aplicado de la conducta. Los investigadores deben tener en cuenta las reglas básicas de construcción de pruebas y estadísticas al desarrollar evaluaciones de validez social (Baer, 1987). Los cuestionarios sobre la aceptabilidad de los programas deberían (a) utilizar escalas que permitan una amplia variación en las respuestas de los consumidores (e.g., una escala de 7 puntos en lugar de 3), (b) requerir una respuesta diferencial por parte del consumidor (i.e., deberían requerir que incluso el consumidor más satisfecho utilice todo el rango de la escala de valoración) (Cone, 1981), (c) especificar el periodo de tiempo que se está valorando (e.g., pedir a los consumidores que valoren sólo los servicios

prestados durante los últimos 3 meses, en lugar de todos los servicios) (McMahon, 1984), (d) abordar todas las dimensiones pertinentes para la aceptabilidad y viabilidad de un programa (e.g., si está interesado en las opiniones de los consumidores sobre los observadores en el aula pregunte directamente) y (e) ser lo más específico posible, ya que una mayor especificidad puede aumentar la utilidad de la información recogida en los cuestionarios de validez social (Mash & Terdal, 1981).

Si, además, se desarrollan técnicas más objetivas para evaluar la validez social, estas evaluaciones podrían resultar aún más útiles. Podrían ampliarse cuatro clases de técnicas para evaluar la validez social de forma más funcional: (a) prueba de que los objetivos y resultados de un programa son válidos en sí mismos, (b) medición discreta, (c) identificación y medición de los precursores conductuales de los tipos de satisfacción e insatisfacción que culminan en el rechazo del programa, y (d) proporcionar a los consumidores experiencias con diferentes opciones de programas, permitiéndoles después elegir la opción más satisfactoria.

*Objetivos y resultados.* Preguntar sobre los objetivos y resultados de un programa es la forma de validez social que se evalúa con más frecuencia (Schwartz, 1991). Van Houten (1979) sugirió que se utilizara la conducta de los modelos normales y de los ejecutores competentes para ayudar a identificar las conductas objetivo y determinar los estándares de un desempeño competente. Sin embargo, la identificación de desempeños competentes plantea cuestiones éticas difíciles. Por ejemplo, ¿quiénes son los modelos competentes para las personas que viven en una institución estatal? ¿Debemos establecer normas comunitarias para estos consumidores, porque suponemos que llegarán allí, o planteamos

normas más bajas adecuadas a su entorno actual, porque suponemos que se van a quedar? En el primer caso, las encuestas sobre el entorno futuro (Anderson & Schwartz, 1986; Fowler, 1982) pueden responder a estas preguntas. Estas encuestas identifican el próximo entorno de los clientes y establecen objetivos de conducta a niveles apropiados para el entorno posterior. Esta técnica puede facilitar la transición de los clientes de un programa a otro (e.g., de un centro residencial a vivir en casa, o de un centro preescolar especializado a una guardería regular).

La conducta de los miembros de la comunidad puede servir de modelo para establecer objetivos de intervención. Los estudiantes de secundaria y universitarios pueden servir de modelo de los niveles adecuados de habilidades conversacionales para enseñar a los adolescentes delincuentes y predelincuentes (Minkin et al., 1976). Se puede observar a compañeros normales en aulas de preescolar para determinar las habilidades de transición deseables dentro del aula y los índices de interacción verbal (Osnes, Guevremont & Stokes, 1986; Sainato, Strain, Lefebvre & Rapp, 1987). Los jueces pueden puntuar la fluidez, la "naturalidad" y el ritmo del habla de estudiantes universitarios tratados por tartamudez (Jones & Azrin, 1969). Se considera que el ritmo óptimo para que los preescolares hagan ofertas de compartir a sus compañeros es el ritmo más frecuentemente aceptado por esos compañeros; los compañeros se convierten en los expertos simplemente por responder consistentemente más a menudo a algunos ritmos que a otros (Kohler & Fowler, 1985; Warren et al., 1976). Los profesores y los compañeros pueden calificar el estatus social de sus alumnos o amigos y, a continuación, identificar habilidades objetivo para los peor valorados. Tras el establecimiento de esas

habilidades, los profesores o los compañeros pueden calificar a los alumnos o amigos de nuevo, para ver si esas habilidades eran realmente cruciales para mejorar su estatus con esos jueces (Hoier & Cone, 1987; Plienis et al., 1987).

Sin embargo, a la hora de seleccionar a los expertos, hay que demostrar que son expertos apropiados para la intervención (Greenwood, 1990). La pertenencia a la comunidad no califica por sí solo a alguien como modelo de conducta adecuada o para juzgar el resultado de una intervención. En general, el consumidor indirecto que contrató el programa puede validar los objetivos del cambio de conducta como casi nadie más puede hacerlo, especialmente si esa persona es la única que se queja (Baer, 1986). Si es su queja la que define el problema, el programa aplicado tiene validez social si su queja se ve satisfecha por él, ellos son la "comunidad" que define la validez de este problema. Sin embargo, si el problema lo define una comunidad más amplia, debemos identificar a los miembros representativos y apropiados de esa comunidad a evaluar la validez social de la solución propuesta.

*Medición no intrusiva.* La medición intrusiva es a menudo una amenaza propia para la validez social, porque a muchas personas no les gusta que se evalúe su conducta. Por lo tanto, el desarrollo de medidas no intrusivas puede ser en ocasiones crucial para la validez social de un programa aceptable en todas las demás dimensiones. La medición no intrusiva se produce cuando el sujeto no es consciente de las observaciones en curso o de la conducta objetivo (Kazdin, 1979). Sin embargo, este tipo de medición rara vez es ética y a menudo es ilegal (American Psychological Association, 1979). Y cuando es ética y legal, a menudo genera algunos problemas logísticos de cómo

hacerla, especialmente fuera de los entornos de laboratorio. Aun así, algunas formas de medición discreta pueden escapar a estas desventajas: por ejemplo, el uso de registros de archivo. Se pueden buscar documentos para recuperar datos sobre tasas de reincidencia, quejas (véase Baer, 1988), accidentes industriales, consumo de energía (Winett, Neale & Grier, 1979) y similares. Es importante recordar sin embargo, que aunque los datos procedentes de fuentes archivísticas escapan a muchas dificultades en el ámbito de la validez social, su validez como datos de resultados es a menudo dudosa. Por ejemplo, la reducción de la reincidencia de los delincuentes juveniles puede reflejar una mayor conducta policial, un cambio en las normas judiciales o la sobrecarga de los registros, en lugar de una menor actividad delictiva (Kazdin, 1979). Del mismo modo, el aumento de la reincidencia de los niños con autismo en centros de tratamiento residencial puede reflejar falencias en los sistemas familiares en lugar de una falta de generalización y mantenimiento de las mejoras de conducta de los niños (Anderson, Christian & Luce, 1986).

*Correlatos conductuales de la satisfacción.* La técnica más básica y crucial para realizar evaluaciones precisas de la validez social sigue siendo la identificación y medición de las conductas observables que se correlacionan con la satisfacción o el rechazo del programa. Las conductas que muestran claramente el apoyo a un programa incluyen la participación continuada en el mismo, la asistencia regular y la llegada puntual a las sesiones del programa, la recomendación del programa a los amigos, la obtención de muchos de los puntos que pueda ofrecer su sistema de incentivos, la defensa del programa frente a los ataques de otros, y un afecto agradable y un gran entusiasmo durante las sesiones

del programa. Las conductas que demuestran claramente la invalidez social de un programa incluyen la retirada, la exigencia de que el personal del programa se marche (e.g., que el comprador del programa "despida" al programa), las quejas a amigos, funcionarios y medios de comunicación, la escasa asistencia y/o la impuntualidad crónica a las sesiones del programa, la negativa a participar o el bajo desempeño en cualquier sistema de incentivos que pueda ofrecer el programa, el incumplimiento generalizado o selectivo de las rutinas del programa, y el afecto negativo y el bajo entusiasmo durante las sesiones del programa.

Muchos de estos índices de validez social son relativamente fáciles de medir y, de hecho, es posible que ya estén incluidos en los sistemas de recopilación de datos en curso. Su desventaja, que a menudo puede superar su utilidad, es su tardanza en emerger. En el momento en que aparecen estos signos de validez o invalidez social, especialmente en el caso de la invalidez social, a menudo ya es demasiado tarde para responder de manera adaptativa a ellos.

Aun así, estas medidas pueden tener alguna utilidad en los programas para personas con discapacidades del desarrollo, porque muchos no requieren mucha habilidad verbal, y las personas con discapacidades a menudo expresan su descontento con un programa bastante temprano en su desarrollo, a menudo de esta manera. De hecho, son más a menudo los consumidores con capacidad de desarrollo los que cargan con una cortesía que les impide molestar a los aplicadores del programa con sus quejas hasta que esas quejas son bastante graves. Por ejemplo, Koegel y sus colegas (p. ej., Dunlap y Koegel, 1980; Koegel y Egel, 1979) han desarrollado un sistema de codificación confiable para evaluar el afecto y el entusiasmo en niños

con autismo. Estas medidas abiertas de satisfacción son potencialmente relevantes para medir la validez social de los programas que atienden a personas con discapacidades graves del desarrollo. Tenga en cuenta nuevamente que este sistema de medición es una medida complementaria, no una evaluación de la eficacia del programa.

Un requisito previo para medir conductas correlacionadas con la satisfacción o insatisfacción del programa es la identificación de estas conductas. Los investigadores pueden identificar los correlatos conductuales de la satisfacción mediante el estudio de programas de tratamiento exitosos (ver Fuoco & Christian, 1986; Paine, Bellamy & Wilcox, 1984, para descripciones de programas exitosos). Identificar correlatos de insatisfacción, especialmente correlatos tempranos y quizás predictivos, es más difícil de alcanzar. Una razón puede ser que los investigadores no difundan información sobre fallas en los programas o consumidores insatisfechos. Tal vez los analistas de conductas de conducta deberían actuar como epidemiólogos y realizar análisis post mortem cuidadosos de los programas que fracasan. Antes de la implementación del programa, los analistas de conducta realizan evaluaciones exhaustivas de las necesidades y, a lo largo de una intervención, evalúan sistemáticamente la eficacia del programa. Este tipo de evaluación exhaustiva debe extenderse a los consumidores insatisfechos y las fallas del programa, y los resultados de estas evaluaciones deben difundirse a través de revistas profesionales y presentaciones en conferencias. Un estudio cuidadoso de los elementos comunes a las fallas de los programas puede conducir a una planificación de programas más efectiva y evaluaciones de validez social más sensibles.

*Oferta de opciones a elegir.* La última medida de la validez social de un programa es el rango de alternativas que un consumidor rechazará para elegir el programa. Elegir un programa entre varias alternativas es la definición conductual fundamental de la preferencia, y ¿qué es más socialmente válido que eso? Sin embargo, organizar tal medida de validez social es simple solo en lógica: brinde a los consumidores experiencia con dos o más programas y déjelos elegir. La probabilidad de que luego rechacen lo que han elegido frente a un rango de alternativas debe ser baja, en la medida en que el rango sea amplio. Sin embargo, a medida que surgen nuevas alternativas a su elección actual, a medida que se amplía el rango, la elección debe ofrecerse nuevamente para garantizar la estabilidad de su preferencia y la durabilidad de la validez social evaluada anteriormente.

La validez de la elección es fuerte y la lógica de la elección es simple, pero la implementación de la elección no lo es. La elección válida requiere que el consumidor tenga una experiencia extensa y casi simultánea con todas las alternativas, que todas las alternativas sean igualmente efectivas y que todas las alternativas estén fácilmente disponibles. Además, los evaluadores deben seleccionar cuidadosamente las opciones de tratamiento, para no sesgar la evaluación al ofrecer el programa objetivo frente a alternativas mucho menos deseables. Organizar esas condiciones resultará costoso en tiempo, esfuerzo, recursos y sofisticación, especialmente si el número de opciones es superior a dos. Y aunque todas estas barreras son formidables, el tiempo puede ser especialmente problemático. La función de las evaluaciones de validez social es proporcionar una alerta temprana del rechazo del programa en cualquiera de sus

formas; por lo tanto, cualquier medida que requiera mucho tiempo para ser válida puede ser inútil.

Como consecuencia, siempre nos vemos tentados a buscar medidas rápidas, baratas, fáciles y graciosamente repetitivas que predigan tempranamente la validez o invalidez social, y nuestro problema intrínseco es la sospecha de que cuanto más rápido, más barato, más fácil, más temprano y más Graciosamente repetitivas son estas medidas cuanto menor es su correlación con los hechos reales que configuran la validez y la invalidez sociales. En otras palabras, la preferencia verbal expresada fácil y tempranamente y, más tarde, las elecciones reales de conducta no siempre se correlacionan mucho, incluso con adultos sin discapacidades (e.g., Lockhart, 1979), y quizás especialmente con adultos sin discapacidades. Aún así, puede valer la pena recomendar estudios de la correlación entre varias formas de preferencia verbal, algunas de ellas muy parecidas a los métodos actualmente en uso, para ver si alguna de ellas se correlaciona especialmente bien con las costosísimas técnicas de elección.

Ofrecer opciones, sin embargo, no es una técnica imposible, sobre todo si no insistimos en la perfección en su ejecución. Se ha utilizado con estudiantes universitarios para evaluar diferentes métodos de enseñanza (Lockhart, 1979), entre adultos con retraso mental severo para evaluar la preferencia de tareas en entornos vocacionales (Mithaug & Hanawalt, 1978; Mithaug & Mar, 1980), entre estudiantes con problemas de aprendizaje para seleccionar las estrategias de instrucción preferidas (Harris, 1986), y con estudiantes que demuestran altas tasas de conducta de autoestimulación para evaluar su preferencia por formas de estimulación más apropiadas (Buyer, Berkson, Winnega &

Morton, 1987). Hasta ahora, no hemos sabido que ninguna de esas aplicaciones menos que perfectas resultara desastrosamente engañosa.

Aunque la lógica de la elección como el método más válido para evaluar la validez social está bien desarrollada (e.g., Fuqua & Schwade, 1986; McMahon & Forehand, 1983), los métodos para implementar la elección en entornos más naturales necesitan trabajar para facilitar la integración en las evaluaciones de programas. Además, la toma de decisiones no es una técnica de evaluación apropiada para todos los programas. No todos los programas permiten comparaciones lado a lado con programas alternativos, pero para programas seleccionados esta técnica puede ser una técnica eficiente y objetiva para evaluar la validez social.

### Conclusiones

La noción de evaluar la validez social de los programas conductuales y luego usar estos datos en las decisiones programáticas ha sido controvertido desde que se introdujo (Wolf, 1978). Sin embargo, en lugar de desaparecer, las evaluaciones de validez social se están volviendo más comunes en la práctica y la investigación de la conducta. Sin embargo, se necesitan cambios en la práctica actual de evaluación de la validez social antes de que estas evaluaciones puedan usarse como predictores precisos y confiables del éxito o fracaso del programa. Como analistas de conducta, necesitamos que los procedimientos relevantes para los programas de cambio de conducta se especifiquen y definan minuciosamente (Baer et al., 1968); deberíamos juzgar las técnicas para evaluar la validez social con los mismos criterios. La mayoría de los procedimientos actuales de evaluación de la validez social no cumplen con esos criterios lo suficiente

como para ser considerados una herramienta tecnológica utilizada para asegurar el mantenimiento de los programas aplicados y avanzar en el estado de nuestra ciencia.

Los procedimientos para realizar evaluaciones de validez social deben incluir las siguientes pautas:

1. Las evaluaciones de validez social deben ser una parte estándar de la aplicación del programa y la metodología de investigación aplicada. La inducción de una evaluación exhaustiva de la validez social debería ser un requisito mínimo, similar a la evaluación de la confiabilidad entre observadores, en todas las investigaciones y prácticas conductuales aplicadas.

2. Las evaluaciones de la validez social deben realizarse de forma prospectiva y durante una intervención, así como al final. De lo contrario, las preocupaciones de los consumidores sobre el programa no pueden responderse de manera que defiendan a los consumidores, el programa y la disciplina.

3. Los analistas aplicados de conducta deben reconocer que existen consumidores más relevantes y poderosos de los que han identificado y consultado hasta ahora. Por lo tanto, se debe incluir un espectro más amplio de consumidores en la evaluación de la validez social en curso, con especial consideración a los consumidores que controlan la viabilidad de los programas comunitarios. Cuando los consumidores no pueden responder a las formas estándar de evaluación de la validez social (e.g., personas con discapacidades del desarrollo), se deben desarrollar técnicas especiales para ellos.

4. Las evaluaciones de validez social, aunque no miden los resultados del programa, merecen el mismo rigor psicométrico que cualquier medida conductual. Rigor no es lo mismo que estandarización,

pero luchar por una medición válida y confiable probablemente conduciría a una metodología más uniforme que la que caracteriza la práctica actual de evaluación de la validez social. Una mejor atención a la validez y la confiabilidad haría que estas medidas fueran mejores predictores de la aceptación o el rechazo del programa.

5. Si se va a lograr la cuarta recomendación, el campo requerirá el desarrollo de algunas medidas objetivas y claramente válidas de validez social e invalidez social (i.e., aceptación y rechazo del programa). De ello se deduce que tendremos que estudiar la aceptación y el rechazo reales del programa, por doloroso que sea, para saber algo acerca de sus antecedentes verdaderamente fiables y válidos. Cuando hayamos identificado precursores válidos y confiables de la aceptación y el rechazo del programa, podemos estudiar qué tan bien se relacionan con ellos estas técnicas de cuestionario o, lo que es más característico para nosotros, qué tan bien se pueden relacionar con ellos. Los procedimientos objetivos para medir la validez social (e.g., observaciones, informes de tasas de adopción de programas) serán invariablemente más costosos de realizar que las medidas subjetivas estándar (e.g., cuestionarios). Sin embargo, este gasto se minimiza si se considera en el contexto de predecir con precisión la viabilidad del programa.

6. Finalmente, los analistas de aplicados de conducta deben ofrecer a sus consumidores y miembros relevantes de la comunidad más información sobre los programas de tratamiento en curso, los programas de tratamiento potenciales y otras opciones de tratamiento. Esta información probablemente representará una combinación de lo que descubrimos que nuestros consumidores quieren saber y lo que creemos que necesitan saber. Esta información no debe utilizarse como

propaganda. Por lo general, y no correctamente, no estamos en el negocio de dar forma a los valores de nuestros consumidores; en cambio, necesitamos saber qué información se requiere para permitir que los consumidores tomen decisiones informadas. En el análisis de la conducta, esta tarea es especialmente importante. El análisis de la conducta a menudo se tergiversa en los medios populares y científicos (Morris, 1985; Todd & Morris, 1983); será mejor que no intensifiquemos el problema. Los consumidores tienen derecho a una información precisa presentada de manera clara e inteligible. No se les debe pedir que tomen decisiones sobre ningún servicio sin la información adecuada sobre ellos y sobre todas las opciones de tratamiento relevantes y disponibles.

Desde la introducción de la validez social, el análisis aplicado de la conducta se ha movido más hacia la prescripción implícita de Wolf (1978): la búsqueda de su corazón. Ha aumentado el uso de evaluaciones de validez social en la investigación de la conducta, y tanto los investigadores como los profesionales (al menos a veces) intentan mejorar su técnica. Mejores evaluaciones de validez social son vitales para la supervivencia del análisis aplicado de conducta; sin embargo, no son la disciplina, solo una herramienta defensiva de la misma. Todavía necesitamos ser aplicados, conductuales, analíticos, tecnológicos, conceptuales, efectivos y tan generalizados como lo requiera el problema (Baer et al., 1968, 1987). La evaluación de validez social de vanguardia debe abordar las siete dimensiones, y cuando lo haga, la disciplina estará un paso más cerca de encontrar su corazón.

## REFERENCIAS

- Aaker, D. A., & Day, G. S. (1971). Introduction: A guide to consumerism. In D. A. Aaker & G. S. Day (Eds.), *Consumerism: Search for the consumer interest* (pp. 1- 19). New York: Free Press.
- American Psychological Association. (1979). *Ethical standards of psychologists*. Washington, DC: American Psychological Association.
- Anderson, S. R., Christian, W. P., & Luce, S. C. (1986). Transitional programming for autistic individuals. *Behavior Therapist*, 9, 205-211.
- Anderson, S. R., & Schwartz, I. S. (1986). Transitional programming. In F. J. Fuoco & W. P. Christian (Eds.), *Behavior analysis and therapy in residential programming* (pp. 76-100). New York: Van Nostrand Reinhold.
- Azrin, N. H., Holz, W., & Goldiamond, I. (1961). Response bias in questionnaire reports. *Journal of Consulting Psychology*, 25, 324-326.
- Baer, D. M. (1986). In application, frequency is not the only estimate of the probability of behavioral units. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 117-136). Hillsdale, NJ: Erlbaum.
- Baer, D. M. (1987, March). A behavior-analytic query into early intervention. Paper presented at the Banif International Conference on Behavioral Science, Banff, Canada.
- Baer, D. M. (1988). If you know why you're changing a behavior, you'll know when you've changed it enough. *Behavioral Assessment*, 10, 219-223.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91-97.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 20, 313-327.
- Barrett, B. H. (1987). Drifting? Course? Destination?: A review of Research in applied behavior analysis: Issues and advances. *The Behavior Analyst*, 10, 253-276.
- Bearden, W. O., & Oliver, R. L. (1985). The role of public and private complaining in satisfaction with problem resolution. *Journal of Consumer Affairs*, 19, 222-240.
- Blase, K., Fixsen, D., & Phillips, E. (1984). Residential treatment for troubled children: Developing service delivery systems. In S. C. Paine, G. T. Bellamy, & B. Wilcox (Eds.), *Human services that work: From innovation to standard practice* (pp. 149-166). Baltimore: Paul Brookes.
- Bornstein, P. H., & Rychtarik, R. G. (1983). Consumer satisfaction in adult behavior therapy: Procedures, problems, and future perspectives. *Behavior Therapy*, 14, 191-208.
- Buyer, L. S., Berkson, G., Winnega, M. A., & Morton, L. (1987). Stimulation and control as components of stereotyped body rocking. *American Journal of Mental Deficiency*, 91, 543-547.
- Cone, J. (1981). Psychometric considerations. In M. Hersen & A. S. Bellack (Eds.), *Behavioral assessment: A practical handbook* (pp. 38-68). New York: Pergamon.
- Deitz, S. M. (1978). Current status of applied behavior analysis: Science vs. technology. *American Psychologist*, 33, 805-814.
- Dunlap, G. (1984). The influence of task variation and maintenance tasks on the learning and affect of autistic children. *Journal of Experimental Child Psychology*, 37, 41-64.
- Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis*, 13, 619-627.
- Engel, J. F., Kollat, D. T., & Blackwell, R. D. (1973). *Consumer behavior*. New York: Holt, Rinehart, & Winston.
- Fowler, S. A. (1982). Transition from preschool to kindergarten for children with special needs. In K. E. Allen & E. M. Goetz (Eds.), *Early childhood education: Special problems, special solutions* (pp. 309-334). Rockville, MD: Aspen.
- Fox, D. K., Hopkins, B. L., & Anger, W. K. (1987). The long-term effects of a token economy on safety performance in open-pit mining. *Journal of Applied Behavior Analysis*, 20, 215-224.
- Friman, P. C., & Hove, G. (1987). Apparent covariations between child habit disorders: Effects of successful treatment for thumb sucking on untargeted chronic hair pulling. *Journal of Applied Behavior Analysis*, 20, 421- 425.
- Fuoco, F. J., & Christian, W. P. (Eds.). (1986). *Behavior analysis and therapy in residential programs*. New York: Van Nostrand Reinhold.
- Fuqua, R. W., & Schwade, J. (1986). Social validation of applied behavioral research: A selective review and critique. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 265-292). New York: Plenum.
- Geller, E. S. (1989). Applied behavior analysis and social marketing: An integration for environmental preservation. *Journal of Social Issues*, 45, 17-36.
- Geller, E. S. (1990). Behavior analysis and environmental protection: Where have all the flowers

- gone? *Journal of Applied Behavior Analysis*, 23, 269-273.
- Giordano, P. (1977). The client's perspective in agency evaluation. *Social Work*, 22, 34-39.
- Greenwood, C. R. (1990, October). Social validity. In S. Warren (Chair), *Innovations in research standards and methods*. Symposium presented at the Annual Meeting of the Division of Early Childhood, Council for Exceptional Children, Albuquerque, NM.
- Harris, K. R. (1986). Self-monitoring of attentional behavior versus self-monitoring of productivity: Effects on on-task behavior and academic response rate among learning disabled children. *Journal of Applied Behavior Analysis*, 19, 417-423.
- Hoier, T. S., & Cone, J. D. (1987). Target selection of social skills for children: The template-matching procedure. *Behavior Modification*, 11, 137-163.
- Jones, R. J., & Azrin, N. H. (1969). Behavioral engineering: Stuttering as a function of stimulus duration during speech synchronization. *Journal of Applied Behavior Analysis*, 2, 223-229.
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1, 427-452.
- Kazdin, A. E. (1979). Unobtrusive measures in behavioral assessment. *Journal of Applied Behavior Analysis*, 12, 713-724.
- Koegel, R. L., & Egel, A. L. (1979). Motivating autistic children. *Journal of Abnormal Psychology*, 88, 418-426.
- Kohler, F. W., & Fowler, S. A. (1985). Training prosocial behaviors to young children: An analysis of reciprocity with untrained peers. *Journal of Applied Behavior Analysis*, 18, 187-200.
- Larsen, D. L., Attkisson, C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning*, 2, 197-207.
- Lebow, J. (1982). Consumer satisfaction with mental health treatment. *Psychological Bulletin*, 91, 244-259.
- Lockhart, K. A. (1979). Behavioral assessment of human preference. *The Behavior Analyst*, 2, 20-29.
- Mash, E. J., & Terdal, L. G. (1981). Behavioral assessment of childhood disturbances. In E. J. Mash & L. G. Terdal (Eds.), *Behavioral assessment of childhood disorders* (pp. 3-76). New York: Guilford.
- Mathews, R. M., & Fawcett, S. B. (1979). Assessing dissemination capability: An evaluation of an exportable training package. *Behavior Modification*, 3, 49-62.
- McMahon, R. J. (1984). Behavioral checklists and rating scales. In T. H. Ollendick & M. Hersen (Eds.), *Child behavioral assessment: Principles and procedures* (pp. 80-105). New York: Pergamon.
- McMahon, R. J., & Forehand, R. L. (1983). Consumer satisfaction in behavioral treatment of children: Types, issues, and recommendations. *Behavior Therapy*, 14, 209-22.
- Michael, J. (1980). Flight from behavior analysis. *The Behavior Analyst*, 3, 1-22.
- Minkin, N., Braukman, C. J., Minkin, B. L., Timbers, G. D., Timbers, B. J., Fixsen, D. L., Phillips, E. L., & Wolf, M. M. (1976). The social validation and training of conversational skills. *Journal of Applied Behavior Analysis*, 9, 127-139.
- Mithaug, D. E., & Hanawalt, D. A. (1978). The validation of procedures to assess prevocational task preferences in retarded adults. *Journal of Applied Behavior Analysis*, 11, 153-162.
- Mithaug, D. E., & Mar, D. K. (1980). The relation between choosing and working prevocational tasks in two severely retarded young adults. *Journal of Applied Behavior Analysis*, 13, 177-182.
- Morris, E. K. (1985). Public information, dissemination, and behavior analysis. *The Behavior Analyst*, 8, 95-110.
- Osnes, P. G., Guevremont, D. C., & Stokes, T. F. (1986). If I say I'll talk more, then I will: Correspondence training to increase peer-directed talk by socially withdrawn children. *Behavior Modification*, 10, 287-299.
- Paine, S. C., Bellamy, G. T., & Wilcox, B. (Eds.). (1984). *Human services that work: From innovation to standard practice*. Baltimore: Paul Brookes.
- Pennypacker, H. S. (1986). The challenge of technology transfer: Buying in without selling out. *The Behavior Analyst*, 9, 147-156.
- Phillips, E. L., Phillips, E. A., Fixsen, D. L., & Wolf, M. M. (1972). *The teaching-family handbook*. Lawrence, KS: Bureau of Child Research.
- Pierce, W. D., & Epling, W. F. (1980). What happened to analysis in applied behavior analysis. *The Behavior Analyst*, 3, 1-9.
- Plienis, A. J., Hansen, D. J., Ford, F., Smith, S., Stark, L. J., & Kelly, J. A. (1987). Behavioral small group training to improve the social skills of emotionally-disordered adolescents. *Behavior Therapy*, 18, 17-32.
- Sainato, D. M., Strain, P. D., Lefebvre, D., & Rapp, N. (1987). Facilitating transition times with handicapped preschool children: A comparison between peer-mediated and antecedent prompt

- procedures. *Journal of Applied Behavior Analysis*, 20, 285-292.
- Scheirer, M. (1978). Program participants' positive perceptions: Psychological conflict of interest in social program evaluation. *Evaluation Quarterly*, 2, 53-70.
- Schepis, M. M., Reid, D. H., & Fitzgerald, J. R. (1987). Group instruction with profoundly retarded persons: Acquisition, generalization, and maintenance of a remunerative work skill. *Journal of Applied Behavior Analysis*, 20, 97-105.
- Schwartz, I. S. (1991). A review of current practice of methods used to assess social validity. Manuscript submitted for publication.
- Seekins, T., & Fawcett, S. B. (1984). Planned diffusion of social technologies for community groups. In S. C. Paine, G. T. Bellamy, & B. Wilcox (Eds.), *Human services that work: From innovation to standard practice* (pp. 247-260). Baltimore: Paul Brookes.
- Seekins, T., Fawcett, S. B., & Mathews, R. M. (1987). Effects of self-help guides on three consumer advocacy skills: Using personal experiences to influence public policy. *Rehabilitation Psychology*, 32, 29-38.
- Seekins, T., Mathews, R. M., & Fawcett, S. B. (1984). Enhancing leadership skills for community self-help organizations through behavioral instruction. *Journal of Community Psychology*, 12, 155-163.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press.
- Skinner, B. F. (1963). Behaviorism at fifty. *Science*, 140, 951-959.
- Strouse, M. C. (1988). The client satisfaction evaluation. In J. A. Sherman, J. B. Sheldon, & M. C. Strouse (Eds.), *Evaluation programs using the Teaching-Family Model for people with developmental disabilities* (pp. 1-36). Topeka, KS: Kansas Planning Council on Developmental Disabilities.
- Todd, J. T., & Morris, E. K. (1983). Misconceptions and miseducations: Presentations of radical behaviorism in psychology textbooks. *The Behavior Analyst*, 6, 153-160.
- Troelstrup, A. W. (1974). *The consumer in American society*. New York: McGraw Hill.
- Van Houten, R. (1979). Social validation: The evolution of standards of competency for targets. *Journal of Applied Behavior Analysis*, 12, 581-591.
- Ware, J. E. (1978). Effects of acquiescent response set on patient satisfaction ratings. *Medical Care*, 16, 327-336.
- Ware, J. E., Davies-Avery, A., & Stewart, A. L. (1978). The measurement and meaning of patient satisfaction. *Health and Medical Care Services Review*, 1, 1-15.
- Warfel, D. J., Maloney, D. M., & Blase, K. (1981). Consumer feedback in human service programs. *Social Work*, 26, 151-156.
- Warren, S. F., Rogers-Warren, A., & Baer, D. M. (1976). The role of offer rates in controlling sharing by young children. *Journal of Applied Behavior Analysis*, 9, 491-497.
- Whang, P. L., Fawcett, S. B., & Mathews, R. M. (1984). Teaching job-related social skills to learning disabled adolescents. *Analysis and Intervention in Developmental Disabilities*, 4, 29-38.
- Winett, R. A., Neale, M. S., & Grier, H. C. (1979). Effects of self-monitoring and feedback on residential electricity consumption. *Journal of Applied Behavior Analysis*, 12, 173-184.
- Wolf, M. M. (1978). Social validity: The case for subjective measurement, or how behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203-214.

*Recibido el 8 de agosto de 1990*

*Decisión editorial inicial 7 de enero de 1991*

*Revisión recibida el 12 de febrero de 1991*

*Aceptación definitiva 17 de febrero de 1991*

*Editor, E. Scott Geller*