

La confiabilidad de la entrevista en el proceso de selección de estudiantes al posgrado

Alvaro J. Arce Ferrer, Eduardo Backoff E.

Resumen

Hay pocos estudios que demuestren cómo mejora la calidad de las decisiones cuando se toman con base en la entrevista. En este trabajo se utilizó la Teoría de la Generalización para investigar las distintas facetas que definen el proceso de medición en la entrevista de admisión a un posgrado mexicano. Estas facetas son: los aspirantes, los reactivos, los entrevistadores y sus interacciones. Asimismo, se investigó la variación que sufre el error relativo cuando se aumenta o disminuye el número de reactivos de la entrevista y el número de entrevistadores. Los resultados muestran un alto grado de confiabilidad de la entrevista, y que el factor que más contribuyó a la varianza observada fue el aspirante, seguido de la interacción del aspirante con los reactivos. También se encontró el número adecuado de preguntas y entrevistadores que mejoran la calidad de la entrevista.

Palabras clave: entrevista, ingreso al posgrado, confiabilidad, Teoría de la Generalización.

Abstract

The interview is frequently used as a selection and evaluation procedure for working and educational settings. Specially, this is true in the case of postgraduate studies in the United States. Regarding to the reliability, research mostly focused on studying interviewer effects on reliability. In this work, a framework from the Generalizability Theory was used to investigate the role of different facets, other than the interviewer. The underlying facets were students, interview items, interviewers, and their interactions. The quality of the information for making admission decisions in a Mexican postgraduate program was then established. Results showed that information derived from interviews does not depend on the particular instances used in the measurement procedure. That is, it was found the students variance component to be the main contributor to the observed score variance, followed by the interaction between students and items. Finally, it was found an optimal number of items and interviewers that maximize the quality of information from interviews.

Key words: interview, postgraduate admission, reliability, Generalizability Theory.

Introducción

Prácticamente, a nivel mundial, la entrevista se utiliza como una técnica complementaria en el proceso de selección del nuevo personal (Maddux, 1991; Campion, Pursell y Brown, 1988). En los Estados Unidos de Norteamérica, la entrevista se considera como el segundo elemento de importancia, después del curriculum vitae, para la obtención de un empleo. Dipboye (1992) y Campion, Palmer y Campion (1997) mencionan que más del 70% de las organizaciones en los Estados Unidos usan la entrevista para tomar decisiones, y que en Inglaterra y Francia los porcentajes aún son mayores: del orden de 90% y 94%, respectivamente.

En el ámbito académico, la entrevista se ha venido utilizando en los procesos de selección de estudiantes, principalmente en los niveles de posgrado. Recientemente, la entrevista se empieza a utilizar con este propósito en los estudios de pregrado. Según Dipboye (1992) en los Estados Unidos, aproximadamente, el 14% de los programas de licenciatura utilizan la entrevista. Un caso que llama la atención son las escuelas de medicina. Puryear y Lewis (1981) y Scott, Chase, y Lefkowitz, (1995) mencionan que en los EU y Canadá cerca del 99% de estas escuelas utilizan la entrevista como parte de sus procesos de admisión.

Asimismo, en el campo de la educación superior en México, la entrevista forma parte de los procesos de selección de estudiantes de posgrado. Esto es cierto, particularmente, para los programas que pertenecen al padrón de excelencia del Consejo Nacional de Ciencia y Tecnología (CONACyT). Y al igual que en Estados Unidos, la entrevista se empieza a utilizar en los niveles de licenciatura (Arce-Ferrer, Burgos-Fajardo y Esquivel-Alcocer, 1999).

La popularidad de la entrevista lleva a pensar que su contribución a la calidad de los procesos de selección ha sido ampliamente estudiada y validada. Sin embargo, Roberts y Porter (1989) y Harasym, Woloschuck, Mandin, y Brundin-Mather (1996) opinan lo contrario. Por su parte, Posthuma y Sommerfreund (1985) y Posthuma y Noh (1990) reconocen que el uso de la entrevista en los procesos de admisión de estudiantes genera muchas controversias. Por un lado, hay quienes reconocen que la entrevista es el procedimiento más obvio para medir atributos no cognoscitivos, relevantes para las decisiones de admisión (Elam, Johnson y Lenhoff, 1994; Shahani, Dipboye y Gehrlein, 1991). Por el otro lado, autores como Markert y Shores (1981) sugieren la necesidad de reemplazar las entrevistas por pruebas estandarizadas, dada sus limitaciones.

Fortalezas y debilidades de la entrevista

La entrevista posee características que resultan atractivas para quienes toman las decisiones de admisión en las universidades. Como parte de estos procesos, la entrevista tiene como propósitos obtener información, facilitar la toma de decisiones, verificar información y reclutar futuros aspirantes (Edwards, Johnson, y Molidor, 1990). Así, se reconocen en la entrevista las siguientes fortalezas:

- Provee información adicional a la obtenida en las solicitudes de admisión. Constituye un medio valioso de acopio de información de variables no cognoscitivas con valor predictivo del éxito académico y profesional (McGaghie, 1990; Vargo, Madill, y Davidson, 1986).¹
- Identifica a los aspirantes idóneos. Taylor (1990) menciona que las entrevistas tienen un doble propósito: identificar a los estudiantes idóneos y a los menos deseables para los distintos programas académicos.
- Permite validar la información contenida en las solicitudes de admisión. Algunos autores reconocen que la entrevista es una ayuda valiosa para cotejar la información en las solicitudes de admisión.

Por otra parte, la entrevista también posee limitaciones (Gabard, Porzio, Oxford y Braun, 1997; Dipboye y Gehrlein, 1991). Entre las más importantes se encuentran las siguientes:

- Su costo es elevado. El proceso requiere de una inversión considerable de tiempo del personal docente para el entrenamiento e implementación de la entrevista, así como el reporte de resultados.
- Genera gastos para quienes la toman. Los entrevistados invierten recursos de traslado, hospedaje y alimentación.
- La información de la entrevista, generalmente, no se valida. Se desconoce si la entrevista aumenta la validez predictiva de las puntuaciones de las pruebas estandarizadas de admisión, del historial académico, de las cartas de recomendaciones y de la carta de exposición de motivos.
- La información de la entrevista, generalmente, es de baja calidad. Los niveles de confiabilidad de la información derivada de las entrevistas se encuentran por debajo de los obtenidos en pruebas estandarizadas que miden atributos equivalentes (Markert y Shores, 1981; Edwards, Johnson y Molidor, 1990). Mientras que su valor típico de confiabilidad, medida por el acuerdo entre jueces, es de 0.70, las pruebas estandarizadas comerciales reportan índices de confiabilidad mayor a 0.85.

¹ Los atributos no cognoscitivos que se evalúan, generalmente, en los posgrados de medicina estadounidenses, son: a) habilidad para interactuar en un grupo; b) preocupación por otros; c) compromiso con la medicina; d) curiosidad intelectual; e) madurez; f) motivación y g) auto confianza (Elam, Johnson y Lenhoff, 1994).

- La entrevista está influenciada por un número mayor de fuentes de error. Típicamente en una entrevista convergen los elementos siguientes: número y formato de preguntas, contenido de la entrevista, número de entrevistadores, características de los entrevistadores, entrenamiento de los entrevistadores, uso de guías de entrevistas y formatos para valorar las respuestas (Nayer, 1992).

Confiabilidad de la información derivada de la entrevista

Markert y Shores (1981) afirman que los estudios de confiabilidad de las entrevistas no han permitido establecer de manera conclusiva la aportación de la información y su valor en los procesos de selección de estudiantes. De una muestra de estudios ² realizados hasta el momento, podemos derivar las siguientes conclusiones:

- Los niveles de confiabilidad varían considerablemente. En los estudios que investigan la estabilidad de los entrevistadores en el tiempo, el valor mínimo fue de 0.14 y el máximo de 0.95. Para los estudios que evalúan la consistencia entre los entrevistadores, los coeficientes de confiabilidad variaron de 0.47 a 0.91.
- Se estima sólo la confiabilidad de los entrevistadores. Los métodos para estimar la confiabilidad se centran únicamente en los errores que surgen de las inconsistencias entre los entrevistadores como del mismo entrevistador, y se asume que este tipo de error es fortuito.
- Son necesarias la preparación del entrevistador y las guías de entrevista. El entrenamiento de los entrevistadores y el uso de guías, ejemplos y escalas para la calificación contribuyen a mejorar la calidad de la información derivada de la entrevista.

Es importante destacar que la confiabilidad de la entrevista básicamente se ha investigado a través del comportamiento del entrevistador, dejando a un lado otros factores importantes que la definen, tales como los reactivos de la entrevista, las ocasiones de su realización, las escalas para calificarlas, etcétera. Asimismo, podemos señalar que se ha carecido de un modelo que permita identificar, describir y cuantificar los principales factores del proceso de evaluación de la entrevista, lo cual ha repercutido, posiblemente, en la escasa o nula investigación del comportamiento de estos factores.

La Teoría de la Generalización (Brennan, 1992; Cronbach, Glaser, Nanda y Rajaratnam, 1972) permite construir un modelo de medición de la entrevista. Esta teoría es una extensión de la Teoría Clásica de la Medida, con la cual se puede estudiar la contribución de múltiples fuentes de variabilidad (i.e., entrevistados, reactivos, entrevistadores y sus interacciones, etcétera) en los procesos de medición. Asimismo, permite modelar las puntuaciones observadas para un examinado en términos de la puntuación asociada al constructo medido, de los errores aleatorios de medición y de los errores sistemáticos (Nunnally y Bernstein, 1995, Lane y Sabers, 1989). Con este modelo se estructuran los factores (llamadas facetas) del proceso de medición y se cuenta con un valioso marco conceptual para entender el origen de la variabilidad de las puntuaciones observadas. Igualmente, permite aislar dichos componentes y proporciona información acerca de sus niveles de variabilidad y su contribución a la confiabilidad. Este marco teórico provee métodos para estimar el impacto de dichas fuentes y para predecir sus efectos en la confiabilidad de la información derivada de la entrevista.

⁰² En el Anexo I se presenta un resumen de la investigación sobre la confiabilidad de la información derivada de la entrevista en los procesos de selección de estudiantes en los Estados Unidos.

Propósito del estudio

El objetivo de esta investigación fue conocer la confiabilidad de la información derivada de la entrevista estructurada, cuando ésta se utiliza como parte del proceso de selección de aspirantes a programas de posgrado en México. El marco de la Teoría de la Generalización se utilizó para definir un modelo que describa y estructure las fuentes de error más importantes. Especial atención se le dio a los factores del número de entrevistadores y número de reactivos que se deben incluir en la entrevista para mejorar la calidad de la información.

Método

Población estudiada

Se estudió el proceso de admisión de la Maestría en Educación Superior, de la Universidad Autónoma de Yucatán (UADY), cuya orientación curricular es la formación de investigadores. El criterio por el cual se seleccionó este programa mexicano fue la accesibilidad al proceso de entrevista y a los resultados de la misma.

La planta docente de la Maestría consta de: 12 profesores de tiempo completo, seis profesores de tiempo parcial y diez profesores invitados (cinco de otras facultades de la UADY y cinco de universidades nacionales y extranjeras). Del total, se seleccionaron 15 profesores para participar en el proceso de la entrevista. El criterio de inclusión de entrevistadores se basó en la disponibilidad de tiempo e interés para participar en el estudio. Las características de los entrevistadores se muestran en la Tabla I, donde se observa que: el 67% son masculinos y el 33% femeninos; el 67% son profesores investigadores y 33% son profesores de carrera; el 80% tienen estudios de maestría y el 20% estudios de doctorado; y, el 93% están contratados de tiempo completo.

Tabla 1
Características de los entrevistadores

Indicador	Genero		Grado Escolar		Estatus académico		Contratación	
	Masc.	Fem.	Maestría	Dr.	Invest	Docente	T.Com	Horas
N=	10	5	12	3	10	5	14	1

Por su parte, la totalidad de aspirantes que solicitaron admisión a este programa en el ciclo escolar 1998-1999 constituyeron nuestra población de estudio. En total participaron 50 estudiantes, cuyas características de género y edad se muestran en la Tabla II. En años anteriores se han rechazado siete aspirantes en promedio.

Tabla 2
Características de los aspirantes

Indicador	Género		Rangos de edad					
	Masc.	Fem.	25-30	31-35	36-40	41-45	46-50	> 50
N =	16	34	12	12	10	9	6	1

Instrumento

Se diseñó una guía de entrevista para ser utilizada por los profesores entrevistadores. La entrevista se planeó para recabar información de cinco atributos no cognoscitivos del aspirante, juzgados deseables para un buen desempeño en el programa y un buen ajuste en la profesión, siendo éstos: 1) la motivación hacia el posgrado; 2) la actitud hacia la investigación; 3) las expectativas hacia el programa; 4) la disponibilidad de tiempo y 5) los factores de riesgo que podrían inducir al abandono del programa. Para evaluar cada atributo se elaboró una pregunta, la cual se califica en una escala tipo Likert, con cinco valores, que describen los diferentes niveles de calidad de las respuestas a cada pregunta.

La guía de entrevista está estructurada en tres secciones: la primera contiene cinco preguntas, una por cada atributo; la segunda está reservada para registrar algunas observaciones adicionales a los atributos medidos; la tercera se reserva para registrar la puntuación obtenida en la entrevista. Las tablas 3 y 4 describen la estructura y contenidos de la entrevista, y la escala de calificación, respectivamente.

La entrevista se estructuró para cubrir la presentación de los entrevistadores, la explicación de los propósitos de la entrevista, la formulación de preguntas relacionadas con los atributos no cognoscitivos evaluados, la aclaración de dudas y respuestas a preguntas formuladas por el entrevistado y la calificación.

Tabla 3
Atributos y reactivos de la entrevista

Atributos	Reactivos *
Motivación hacia el posgrado	1. ¿Mencione las razones por las que Ud. desea estudiar un posgrado en educación?
Actitudes hacia la investigación	2. ¿Cuál es su percepción de la investigación en la educación?
Expectativas hacia el programa	3. ¿Qué espera obtener del programa?
Disponibilidad de tiempo	4. ¿Cómo tiene pensado organizar su tiempo?
Factores de riesgo	5. ¿Describa las situaciones desfavorables para su permanencia en el programa?

** Los reactivos en esta Tabla contienen modificaciones para asegurar la equidad en futuros procesos de selección al posgrado.*

Tabla 4
Ejemplo de la escala de calificación para el atributo no cognoscitivo que se relaciona con las actitudes hacia la investigación

Atributo	Descripción de la escala de calificación *
Actitudes hacia la investigación	El puntaje de uno se asigna a repuestas que reflejan la limitada comprensión del papel de la investigación en la educación. El estudiante no demuestra una actitud valorativa con respecto a la investigación. El puntaje de cinco se otorga a las respuestas que indican la complejidad del análisis que realiza el aspirante para deducir la relevancia de la investigación para la educación. Para el puntaje de cinco, las personas llegan a presentar ejemplos de prácticas o reformas educativas que fueron el resultado de la investigación. El estudiante demuestra una actitud valorativa con respecto a la investigación e incluso puede llegar a formular enunciados en los que indique su deseo de incorporarse lo más pronto posible a trabajos de investigación. Puntajes intermedios son asignados con base en su cercanía a los puntos que definen los extremos de la escala.

** La descripción de la escala contienen modificaciones para asegurar la equidad en futuros procesos de selección al posgrado.*

Procedimiento

Se conformaron cinco comités de entrevista con igual número de integrantes (tres por comité). La asignación de los profesores a cada comité fue aleatoria. La decisión de contar con tres entrevistadores por comité se tomó con base en el número de entrevistados, la disponibilidad en tiempo de los entrevistadores y los requerimientos de la entrevista estructurada. La aleatoriedad fue esencial para inducir homogeneidad entre los comités en cuanto a las características de sus integrantes.

Previo a la entrevista, todos los profesores participaron en un programa de entrenamiento impartido por el personal que diseñó la guía de entrevista. El entrenamiento se dirigió a clarificar preguntas respecto del uso de la guía de entrevista y de la escala de calificaciones y, también, ofreció la oportunidad para modelar las habilidades requeridas para conducir la entrevista.

Los aspirantes fueron asignados aleatoriamente a cada uno de los comités. Cada comité entrevistó a diez aspirantes. Las entrevistas fueron programadas para ser realizadas en un periodo de dos semanas en las que cada aspirante estaba informado de la fecha, lugar y hora a la que debía de asistir a la entrevista.

El tiempo promedio de duración de cada entrevista fue de aproximadamente 25 minutos, en el cual los integrantes de los comités se alternaron para formular las preguntas. Esto fue hecho para reducir la monotonía que pudiera resultar de utilizar sólo a un entrevistador para leer todas las preguntas. Se procuró mantener constante, tanto el contenido de la pregunta como la forma de preguntar.

Cada miembro del comité calificó independientemente las respuestas a cada una de las preguntas, utilizando la escala de calificación diseñada para la entrevista; la calificación se registró individualmente para cada pregunta. Cada entrevistador, reportó la calificación total de cada aspirante, la cual se calculó mediante la suma de las calificaciones otorgadas a cada pregunta. Finalmente, se integró la calificación global del comité para cada aspirante. Ésta fue calculada promediando las calificaciones de cada uno de los tres entrevistadores en el comité para cada estudiante.

Con base en la Teoría de la Generalización, antes mencionada, se construyó un modelo para recrear las condiciones en las que se llevó a cabo la entrevista. Formalmente, el modelo que dio sustento a la entrevista conceptualmente proviene de un diseño cruzado con dos facetas de medición y una población. Este diseño comúnmente se define como **a x r x e** y representa la manera como las puntuaciones de la entrevista fueron recolectadas por cada comité. La Figura 1 describe la estructura del diseño.

Este diseño indica que los aspirantes (a) contestaron todos los reactivos (r) y sus respuestas fueron evaluadas por todos los entrevistadores (e). El modelo anterior posee siete componentes que en conjunto explican la variabilidad de las puntuaciones observadas. La contribución de cada componente fue estimada a partir del algoritmo de los valores esperados de los componentes de varianza para diseños cruzados con dos facetas aleatorias (Brennan, 1992; y Kirk, 1982). A partir de los estimados de la varianza de cada faceta del modelo, se calcularon los coeficientes de generalización (i.e., coeficientes de confiabilidad). Tanto los componentes de la varianza como los coeficientes de confiabilidad fueron estimados para cada comité de entrevista por separado.

Resultados

El modelo hipotético que se propone describe el proceso de medición a partir de sus componentes. Para un sistema de medición con dos facetas cruzadas (**a x r x e**), la varianza de las puntuaciones observadas (s^2_y) puede ser partida en siete componentes como se indica a continuación:

$$\sigma^2_y = \sigma^2_a + \sigma^2_r + \sigma^2_e + \sigma^2_{axr} + \sigma^2_{axe} + \sigma^2_{rxe} + \sigma^2_{axrxe}$$

El significado y la interpretación de cada uno de los términos del lado derecho de la ecuación se presentan en la Tabla 5.

Tabla 5
Fuentes de variabilidad para el modelo teórico de la entrevista

Fuente de varianza	Descripción del tipo de variabilidad	Notación	Media esperada de cuadrados
Aspirantes (a)	Diferencia sistemática entre los aspirantes atribuida a niveles diferentes de desempeño en los atributos evaluados por la entrevista.	σ^2_a	$s^2_{axrxe} + n_c s^2_{axr} + n_r s^2_{axe} + n_r n_c s^2_a$
Reactivos (r)	Diferencia sistemática en la dificultad de los reactivos.	σ^2_r	$s^2_{axrxe} + n_a s^2_{rxe} + n_c s^2_{axr} + n_a n_c s^2_r$
Entrevistadores (e)	Diferencia sistemática en el nivel de exigencia de los entrevistadores.	σ^2_e	$s^2_{axrxe} + n_a s^2_{rxe} + n_r s^2_{axe} + n_a n_r s^2_e$
a x r	Diferencia aleatoria en la inconsistencia entre un reactivo y otro en conductas específicas de los aspirantes.	σ^2_{axr}	$s^2_{axrxe} + n_c s^2_{axr}$
a x e	Diferencia aleatoria en la inconsistencia entre los evaluadores en conductas específicas de los aspirantes.	σ^2_{axe}	$s^2_{axrxe} + n_r s^2_{axe}$
r x e	Diferencia sistemática debida a la diferencia de los niveles de exigencia utilizados para evaluar cada reactivo.	σ^2_{rxe}	$s^2_{axrxe} + n_a s^2_{rxe}$
a x r x e	Residuo que contiene el error de medición asociado a eventos aleatorios, facetas no medidas e interacción entre aspirantes, reactivos y evaluadores.	σ^2_{axrxe}	s^2_{axrxe}

La Tabla 6 exhibe los valores estimados y sus porcentajes para cada una de las fuentes de varianza por comité de entrevista. Esta tabla también contiene los valores de los coeficientes de confiabilidad para cada uno de los comités de entrevista.

Tabla 6
Valores estimados de los componentes de varianza y confiabilidad para cada comité

Comités Fuente	I		II		III		IV		V	
	σ^2	%	σ^2	%	σ^2	%	σ^2	%	σ^2	%
Aspirantes (a)	0.493	45	0.340	42	0.520	50	0.425	42	0.373	46
Reactivos (r)	0.058	5	0.033	4	0.035	3	0.030	3	0.028	3
Entrevistadores (e)	0.023	2	0.015	2	0.000	0	0.002	2	0.017	2
a x r	0.266	24	0.043	5	0.217	21	0.255	25	0.010	1
a x e	0.010	1	0.064	8	0.061	6	0.026	2	0.087	11
r x e	0.033	3	0.001	0	0.005	.06	0.035	3	0.004	.05
a x r x e	0.213	19	0.309	39	0.192	19	0.233	24	0.285	36
Confiabilidad	0.875		0.876		0.876		0.849		0.845	

El componente de la varianza que se asocia con los aspirantes (s_{2a}) representa un estimado de la variación de las puntuaciones en la entrevista atribuida al componente que ésta mide. En la Tabla 6 se observa que este componente resultó ser el más grande en todos los comités. Las diferencias en los niveles del atributo puro explicaron entre el 42% y el 50% de la varianza de las puntuaciones observadas. Estos valores se encuentran reflejados en los coeficientes de confiabilidad los cuales variaron entre 0.845 y 0.876.

La magnitud del componente de la varianza asociado con los reactivos (s_{2r}) es menor ya sea cuando se compara relativamente con otros componentes o cuando se interpreta en términos absolutos. Comparándola con otros componente, los varianza de los reactivos explican entre el 3% y el 5% de la varianza de las puntuaciones observadas en la entrevista. En términos absolutos, si se toma la desviación estándar (i.e., la raíz cuadrada de la varianza de los reactivos), los valores mínimo y máximo de las medias de los reactivos son de aproximadamente 2.25 y 2.75, respectivamente. Este rango es pequeño comparado con el posible rango de las medias de los reactivos, el cual es de 5. Estos resultados son muy similares entre los comités.

De manera similar, el componente de la varianza para los entrevistadores (s_{2e}) resultó pequeño en términos relativos y absolutos. En la Tabla 6 se observa que, comparado con otros componentes de varianza, los entrevistadores explican no más del 2% de la varianza de las puntuaciones observadas. Al considerar la desviación estándar se observa que el valor esperado del rango de las medias es aproximadamente 2.35 a 2.65. Este rango es pequeño comparado con el posible rango de las medias de los entrevistadores, el cual es de 5. Los resultados anteriores son muy similares entre los comités.

Similarmente, la magnitud de las varianzas para las interacciones entre aspirantes y entrevistadores (s_{2axe}) y reactivos y entrevistadores ($s_{2rx e}$) fue pequeña, siendo de 1% y 3%, respectivamente, para el comité I. Los entrevistadores en este comité estuvieron bien calibrados dado que mantuvieron el mismo criterio para juzgar las respuestas y lo aplicaron consistentemente para todos los aspirantes. Los resultados son similares para la mayoría de los comités, exceptuando el comité V. En éste, s_{2axe} explica el 11% de la varianza de las puntuaciones observadas. Este componente se asocia con la inconsistencia entre entrevistadores. Para el caso del comité V, la inconsistencia puede, en parte, ser el resultado de las diferencias en la severidad con la que los entrevistadores evalúan las respuestas de los aspirantes. La posibilidad de que lo anterior se deba a la inconsistencia en el entendimiento de la escala para medir la calidad de las respuestas fue rechazada, debido a que la interacción entre reactivos y entrevistadores fue baja.

En la Tabla 6 se observa que el componente asociado con la interacción entre los aspirantes y los reactivos (IMAG) explicó entre el 24% y el 25% de la varianza de las puntuaciones, para tres de los cinco comités, mientras que para los restantes dos comités este componente explicó entre el 1% y el 5% de la varianza de las puntuaciones. La presencia de una interacción alta es consistente con la naturaleza y dificultad de las preguntas formuladas para cada atributo no cognoscitivo explorado en la entrevista. Por ejemplo, no es lo mismo dar respuesta a una pregunta que explora la motivación para estudiar un posgrado, que dar respuesta a la pregunta que explora experiencia en investigación. La primera pregunta puede ser contestada en términos generales y no requiere de contar con vivencias pasadas; mientras que para dar una respuesta a la segunda, el aspirante debe contar con experiencia previa en investigación y a partir de ella articular su respuesta. Por otro lado, la presencia de una interacción baja entre aspirantes y reactivos indica una mayor consistencia en la forma de responder a cada una de las preguntas por cada uno de los entrevistados.

Finalmente, el componente de la varianza para el residuo se encontró que explicaba entre el 19% y el 39% de la varianza de las puntuaciones. La presencia de residuos cuya magnitud es grande sugiere: a) que la interacción entre los aspirantes, los reactivos y los evaluadores es grande; b) la presencia de otras fuentes de errores que el proceso de medición basado en el diseño de dos facetas no capturó, o c) ambos.

Discusión

Está ampliamente documentado que los exámenes de conocimientos y las pruebas de aptitudes proporcionan información importante en el proceso de selección al posgrado. Sin embargo, este proceso mejora cuando se incluye información de variables no cognoscitivas del estudiante, como la motivación, habilidades de comunicación, curiosidad intelectual, manejo del estrés, entre otras. Por lo común, esta información se obtiene a través de la entrevista.

Sin embargo, la escasa investigación sistemática del valor real que aporta la entrevista es un aspecto que ha captado la atención (Fruen, 1980; Graham y Boyd 1981; y Dipboye y Gehrlein, 1991). El desconocimiento de las propiedades de la entrevista puede inducir a subestimar el impacto de una selección inapropiada de estudiantes de posgrado en el futuro de las instituciones de educación superior.

La Teoría de la Generalización permite entender con mayor profundidad el comportamiento de los errores que intervienen durante la entrevista, para lo cual: 1) proporciona un modelo que estructura las fuentes principales de variabilidad; 2) ayuda a explorar el origen de la variación en las puntuaciones observadas más allá de lo que permiten los índices tradicionales de confiabilidad entre evaluadores y 3) estimula la búsqueda de argumentos teóricos que expliquen la inconsistencia entre lo que el modelo predice y lo que se observa en la realidad

Los resultados mostraron que la varianza asociada con los aspirantes es la fuente que más contribuye a explicar la variación en las puntuaciones dadas por los comités de entrevistadores. También se observó que los entrevistadores y los reactivos no introdujeron cantidades de error significativas.

Otro resultado interesante fue la magnitud de la confiabilidad, la cual varió entre 0.84 y 0.87 para los cinco comités estudiados. Estos valores pudieran ser vistos como aceptables si se toma en cuenta que en estudios previos el rango de valores de los estimados de confiabilidad entre evaluadores, se distribuyó entre 0.47 y 0.91 (Shaw, Martz, Lancaster y Sade, 1995; Carpio y Brown, 1993).

El contar con un entrenamiento y guías estructuradas para llevar a cabo la entrevista son condiciones necesarias más no suficientes para una entrevista de calidad (Arce Ferrer, 2000). La presencia de interacciones altas entre los aspirantes y los reactivos nos lleva a pensar en la necesidad de prestar mayor atención a la decisión de medir cada atributo no cognoscitivo con una pregunta. La guía de entrevista como se diseñó contiene cinco preguntas en la que cada una explora uno de los cinco atributos no cognoscitivos.

La combinación de múltiples atributos en una sola guía puede, en parte, ser responsable de la presencia de una interacción alta entre aspirantes y reactivos, así como de su variación a través de los comités. Dada la naturaleza multidimensional de la cualidad medida en la entrevista, la interacción refleja la inconsistencia de los aspirantes de una pregunta a otra y también la inconsistencia de un atributo a otro. En tres de los cinco comités se observó la presencia de interacciones altas para el componente mencionado. Esta variación pudo deberse a que algunos de los comités formularon preguntas adicionales a las establecidas en la guía, o bien, formularon preguntas para dar seguimiento a ciertas respuestas.

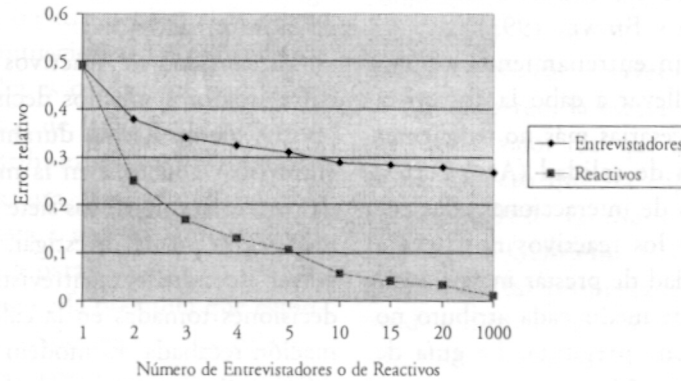
La disponibilidad de información acerca de las fuentes de error y sus magnitudes permite también evaluar las decisiones tomadas durante la planificación de la entrevista y compararlas con posibles alternativas. El contar con estimaciones de la contribución de cada faceta al proceso de la entrevista ayuda a identificar aquellas que están afectando la calidad de la información. Con esta información se puede ponderar la necesidad de intervenir el proceso de entrevista con el propósito de mejorarlo. Es decir, se está en una mejor posición de formular posibles acciones, que de implementarse disminuirían el efecto de los errores y por consiguiente redundarían en una mayor calidad de la información.

La cantidad de reactivos y el número de entrevistadores son dos decisiones importantes que fueron hechas durante el diseño de la entrevista. Con base en la magnitud observada para cada uno de los siete componentes del modelo, se puede investigar, sin necesidad de volver a conducir la entrevista, el efecto de las decisiones tomadas en la calidad de la información recabada. El modelo permite explorar de manera sistemática el efecto de variar el número de reactivos o variar el número de entrevistadores en el estimado de la confiabilidad de la información. Por ejemplo, de haberse utilizado un proceso de entrevista en la que cada comité estuviera integrado por un entrevistador quien evaluara los cinco atributos con una pregunta, la varianza del error sería igual a la suma de las varianzas de las interacciones que afectan la posición relativa de los aspirantes (Brennan, 1992). Por ejemplo, para el comité I, el valor del error sería igual a 0.4893, con un valor esperado de la varianza igual a 0.9886 y un valor del coeficiente de confiabilidad de 0.505.

La Figura 2 describe los resultados de explorar dos alternativas para mejorar el proceso de la entrevista. La alternativa "A" consistente en variar el número de entrevistadores en cada comité de 1 a 2, 3, 4,5,10, 15, 20 o 1000. Con esta alternativa se busca disminuir el impacto de la varianza en el error de las fuentes en la que intervenga la faceta del entrevistador. La segunda alternativa investigada, la alternativa "B", explora lo que sucedería con la magnitud del error cuando el número de reactivos en la guía de entrevista se variara de 1 a 2, 3, 4,5,10, 15, 20 o 1000.

Note que las condiciones con las que se llevó a cabo la entrevista (3 evaluadores por comité y cinco preguntas) forman parte del repertorio de variaciones y las restantes son nuevas condiciones para predecir lo que sucedería si se modificasen las condiciones originales de la entrevista. La Figura 2 describe la magnitud de la varianza del error para las dos alternativas exploradas.

Figura 2
Relación entre el error relativo y las condiciones



Con respecto al número de entrevistadores (alternativa A), se observa una desaceleración de la curva la cual se estabiliza cuando el número de entrevistadores es igual a cuatro. Se observa que las reducciones mayores ocurren en el intervalo de uno a cuatro entrevistadores, en el cual el error se reduce de 0.50 a 0.31, aproximadamente. A partir de cinco entrevistadores la reducción en el error es mínima y no justificaría la inversión de recursos.

En cuanto al número de reactivos (alternativa B), cuando éste se triplica, de 1 a 3, el error se reduce de 0.50 a 0.185, aproximadamente. También se nota que el máximo descenso en el error ocurre en el intervalo de 1 a 10 reactivos. Después de 10 reactivos, el comportamiento del error es estable y una mayor reducción no justificaría la inversión en tiempo y recursos. Este resultado corrobora la práctica común de seleccionar un número reducido de preguntas para formular en una sesión de entrevista (Campion, Palmer y Campion, 1997).

Finalmente, las dos alternativas pueden ser contrastadas. Aumentar el número de reactivos reduce más el error que aumentar el número de entrevistadores. Con la primera se puede, prácticamente, desaparecer el error; mientras que con la segunda el valor mínimo que puede alcanzar el error es de 0.25, aproximadamente. Es decir, el rango efectivo en la reducción del error es casi del doble para la alternativa "B" que para la alternativa "A".

Los resultados de esta investigación pueden ser utilizados para proporcionar lineamientos para el uso de la entrevista en los procesos de selección de aspirantes al posgrado. Primero, se recomienda que las instituciones de educación superior investiguen la confiabilidad de sus procesos de entrevista. El modelo propuesto en este estudio puede constituir el punto de partida para dichos estudios. Segundo, se recomienda investigar los estimados de los componentes del modelo y utilizarlos para realizar predicciones en cuanto a la reducción del error a través de las modificaciones del número de reactivos y de entrevistadores. Tercero, sería deseable contar con definiciones explícitas de las variables no cognoscitivas que se miden en la entrevista. Particularmente, se recomienda formular más de una pregunta por cada una de los atributos medidos y no excederse de diez preguntas por atributo. Finalmente, se recomienda investigar el grado de intercambio que puede haber entre grupos de entrevistadores. La formación de grupos de entrevistadores mediante asignación aleatoria es insuficiente para garantizar la equivalencia de los grupos. Esta falta de equivalencia entre grupos de entrevistadores necesita verificarse, no obstante el proceso de entrenamiento en el que éstos hubiesen participado.

Referencias

- ARCE FERRER, A., Burgos, Fajardo, R. y Esquivel Alcocer, L. (1999). *La entrevista como herramienta de selección en la Maestría en Educación* . Proyecto financiado por la Universidad Autónoma de Yucatán, México con clave FEDU-99-002.
- ARCE FERRER, A. (2000). *Diseño de entrevistas fiables a partir de resultados de la teoría de la generalización* . IV Reunión Internacional de Pensamiento y Lenguaje, Guanajuato, México.
- BRENNAN, R. (1992). *Elements of generalizability theory* , Iowa City, IA: The American College Testing Program.
- CAMPION, M, Pursell, E y Brown, B. (1988). “Structured interviewing: Raising the Psychometric Properties of the Employment Interview”, *Personnel Psychology* 41: 25-42.
- CAMPION, M, Palmer, D. y Campion, J. (1997). “A review of structure in the selection interview”, *Personnel Psychology* , 50: 655-702
- CARPIO, B. y Brown, B. (1993). “The admission process of a bachelor of science in nursing program: initial reliability and validity of the personal interview”, *Canadian Journal of Nursing Research* , 25 (3), pp. 41-52.
- CRONBACH L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements* , New York: Wiley.
- DIPBOYE, R. (1992). *Selection Interviews: Process Perspectives* , Cincinnati: Southwestern Publishing Co.
- DIPBOYE, R. y Gehrlein, D. (1991). “The graduate admission interview”, *College Admissions* , Winter, pp. 50-59.
- EDWARDS, J., Johnson, E. y Molidor, J. (1990). “The interview in the admission process”, *Academic Medicine* , March, pp. 167-175.
- ELAM, C., Johnson, M. y Lenhoff, K. (1994). “The medical admission interview: conduct and content”, *College and University* , Fall, pp. 20-27.
- FRUEN, M. (1980). “Medical school admission interview: pro and con”, *Journal of Medical Education* , 55, pp. 630-631.
- GABARD, D., Porzio, R., Oxford, T. y Braun, R. (1997). “Admission interviews: questions of utility and cost in masters of physical therapy programs in the United States”, *Physiotherapy Research International* , 2 (3), pp. 135-149.
- GRAHAM, J. y Boyd, M. (1981). “A structured interview for dental school admissions”, *Journal of Dental Education* , 46 (2) pp. 78-82.
- HARASYM, P., Woloschuck, W., Mandin, H. y Brundin-Mather, R. (1996). “Reliability and validity of interviewers’ judgements of medical school candidates”, *Academic Medicine* (Supplement), 71 (1) pp. S40-S42.
- KIRK, R.E. (1982). *Experimental design: Procedures for the behavioral sciences* Monterey, CA: Brooks/Cole.
- LANE, S., y Sabers, D. (1989). “Use of generalizability theory for estimating the dependability of a scoring system for sample essays”, *Applied Measurement in Education* , 2, 195-205.
- MADDUX, R. (1991). *Entrevistas Efectivas* , México: Trillas.
- MARKERT, R. y Shores, J. (1981). “Assessing fairness in the medical school admission interview”, *College and University* , Winter, pp. 160-166.
- MCGAGHIE, W. (1990). “Qualitative variables in medical school admission”, *Academic Medicine* 65 (3), pp. 145-149.
- NAYER, M. (1992). “Admission criteria for admission to physiotherapy schools: how to choose among many

- applicants”, *Physiotherapy* , Canada, 44(3), pp. 41-46.
- NUNNALLY, J. y Bernstein, I. (1995). *Psychometric Theory* . USA: McGraw-Hill Inc.
- PURYEAR, J. y Lewis, L. (1981). “Description of the interview process in selecting students for admission to US medical schools”, *Journal of Medical Education* , 56, pp. 881-885.
- POSTHUMA, B. y Noh, S. (1990). “Interview scores and academic grades as selection criteria for admission to an occupational therapy program”, *Canadian Journal of Educational Therapy* , 57 (5), pp. 285-291.
- POSTHUMA, B. y Sommerfreund, J. (1985). “Examination of selection criteria for a program in occupational therapy”, *American Journal of Occupational Therapy* , 39 (7), pp. 441-445.
- RICHARDS, P., McManus, I. y Maitlis, S. (1988). “Reliability of interviewing in medical school selection”, *British Medical Journal* , 296, pp. 1520-1521.
- ROBERTS, G. y Porter, A. (1989). “Medical student selection-time for change: discussion paper”, *Journal of the Royal Society of Medicine* , 82, pp. 288-291.
- SCOTT, A., Chase, L. y Lefkowitz, R. (1995). “A national survey of admission criteria and processes in selected allied health professions”, *Journal of Allied Health* , 24(2), pp. 95-107.
- SHAHANI, C., Dipboye, R. y Gehrlein, D. (1991). “The incremental contribution of an interview to college admissions”, *Educational and Psychological Measurement* , 51(4), pp. 1049-61.
- SHAW, D., Martz, D., Lancaster, C. y Sade, R. (1995). “Influence of medical school applicants” demographic and cognitive characteristics on interviewees” ratings of non cognitive traits”, *Academic Medicine* , 70 (6), pp. 532-536.
- TAYLOR, T. (1990). “The interview: one more life?”, *Academic Medicine* , March, pp. 177-178.
- VARGO, J., Madill, H. y Davidson, P. (1986). “The pre-admission interview as a predictor of academic grades and fieldwork performance”, *Canadian Journal of Occupational Therapy* , 53 (4), pp. 211-215.
- YODAS, J., Bogaro, C. y Suman, V. (1996). “Level of intrateam agreement on candidate scores in a structured interview process”, *Journal of Allied Health* , Fall, pp. 303-312.

Anexo 1

Resumen de una muestra de estudios de confiabilidad de las entrevistas

Autor	Características de la entrevista	Atributos medidos	Hallazgos
Youdas <i>et al.</i> (1996)	Entrevistadores calificaron un grupo de entrevistas estructuradas filmadas. Un año después, el mismo grupo de entrevistadores se reunió y calificó las entrevistas de nuevo. Se proporcionó entrenamiento a los entrevistadores en la primera ocasión más no en la segunda. La confiabilidad fue estimada mediante el grado de acuerdo en las ocasiones	Cuatro habilidades no cognitivas	Los niveles de correlaciones variaron ampliamente. El rango de las correlaciones fue de 0.14 a 0.95. Las puntuaciones en la segunda ocasión variaron más que las de la primera ocasión. Los autores piensan que este resultado se pudo deber a la falta de discusión previa del criterio a utilizar
Shaw, Martz, Lancaster y Sade (1995)	Las calificaciones de una entrevista estructurada fueron analizadas. Se proporcionó entrenamiento a los entrevistadores. La confiabilidad fue estimada mediante el grado de acuerdo entre entrevistadores.	Seis habilidades cognitivas.	La mediana de los coeficientes de confiabilidad fue de 0.47. No obstante el entrenamiento recibido, el estimado de la confiabilidad fue bajo.
Carpio y Brown (1993)	Se utilizó entrevista semi estructurada en la que se proveyó de lineamientos, muestra de preguntas y escala de calificaciones. Se proporcionó entrenamiento a los entrevistadores.	Cuatro habilidades no cognitivas.	El rango del coeficiente de confiabilidad varió de 0.79 a 0.91
Richards, McManus y Maitlis (1988)	La confiabilidad fue estimada mediante el grado de acuerdo entre entrevistadores. Los mismos entrevistadores calificaron en una segunda ocasión las entrevistas. El tiempo entre la primera y segunda ocasión fue de tres años. La entrevista fue filmada. Se proporcionó entrenamiento a los entrevistadores en las dos ocasiones. La confiabilidad fue estimada mediante el grado de acuerdo en las dos ocasiones.	Cinco habilidades no cognitivas	Los niveles de correlaciones variaron ampliamente. El rango de las correlaciones fue de 0.42 a 0.88. Los autores observaron que la cantidad de experiencia de los entrevistadores mejora la confiabilidad. Entrenamiento previo a la segunda ocasión mejora la confiabilidad de las puntuaciones.