



Centro UC
Medición - MIDE

Pilotajes cualitativos basados en entrevistas cognitivas



Contenido

Presentación	3
I. Descripción de las técnicas de entrevista cognitiva y think aloud.....	5
¿Qué tipo de reportes verbales se utilizan?	5
Reportes concurrentes – <i>think aloud</i>	5
Reportes retrospectivos – <i>entrevista cognitiva</i>	6
II. Consideraciones metodológicas para el diseño de un estudio cualitativo basado en entrevistas cognitivas	8
A. Diseño muestral	8
Acerca de quienes componen la muestra	8
Acerca del tamaño muestral	10
B. Definición de protocolos.....	11
Selección de ítemes.....	11
Definición de la lógica de la tarea (Howell et al, 2013)	12
Pauta de entrevista.....	12
C. Análisis de datos obtenidos	14
III. Recomendaciones para la implementación de un estudio cualitativo basado en entrevistas cognitivas	17
A. Perfil y rol del entrevistador	17
B. Capacitación del entrevistador.....	19
C. Métodos de registro	21
IV. Bibliografía citada y bibliografía de consulta	22

Presentación

En el presente documento se expone una síntesis de los aprendizajes y conclusiones a las que llegó el grupo de estudios que se conformó en MIDE UC, en el marco de la implementación de su Plan de Formación del año 2016.

En este trabajo participaron alrededor de 15 profesionales de distintas áreas del Centro. El objetivo fue generar una instancia de indagación y discusión de literatura que abordara el diseño e implementación de estudios cualitativos basados en entrevistas cognitivas a propósito de los procesos de construcción de ítemes y pruebas.

En el marco de la realización de proyectos de este tipo en MIDE UC hay ocasiones en que no es posible llevar adelante una aplicación de campo para pilotear los ítemes y estimar así sus propiedades psicométricas para poder hacer luego el ensamblaje de las pruebas definitivas¹; esto se debe habitualmente a la dificultad de acceder a una muestra que permita hacer ese estudio de campo (por ejemplo, cuando la población que será evaluada es tan reducida, como en el caso de algunas especialidades de profesores de la educación técnico profesional, que no hay casos disponibles para conformar una muestra). En otras ocasiones, se introducen innovaciones en la manera de preguntar o en el tipo de contenidos o constructos que desean ser evaluados, por lo que se torna necesario comprender si el diseño de esos ítemes permite acceder a lo que un evaluado sabe o es capaz de hacer. En ambos tipos de situaciones se ha recurrido a la estrategia de estudios cualitativos basados en entrevistas cognitivas.

Estas técnicas basadas en entrevistas cognitivas aportan información distinta que una aplicación experimental y, ya sea porque no hay opción de implementar una aplicación de campo a una muestra de evaluados que represente bien a la población, o porque interesa recoger evidencia de validez de un instrumento basada en los procesos de respuesta, la información que se obtiene no es “convalidable” o intercambiable con aquella que se obtiene en una prueba de campo.

¹ En este documento se hará referencia al término “pilotaje tradicional” para señalar estas aplicaciones de campo, en que se prueban los ítemes con una muestra de sujetos, se hace análisis de ítemes (basados en TCM y/o IRT y se seleccionan ítemes para ensamblar los instrumentos definitivos).

A partir de la experiencia en MIDE UC, y de lo que el Grupo de Estudios aprendió durante su trabajo, se genera este documento para compartir ciertos criterios básicos que contribuyan a orientar futuros estudios en el marco de los proyectos y asesorías que MIDE UC desarrolla. De este modo, el presente documento constituye una guía y un referente para el diseño e implementación de esta forma de acceder a información de los ítemes contruidos para contribuir a la generación de instrumentos de medición de calidad.

I. Descripción de las técnicas de entrevista cognitiva y think aloud

Los Estándares para el diseño y uso de pruebas en Educación y Psicología dictados por la Asociación Americana de Investigación en Educación (AERA por su sigla en inglés), la Asociación Americana de Psicología (APA) y el Consejo Nacional de Medición en Educación (organismo estadounidense cuya sigla es NCME) establecen la búsqueda de evidencias de validez como un requisito en la elaboración de un test. La validez se refiere al grado con que la evidencia y la teoría dan sustento a las interpretaciones de los puntajes de un test considerando el uso que se le quiere dar. El proceso de validación implica la acumulación de evidencia relevante para dar base científica y sólida a la interpretación de los puntajes que se quiere hacer.

En este contexto, una de las fuentes de validez a utilizar es la información proporcionada a partir de reportes verbales, entendida como verbalizaciones que realizan los evaluados acerca de la tarea en la que han participado. Dichos reportes verbales aportan evidencia importante sobre el ajuste entre el constructo que se quiere medir y la tarea que en efecto realizan los evaluados (Padilla y Benítez, 2014).

Las técnicas de reporte verbal se realizan con participantes que forman parte de la población que se quiere evaluar con un instrumento: puede tratarse de un grupo representativo de la población, o bien, de un grupo específico en atención a alguna característica especial a la que se quiera atender (Howell, Phelps, Croft, Kirui y Gitomer, 2013).

¿Qué tipo de reportes verbales se utilizan?

Los reportes verbales se pueden categorizar en dos tipos, de acuerdo al momento en que son solicitados: reportes concurrentes (think aloud) y reportes retrospectivos (entrevista cognitiva).

Reportes concurrentes – *think aloud*

Son los reportes verbales que solicitan al evaluado *verbalizar los pensamientos mientras se lleva a cabo la tarea solicitada*. Esto es lo que tradicionalmente se ha denominado *think aloud*. En estos reportes, se le pide al evaluado que verbalice la información a la que pone atención mientras genera una respuesta. Esta estrategia de

investigación psicológica es especialmente útil para acceder a los conocimientos y procesos cognitivos que se ponen en marcha cuando se resuelven problemas (Flotts, 2004).

El objetivo es mantener la secuencia del pensamiento lo más inalterada posible, tratando de evidenciar el proceso cognitivo que subyace a la resolución de una tarea. Para esto, se les solicita a los participantes pensar en voz alta mientras resuelven las tareas encomendadas, tareas que por cierto pueden incluir desde ecuaciones matemáticas, pasando por rompecabezas, problemas de comprensión lectora, o *casi* cualquier tarea a nivel cognitivo que se requiera testear (Charters, 2003).

Según una revisión realizada por Charters (2003), una limitación que es necesario tener en cuenta al utilizar este tipo de técnicas, es que los procesos cognitivos más profundos podrían no ser revelados a través de pensamiento hablado, pues se encontrarían fuera del alcance del lenguaje verbal. Incluso, lo que a nivel de pensamiento puede ser resuelto o revelado en un solo instante y de una sola vez, a nivel de lenguaje debe revelarse paso a paso, lo que genera un desfase, algo así como un cuello de botella, entre el proceso cognitivo y el lenguaje en sí.

Reportes retrospectivos – *entrevista cognitiva*

En este tipo de reportes verbales, el evaluado resuelve la tarea solicitada de manera “normal”, y solo al finalizar dicha tarea se le solicita información acerca de su proceso de respuesta. En contraposición a la técnica del *think aloud*, los reportes verbales retrospectivos implican una elaboración activa por parte del sujeto y, en consecuencia, arrojan como resultado un reporte de actividad cognitiva distinto que los que entrega la primera técnica mencionada. Dentro de este tipo de técnicas se encuentran las entrevistas cognitivas.

De acuerdo a una revisión realizada por Young et al (2014), las técnicas retrospectivas como la entrevista cognitiva son más utilizadas que las herramientas concurrentes, ya que proveen información específica sobre los aspectos de la tarea que requieren ser analizados, entre otras cosas, porque son más “naturales” de llevar adelante por parte del evaluado y porque suelen estar guiadas por una pauta que orienta de manera clara y focalizada el levantamiento de la información.



A partir de lo expuesto anteriormente, se puede deducir que los métodos de reportes verbales concurrentes, como el *think aloud*, constituyen una fuente muy rica para indagar en procesos de resolución de problemas a nivel cognitivo. Por otro lado, los métodos de reportes de variables retrospectivos, como la entrevista cognitiva, son muy adecuados, por ejemplo, para analizar la calidad de los ítems que componen un instrumento, pues entregan información precisa y concisa sobre los aspectos que se quieren evaluar (Howell, Phelps, Croft, Kirui y Gitomer, 2013).

En lo que sigue del presente documento, la presentación se focalizará en la técnica de las entrevistas cognitivas (y no en la técnica del *think aloud*), dado que es una técnica más cercana a las necesidades y prácticas que se necesita implementar en el marco de los proyectos que MIDE UC lleva a cabo. En cualquier caso, se recomienda consultar la bibliografía que se presenta al final de este documento si se desea o necesita diseñar e implementar un estudio basado en la técnica de *think aloud*.



II. Consideraciones metodológicas para el diseño de un estudio cualitativo basado en entrevistas cognitivas

A. Diseño muestral

Acerca de quienes componen la muestra

Los criterios en base a los cuales se escoge a los sujetos para el estudio cognitivo dependen del propósito de este estudio y del tipo de representatividad que se requiere. Así, un criterio para seleccionar a los participantes podría ser el buscar la conformación de un grupo que sea representativo de la población evaluada; pero podría darse el caso también de que interese poner un foco especial en una característica específica de esta población (en tanto exista la hipótesis que dicha característica podría estar asociada a una cierta contaminación del constructo) y así, lo que se debe hacer es sobrerrepresentar personas con esa característica en el estudio cognitivo (Howell, Phelps, Croft, Kirui & Gitomer, 2013). Por ejemplo, se podría querer saber si la comprensión de un texto literario (una narración breve) en un curso es muy distinta entre quienes son hablantes nativos del español versus aquellos que han aprendido dicho idioma como una segunda lengua; si este fuera el propósito del estudio cognitivo, entonces habría que conformar el grupo de participantes teniendo un subgrupo de ellos que no sea hablante nativo del español, que esté más ampliamente representado en la muestra que lo que correspondería si se quisiera tener una representación de la población escolar.

Cuando se hacen entrevistas cognitivas para estudiar la validez de un instrumento a partir de evidencia sobre los procesos de respuesta a posteriori, es frecuente que los sujetos que se escogen para participar del estudio cualitativo formen parte de la población que fue evaluada con el instrumento en cuestión. Producto de dicha evaluación, se cuenta con información respecto de su desempeño, la que suele utilizarse como criterio para la selección: si el propósito del estudio es tener una muestra representativa, entonces se puede usar la información del desempeño de los evaluados para replicar la distribución de esos desempeños; si el propósito fuera entender qué hay a la base de ciertos errores que se cometen al responder las preguntas del instrumento, entonces la selección se focalizaría en quienes muestran empíricamente esos errores.

En cualquier caso, la mayor parte de los estudios cognitivos que se hacen para pilotear preguntas (porque no se cuenta con muestras suficientes para hacer un pilotaje “tradicional”) o para estudiar los procesos de respuesta, van a requerir contar con muestras heterogéneas, que si bien no pretenden ser plenamente representativas de la población, sí están compuestas por personas que pueden aportar información variada a las preguntas que orientan la indagación.

Cabe señalar que la literatura revisada y la experiencia en la realización de estudios cognitivos en MIDE, fundamentalmente aquella que tiene que ver con la construcción de pruebas para profesores, muestra que se suele excluir de la muestra a quienes tienen niveles muy bajos de desempeño (información que puede ser inferida de antecedentes académicos o laborales de las personas, o bien, de sus resultados en evaluaciones previas), existiendo para ellos razones de distinto orden:

- Personas de muy bajo desempeño en pruebas de conocimiento entregan poca información en las entrevistas, porque su desconocimiento del contenido evaluado limita su posibilidad de analizar o reflexionar en torno a la formulación de la pregunta y sus distintos componentes: al saber poco, es poco lo que puede decir acerca de las preguntas. En la misma línea, hay veces en que los niveles más bajos de desempeño pueden estar asociados a haber respondido la prueba al azar o a un bajo compromiso con la tarea y, en ese sentido, la información que podrían aportar en una entrevista cognitiva sería muy escasa o de baja calidad.
- Si se sabe poco, participar de una entrevista en que ese bajo conocimiento o desempeño queda absolutamente de manifiesto en una situación de contacto interpersonal, es muy ansiógeno, lo que puede repercutir de manera importante en el devenir de la entrevista.

En suma, la pregunta por las características de los sujetos que componen la muestra de un estudio basado en entrevistas cognitivas se responde a partir del objetivo que se tiene para hacer el estudio. Si hay hipótesis respecto del funcionamiento diferencial que los ítemes podrían tener respecto de algunas variables individuales, entonces se deben escoger sujetos que representen adecuadamente dicha variabilidad (por ejemplo, si hay información que permita suponer que, en el caso de una prueba de conocimientos pedagógicos, profesores que se desempeñan en un determinado contexto podrían tener más dificultades para contestar las preguntas, entonces ese tipo de profesores debería estar especialmente representado en la muestra del

estudio); si lo que se quiere es probar los ítemes para tener evidencia respecto de cómo funcionaría en la población general, entonces se debe tener una muestra heterogénea que garantice que estén representadas las principales diferencias de interés de esa población, teniendo como restricción el que se trate de personas que efectivamente pueden dar información relevante.

Acerca del tamaño muestral

La literatura revisada no entrega una regla clara ni específica respecto del tamaño que deben tener las muestras con las que se llevan adelante los estudios. A modo de ejemplo, se pueden citar diseños con 18 sujetos para estudiar 40 ítemes de conocimientos pedagógicos (Howell, et al, 2013), otros con 36 estudiantes para estudiar 11 ítemes de ciencias (Noble, Suarez, Rosebery, O'Connor, Warren & Hudicourt-Barnes, 2012), y en la experiencia de MIDE en la construcción de pruebas para profesores de modalidad técnico profesional y educación especial, 5 sujetos para estudiar 20 ítemes.

Debido a esta ausencia de reglas, una vez más parece razonable tener en consideración un criterio (en vez de un número fijo), que es de hecho el que subyace a parte importante de la investigación basada en métodos cualitativos: la saturación de la información. Esto significa entrevistar sujetos hasta que la información que un nuevo participante entregue no aporte algo nuevo y sea redundante con lo ya levantado.

A partir de la experiencia en el área Diseño de Pruebas, cuando se han hecho estudios cognitivos como única técnica de pilotaje para probar ítemes, en reemplazo de un piloto tradicional, se ha entrevistado a una muestra de 5 sujetos para estudiar un set de 20 ítemes. Es posible tomar esta práctica como referencia, atendiendo siempre a que el criterio general que se debe satisfacer es el de la saturación de la información señalado en el párrafo anterior.

B. Definición de protocolos

El protocolo para la entrevista cognitiva se relaciona con los siguientes componentes: selección de ítemes, definición de la lógica de la tarea² y pauta para conducir la entrevista.

Selección de ítemes

La selección de ítemes a estudiar bajo esta modalidad depende, obviamente, del motivo por el cual se lleva adelante el estudio: si se quiere pilotear todos los ítemes disponibles para hacer un buen ensamblaje de una prueba definitiva (y no hay muestra para hacer un pilotaje tradicional), es distinto de si se harán las entrevistas para probar nuevos formatos de preguntas o si se quiere medir un constructo que se ha definido de manera muy específica para efectos de una medición particular. En el primer caso, no hay nada que seleccionar (se estudiarán todos los ítemes), mientras que en el segundo, los ítemes que de modo preferencial serán objeto de estudio podrían ser aquellos que merecen más dudas desde su construcción³ (por ejemplo, cuando no se tiene claridad respecto de si permiten capturar de buena manera el constructo que se quiere medir).

En uno de los artículos revisados (Howell et al, 2013), se presenta un estudio basado en las pruebas del proyecto MET⁴ para evaluar conocimiento pedagógico, que tiene como objetivo indagar a posteriori evidencia de validez a partir de los procesos de respuesta de los sujetos. Para ello, se somete a entrevista cognitiva un set de ítemes de las pruebas, que ya habían sido aplicados, y considera para su selección los siguientes criterios:

- Buena capacidad discriminativa
- Distintos grados de dificultad
- Que aborden diversidad de contenidos
- Que involucren diversidad de tareas de enseñanza
- Que tengan diversidad de formatos

² *Task rationale* en la versión original de la bibliografía revisada.

³ Estos ítemes deberían ser los “primeros candidatos” para el estudio, pero dado que la responsabilidad del desarrollador de pruebas incluye recolección de evidencia de validez, es razonable que la selección sea más amplia.

⁴ MET (Measures of Effective Teaching) es un estudio a gran escala realizado en EEUU que analiza la relación entre 4 medidas de calidad de la enseñanza: logros aprendizajes de estudiantes, observación de clases, evaluaciones de los alumnos acerca de sus ambientes de clase y conocimientos de los docentes

Definición de la lógica de la tarea (Howell et al, 2013)

Este es un trabajo de análisis del equipo técnico a cargo del estudio que genera como resultado una explicitación de la racionalidad que hay detrás de cada ítem, es decir, del tipo de conocimiento que se debe emplear para responderlo y del tipo de procesamiento cognitivo que se debe hacer de la información.

Esto permite construir hipótesis de la lógica que hay detrás de la selección o construcción de la respuesta correcta e hipótesis para explicar por qué no se escogen las respuestas incorrectas. A su vez, estas hipótesis son las que permiten diseñar un protocolo de entrevista que ayude a indagar en los modos precisos de razonar acerca de los ítems y aportar así evidencia relevante.

Este proceso de análisis previo a la realización de las entrevistas cognitivas es altamente deseable, toda vez que va a permitir dirigir de modo más provechoso esas entrevistas (porque permitirá focalizar las preguntas en aquellos aspectos que son propios y específicos del ítem), y va a contribuir también a codificar las respuestas y organizar la información que se levanta durante el proceso. Es probable que esta especificación esté disponible desde el proceso de construcción de las preguntas, tanto por el indicador de la tabla de especificaciones que da origen al ítem, como por información específica que se registra en los sistemas de gestión de los bancos de ítems con los cuales se trabaja en MIDE (por ejemplo, ficha del ítem o campos que quedan guardados en el SW de preguntas).

Cuando se trata de ejercicios que implican la resolución de un problema, o la producción de una respuesta, esta descripción de la lógica de la pregunta debiese dar cuenta de los pasos o fases que están involucrados en esa resolución o producción y, en caso de haber claridad al respecto, de los procedimientos más apropiados y los menos apropiados.

Pauta de entrevista

Todas las experiencias de estudios basados en entrevistas cognitivas revisados contaban con una pauta de entrevista estructurada o semiestructurada. También es frecuente que estas pautas tengan un componente común para todos los ítems, y otras preguntas específicas para cada uno de los ítems testeados (y es para estas

preguntas específicas que es de utilidad contar con la descripción de la lógica que subyace a la resolución del ítem, según vimos en la sección anterior).

La conformación de la pauta de entrevista va a estar siempre determinada por las preguntas que guían la realización del estudio. No hay por tanto una “pauta estándar” que se pueda recomendar. Sin embargo, a modo de referencia, se presenta a continuación la pauta de entrevista del estudio basado en preguntas de MET (Howell, Phelps, Croft, Kirui & Gitomer, 2013).

1. ¿Cuál fue su respuesta en este ítem?
2. El contexto presentado, ¿le es familiar en relación con su experiencia?
3. ¿En qué se parece?
4. ¿Cree que este es un contexto que podría ser familiar para otros profesores?
5. Usted seleccionó ___ como respuesta. ¿Puede decirme por qué decidió que esta era la mejor respuesta? (Se agregan preguntas específicas por ítem)
6. Vamos a las otras opciones. ¿Por qué no seleccionó ___? (Se agregan preguntas específicas por ítem)
7. ¿Hay algo que le parezca poco claro en la pregunta? O bien, Usted mencionó que ___ le parecía poco claro, ¿puede decirme por qué? ¿Qué supuesto terminó haciendo para responder a la pregunta?
8. Mirando hacia atrás, ¿hay algún aspecto de la pregunta que le “saltó” al leerla por primera vez?

Otro estudio, esta vez para indagar en diferencias en la comprensión de ciertos ítemes entre estudiantes que hablan inglés como lengua materna, y aquellos que lo manejan como segunda lengua (Young, King, Cogan, Ginsburgh, Kotloff, Cabrera, Cavalie, 2014), utilizó la siguiente pauta para conducir las entrevistas cognitivas:

1. ¿Cree que esta pregunta tiene una respuesta correcta?, ¿por qué?
2. ¿Fue esta pregunta difícil, más o menos, o fácil para usted?, ¿por qué?
3. ¿Es esto algo que han aprendido en su clase de matemáticas (ciencia) antes?
4. ¿Había palabras, ideas, o cualquier otra cosa que hacen que sea fácil responder a esta pregunta?
5. ¿Hay algunas palabras, ideas, o cualquier otra cosa que se hizo esta pregunta más difícil o confusa para contestar?

Finalmente, y teniendo como único propósito ilustrar ejemplos de pautas para conducir entrevistas cognitivas, se presenta el cuestionario que durante 2016 el área



Desarrollo de Pruebas desarrolló y empleó para el pilotaje de las preguntas de educación especial y 10 especialidades de la educación técnico profesional.

1. ¿Qué solicita este ítem? Explique con sus palabras lo que se le preguntó.
2. Ud. escogió la opción Y como su respuesta. ¿Me podría decir por qué decidió que esta era la mejor opción?
 - a. (Solo para quienes dicen que no hay una opción correcta)
De entre las opciones: ¿Cuál le parece más atractiva o está más cerca de escoger? ¿Por qué?
3. Veamos las otras opciones. ¿Por qué no escogió X1, X2, X3 como su respuesta?
4. Hubo alguna palabra, idea, frase o imagen que le hizo más fácil o difícil responder la pregunta? ¿Dónde? ¿Por qué?
5. El conocimiento que se evalúa en este ítem ¿es relevante en la especialidad? ¿Por qué?

C. Análisis de datos obtenidos

Dada la naturaleza de estos estudios, la estrategia de análisis de datos a usar está fuertemente determinada por el propósito del estudio y las hipótesis que se esperan contrastar. En ese sentido, al explorar la calidad de los ítemes, los resultados de las entrevistas cognitivas entregan información sobre aspectos como: qué tan comprensible es una pregunta, cómo se perciben los distractores (en caso de tratarse de un ítem de selección múltiple), qué secuencia de acciones está a la base del proceso de respuesta, etc. En términos generales, el análisis de datos debe orientarse a agrupar información que permita responder a la pregunta sobre las distintas maneras de enfrentar un ítem o responderlo, y si estas difieren significativamente de lo esperado y, consecuentemente, los datos obtenidos deben interpretarse como evidencia a favor o en contra de aquello que se espera. Adicionalmente las entrevistas cognitivas aportan valiosa información para el desarrollo de pautas o rúbricas de corrección de preguntas abiertas, especialmente en la creación de créditos parciales.

En cuanto al análisis de resultados propiamente tal, este depende del tipo de preguntas que incluya la entrevista. Para aquellas preguntas de respuesta cerrada se calcula su frecuencia, y al emplear muestras pequeñas, se sugiere presentar porcentajes solo en caso que sea útil para la interpretación de los resultados. Eventualmente, el tamaño muestral podría permitir que se realizaran, por ejemplo, comparaciones entre grupos, en caso que este fuera un elemento relevante según el propósito para el cual el test se quiere usar. No obstante lo anterior, no debe perderse

de vista que estamos frente a estudios de carácter eminentemente cualitativo, y que es a este nivel de análisis que se centra su riqueza.

Para analizar las respuestas de las preguntas abiertas, será necesario que en primer lugar estas sean codificadas, a través de una pauta, derivada directamente del proceso de construcción y que debe incluir las respuestas, líneas argumentativas o secuencias de pasos esperadas, así como aquellas explícitamente no deseadas, en tanto podrían significar una “contaminación” en la medición del constructo. Este último punto es muy relevante, ya que no solo es necesario acumular evidencia que avale que se está midiendo lo que se pretende medir, sino también de que no se están midiendo aspectos no deseados. Una vez codificadas las respuestas, estas se pueden resumir cuantitativa o cualitativamente en términos de aquello que va a favor y en contra de lo esperado.

Sin embargo, lo central es el enfoque cualitativo, es decir, la realización de un juicio respecto de la centralidad de los hallazgos más allá de la frecuencia con que se presentan, en función de la información que aportan. Por ejemplo, si en un ítem de geometría, 1 participante de 10 lo resuelve correctamente sin hacer lo que el enunciado indica, sino solamente mirando el dibujo, sería importante examinar la representatividad de este examinado en la población a evaluar. E incluso si no fuera un sujeto especialmente representativo, habría evidencia de la existencia de un problema en el ítem que podría afectar la validez del instrumento y que requiere revisión.

Es importante agregar que la interpretación de los datos debe realizarse atendiendo a las características de los sujetos que componen la muestra en términos de su representatividad, en caso que esta sea conocida. Respecto del nivel de habilidad de los respondentes, en términos generales, se esperaría que cada una de las respuestas obtenidas tenga una cierta coherencia con el nivel de habilidad de quien responde, sin embargo no siempre se dispone de esta información, salvo que se trate de estudios de validación realizados después de la aplicación del instrumento. En general, no se recomienda usar el propio instrumento en desarrollo como referente, dado que este mismo es el que está siendo objeto de escrutinio del estudio. En aquellos casos en que los resultados parciales sean la única fuente de información disponible, se sugiere considerar este dato con extrema cautela y como un antecedente preliminar, ya que el instrumento pudiera presentar algún problema o tener alguna distorsión que aun requiera ser corregida.

Respecto de la sistematización de resultados, se sugiere realizar un resumen de la información obtenida en cada una de las preguntas para cada ítem, indicando si se consiguen resultados a favor o en contra de sus hipótesis a la base y qué tan concluyente es la evidencia obtenida en una u otra dirección. Cómo se indicó anteriormente, qué tan concluyente es una evidencia, no se sustenta necesariamente en una alta frecuencia de cierta respuesta, sino en la relevancia que tiene un hallazgo en tanto evidencia de validez a favor o en contra de lo esperado, de acuerdo con lo que el ítem pretende medir. Ejemplos de hallazgos relevantes:

- Evidencia que es posible responder el ítem usando un razonamiento que difiere de lo que este pretende medir.
- Evidencia que existen respuestas o líneas argumentativas correctas y distintas de las contempladas en el proceso de construcción.
- Evidencia que alguna palabra del enunciado o elemento del estímulo que no es relevante para el constructo o atributo que se pretende medir confunde o, por el contrario, ayuda a los respondentes a llegar a la respuesta correcta en caso de preguntas de selección múltiples.
- Evidencia que en preguntas de selección múltiple se eligen distractores por motivos distintos de los que orientaron su construcción.
- Evidencia de errores conceptuales subyacentes a una respuesta o línea argumentativa que no fueron considerados en la pauta de corrección o para construir distractores
- Evidencia que los ítems no elicitaban respuestas o líneas argumentativas no deseadas.

Los resultados de una entrevista cognitiva podrían conducir a:

- Modificación de un ítem, su objetivo o indicador de evaluación, lo que, eventualmente conlleva la aplicación de nuevas entrevistas;
- la aprobación o rechazo de un ítem en tanto existe información robusta en una dirección;
- la realización de nuevas entrevistas, aumentando por ejemplo la muestra de evaluados de ciertas características para examinar la prevalencia de una respuesta o línea argumentativa en particular.



III. Recomendaciones para la implementación de un estudio cualitativo basado en entrevistas cognitivas

Esta sección del documento aborda ideas vinculadas a la ejecución misma del pilotaje basado en entrevistas cognitivas. Se presentan algunas recomendaciones específicas que pueden facilitar el proceso; es un capítulo de orden más práctico y aplicado que los precedentes.

Para llevar adelante este tipo de pilotaje se requiere de un equipo de personas que hagan las entrevistas, quienes las codifiquen y quienes analicen la información levantada. Estas funciones pueden ser responsabilidad de las mismas personas o pueden asignarse a personas diferentes (por ejemplo, quien entreviste puede ser quien luego codifique, o bien puede tener personas/su equipo especialmente dedicados a cada una de estas funciones).

A. Perfil y rol del entrevistador

Antes de comenzar a levantar el perfil del entrevistador, se sugiere determinar si será un equipo externo el encargado de recolectar los datos, es decir, de la aplicación de cuestionarios, la conducción de entrevistas y la categorización de las respuestas de los participantes (esto, en lugar de los equipos responsables de la construcción de las pruebas); y si este equipo será responsable de todas estas tareas o solo de algunas. Los pros y contras de cada una de estas opciones serán materia de análisis de los responsables del proyecto, pero en cualquier caso, se debe cautelar que quienes lleven adelante el proceso sean personas idóneas y capacitadas para cumplir sus funciones a cabalidad.

Asimismo, independiente de esta decisión, se requiere que la recolección de datos se realice mediante procedimientos estandarizados, descritos en los protocolos levantados por las áreas responsables. Además, durante la realización de todos estos procedimientos se debe cautelar el resguardo y confidencialidad de todos los instrumentos y datos que emanan de esta tarea, tal como se trabaja en el marco de cualquier pilotaje y aplicación.

Luego de delimitar los alcances generales de la labor del entrevistador se puede comenzar a listar las labores más específicas que le conciernen, dentro de las cuales podemos mencionar:



1. Presentarse al entrevistado, describir la tarea y hacerle firmar la cláusula de confidencialidad.
2. Aplicar el set de preguntas.
3. Conducir la entrevista en base a un cuestionario, procurando obtener del entrevistado información de la más alta calidad posible.
4. Registrar las respuestas del entrevistado con el mayor detalle posible, sin desatender la situación de entrevista.
5. Categorizar las respuestas del cuestionario, apropiándose de los criterios establecidos.
6. Analizar las respuestas o evidencias basándose en los criterios de categorización establecidos previamente.
7. Categorizar las respuestas entregadas de acuerdo con los criterios establecidos con anterioridad, sistematizando y registrando en forma fidedigna la información recibida.

Respecto a las habilidades que se espera que posean y considerando la importancia de la tarea, el entrevistador debe ser un profesional capacitado, idóneo y bien entrenado en la conducción de entrevistas. Debe ser capaz de extraer información relevante, necesaria, suficiente y de calidad a partir de las respuestas de los entrevistados, de distinguir lo relevante de lo accesorio y orientar las entrevistas en esa dirección. Debe poseer un alto grado de compromiso con la tarea que realiza y estar consciente del impacto de la misma en todo el estudio.

Es recomendable que sea una persona empática, capaz de crear ambientes amables para el desarrollo de las entrevistas; que transmita confianza y seguridad al entrevistado y que vele por la confidencialidad de las respuestas obtenidas.

Tan importante como la capacidad de dar instrucciones claras y precisas a los entrevistados es que se ciña estrictamente a los protocolos determinados para las entrevistas. En el caso que se presente alguna situación, fuera de protocolo, que requiera la toma de decisiones en forma inminente, debe quedar registrada tanto la decisión tomada como su causa u origen y se debe informar en forma inmediata, tan pronto concluya la entrevista, a los responsables del estudio.

La formación profesional en ciencias sociales o ciencias de la educación, como psicólogos, sociólogos o profesores resulta adecuada para esta tarea; también se puede incluir estudiantes universitarios o egresados de las áreas mencionadas. La

experiencia previa en entrevistas cualitativas resulta beneficiosa aunque no debe ser considerada un requisito.

Todos los profesionales que formarán parte del *equipo de entrevistadores* para un pilotaje, cualquiera este sea, deben pasar por un proceso intenso y exhaustivo de capacitación hasta demostrar que han comprendido y se han comprometido con los principios fundamentales de la tarea.

Dentro de las competencias deseables se pueden mencionar: responsabilidad, ajuste a procedimientos, orden, flexibilidad, capacidad de aprendizaje, capacidad de establecer adecuadas relaciones interpersonales, orientación a la calidad y capacidad de atención y concentración.

B. Capacitación del entrevistador

Para llevar adelante un pilotaje basado en entrevistas cognitivas, no solo se debe contar con un equipo de entrevistadores idóneos para llevar a cabo las entrevistas, sino también debe considerarse una fase de capacitación o entrenamiento. Es muy importante que la metodología a utilizar se encuentre bien descrita y documentada, de modo de traspasar de manera adecuada la información a los miembros del equipo externo.

A partir de la experiencia del Área de Desarrollo de Pruebas de MIDE UC, los temas que se recomienda incluir en esta capacitación son los siguientes:

- **Contextualización:** se requiere que los encargados de recopilar los datos manejen información general que pueda ayudarlos a entender la importancia de su tarea y cómo esta se relaciona con los objetivos del estudio. Específicamente, información sobre la metodología escogida (características generales, objetivos, muestra, etc.) y el instrumento a utilizar (propósito, reactivos, categorías de respuestas, etc.).
- **Procedimientos y rol del entrevistador:** es muy importante que en la capacitación se presenten los materiales, tareas y responsabilidades que requiera la metodología escogida, y que deben tener en cuenta los miembros del equipo externo encargados de recabar la información, desde el recibimiento del evaluado hasta el último registro de la información que deben realizar.

- Cuestionario: si se ha decidido utilizar un cuestionario como medio de recolección de la información, el foco de la capacitación debe ser su utilización. Según la experiencias de ADP, algunos puntos importantes son los siguientes:
 - Formato
 - Preguntas y contrapreguntas.
 - Categorías de respuestas para cada pregunta
 - Benchmark de respuestas para cada pregunta
 - Ejemplos de respuestas para cada pregunta
 - Excepciones a las categorías y marcas de respuestas.
 - Ejercicios
- Otros: si la metodología lo requiere, se debiese incluir en esta capacitación ciertos conocimientos relacionados con evaluación (enunciados, contextos, opciones, rúbricas, niveles de desempeño, etc.), específicos para la unidad de análisis del estudio (pregunta cerrada de opción múltiple, pregunta abierta). De la misma manera, si se utiliza un software especial para el registro, codificación y/o análisis de la información, esta instancia de capacitación debe incluir nociones sobre su uso.

La información sobre esta capacitación debe quedar por escrito y además, ser entregada a los miembros del equipo de recolección de datos para su repaso. Cuando incluya información confidencial, se deben tomar los resguardos de seguridad necesarios.

Antes de iniciar el proceso propiamente tal es altamente deseable implementar una *“marcha blanca”* es decir, un ensayo de los procedimientos entre todos los equipos involucrados. Durante esta misma etapa se recomienda realizar un modelamiento del proceso de entrevistas, de modo de detectar posibles problemas con el método y los instrumentos asociados a él y/o dificultades en la integración de la información por parte del equipo externo. Este modelamiento puede ser realizado mediante una situación simulada y/o en la primera entrevista.

Posterior a la capacitación, el área a cargo del entrenamiento debe revisar los registros y categorías utilizadas por los evaluadores, con el objetivo de detectar la adecuada apropiación de estos instrumentos. Y si resulta necesario, volver a modelar e incluso capacitar a los miembros del equipo externo.

C. Métodos de registro

Es de vital importancia manejar un adecuado y exhaustivo registro de la información y los datos recogidos a través de las entrevistas pues esta información será la base de los posteriores análisis del equipo encargado.

Con el propósito de documentar con transparencia y fidelidad los datos, es recomendable que todas las entrevistas sean grabadas y/o transcritas. Esta decisión depende del análisis del tiempo/costo que esto tendría y de los fines del estudio. Si las entrevistas son transcritas, este procedimiento debe hacerse de manera absolutamente fidedigna, es decir, sin editarlas, sin corregir sintaxis, sin resumir, etc.

A partir de estos registros idealmente se sugiere levantar un sistema que permita organizar, compilar y acceder a estos registros, además de posibilitar la realización del análisis de la información de la mejor manera posible. En la experiencia del equipo ADP, el análisis requiere volver a la evidencia (tanto las respuestas del cuestionario, categorización del entrevistador y el audio de la entrevista) de tanto en vez, así que se requiere que este sistema esté disponible para los miembros del equipo, dependiendo de su cargo. Para tales efectos se recomienda el uso de Access, un software que permite gestionar los archivos como bases de datos, hacer validaciones, construir tablas, formularios, consultas y reportes, etc.



IV. Bibliografía citada y bibliografía de consulta

Charters, E. (2003) *The Use of Think-aloud Methods in Qualitative Research An introduction to Think-aloud Methods*. Brock Education, 12, 2, 68-82.

Flotts, M. (2004) *Inteligencia y Conocimiento Claves para Comprender la Experticia*. Tesis para optar al grado de Doctor en Psicología, Pontificia Universidad Católica de Chile.

Howell, H., Phelps, G., Croft, A., Kirui, D. y Gitomer, D. (2013). *Cognitive Interviews as a Tool for investigating the Validity of Content Knowledge for Teaching Assessments*. Educational Testing Service.

Padilla, J., y Benítez, I. (2014) *Validity evidence based on response processes*. Psicothema, 26, 1, 136-144.

Young, J., King, T., Cogan, M., Ginsburgh, M., Kotloff, L., Cabrera, J., Cavalie, C. (2014) *Improving Content Assessment for English Language Learners: Studies of the Linguistic Modification of Test Items*. Educational Testing Service.