

VALIDEZ Y CONFIABILIDAD: CRITERIOS PARA LA ACREDITACIÓN DE LA CALIDAD DE LA EVALUACIÓN ONLINE

Mg. Laura Llull
CIAFIC/ CONICET (Centro de investigaciones en
Antropología Filosófica y Cultura) y Universidad
Nacional de Río Negro. Argentina
laurallull@gmail.com

Resumen:

La evaluación en línea u on line o e-evaluación no se distingue de las otras evaluaciones en sus consideraciones generales y por lo tanto se le aplican todas las premisas que las teorías pedagógicas destinan a este proceso: propósitos, etapas, tipos, condiciones de calidad y hasta estrategias de recolección de datos; pero a su vez tiene una particularidad que viene dada por el entorno (Internet) en donde se desarrolla, el contexto (la era digital, la Web 1.0, la Web 2.0, etc.) en donde se desenvuelve, los participantes del proceso educativo (e- teacher y e-student) y las teorías del aprendizaje y de la enseñanza online. De allí la importancia de contar con estándares que tengan en cuenta estos conceptos considerados de manera integral, abordando lo general y lo particular de la evaluación online, en sus aspectos pedagógicos e informáticos a la vez. De esta manera, se podría superar el recelo que existe para el uso de la e-evaluación, en especial cuando las implicancias de sus resultados son altas tanto para alumnos, como profesores, directivos, etc. y evitar caer en la aplicación única de evaluaciones que tienden a aprendizajes superficiales y declarativos. En este trabajo proponemos los conceptos de validez y confiabilidad del proceso de evaluación entendidos en un sentido amplio e integral, no psicométrico, y confiabilidad del sistema o dependability como criterios para la consideración de la calidad de la evaluación online ya que brindan una mirada de la misma ya no centrada en las herramientas y los resultados, sino en las interpretaciones de lo acontecido, las consecuencias y el proceso evaluativo y de aprendizaje de manera integral, garantizando evaluaciones más útiles, significativas y creíbles.

Palabras claves:

evaluación online, e-evaluación, validez, confiabilidad, calidad educativa, calidad de la evaluación.

Introducción:

La educación a distancia ha ido ganando lugar como alternativa de modalidad educativa, y con este crecimiento ha ido aumentando la preocupación por su calidad para validar la acreditación de los aprendizajes logrados en los programas educativos bajo esta modalidad, en especial aquellos online. Un aspecto que no falta en los criterios, estándares e indicadores para acreditar programas online es la evaluación, llamada frecuentemente e-evaluación o evaluación online. Ésta se define como aquella que involucra el uso de un sistema Web (Crisp, 2007) en línea y que incluye al menos tres componentes (Crisp, 2007):

1. Creación, almacenamiento y reparto de las evaluaciones entre los alumnos.
2. Captura, calificación o puntuación, almacenamiento y análisis de las respuestas de los alumnos.
3. Compaginación, regreso y análisis de los resultados.

Al considerar la calidad de un proceso evaluativo online, consideramos que hay dos grandes áreas a tener en cuenta. Por un lado el lado pedagógico y por otro aquellos aspectos referidos al soporte informático. Generalmente, el uso tradicional que se le ha dado a las tecnologías en la evaluación estaba referido a la colección de datos y su manipulación especialmente en pruebas objetivas y estandarizadas y en evaluaciones cuantitativas. Con los años se han agregado múltiples funciones que realizan las tecnologías, sobre todo con el advenimiento de Internet y la web2.0.

Estos aportes significan buenas y nuevas posibilidades que llevaron a algunos a considerar el aprendizaje online y las evaluaciones como panaceas capaces de modificar los vicios de las viejas evaluaciones presenciales y promover un nuevo modelo (Tobin, 2004; Palloff & Pratt, 2009, Crisp, 2007).

Sin embargo, sobre la e-evaluación pesa una gran preocupación que impacta en las decisiones que de ella se desprenden y sus consecuencias.

Todos los tipos y modelos de evaluación tienen un propósito final que es provocar algún impacto en la realidad. En este sentido, las evaluaciones pueden tener consecuencias que variarán según la utilización que le den los destinatarios. Éstas pueden ser, por un lado, más formales y explícitas (high stakes) o por el otro, ser formativas y sin consecuencias formales explícitas (low stakes). En el caso de las primeras se incluye aquellas que tienen altas implicancias o consecuencias directas importantes para los individuos o instituciones. Como ejemplo podemos encontrar a los exámenes para aprobar un curso, las pruebas para hacer una selección de personas, determinar el ganador de un premio o una beca, etc. El segundo caso implica aquellas que tienen como propósito fundamental provocar consecuencias “blandas” o de bajas implicancias para los interesados. Este tipo de evaluación no tiene una consecuencia formal y principalmente buscan comprender mejor una realidad y promover su mejora y generalmente se las llama evaluación formativa.

En el caso de la e-evaluación a pesar de todas las ventajas y aportes generalmente es aceptada para las evaluaciones de bajo impacto, optando para las pruebas que llevan a decisiones de alto impacto o high stake por instancias presenciales.

La razón de esta situación está basada, la mayor parte de las veces, en sus limitaciones, pero sobre todo en dos de ellas: los problemas respecto a la distribución, acceso, administración y realización de las evaluaciones en lugares remotos y a su vez, para garantizar la autoría de esos trabajos.

Como consecuencia, se ha popularizado el uso de pruebas con respuestas predeterminadas y algorítmicas, preguntas al azar y limitación de tiempo para su realización. Esto ha llevado a que se acuse a las evaluaciones on line de ser más proclives a aprendizajes más superficiales o de tipo declarativo (Tobin, 2004; Palloff & Pratt, 2009, Crisp, 2007) como los exámenes de opciones múltiples o multiple choice y de respuestas breves, en lugar de evaluaciones profundas basadas en preguntas abiertas que tienden al aprendizaje más funcional en lugar de factual.

Como respuesta a esta situación y especialmente con la aparición de la web 2.0 han surgido nuevas lógicas de aprendizaje y enseñanza que conllevan a nuevas formas o modalidades de evaluación que buscan captar aquellos aprendizajes generales profundos comunes a otras modalidades de evaluación y los específicos que surgen como resultado de la lógica de aprendizaje colaborativo, participativo y activo que se desprende del uso de la web 2.0.

Los procesos acreditadores de programas educativos a distancia, especialmente aquellos que llevan a cabo las agencias acreditadoras que como resultado otorgan validez a los títulos y certificaciones, se proponen dar cuenta de este escenario, de sus puntos robustos y de sus limitaciones, y determinar criterios e indicadores que garanticen la calidad de las evaluaciones de aprendizaje y por lo tanto la calidad de los procesos formativos y los aprendizajes logrados.

En este trabajo nos proponemos dar cuenta de ciertos criterios que abarcan las dos áreas de consideración: la pedagógica y la informática. Para el primer caso proponemos los conceptos de validez y la confiabilidad y para el segundo caso el concepto de confiabilidad del sistema o dependability.

Estos conceptos se interrelacionan y se condicionan mutuamente. Por lo tanto, su consideración por separado es a fines favorecer su análisis.

Calidad de las evaluaciones en entornos online de aprendizaje

Para hablar de calidad de la evaluación es preciso que recordemos el valor de la evaluación como herramienta para lograr el aprendizaje y no sólo para medirlo. La evaluación tiene un poder importantísimo para alumnos, profesores y todo el sistema educativo porque determina el destino de los alumnos, programas, trayectorias profesionales, académicas, etc. Por eso es preciso detenerse y reflexionar sobre el fin de la evaluación como una herramienta para igualar las oportunidades educativas y no para discriminar o distinguir aquellos que alcanzan cierto dominio de las competencias requeridas de los que no. Por ello, por este fin igualador de

oportunidades, es que la evaluación tiene que cumplir con todos los requerimientos necesarios para que sea justa y equitativa (Camilloni, 1998). Para que las inferencias producto de una evaluación gocen de credibilidad, ésta debe válida y confiable (Wijekumar, Fergunson y Wagoner, 2006). Todos los aspectos que integran un programa de evaluación condicionan a éste en sus aspectos de validez y confiabilidad y por lo tanto en su credibilidad y utilidad para evaluar justa y equitativamente a los alumnos; por ello es que, como generalmente se hace, las asociamos al concepto de calidad (Camilloni 1998, Ravela, 2006; Messick, 1995; Lafourcade, 1969; Wijekumar, Fergunson y Wagoner, 2006) y deben ocupar especial consideración a la hora de establecer criterios e indicadores para la acreditación de programas educativos a distancia.

Validez y

confiabilidad de la e-evaluación

En cuanto a la forma de concebirlas, podemos identificar dos posturas. La primera pone el énfasis en las mediciones, los cálculos objetivos y los resultados replicables y predictivos. En el caso de la educación, esta mirada de la validez y la confiabilidad es frecuente al hablar de evaluaciones estandarizadas, a gran escala y con fines sumativos y acreditadores. Por otro lado, las visiones menos psicométricas de la validez y la confiabilidad, consideran la integración con el acto formativo, el contexto psicosocial del proceso de enseñanza y aprendizaje, y están ordenadas a resultados que permitan explicar e interpretar los procesos y resultados y lograr mejoras. Se suele utilizar en relación a evaluaciones en un contexto, unas personas y unas relaciones particulares (un curso, una clase, una institución determinada). Claro, estas dos miradas no son excluyentes y se pueden combinar y enriquecer mutuamente. Esta postura entiende que la importancia de la validez y la confiabilidad no reside en el valor predictivo y replicable de los resultados, sino en su capacidad para explicar e interpretar y lograr mejoras en los procesos de enseñanza y aprendizaje (Wijekumar, Fergunson y Wagoner, 2006; Bookhart, 2003). Por lo tanto, lo que necesita ser válido y confiable no son los instrumentos sino la interpretación o el significado de los resultados y de todo el proceso evaluativo, así como las implicancias del mismo (Messick, 1995).

Validez

La validez se adjudica a los resultados de una prueba si realmente se refieren a la conducta que se pretende medir y no otra. Los resultados pueden ser válidos para un propósito y no para otro. Las medidas que dan cuenta de la validez, por lo tanto, nos dicen cuán apropiadas, significativas y útiles pueden ser las inferencias específicas que se puedan realizar a partir de los puntajes (Ruhe, 2002). Es por eso que es clave que el instrumento sea elegido de acuerdo al objetivo o propósito de la evaluación. Si las evaluaciones y sus instrumentos están ordenadas en función del propósito y son válidas, las conclusiones que se hagan en función de esas pruebas serán mucho más útiles y significativas. Por ejemplo, si el objetivo es conocer la capacidad del alumno para argumentar, difícilmente los resultados de una prueba de opción múltiple sean válidos ya que éste tipo de prueba mayormente verifica el recuerdo de hechos, conceptos o datos.

La validez sobre todo en el ámbito educativo no puede ser perfecta porque al evaluar hablamos de inferencias sobre los aprendizajes de los alumnos y éstos están influidos por múltiples factores (Camilloni, 1998). Por lo tanto variará de acuerdo a la situación de aplicación, los propósitos definidos, hasta por el instrumento elegido (especialmente porque ningún instrumento es totalmente perfecto para evaluar cierto aprendizaje). Es decir no hay un ajuste lineal entre instrumento y aprendizaje.

En conclusión, para determinar la validez es preciso conocer los criterios externos a la evaluación misma que orientaron su construcción y administración para poder incluir en la consideración todos esos aspectos que influyen de una manera u otra.

Sin embargo, habrá que validar también los criterios para ver la relación entre éstos con los propósitos de enseñanza del docente, de los alumnos, la institución y la sociedad, y con las teorías que dan fundamento y sostienen el trabajo docente.

Más recientemente (Ruhe 2002; Linn 1993; Messick 1995; Moss, 1992; Ravela, 2006, Gilbert Valverde 2001; Camilloni, 1998) la validez ya no se centra exclusivamente en las pruebas y se ha extendido a considerar tanto las evaluaciones de desempeño, las consecuencias sociales, interpretaciones y usos que se hacen de los resultados de las evaluaciones, el ámbito donde se usarán, los impactos generados en función de su relación con aquello que se desea evaluar y la evidencia empírica.

Gilbert Valverde dice en este sentido “En otras palabras, la validez se refiere a la calidad de las conclusiones que tomamos a partir de las mediciones y a las consecuencias que las mediciones generan en los procesos que se proponen medir” (Gilbert Valverde, 2001:22).

Un aporte interesante es aquel que hace Messick quien propone un concepto unificado de validez que considera tanto la base empírica de todos los datos relevantes para construir la validez y las relaciones implícitas con otros constructos aspectos, llamada la base de evidencia, como las consecuencias de la interpretación y uso de las pruebas (Hube, 2002). Dentro de estas últimas incluye las implicancias y consecuencias actuales y potenciales de la interpretación de las clasificaciones como una base para la acción. Según Messick lo que necesita ser válido no son los instrumentos, etc. sino la interpretación o el significado que se adjudica a los puntajes, así como las implicancias de las acciones que estos significados acarrearán (Messick, 1995). Por lo tanto, para este autor la validez se refiere a evaluar el valor de las inferencias hechas a partir de los puntajes o clasificaciones.

En este sentido un tema que preocupa especialmente a Messick son los casos de invalidez ya que éstos llevan a imparcialidades e injusticias (Messick, 1980; Hube 2002) no porque haya un error en la medición, sino porque pueden ocurrir consecuencias no anticipadas que sean negativas o dañinas para los implicados en las evaluaciones.

En síntesis, la validez según Messick es un constructo multifacético que abarca contenido, la relación costo-beneficio- implicancias y consecuencias no intencionadas.

A continuación presentaremos algunos de tipos de validez que encontramos en los autores consultados:

-

Validez de contenido o curricular

(Lafourcade, 1969; Popham, 1983; Camilloni, 1998): cuando la selección de contenidos incluidos en la evaluación es una muestra representativa de los contenidos del curso, la clase, etc. Es decir, el grado en que la prueba ejemplifica el ámbito de conducta o de contenido sobre

el que se pretende inferir. Este tipo de validez suele ser difícil de indagar por la dificultad que representa la cuantificación de los contenidos de un programa como para realmente verificar la representatividad o ejemplaridad de una prueba de manera cuantitativa.

-

Validez descriptiva

(Popham, 1983): da cuenta del grado en el que un test basado en criterios mide realmente lo que su esquema descriptivo dice medir. Para fijar este tipo de validez, especialmente útil para los test basados en criterios, debemos contar con el esquema descriptivo del test que contiene las especificaciones del mismo y que establece las reglas básicas para crear los ítems. Estas reglas deben carecer de toda ambigüedad porque deben ser claramente entendidas tanto por las personas que construyan los ítems o preguntas y por aquellas que interpretarán el rendimiento de los alumnos o examinados. De esta manera, gracias a la claridad descriptiva de las especificaciones de las pruebas se garantizará que haya cierta homogeneidad en los ítems independientemente de la persona que los confeccionó. Un último paso para confirmar este tipo de validez es pedirle a los examinadores que evalúen si los ítems son congruentes con las especificaciones del test. Un nivel de congruencia de un 90% o más se considera satisfactorio para este tipo de validez.

-

Validez predictiva

(Camilloni, 1998): es la correlación entre el desempeño logrado en la prueba y su desempeño posterior tanto dentro del curso como fuera de él.

-

Validez de construcción o constructo

(Camilloni, 1998): es la coherencia entre el programa de evaluación, el proyecto pedagógico y las teorías que los sostienen.

-

Validez de convergencia

(Camilloni, 1998): relación convergente entre un instrumento de validez ya reconocida y otro que se pretende usar.

-

Validez manifiesta

(Camilloni, 1998): es cuando el programa de evaluación es percibido por sus participantes como válido.

-

Validez de significado

(Camilloni, 1998): es cuando el programa de evaluación tiene significado para los alumnos y los motiva a mejorar sus aprendizajes.

-

Validez de retroacción

(Camilloni, 1998): es el efecto que genera la evaluación en la enseñanza. Es decir aquello que se pretende evaluar, es lo que luego será enseñado.

-

Validez de consecuencias o de uso

(Ravela, 2006): es cuando hay una relación coherente entre el uso que se pretende hacer de los resultados o las consecuencias que generarán ese uso y lo que realmente permiten inferir esos resultados o aquello para lo cual fue diseñada la evaluación. Es decir, la consistencia entre los propósitos de la evaluación y los usos que se la darán a sus resultados.

-

Validez de las condiciones de aplicación

(Ravela, 2006): Las condiciones de aplicación no entorpecen sino que favorecen la obtención de la información pretendida.

Confiabilidad

En cuanto al concepto de confiabilidad o fiabilidad clásico, una evaluación es confiable cuando mide aquello que se busca conocer con precisión y exactitud, así como con la sensibilidad suficiente para identificar los grados de presencia o magnitud (Camilloni, 1998). Por lo tanto, la confiabilidad tiene que ver con la precisión de la evidencia empírica y las mediciones obtenidas. Esta característica implica otra definición que se puede encontrar en la bibliografía (Popham, 1983; Downing, 2004) que establece que es la constancia de la medición (Popham, 1983) o la reproductividad de los resultados de la evaluación. Es decir, es la estimación del grado de consistencia o constancia entre repetidas mediciones efectuadas a los mismos sujetos con el mismo instrumento. Es preciso aclarar que, en cuanto a la medición de conductas, difícilmente se obtendrán dos mediciones exactamente iguales porque hay múltiples factores (cansancio, estrés, nuevos aprendizajes, etc.) que influyen en los individuos y producen diferencias lógicas ya que las personas no se comportan de manera idéntica dos veces. Si el intervalo entre una instancia de prueba y la otra es breve la diferencia entre las mediciones será menor respecto a aquella situación donde hay más tiempo mediando entre las dos. Las diferencias entre ambas instancias son producidas por factores que entran dentro de los márgenes de error posibles. Mientras menores sea ese margen de error, menores diferencias habrá entre las sucesivas oportunidades de prueba y más consistente será el instrumento. En un mundo ideal no habría error en las mediciones y entonces aquello observado sería el dato verdadero. Pero en la realidad esto no sucede y los coeficientes de confiabilidad se utilizan para estimar el error de medición en una evaluación y determinar el grado de consistencia de la prueba (Lafourcade, 1969; Ravela, 2006). La fórmula básica que define esto es: $X = T + e$ Donde: X = es el resultado obtenido; T = la realidad o el resultado verdadero medido; e = el error de medición.

Según la teoría clásica de medición (CMT) es la proporción de la varianza del puntaje verdadero a la varianza de puntaje total.

La confiabilidad no es un valor absoluto y se hablan de grados ya que ninguna medición es perfecta y todas están sujetas a cierto margen de error. Lo fundamental para un proceso evaluativo en términos de confiabilidad es por un lado poder estimar el error para controlarlo y

tenerlo en cuenta en las interpretaciones y por el otro poder establecer el grado de confiabilidad necesaria según el propósito y los objetivos del curso. Dicho de otra manera es preciso establecer el grado de precisión y exactitud que se requiere de una evaluación según el impacto que se busca generar en la realidad, teniendo en cuenta que la confiabilidad es una característica de los resultados obtenidos y no del instrumento en sí (Downing, 2004)¹.

Siendo que la confiabilidad es el grado de constancia de una prueba en sus mediciones (es decir la constancia de una conducta en una prueba) si un examen mide varias habilidades o conductas diferentes se tendrá que considerar la confiabilidad de cada serie de ítems que incluye cada tipo de conductas para hacer una medición correcta.

Podemos determinar la confiabilidad de una evaluación cuando existe (Ravela, 2006; Camilloni, 1998):

- Estabilidad: los resultados permanecen semejantes cada vez que se administra la prueba.
- Exactitud: la evaluación distingue los aspectos que mide de otros irrelevantes.
- Sensibilidad: La evaluación mide con sensibilidad a los cambios de magnitud sin ambigüedades.
- Objetividad: la evaluación no varía según el evaluador.

1

Es importante tener en cuenta que en las pruebas construidas por los docentes el coeficiente aceptable

es un mínimo de .60. y en el caso de las pruebas estandarizadas el mínimo es .90.

Dependability o Confiabilidad del sistema o software

Este concepto proviene del mundo de las computadoras y los sistemas informáticos y se aplica al análisis del software desde una perspectiva funcional. Su consideración es particularmente útil para los casos de evaluación on line (Weippl, 2007). El concepto de dependability o confiabilidad del sistema o software difiere del concepto de reliability el cual corresponde a la confiabilidad de una medición. En el caso de la lengua española esta distinción no es posible porque en ambos casos la traducción sería confiabilidad. Por ello, para diferenciar ambos conceptos en el caso de la dependability se aclarará cada vez que se refiere a la confiabilidad del software.

Este concepto está formado por cuatro elementos (Weippl, 2007):

- Disponibilidad: Es la capacidad de un sistema de estar listo para brindar un servicio correcto. Es decir que el servicio sea posible de acceder por todos sus usuario de manera correcta. La concurrencia es la capacidad de los sistemas para ser usados por todos los usuarios a la vez.

- **Fiabilidad:** Es la continuidad del buen servicio. Un servicio no sólo debería estar disponible y funcionar correctamente al comienzo sino que también deben proveer un buen servicio durante todo el lapso de tiempo que sea requerido.
- **Seguridad:** Es la ausencia de consecuencias negativas para los usuarios y el ambiente producidas por el sistema.
- **Integridad:** Es la ausencia de alteraciones impropias del sistema. Implica la integridad de la aplicación y de los datos. La integridad de los datos es importante porque garantiza que los datos volcados por los alumnos y profesores no puedan ser modificados ni antes ni después. De ser alterados, la integridad implica evidencia de las modificaciones realizadas y del origen de las mismas. El concepto de “no-repudio del origen” implica la integridad concerniente a la identidad del que envía o modifica los datos. Es decir que no se puede negar el origen o la identidad del que hace estas acciones. En síntesis, la integridad significa que los datos y las repuestas de las e-evaluaciones son guardados de manera que no se puedan modificar ni modificar o que si se modifica se puede saber quién realizó los cambios.
- **Mantenibilidad:** Es la habilidad de sufrir modificaciones y reparaciones o actualizaciones. Estos cambios pueden producirse para adaptarse a nuevos requerimientos o porque fueron integrados a otros sistemas, etc.

Conclusión

Los procesos de acreditación de programas educativos a distancia y virtuales u online incluyen generalmente en los estándares que consideran un apartado especial a la evaluación. Su importancia es evidente ya que a partir de la misma se toman decisiones que impactan en la realidad y los sujetos participantes. De allí que las evaluaciones tengan que ser precisas, confiables, creíbles. En el caso de los entornos online de aprendizaje, la atención suele estar ubicada en las herramientas de evaluación y en aspectos vinculados al funcionamiento del sistema, tanto para la seguridad y la determinación de la autoría para evitar el plagio como la administración y utilización de esas herramientas. Es por ello, un aporte la consideración del concepto de dependability que da cuenta de los aspectos que garantizan el funcionamiento confiable del sistema y de la información que en él circula. Sin embargo, el aspecto informático es solo una cara de la moneda. Además, está la cuestión pedagógica que puede ser abordada en forma completa a partir de los conceptos de validez y confiabilidad. A partir de un abordaje más holístico y menos psicométrico de la validez y la confiabilidad, que las considera como una configuración resultante de todo el proceso evaluativo y no sólo de los instrumentos y que tiene en cuenta sobre todo la integración de la evaluación con el acto formativo y el contexto psicosocial del proceso de enseñanza y aprendizaje, podemos concluir que la importancia de la validez y la confiabilidad no reside en garantizar la predictividad y replicabilidad de los resultados, sino en su capacidad para explicar e interpretar y lograr mejoras en los procesos de enseñanza y aprendizaje (Wijekumar, Fergunson y Wagoner, 2006; Brookhart, 2003).

En conclusión, el uso de las tecnologías en si mismas no garantiza que las evaluaciones generen más aprendizajes o sean de mayor calidad, sino que el factor determinante es que todo el

proceso de evaluación sea válido y confiable, en su aspecto pedagógico e informático, para que las decisiones y las consecuencias de las evaluaciones sean justas y equitativas en cuanto a generación de posibilidades de aprendizaje, que al fin y al cabo es lo que buscan garantizar los procesos de acreditación de programas educativos. De allí, el aporte de estos conceptos (validez, confiabilidad del proceso evaluativo y confiabilidad del sistema) para la determinación de criterios, estándares e indicadores de calidad para programas educativos online.

Bibliografía

- Bookhart, M. Susan M. (2003) Developing measurement theory for classroom assessment purposes and uses. Educational Measurement: Issues and practice. Winter 5-12
- Camilloni, A. (1998) La calidad de los programas de evaluación y de los instrumentos que lo integran, en Camilloni, A.R.W., Celman, S., Litwin, E. & M. del C. Palou de Maté. (1998) La evaluación de los aprendizajes en el debate didáctico contemporáneo. Buenos Aires: Paidós.
- Crisp, G. (2007) The e- Assessment handbook. N.Y.: Continuum.
- Downing, S. M. (2004) Reliability: on the reproducibility of assessment data. Medical Education, 38: 1006–1012 - Gilbert Valverde (2001); “La interpretación justificada y el uso apropiado de los resultados de las mediciones”. En Ravela, P. (editor); Los Próximos Pasos: ¿Hacia dónde y como avanzar en la evaluación de aprendizajes en América Latina?. PREAL/GTEE - Lafourcade, P. D. (1969) Evaluación de los aprendizajes. Buenos Aires: Kapelusz - Linn, R. L. (1993) Educational Assessment: Expanded expectations and challenges. National Center for Research on Evaluation. Los Ángeles, USA.
- Messick, S. (1980) Test validity and the ethics of assessment. American Psychologist, 35, 1012 – 1027 En: Ruhe, V. (2002) Issues in the validation of assessment in technology- based distance and distributed learning: What can we learn from Messick´s Framework? International Journal of testing, 2 (2), 143-159 - Messick, S. (1995) Standards of validity and the validity of standards in performance assessment. Educational Measurement: Issues and Practice, 14 (4), 5-8
- Moss, P. (1992) Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research,

62(3), 229-258 - Palloff, R.M. & Pratt, K. (2009) Assessing the online learner. San

Francisco: Jossey Bass. - Popham, W. J. (1983) Evaluación basada en criterios. Madrid: Ed.

Magisterio Español. - Ravela, P. (2006) Para comprender las evaluaciones educativas. Fichas didácticas. PREAL/ Grupo de trabajo sobre estándares y evaluación.

http://www.preal.org/Biblioteca.asp?Pagina=2&Id_Carpeta=225&Camino=315%7CGrupos%20de%20Trabajo/38%7CEvaluaci%F3n%20y%20Est%E1ndares/225%7CPublicaciones [Consultado: Octubre 2010].

- Ruhe, V. (2002) Issues in the validation of assessment in technology-

based distance and distributed learning: What can we learn from Messick 's Framework? International Journal of testing, 2 (2), 143-159 - Tobin, T.J. (2004). Best Practices for Administrative Evaluation of Online

Faculty. Online Journal of Distance Learning Administration, Vol.VII, No II.

<http://www.westga.edu/~distance/ojdla/summer72/tobin72.html>. [Consultado 10/08/09] - Weippl, E. (2007) Dependability in Assessment. International Journal on

ELearning. 6(2) 293-302 - Wijekumar, K.; Ferguson, L. & Wagoner, D. (2006). Problems with

Assessment Validity and Reliability in Web-Based Distance Learning environments and solutions, Journal of Educational Multimedia and Hypermedia; 15, 2: 199 -215

Laura Llull

laurallull@gmail.com <http://ar.linkedin.com/pub/laura-llull/14/444/252> Es Profesora y Licenciada en Ciencias de la Educación, recibida en la Universidad Católica Argentina. Además es Mg. En Educación a Distancia en la Universidad Tecnológica Metropolitana de Chile, título que obtuvo con la tesis Validez y confiabilidad en la evaluación on line: modelos, criterios y estrategias. Actualmente está cursando su segundo año del Doctorado en Ciencias de la Educación en la Universidad Nacional de Cuyo (Mendoza, Argentina). Para desarrollar sus estudios de doctorado cuenta con una beca del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y se desempeña en el Centro de Investigaciones en Antropología Filosófica y Cultural (CIAFIC) en el departamento TICs.

Sus estudios están orientados principalmente a la evaluación y acreditación de los procesos de formación online, en los cuales utiliza la metodología de investigación en la red o e-research. En particular sus esfuerzos están abocados al análisis de los modelos, criterios e instrumentos de evaluación de los aprendizajes en entornos virtuales de educación superior, de tal modo de poder generar estándares de calidad orientadores del diseño de cursos y programas online que permitan perfilar las competencias didácticas del e-teacher para formular y aplicar estrategias de evaluación comprensivas u holísticas.

Además desarrolla tareas docentes en la Universidad Nacional de Río Negro como profesora a cargo de la materia Teoría e Historia Pedagógica en el profesorado virtual y presencial de Lengua y Literatura y en los profesorados presenciales de Teatro, Química y Física.

Previo a sus estudios de doctorado se ha dedicado al asesoramiento y desarrollo de procesos de formación a través del uso de tecnologías, especialmente utilizando herramientas online como e-learning y mobile learning en el ámbito de la educación superior y empresarial.