

Uso de Internet para la extracción de muestras probabilísticas en la formación de investigadores

ANTONIO BARTOLOMÉ PINA
Universidad de Barcelona

Contacto:
abartolomepina@gmail.com

FRANCISLÊ NERI DE SOUZA
Universidade de Aveiro

Contacto:
fns@ua.pt

MARCELO CARNEIRO LEAO
Universidad Federal Rural de Pernambuco

Contacto:
mbcleao@terra.com.br

RESUMEN

El objetivo es analizar y proponer estrategias de investigación sobre datos procedentes de la información latente en Internet. No se trata de una investigación ni una revisión de investigaciones sino un análisis de las posibilidades de estas estrategias a partir de la experiencia personal y algunos trabajos.

Estos trabajos se distinguen de aquellos que recogen los datos "a través de Internet", por ejemplo mediante cuestionarios, y de aquellos que utilizan datos recogidos en Internet (datos estadísticos por ejemplo). Analizan documentos contenidos en Internet, extrayendo los datos a partir del análisis de esos documentos. Esos documentos fueron depositados en su momento en la Red ajenos al futuro uso que se haría de ellos como objeto de investigación. Se puede decir que se trata de un modelo de investigación de "trazos" o "restos dejados por los usuarios de Internet".

Se han revisado investigaciones realizadas a partir de este corpus latente, centrándose en el objeto de la investigación (restringido al ámbito educativo), el contexto, la metodología y las consideraciones éticas.

Las conclusiones resaltan las posibilidades y límites de este tipo de investigaciones.

PALABRAS CLAVE: Internet, muestra, investigación educativa.

ABSTRACT

This study aims to propose and analyse strategies for using latent information at Internet, in Educational research. So, this is not a research work but an exploratory study.

Such work is distinguished from those that collect the data "through Internet" with questionnaires or forms, and from those using databases at Internet. These studies analyse documents in Internet, extracting data from them. These documents, video, photos, texts, ... were some day distributed through the Web not knowing this future use as research object. We can say that it is a research model based on trails left by Internet users.

To this work we have revised the literature with research work based on this latent corpus, addressing four questions: What can we to investigate with these data? Which is the context where to find these data? How to select samples? And what ethics aspects to consider?

Conclusions of this work detail possibilities and opportunities as well as the limits of this research method.

PALABRAS CLAVE: Internet, sample, educational research.

Introducción

La investigación educativa basada en datos observados y recogidos de la realidad se ha enfrentado tradicionalmente a dos procesos que consumen tiempo y recursos: la obtención de los datos y el análisis de esos datos. A mayor volumen de la muestra, más tiempo y recursos debían ser invertidos en esas dos etapas. Un colectivo que se veía especialmente perjudicado era el de los estudiantes que debían realizar sus prácticas investigadoras sobre datos ya existentes y en una dimensión manejable.

Desde hace años el segundo problema se ha ido solucionando gracias a la introducción de las Tecnologías de la información, tanto en el manejo de datos cuantitativos como cualitativos. Lejos quedan los años en que una parte importante del tiempo de formación del investigador se dedicaba a la aplicación manual de complejas fórmulas estadísticas. Y aunque el camino por recorrer en el análisis de la información cualitativa es aún largo, el análisis de textos o registros audiovisuales se ve facilitado mediante nuevas herramientas que en mayor o menor medida facilitan esta tarea.

El primer problema también está encontrando fórmulas de solución. Por un lado, existen grandes bases de datos sobre las que podemos trabajar. Y por otro, día a día Internet (la Web) ha ido acumulando más y más información en forma de textos, imágenes, vídeos, etc. Es lo que podríamos llamar un corpus latente de contenidos, disponibles en Internet para quien desee y posea las habilidades necesarias para extraerlo.

Ya hace más de 10 años, Robinson (2001) señalaba que Internet se había convertido en un fórum de comunicación informal, recogiendo un retrato de las convergencias y divergencias de las personas en diversas áreas de la sociedad humana. En Internet podemos encontrar historias, blogs, redes sociales, foros, agregadores, sitios web y otras muchas herramientas abiertas a las personas comunes. En Internet están presentes datos no numéricos y datos no estructurados como textos, sonidos, vídeos, imágenes, y también existe una gran cantidad de datos numéricos que reflejan el número de accesos, localizaciones, frecuencias, valoraciones rápidas, etc. Además de los datos ya disponibles, que Robinson (2001) llama “narrativas no solicitadas” una gran cantidad de datos pueden ser producidos a través de cuestionarios, encuestas y páginas dinámicas.

Diversos investigadores (Birnbaum, 2004; Neri de Souza y Almeida, 2009) afirman que es posible obtener fácilmente una muestra de gran dimensión a partir de los datos contenidos en Internet. Esta muestra es heterogénea por lo que se refiere a la edad, educación, raza, nacionalidad y clase social. Así, a pesar de que Internet tenía una tasa de penetración global escasa, el 15,4% según Neri de Souza (2010) su rápido crecimiento garantiza la relevancia y representatividad en todas las sociedades y niveles sociales. Como dice Birnbaum (2004), Internet es un medio conveniente para la investigación internacional o intercultural.

El concepto de “*crowdfcasting*” está cada vez más difundido y tiene su base en la investigación en plataformas como Twitter o Facebook o Google+ (*crowdsourcing*). Con esta técnica es posible estudiar el comportamiento de millones de usuarios de estos medios de interacción social (Barbosa y O’Reilly, 2011). El análisis de estos datos (*datamining*) se ha convertido en objeto de trabajo para diversas empresas de prospección de mercado y previsión de comportamiento de los consumidores.

La investigación educativa no ha entendido todavía completamente la riqueza de información y el desafío que supone este contexto. Necesitamos de mejores modelos, técnicas y una reflexión metodológica que construya una epistemología de la investigación sobre datos contenidos en Internet. Por ejemplo, no está claro bajo qué condiciones las encuestas en Internet pueden ser efectivas, qué factores pueden influir en su validez, cómo la implementación de algunas técnicas puede mejorar la ratio de respuestas y la calidad de los datos, y cómo reaccionan los entrevistados a las entrevistas en línea (Zhang, 1999).

En este texto realizaremos una aproximación a los temas planteados, dentro de una perspectiva amplia referida al campo de las Ciencias Humanas y Sociales, aunque con referencias directas al campo educativo, y atendiendo principalmente a los problemas derivados de la obtención de la muestra.

Propósito de la investigación

En este texto se estudiarán en estas cuatro cuestiones:

- Qué se puede investigar en Internet
- Cual es el contexto de los datos en Internet
- Cómo seleccionar muestras en Internet
- Cuáles son las consideraciones éticas a considerar.

Qué se puede investigar en internet

Como ya se ha indicado, a pesar del creciente nivel de penetración de Internet, no todos los grupos y clases sociales están representados. Por ejemplo, no podemos hacer un estudio con indios de una tribu del Amazonas con datos exclusivamente contenidos en Internet. Pero sí podremos hacerlo sobre la opinión de antropólogos que han visitado estas tribus y que exponen sus trabajos académicos en foros en línea.

En este ejemplo no consideramos los artículos que estos antropólogos han publicado en revistas científicas, lo que sería una revisión de textos sino que trabajamos sobre los datos producidos por etnógrafos, expertos ambientales y otros agentes sociales que construyen opinión sobre los indios a través de sus blogs, foros, redes sociales y otros medios informales en Internet. Es decir, datos que tienen un potencial latente de investigación en ciencias sociales y que no fueron construidos para un propósito de investigación específico.

En nuestro ejemplo se ve que aunque no tengamos una representación directa de todos los niveles sociales, podemos tener una representación indirecta.

Otro ejemplo de contenidos analizables modificando los procesos nos lo proporcionan los estudios económicos para calcular la inflación. Ésta se estima mediante llamadas telefónicas, visitas a tiendas y mercados, recogiendo los precios de productos y servicios. Después, estos datos son confirmados y analizados y a partir de ellos se calcula la inflación. Pues bien, dos investigadores del MIT, Alberto Cavallo y Rigoberto Rigobon, pusieron en marcha el proyecto Billion Prices (bpp.mit.edu). Con este software en la red es posible seguir los precios de más de cinco millones de mercancías en más de

70 países, obteniendo tasas de inflación en tiempo real (Easter, 2011). Evidentemente este proyecto no sustituye a las estadísticas oficiales ya que hay precios de prestación de servicios, como peluquería o dentista, que no son fácilmente accesibles a través de Internet. Pero proporcionan un valor añadido que antes no existía.

De modo similar centenares de empresas prestan un servicio de estudios de mercado a través del comportamiento de sus consumidores en la Red. Empresas de venta por Internet toman nota de los patrones de conducta de sus clientes, por ejemplo Amazon, para poder ofrecerles ofertas ajustadas a sus necesidades.

Veamos ahora cómo estos ejemplos pueden trasladarse al ámbito educativo. Si deseamos conocer los hábitos lingüísticos de los jóvenes de Cataluña podríamos hacer una encuesta sobre el uso de la lengua a partir de los centros educativos. Pero también podríamos analizar la lengua utilizada en los mensajes en una red social (o también en foros, etc.) en los que estén activos la localización geográfica y la identificación de edad. Más adelante trataremos el tema de la validez y fiabilidad de estos datos. Lo que nos interesa aquí es señalar que podemos analizar no sólo qué lengua utilizan sino el nivel de vocabulario empleado, la construcción sintáctica, y otros elementos de la redacción de los mensajes. Otro ejemplo, esta vez real, al que haremos referencia más adelante: la imagen que tienen de sí mismas las adolescentes españolas a partir del análisis de las imágenes contenidas en fotologs.

El segundo ejemplo nos muestra que no estamos hablando únicamente de investigaciones de carácter cuantitativo: imágenes y vídeos resultantes de innumerables interacciones sociales en el mundo virtual nos permitirán recoger datos para nuestros estudios.

Algunos investigadores (Battelle, 2005; Neri de Souza, 2010) coinciden en afirmar que en el interior de las bases de datos de Google encontramos un potencial campo de investigación para miles de tesis doctorales en Antropología Cultural, Psicología, Historia, Sociología, Economía o Educación entre otros.

Justificación

Contexto de los datos en Internet en función del tipo de estudio

Podemos considerar que los estudios sobre el corpus latente en Internet puede consistir básicamente en estudios sobre el contenido y estudios sobre la interacción.

Los estudios sobre el contenido son aquellos que buscan datos en los documentos localizados en páginas y sitios web públicos en Internet. Son ejemplos de fuentes de datos los repositorios de documentos textuales, vídeos o música, periódicos, sitios web institucionales, páginas home personales, blogs, wikis, etc.

Los estudios sobre la interacción son aquellos que recogen los datos a partir de las interacciones de los usuarios a través de esos sitios web. Son ejemplos de fuentes de datos los foros, los mensajes en las redes sociales, los comentarios en blogs y en servicios de noticias, los mensajes de correo y las listas de distribución, etc.

Estos estudios sobre la interacción tienen una larga historia en Internet. En el ámbito educativo son numerosas las investigaciones sobre la actividad en los foros (Silverman, 1995; Alavi y Leidner,

2001; Beldarrain, 2006). Un objeto característico de estudio dentro de la Educación y el uso de TICs ha sido el trabajo colaborativo en cursos y programas formativos. Actualmente ha cobrado importancia el análisis de redes (Hiltz, 2005), habiéndose convertido en un campo prometedor y sugerente para la evaluación de los procesos de aprendizaje, desde una perspectiva Conectivista. (Siemens, 2008). Los análisis de la interacción en las redes también se están utilizando como herramienta de evaluación, como ya se hizo en su momento con los foros. En este trabajo no consideramos los estudios sobre la interacción, centrándonos en los estudios sobre el contenido.

En sentido estricto también tendríamos que considerar la búsqueda de información a partir de bases de datos en la red. Pero este tema no es el objeto de nuestro estudio. Las bases de datos presentan características propias por cuanto son repositorios cerrados y estructurados, es decir, de población conocida y con sistemas de selección de datos propios y definidos.

En función del carácter del sitio

Dentro de los sitios y páginas web con contenido a analizar podemos realizar otra distinción importante entre las Noticias (*News oriented*) y los Documentos (*Content oriented*).

En el primer caso encontramos todos los sitios de diarios, sitios institucionales, sitios con noticias en la primera página, y, notablemente, todos los blogs. En el segundo caso todos los documentos, archivos de documentos, y en particular los sitios wiki.

Es importante notar cómo se altera el resultado en el tiempo en función de si seleccionamos “sitios web” o “páginas “. Los sitios web (por ejemplo, un blog) se compone de páginas (en este caso “entradas o noticias”). Lo mismo pasa con un diario, con un documento de texto, con los apuntes de un curso, etc. Las páginas en la Web, a diferencia de lo que sucede con las páginas en papel, no son meras divisiones mecánicas por razones técnicas, sino que suponen la organización de contenidos diferentes. Una única página web puede equivaler a numerosas páginas en papel o incluso a un documento completo, por ejemplo, un artículo científico. Este artículo sería una única “página web” dentro de un “sitio web” que podría ser una revista o una biblioteca.

Veamos que sucede si nuestra selección se refiere a “sitios web”. Si seleccionamos la página principal de un sitio web de actualidad (*News oriented*), el contenido de la página cambiará de un día para otro: recogerá otra noticia, otro hecho u otro comentario. En el caso de un blog puede ser que no tenga nada que ver con el que se había seleccionado la semana anterior.

Si escogemos la página principal de un documento (*Content oriented*), ésta siempre hará referencia al mismo tema y contendrá básicamente el mismo contenido.

La situación se invierte, curiosamente, si la selección de datos se realiza en base a “páginas web”. En un sitio web de actualidad, cada página corresponde a una noticia, y esta no cambia en el tiempo (aunque ocasionalmente se hagan correcciones para subsanar errores de bulto). En una noticia o una entrada de un blog, los únicos contenidos que cambian (por adición) son los comentarios de los lectores, los cuales son trabajados en los “estudios de la interacción” ya señalados. Es decir, la página correspondiente a una noticia determinada no cambia en el tiempo.

En cambio en muchos sitios web que no recogen noticias (*content oriented*) los documentos que corresponden a las “páginas” del sitio están en continuo proceso de actualización y cambio. Por supuesto, en el caso de revistas científicas, repositorios tipo bibliotecas, etc. estos cambios no se

producen o están limitados, aunque tampoco es una norma general (recordar el caso del artículo de Nature del año 2005 sobre la Wikipedia).

No necesitamos decir que estas consideraciones afectan directamente al “fiabilidad” de los datos. Así, las primeras decisiones que debe tomar el investigador se refieren a:

- ¿Voy a hacer un estudio sobre el contenido o sobre la interacción?
- En el primer caso, ¿estoy ante sitios informativos o sitios documentales?
- Y ¿mi muestra se refiere a “sitios web” o a “páginas web”?

Aspectos metodológicos

Análisis de poblaciones y técnicas de selección de muestras

Repositorios cerrados/finitos

Los repositorios cerrados, finitos, son colecciones de documentos que poseen un número determinado de objetos, aunque éste pueda ser muy grande. Un ejemplo sería el repositorio de recursos de aprendizaje MERLOT. Casi todos los repositorios de Objetos de aprendizaje, recursos docentes, etc. asociados a instituciones educativas poseen este carácter.

Un ejemplo característico son los diferentes estudios realizados sobre contenidos de un sitio determinado para establecer su validez y fiabilidad. El año 2005 Giles (2005) realizó un estudio comparativo entre los contenidos de la Wikipedia y de la Enciclopedia Británica encontrando niveles similares de error. Greenemeier (2007) comparó dentro de la misma Wikipedia las aportaciones anónimas con las realizadas por usuarios registrados sin encontrar diferencias significativas. Otro modelo es aquel que contrasta la validez de un sitio a partir del juicio de expertos sobre los contenidos del mismo. También aquí la Wikipedia es un objeto de estudio continuo en diferentes áreas como la Medicina al haberse convertido en la principal fuente de información en la red sobre salud (Leithner et Al, 2010).

No debe extrañar que este sea también un procedimiento habitual para evaluar la calidad de los sitios educativos y de los cursos en línea. Aunque existe una larga tradición de valoración de la calidad de estos cursos a través de su efectividad y no de los contenidos (Strother, 2022), los estudios sobre la calidad de contenidos en los cursos son muchas veces publicados en informes internos.

El carácter finito del repositorio no implica que el número de documentos contenido sea pequeño ya que pueden ser miles o cientos de miles. Tampoco implica que el número de documentos sea fijo o estable: continuamente se van añadiendo nuevos documentos, y, en muchos casos, el paso del tiempo lleva a eliminar algunos de los documentos por razones de obsolescencia, inadecuación, contenido inapropiado, y otras. Como consecuencia tampoco permite suponer que en un momento dado podamos necesariamente decir con precisión el tamaño exacto de la muestra. Sin embargo, el aspecto más interesante de estos repositorios, desde el punto de vista que nos ocupa, es que, en el momento de extraer la muestra utilizando mecanismos propios del repositorio (e.g. “buscar materiales de evaluación sobre ciencias para el nivel de secundaria”), ésta se obtiene a partir del total de elementos contenidos en el repositorio.

Repositorios abiertos/infinitos

Los repositorios abiertos, infinitos parecen en sí mismos una contradicción: todo repositorio de

documentos contiene un número limitado de objetos en un momento dado, sea este número conocido o no. Quizás sería más adecuado hablar de repositorios de fronteras no precisas, o límites que cambian muy rápidamente. Pensemos en el caso de Youtube. El procedimiento de recogida de documentos en Youtube (como en Flickr, etc.) hace que el número de elementos contenidos esté cambiando continuamente, y su número es tan grande que cada selección de una muestra generará un resultado diferente.

Quizás una diferencia conceptual importante sería considerar que las muestras de los repositorios cerrados son en realidad, subpoblaciones definidas, sobre las que a su vez podemos establecer procedimientos aleatorios de selección, en tanto que las muestras de los repositorios abiertos son siempre muestras o subconjuntos de la subpoblación que habríamos definido. La tabla del siguiente ejemplo puede aclararlo. En el primer caso la extracción repetida en el tiempo daría razonablemente conjuntos similares. En el segundo caso, cada nueva extracción prácticamente nos proporcionaría una nueva muestra.

Estos repositorios proporcionan herramientas que facilitan la extracción de muestras, tanto aleatorias como arbitrarias. Es el caso del estudio sobre los vídeos promocionales de la Universidades en YouTube de Inês Silva y Ot. (2010). Frigola (en preparación) construye una propuesta de géneros audiovisuales en Internet a partir de muestras de vídeos en repositorios específicos. Tortajada, I.; Willem, C.; Crescenzi, L.; Araüna, N.; Tellado, I. (2012) analizan la imagen que los adolescentes ofrecen de sí mismos también repositorios específicos de fotologs.

La Web

La tercera opción para la obtención de los datos es realizar una selección a partir del conjunto de la Web. El carácter no estructurado, no controlado, distribuido y heterogéneo de la Web hace imposible escoger datos a partir de toda la Web. Es decir, siempre escogeremos muestras a partir de un subconjunto de datos, es decir, de una subpoblación. El caso más notorio es la división idiomática: generalmente nuestras búsquedas, se limitarán a uno o varios idiomas. Así pues, deberíamos hablar de “la Web de habla inglesa”, etc.

Para la selección de los datos requerimos los servicios de un indexador, por ejemplo, un buscador como Google. Esta es otra fuente de reducción poblacional, ya que automáticamente nos estamos limitando a las páginas sobre las que trabaja el buscador.

Nuestro estudio raramente pretenderá indagar sobre todos los documentos contenidos en la Web. En general buscaremos documentos con algunas características como por ejemplo, “páginas web que tratan el tema de la violencia doméstica”. De nuevo aquí, el proceso pasa por generar o definir una subpoblación de la que luego extraer la muestra a estudiar. En este proceso introducimos varias fuentes de error como las generadas por los universos lingüísticos, la distinción en páginas y sitios web, la redundancia informativa en la Web, la limitación a los contenidos en abierto, etc.

Consideraciones éticas

Las cuestiones éticas relacionadas con la recogida y utilización de datos obtenidos en Internet son objeto hoy de gran preocupación (Bassett y O’Riordan, 2002). Robinson (2001) propone un modelo que describe el proceso para decidir cuando los datos son directamente publicables o necesitan de una autorización.

Además de la autorización de las personas afectadas por los datos, debemos considerar la posible autorización de quien colocó esos datos en Internet. De alguna manera nos estamos beneficiando de su trabajo y esto puede ser resultado de una actitud generosa y activa del “productor” de los datos o simplemente puede ser fruto de nuestra habilidad para encontrar los datos.

A pesar de la escasez de trabajos específicamente sobre la ética en el uso de datos del cuerpo latente en Internet, podemos encontrar algunas propuestas como las de Nosek, Banaji, y Greenwald (2002). Estos investigadores señalan las posibles consecuencias de ausencia de un investigador durante la creación de los datos, la publicidad de datos confidenciales o referidos a la identidad de los autores, la seguridad del almacenamiento, transmisión o monitorización de esos datos.

Reflexiones finales

Como conclusión de este estudio podemos señalar las enormes potencialidades para la investigación sobre temas educativos y sociales que ofrece los contenidos latentes, tanto por lo que se refiere a contenidos de difícil acceso como a amplitud de muestras.

También las posibilidades para la formación de investigadores noveles que ven la posibilidad de realizar estudios sobre muestras difíciles de generar a partir de observaciones directas por sus limitaciones de medios.

Por otro lado hemos visto las precauciones que imponen esta metodología en relación a la validez y a la fiabilidad de la muestra. El carácter de “aleatoriedad intervenida” que posee alguno de los procesos explicados puede plantear problemas si se desea utilizar técnicas inferenciales basadas en una teoría probabilística.

Y hemos hecho mención a una todavía incipiente reflexión sobre los aspectos éticos, en muchos casos condicionados por los cambios sociales generados desde Internet en relación a aspectos como privacidad, transparencia y ubicuidad.

Quizás un aspecto que merece recogerse aquí es la posibilidad de generar procesos híbridos (datos latentes en Internet y datos recogidos en muestras observadas).

Referencias

- Alavi, M., y Leidner, D. E. (2001). Research Commentary: Technology-mediated Learning - A call for greater depth and breadth of research. *Information Systems Research*, 12 (1), 1-10.
- Barbosa, P., Y O'reilly, A. S. (2011). *Harvard Trends: Tendências de Gestão* (1ª ed.). Porto: Vida Económica.
- Battelle, J. (2005). *The Search: Como o Google Mudou as Regras do Negócio e Revolucionou a Cultura* (1ª ed.). Lisboa: Casa das Letras.
- Beldarrain, Y. (2006). Distance Education Trends: Integrating new technologies to foster student interaction and collaboration. *Distance Education*, 27 (2), 139-153.
- Birnbaum, M. H. (2004). Human Research and Data Collection Via the Internet. *Annu. Rev. Psychol.*, 55, 803–832. doi: 10.1146/annurev.psych.55.090902.141601
- Easter, M. (2011). The Prices Are Right: Economists Find a Faster, Cheaper Way to Measure Inflation. *Scientific American*, 305 (4), 13.
- Giles, J. (2005). Internet Encyclopaedias Go Head to Head. *Nature*, 438 (7070), 900-901. doi: 10.1038/438900a
- Greenemeier, L. (2007). Wikipedia “Good Samaritans” Are on the Money. *Scientific American*. *Scientific American*, 19 Octubre, 2007. Extraída el 15 de agosto de 2013 desde <http://www.sciam.com/article.cfm?id=good-samaritans-are-on-the-money>.
- Hiltz, S. R., y Goldman, R. (2005). *Learning Together Online: Research on Asynchronous Learning Networks*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Leithner, A., Maurer-Ertl, W., Glehr, M., Friesenbichler, J., Leithner, K., y Windhager, R. (2010). Wikipedia and osteosarcoma: a trustworthy patients' information? *Journal of the American Medical Informatics Association*, 17 (4), 373–374.
- Neri De Souza, F. (2010). Internet: Florestas de Dados ainda por Explorar. *Internet Latent Corpus Journal*, 1 (1), 2-4.
- Neri De Souza, F., y Almeida, P. (2009). Investigación em Educação em Ciência baseada em dados provenientes da internet. Trabajo presentado en el *XIII Encontro Nacional de Educação em Ciências*, Castelo Branco.
- Robinson, K. M. (2001). Unsolicited Narratives from the Internet: A Rich Source of Qualitative Data. *Qualitative Health Research*, 11 (5), 706-714 doi: 10.1177/104973201129119398
- Siemens, G. (2008, 27 January). Learning and knowing in networks: Changing roles for educators and designers. Trabajo presentado en el *ITFORUM for Discussion*, University of Georgia.
- Silva, I., Martins, S., y y, T. (2010). Vídeos promocionais das Universidades no YouTube. *Internet Latent*

Corpus Journal, 1 (1), 34-46.

Silverman, B. G. (1995). Computer Supported Collaborative Learning (CSCL). *Computers & Education*, 25 (3), 81-91.

Strother, J. (2002). An Assessment of the Effectiveness of e-learning in Corporate Training Programs. *International Review of Research in Open and Distance Learning*, 3 (1), 1-16.

Zhang, Y. (1999). Using the Internet for Survey Research: A Case Study. *Journal of the American Society for Information Science*, 51 (1), 57-68.