Abstract
One main challenge that has faced distance education since its inception has been a relative lack of knowledge concerning how students actually interact with the materials. This has made it difficult to decide if changes in content and/or methods make a positive or negative impact on learner behaviour and overall outcomes. Simply because students are at a distance, educators do not get the same kinds of immediate explicit and implicit feedback that comes when face to face. Web analytics provide an incredible opportunity for educators to receive helpful information regarding their students’ usage and behaviour patterns — on a scale that has the potential to transform the entire industry. Utilized to date primarily in business to track the online behaviour of consumer groups and to test related marketing efforts, web analytics can also be used in distance education to improve the tracking of learner behaviour and to test the impact made by changes in content or presentation. In this chapter, we introduce web analytics, discuss its impact through a case study, and offer a vision of what impact data-driven decision making through the use of web analytics can make on distance education now and in the future.

Introduction
As distance education changes and grows, so too should the tools and techniques used to design and evaluate it. One of the major challenges facing those who are designing and evaluating distance education is to...
better understand if and how people actually utilize the educational experiences they create (Cadez, Heckerman, Meek, Smyth, & White, 2003). When do users access their educational materials? What type of browser, operating system, screen size, and connection speed are they using? How do they navigate through what is presented to them? How long do they take to complete certain activities? How does their behaviour relate to their results? Interestingly, as distance education becomes more Internet-based, data can be collected that helps answer these types of questions (Rogers, Flores, & Matthews, 2007). This chapter describes the potential for using web analytics in the design and evaluation of distance education. We describe how current web analytics technology can be used to track and understand the behaviour of students in order to improve their overall experience through a case study outlining the implementation and initial utilization of web analytics with one of the largest distance education providers in the U.S. We then suggest ways in which emerging technologies might enhance this understanding, and how distance educators can take an active role in shaping the future of web analytics.

According to the Web Analytics Association (2005), web analytics is defined as “the measurement, collection, analysis, and reporting of Internet data for the purpose of understanding and optimizing Web usage.” Typically, web analytics have been used in marketing and business settings to monitor and test the best ways to drive traffic to a website, as well as to track user behaviour while on the website, in order to maximize the conversion of visitors into customers. Baker (2007) insightfully observes that “the rapidly growing volumes of computerized data has keyed the need for development of more automated ways of extracting actionable knowledge.” Web analytics is one powerful way of extracting actionable knowledge in both business and in the field of distance education. We will begin by outlining the key terms and general process of utilizing web analytics. Next, we will use a case study to explain the practical implications in applying web analytics in distance education. We conclude with our vision of what the future may hold and with some questions for future research.
Web Analytics Terms and Processes

When adhering to strict ethical standards, many websites can currently use Internet-based tools to track the behaviour of users on their own site in order to improve the site and the experience for their users. Standard ethical practice includes not tracking a user when he or she is not on your site, as well as keeping all individual user data anonymous. Data of individual users is not considered nearly as valuable as the conglomeration of data from hundreds, thousands, and even millions of users. As data is collected and analyzed, trends are noted, hypotheses are formed, and alterations to the website based on those hypotheses can be implemented and tested.

For web analytics to be useful, organizations must first determine what types of outcomes they desire from users. There are so many metrics that could be tracked that it is absolutely essential for stakeholders to identify the metrics most meaningful to them — the ones they want tracked and monitored on a regular basis. These are called the Key Performance Indicators (KPIs). KPIs are determined from the Key Business Requirements (KBRs) or, in the case of distance education, the Key Educational Requirements (KERs). What are the main objectives that you want your website to accomplish? Which metrics will provide you the most meaningful information about how well you are accomplishing them? Take the dashboard of your car as an analogy. Many metrics could be tracked and reported about the current state of your car: combustion chamber temperature, fuel/oxygen mixture, fan belt RPM, and many more. But only a few are displayed on your dashboard, and most of us actively use only a selection of those. So, why are fuel level, speed, and turn signal functionality displayed on your dashboard while coolant level and spark plug efficiency are not? First, because these former metrics have direct, observable consequences to you, the operator (we all know how costly speeding tickets can be). Second, because you, the operator, can do something about them (the pressure of your foot on the accelerator directly affects your speed). Similarly, certain web analytics metrics become KPIs because of their impact on the ultimate outcome, as well as your ability to make actionable decisions based on them.
Depending on which vendor you decide to use (which will be discussed later), data is collected and reported in slightly different ways. Almost all web analytics providers, however, report the following types of data (although they may define them slightly differently):

> unique visitors (either from a unique IP address or an instance of a cookie),
> visits (valuable because each visit represents an opportunity for meeting the relevant KPI),
> page duration (as well as visit duration and bounce rates),
> pathing (including entry and exit pages), and
> visitor demographics (geographic location, time of day/week/month, technology used, etc.).

Data like this makes it fairly easy to discover which pages or sections of your site are the most popular and effective in helping users accomplish any objectives that can be tracked and measured. Data segmentation afforded by more sophisticated tools allows for more nuanced insights to arise regarding particular user groups. Additionally, options are even available for combining data from non-Internet sources with web analytics in order to infer relationships between web usage patterns and information from other user contact points (e.g., in a brick-and-mortar store or classroom).

With all the powerful metrics at your fingertips from even basic web analytics services, it is easy to feel overwhelmed by the amount of data. New users have often described their initial experience as “drowning in data” (Snibbe, 2006). It is also easy to lose track of what you are actually trying to discover, and what you would do with that information if you knew it. Here, it may be helpful to point out the critical yet often subtle distinction between outputs and outcomes: many analytics metrics simply (and valuably) track the outputs of our distance education efforts — how many people visited the site, where they came from, how long they stayed on a particular page, which of our expensive simulations they actually interacted with, their scores on our mini-assessments, and so on.
Interesting as they may be, many of these metrics become meaningful and actionable (and could therefore be termed KPIs) only when they are tied into the overall picture of outcomes, which often goes far beyond clicks and even conversions. What kind of penetration are we getting in our target demographic? What are learners actually taking away from the course, and how has that changed their behaviour?

In helping you understand some of the basics of using web analytics, we now will describe the initial implementation and analysis process of one distance education provider.

Case Study (BYU Independent Study)

Brigham Young University Independent Study (BYU IS) is based at a large, private, western university, and services approximately a hundred thousand distance education students in all fifty states and numerous other countries. Students were enrolled in more than five hundred high school or university courses. In 2008, several members of the BYU IS marketing team attended a presentation given by the primary author of this chapter (who was then faculty in the business school and teaching courses on web analytics). They soon brought him in as a consultant in order to figure out how best to utilize web analytics in their marketing efforts. This naturally led to utilizing web analytics in evaluating user behaviour in their online courses, as well. Some of the major decisions and outcomes of this process to date are described below.

Hiring or training expertise in Web analytics

Using web analytics can be a powerful way to make data-driven decisions, but data can also be overwhelming, confusing, and misleading at times. Having simply a surface-level understanding regarding web analytics is probably not sufficient in many implementation situations or when using the data to make significant, costly decisions. Common mistakes many beginners make can be learned from through painful experience, or avoided through hiring someone with previous experience.

BYU IS chose to hire a consultant who would provide expertise in the key initiation decisions, as well as provide training for selected staff.
(until most of the analysis activities could be performed internally). They also offered certain graduate students, who were being trained by the consultant, access to the data for research purposes in exchange for their analysis and resulting recommendations.

**Determining KERs and KPIs**

> “Not everything that can be measured is important, and not everything that is important can be measured.”
> — Albert Einstein

The mission statement of BYU Independent Study is: “To make quality educational experiences available to all who can benefit from individualized learning.” This statement was broken down into the following two KERs: (1) quality educational experiences, specifically those intended for (2) individualized learning. As a starting place, each of these KERs was assigned related KPIs. In a later section, we will describe those KPIs in relation to the data analysis of a single course.

Obviously not everything that is important can be tracked or measured using data collected from web analytics, so discussion also occurred regarding how to use other research methods to triangulate and enhance the data retrieved from web analytics. In many cases, one research source will provide valuable insights into data collected with a different method. The main goal in defining KERs and determining KPIs, however, is to clarify what indicators can be measured and then select those indicators that are most important to your objectives.

**Choosing a vendor**

As mentioned previously, there are a variety of vendors offering web analytics tools and services (for example, AWStats, Analog, Google Analytics/Urchin, Yahoo/IndexTools, Woopra, Microsoft/Gatineau Project, and Omniture). There are free software and services, and premium software and services. The premium options offer significantly more support and more options for customization, and allow you to find answers to more detailed questions, but the costs could be prohibitive for many distance education providers. Each vendor will offer you
The Use of Web Analytics in the Design and Evaluation of Distance Education

information regarding its service and you can decide which to choose in light of your particular needs and budget.

After some discussion, the initial decision of BYU IS was to use Google Analytics. This was based on ease of implementation and cost (free at the time of use). As BYU IS’ questions and needs become more sophisticated, it is likely that they will upgrade to a premium service (Omniture, for example).

Implementation
In the case of Google Analytics, the basic code is fairly simple to insert in the relevant pages. Additional effort was needed, however, to implement code in a way that data could be collected and reported from groupings of courses into the same report suite (in addition to report suites for individual courses). For example, data can be viewed for all English vs. math and science courses, all university vs. high school courses, and so on. This enables analysts to see patterns more generally, and for decision makers to compare and contrast data from the relevant sections of courses. More detailed implementation would be needed to engage in A/B or multivariate testing (processes that allow users to compare the conversion results of two or more designs in a simultaneous randomized test).

Analysis
BYU Independent Study offers courses in over fifty subject areas, including Anthropology, Organizational Behaviour, Exercise Science, Slavic Languages, and Ancient Scripture. To begin exploring the potential of web analytics in distance education, one of the largest courses (a college algebra course) was chosen for initial analysis. The analysis in this section mainly applies to this course.

The graduate students who performed the analysis were concurrently enrolled in an introductory course on web analytics in online instructional design and evaluation. As part of a class assignment, they attended a presentation regarding the aims of BYU IS and were given access to the Google Analytics data. They were limited in some respects, as they did not have ready access to other data sources, and

Emerging_Technologies_in_Distance_Education_Interior.indd   237
15/06/10   3:40 PM
the data in Google Analytics had only been collected for the three or four months previous to their analysis. Table 12.1 captures the KERs, KPIs, and initial observations from the analysis.

Table 12.1 Summary analysis of college algebra course

<table>
<thead>
<tr>
<th>KER</th>
<th>KPI</th>
<th>Initial Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality educational experiences</td>
<td>Material is engaging</td>
<td>&gt; Average page views per visit are approximately equal to average pages per lesson.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt; As lesson number increases, average page views decrease.</td>
</tr>
<tr>
<td></td>
<td>Educates students</td>
<td>&gt; Page views and time on page metrics for the online evaluation are surprisingly low.</td>
</tr>
<tr>
<td>Individualized learning</td>
<td>Material is clear and easy to navigate</td>
<td>&gt; A noticeable page-viewing trend is consistent throughout the course and across visitor segments.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt; Certain pages are viewed exclusively by a specified visitor segment.</td>
</tr>
<tr>
<td></td>
<td>Anytime, anywhere</td>
<td>&gt; Predictable viewing trends are observed by day of week and time of day.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt; The course has minimal international access, and majority of visits are from two states.</td>
</tr>
</tbody>
</table>

The first KPI that BYU IS chose to explore in connection with their KER of “quality educational experiences” was “student engagement with the course material.” The debate about what constitutes “student engagement” and how it might be reliably measured extends to web analytics. Page-view metrics seem one of several reasonable factors to consider when attempting to measure this elusive construct. The analysts noticed a simple yet interesting correlation between page views per visit and pages per lesson. The averages of each of these metrics were quite close: average page views per visit = 6.5, and average pages per lesson = 5.09, suggesting that on average, a student’s plan is to complete one lesson in a sitting. This is something to consider when determining the amount of content one lesson should cover. Armed with this data, designers and analysts could conduct systematic investigations into the effects of lesson
length on page views, and even tie that data to quiz performance and other evaluative feedback to further optimize their content structure.

Further analysis of page-view metrics revealed several other interesting trends. First and most obvious was that with each subsequent lesson, the average unique pages viewed decreased, falling quite consistently for the initial eight lessons preceding the first midterm. The average unique page views then generally levelled off with only a very minimal decline afterwards (see Figure 12.1).

The data analyzed covered a four-month period from the first of July to the first of November. The observed decrease could have been due to attrition, or to students spending more time on their face-to-face classes and less on independent study courses, or to some other factor altogether. The analysts concluded that it would be easier to justify a particular hypothesis about the decline if there were a year of data to examine, which is the amount of time allotted to finish an independent study course without applying for an extension. Time-based trends like this one can be especially insightful when dealing with user experiences and conversion goals, like many in distance education, that extend over more than a single visit.
Determining the extent and value of the education students receive from a course is unilaterally difficult. This case study proved exceptionally so, as the analysts did not have access to any data about completion or scoring of assignments or midterms. Determining that the data from online course evaluations could prove an appropriate substitute, they analyzed page visits and time-on-page statistics for the course evaluation page. The results were quite surprising: there were only 26 visits from the 1,460 total visitors, and of those few visits, only 8 students spent more than one minute on the evaluation (which contained at least a couple dozen multiple-choice questions and some free-text questions). As it turns out, there was a good reason for this: a written evaluation is handed out with each proctored final. The inclusion of an online evaluation seemed to be a vestigial part of the course buried at the bottom of the menu, and any resulting data simply ignored. This analysis experience illustrates well the importance of critical questioning and communication in dealing with web analytics data. Had the team taken the data from the evaluation page at face value, without any further investigation as to possible causes, their analysis, recommendations, and any resulting adaptations to the site would have been irrelevant (if not an all-out waste).

An additional trend was found when unique page views were analyzed by lesson. The first page of each lesson had many more unique page views than any other page in the lesson. The unique page views fell sharply even between the first and second page of each lesson. Then they remained fairly constant, until the last page of the lesson where there was a very noticeable uptick in unique page views (see Figure 12.2 and Figure 12.3).

One probable explanation for the disproportionate number of first-page views is the navigational structure. Students first navigate by lesson number. Once the lesson is chosen, students are automatically taken to the first page of the lesson. It is only after lesson selection that the next navigational level (pages in the lesson) is displayed for selection. At that point a “course-wise” student who understands that these pages are not necessary for lesson completion may readily skip to the last page, where they are directed to a (third-party) off-site link to complete
The Use of Web Analytics in the Design and Evaluation of Distance Education

Figure 12.2 UPV by lesson page for "All Students" segment

![Graph showing Unique Page Views Lessons 3–8 for "All Students" segment](image)

Figure 12.3 UPV by lesson page for "Engaged Students" segment

![Graph showing Unique Page Views Lessons 3–8 for "Engaged Students" segment](image)
the lesson and assignment. This also could be a plausible explanation for the uptick in views for the last page of each lesson. Here, the team suggested a change that might improve ease of navigation as well as enabling better tracking of student intent: changing navigation of lessons and pages to naming according to content, rather than numbering alone. Often, substantive changes that improve site usability and support student learning can be double-purposed to provide richer analytics data that will help perpetuate the cycle.

A week or so before the presentation, Google rolled out a new analytics suite, including custom segmentation and custom reports. The analytics students took advantage of this, and after a little online research, decided to create two segments: an “all students” segment and an “engaged students” segment. To determine the criteria for inclusion in the “engaged” segment, they simply required the average page views and the average time-on-page to be greater than the average of those metrics for the “all students” segment. The most surprising results came in discovering that though the numbers changed for the two segments, the trends remained the same. These trends become visually apparent when the data is exported to Excel and displayed as line graphs, as seen in Figure 12.2 and Figure 12.3.

So if the trends remained the same, what differed between the two visitor segments? Data sorting and analysis revealed fifteen lesson pages that were viewed exclusively by the “engaged” segment. These pages were all middle pages (neither first nor last pages) of lessons towards the end of the thirty-three-lesson course (lessons 24, 28, 30, and 33). They were also all pages that were fairly complex in the worked examples and/or formulas covered. Is inclusion in the “engaged” segment a predictor of course completion, or is it an indication that the students are actually having a harder time digesting the material (thus spending more time on each page and viewing more pages)? To answer that question, more information and more analysis is needed.

If “anytime” and “anywhere” are important KPIs, then there should be some supportive evidence of that in the analytics data. The client was quite interested in the “where” and “when” data summary. Site access was reported by time of day and by day of week. Not surprisingly, the
biggest day for site access was Monday, with Tuesday through Thursday access being fairly equal, and then falling off on Friday, with Saturday and Sunday being the days with the least traffic. As far as time of day: timing was split fairly evenly between 10 AM and 10 PM, with the hours of noon to 4 PM slightly heavier. It is important to note that often the metrics of initial interest to a client are not necessarily the ones that can yield the most relevant insights or have the most impact on the outcomes.

The geographic distribution of site visitors was also reported. By far the biggest percentage of all site users were from California and Utah. The Independent Study management confirmed that this aligns with registration data. It is interesting to note that although analytics for other college-level distance education courses showed a significant number of international visits, the international visits for the College Algebra course were minimal. It is counterintuitive to think that this could be solely a language translation issue, as mathematics could be considered a symbolic language in and of itself with few cultural or international barriers. Could a lack of international students (from Canada, Japan, China, Germany, and the UK) enrolled in the College Algebra course possibly indicate less of a need because post high-school students in those countries are more proficient in basic algebra? Though there is too little data to draw such a sweeping conclusion, this type of question points out the potential for web analytics to inform us about educational issues and trends that go well beyond the design and evaluation of a particular course.

Additional questions for future analysis include: What else can be determined given more in-depth access to course content and metrics, including third-party content, and test and completion data? Do these other sources of data confirm or refute web analytics data? Were navigation recommendations implemented? Where are the best sections to set up A/B testing to confirm/refute improvement?

Discussion

Effective use of web analytics data in distance education, and most other applications, comprises four basic objectives (Hendricks, Plantz, & Pritchard, 2008), perhaps more aptly termed opportunities:
First, we must **define** the goals/objectives of the interaction. One of the unique challenges of using web analytics in distance education, as opposed to marketing or business applications, is determining how to translate Key Educational Requirements into Key Performance Indicators — essentially tracking learning through clicks. After all, learning the chemical processes of photosynthesis through an online simulation may involve the same physical behaviours (clicking, reading, scrolling, etc.) as purchasing a duvet cover, yet certainly the mental behaviours of these two interactions would be significantly different. The challenge for distance educators using web analytics is to discover how the learning and the cognitive processes and behaviours of their users are manifested in their online processes and behaviours.

Secondly, we must **measure** both the outputs and outcomes of the interaction. Here’s where web analytics services truly shine. Once the KPIs, conversion goals, and funnels are defined, the computer and the user do the rest. Still, no matter how specific we make our KPIs, no matter how precisely we define the funnels that lead to them, in the end we are still trying to get a picture of human learning through online behaviour, and we will always have cause to wonder at times whether we’re really measuring what we think we’re measuring, whether what we think we’re measuring is even what we should be measuring. A little puzzling over those questions can be quite healthy.

Thirdly (perhaps most the crucial objective), we must **use** the resulting data to make improvements in the interaction. A question to be asked is this: “What would I do with this information if I had it?” If you do not know how you are going to use the data and what changes you will make as a result of different possible outcomes, then you should consider exploring other metrics where there is some actionable outcome. Especially in web analytics, where so much data can be gathered with so little effort, distance education professionals must pay particular attention to utilization.
Finally, analytics data thus interpreted and utilized may be shared to the benefit of users, other practitioners, and the distance education community as a whole. Web analytics in general, and especially its application to the field of technology-enhanced and distance education, is an emerging discipline. Valuable lessons that could help everyone include: how to incorporate analytics dashboards into the design workflow of distance education resources; how to translate learning objectives into effective KPIs; and how to create innovative metric mashups that combine metrics to illuminate deeper outcomes, deeper characteristics of the user, and their interactions. Insights like these could prove invaluable in bringing the power and agility of web analytics to education, and the depth and subtlety of education to web analytics.

These challenges are by no means reasons to shy away from the use of web analytics in the design and evaluation of distance education efforts. In fact, integrated web analytics data can help meld evaluation and design processes — and researchers and practitioners in distance education may in fact be uniquely positioned to take the use of analytics data in design process and strategic decision-making to a new level.

Conclusions
We conclude with our vision of what the future may hold, and some questions for future research.

**Future vision and research questions**
It is possible that the future of distance education could be dramatically influenced by information acquired through web analytics. The potential to collect and analyze real-time data from vast numbers of students could teach us a lot about how people interact with and learn in online learning environments. In reaching this potential, certain questions deserve more exploration:

> How do different segments of students (geography, age, gender, education level, major, etc.) interact with online resources?
What are the common KPIs applicable to industry?
What are the most effective ways to implement multivariate testing?
How can web analytics be used to give individual students information regarding the relation of their own use patterns and results of others with similar patterns?
What are the best ways to automate some decisions based upon data indices?
How can the data gained through web analytics be combined with other evaluation methods (e.g., qualitative methods) in order to give a more complete picture of learner intent?
How will data-driven decision-making through the use of web analytics change the processes by which distance education is designed and evaluated?

This chapter provided insight into the way in which web analytics might be used in the design and evaluation of distance education. Cadez, Heckerman, Meek, Smyth, and White (2003) propose that “arguably one of the great challenges … in the coming century will be the understanding of human behaviour in the context of ‘digital environments’ such as the web” (p. 399). While acknowledging that using web analytics in distance education is an emerging endeavour, several strengths and opportunities are apparent. Behaviour and results in online environments can be monitored and analyzed with more ease and agility than ever before. While maintaining an emphasis on high-quality ethical standards, web analytics provides a clear opportunity for monumental contributions in making data-driven decisions in the design and evaluation of distance education. The capability to continually track and monitor learner behaviour on such a large scale, and the insights gained from doing so, could transform the way we think about distance education.

Note: Interested readers can find a list of additional readings and resources about web analytics in education at http://tinyurl.com/wa-resources.
REFERENCES


