Education Tech Research Dev
DOI 10.1007/s11423-013-9317-2

RESEARCH ARTICLE

The role of evaluative metadata in an online teacher resource exchange

Samuel Abramovich · Christian D. Schunn · Richard J. Correnti

Abstract A large-scale online teacher resource exchange is studied to examine the ways in which metadata influence teachers' selection of resources. A hierarchical linear modeling approach was used to tease apart the simultaneous effects of resource features and author features. From a decision heuristics theoretical perspective, teachers appear to rely on complex heuristics that integrate many dimensions when determining whether to download a resource. Most surprisingly, numbers of ratings more strongly predict downloads than do mean rating levels, such that multiple negative ratings appear to attract more downloads than do few positive ratings. Implications for system design are discussed.

Keywords Online · Teaching resources · Metadata · Ratings · Hierarchical linear model · Decision heuristics

Introduction

The intellectual rigor, coherence, and appropriateness of materials used in classrooms can have a large influence on student learning (Ball and Cohen 1996). Prior to the advent of teaching materials being available on the Internet, educators were limited in their ability to acquire teaching resources. Teachers could slowly acquire lesson plans and other resources from fellow teachers, local material distributors, or from publishing companies who would act as *de facto* resource filters, controlling the type and quality of resources available to teachers (McCutcheon 1980).

S. Abramovich (✉)
Department of Learning and Instruction, Department of Library and Information Studies, University at Buffalo – SUNY, 515 Baldy Hall, Buffalo, NY 14260, USA
e-mail: samuelab@buffalo.edu

C. D. Schunn · R. J. Correnti
Learning Research & Development Center, University of Pittsburgh, Pittsburgh, PA, USA

Published online: 02 November 2013

Online teacher resource exchange represents a response to the desire of teachers to find high-quality, ready-to-implement teaching resources (Gray et al. 2010; Recker et al. 2004). These systems continue to grow in popularity, raising many questions about how they function and how they influence teaching and learning. As of 2013, over 8,500,000 sales have taken place on TeachersPayTeachers (TPT), a popular website where teachers can sell resources. Teach for America's (TFA) resource exchange, TFANet, has had 75 % of its 10,000 active corps members download resources (Tucker 2011). ShowMe, a site that allows teachers to share pre-recorded lessons, raised US$800k in venture capital (Schonfeld 2011).

The online sharing of resources may allow for more rapid spread of high quality resources. However, the way in which the sharing takes place may also have a large influence on whether benefits in teaching practice are found. Teachers might ignore critical reflection features of online resource exchanges such as ratings and comments; essentially turning exchanges into indiscriminately used online resource libraries. Even worse, there is the potential that these systems might harm the practice of teaching. For example, as these online exchanges typically have hundreds to thousands of resources of varying quality for a given topic or purpose, a teacher might only find lower quality resources given the limited amount of time any teacher has to prepare for instruction. The potential for online resource exchanges to have weak or negative effects on teaching practice is highlighted by research findings that continue to question the connection between instructional improvement and technology (Hennessy et al. 2005).

To increase our understanding of the potential benefits and challenges of online teacher resource exchanges, we examine one of the most salient features of exchanges, the rating and commenting system. In other contexts, teacher can accurately evaluate their peer's lesson plans and even provide correction to large errors (Ozogul et al. 2008). A number of websites for teacher resources have begun to employ peer or user ratings as an evaluative filter. For example, the National Education Association,[1] the Public Broadcasting System,[2] and the Department of Education of Ohio[3] all provide a variety of lesson plans online targeted for teachers and addressing a specific teaching need. Theoretically, the scope of each website along with the ratings and comments allow users to better select the resource that match teacher and student need.

However, online peer review is influenced by many situational variables (Morrison 2010). A teacher's beliefs about subject-specific pedagogies can hinder accurate evaluations of teaching resources (Remillard 2012). The prior knowledge and goals of the individual giving the rating will affect their rating process, even if clear criteria or rubrics are provided (Tillema 2009). Consequently, a teacher can have an incorrect, negative view of an ostensibly high-quality teaching resource based on a negative view of the associated pedagogy of the resource. Additionally, it is difficult to find consensus amongst educators on what constitutes a high quality teaching resource (Sumner et al. 2003). Even peer review that occurs only online can still be influenced by educational settings (Morrison 2010).

Consequently, the functional behavior of online teacher resource exchanges is challenging to predict and little research has examined teacher exchanges with real-world data (Manouselis et al. 2010). Thus it is imperative to discover how these systems work to

1 http://www.nea.org/tools/BrowseAllLessons.html.

2 http://www.pbs.org/teachers.

3 http://ims.ode.state.oh.us/.

unpack their potential effects on teaching. Understanding which factors influence teachers' selections of resources could guide system improvements.

Theoretical background

Teacher decision making and online resources

It is not surprising that the decisions a teacher makes during instruction can have large impacts on student learning (Colton and Sparks-Langer 1993; Munby 1982). However, even the decisions during planning prior to instruction can be indicative of and influential on teacher behaviors, which in turn influence what students learn (Doyle and Ponder 1977; Maloch et al. 2003; Stern and Shavelson 1983; Weiss et al. 1992). For example, the resources that a teacher selects and how they incorporate them into their instruction relates to their professional competency and teaching beliefs (Gueudet and Trouche 2012). However, the few studies that investigate how Internet technologies directly influence teacher planning either offer the basic conclusion that technology does affect teacher planning (Tubin and Edri 2004) or suggests ways to think about how technology could be used for planning (Harris and Hofer 2009).

Several studies have sought to determine the criteria and beliefs that teachers use during their resource searches. Recker et al. (2007) conducted two teacher professional development workshops for *Instructional Architect*, a web tool for creating online learning activities using learning resources from the National Science Digital Library.[4] They conclude that teachers claim a variety of goals when designing lessons that use online resources Fitzgerald et al. (2003) surveyed users of *The Gateway to Educational Materials*, a website that provides curated links to educational materials (7,000 at the time of the survey). 105 responses to these questions indicated dissatisfaction with how the site supported users to find resources, including missing resources and inaccurately described content. Perrault (2007) conducted an online survey of New York State biology teachers to determine their online information seeking practices. The repot found that teachers were not finding the resources they seek and relying on search engines, such as Google, rather than organized and vetted digital libraries. This pattern is argued to be the result of significant number of low-quality educational resources also found online Clements and Pawlowski (2012) surveyed users of Open Educational Resources on issues of re-use, quality, and trust. They found that instruments such as peer reviews and rankings could improve the quality of resources from the point of view of teachers.

However, these studies are left wanting in several regards. Either the number of participants is so low, 13 in the study by Recker et al. (2007), or that the participants only come from one content area, Biology in Perrault's study (2007), raising questions about the generality of reported findings. Clements and Pawlowski study (2012) only surveyed 146 users, who had self-selected for their interest in Open Educational Resources. The Fitzgerald et al. (2003) study surveyed 1,229 teachers, but suffered from the general challenges of generalizing from survey self-report to actual behaviors; teachers may be unwilling to mention or unaware of the simple heuristics they use to guide their selection or use of online resources (Nisbett and Wilson 1977; Speer 2005).

By comparison, few studies rely on actual data generated by true resource exchanges. A typical example is a study by El-Hani and Greca (2013) that included usage statistics in

[4] NSDL.org.

their analysis of an online community of practice. The number of participants in their community was 87, far less than the number of participants seen in popular online teacher resource exchanges. Other studies are based on systems where the resources provided are either vetted or controlled by a limited group. Consequently, researchers have only begun to explore the research opportunities provided by analyzing large sets of data generated by community-driven efforts.

Finally, much of the research on online educational resources is often focused on the use of resources that are to be used online during class instruction (i.e., not the broader class of resources, such as worksheets, lesson plans, or readings, that can be acquired online but then be used in more traditional ways during instruction). For example, Recker et al. (2012) conducted a study on a teacher professional development program to improve instruction using online resources during instruction. However, while teachers may occasionally use an online resource during instruction, given intermittent Internet access or poor student-to-computer ratios in many classrooms, a more common occurrence will be the search for traditional resources prior to instruction. Teachers want resources that are easy to find and implement—specifically lesson plans, activities, and handouts (Hanson and Carlson 2005).

Teaching resources versus learning resources

In order to better understand how teachers select teaching resources in online resource exchanges, we draw a distinction between a learning resource and a teaching resource. A learning resources is a digital document that can be used for learning (Lau and Woods 2009; Ochoa and Duval 2008). By contrast, a teaching resource might include some of the same information as a learning resource but also includes additional information regarding its implementation (e.g., purposes, contexts, and procedures). For example, math teachers have been known to take learning resources that are designed to induce high-level mathematical thinking and proceduralize the mathematics during instruction, lowering the effectiveness of the resource (Stein et al. 1996). A properly designed teaching resource can provide information on effective pedagogies, subject matter content for the teacher, or suggest ways that a teacher can relate learning to other parts of a curriculum (Davis and Krajcik 2005).

In addition, learning resource exchanges often have different goals than that of teacher resource exchanges. For example, learning resource systems such as MERLOT (Li 2010), ARIADNE (Klerkx et al. 2010), Connexions (Dholakia et al. 2006), and OER Commons (About OER Commons 2013) are designed to be forums where many different types of educators can exchange a wide variety of online learning resources. This object type diversity necessitates the use of standards that are complex. For example, the IEEE Learning Object Metadata standard requires resources be identified according to 4 aggregation levels, three interactive types, and a long list of resource types (Ochoa et al. 2011; Vargo et al. 2003) along with issues of copyright and ownership of resources. This complexity can lead to large difficulty in finding appropriate resources (Diekema and Olsen 2011), and teachers may prefer to use other methods, like evaluative metadata, to guide their search for materials.

Evaluative metadata

A traditional online teaching resource search relies on simple metadata about the content of the material, such as content type, grade level, difficulty level, duration, etc. Teachers will likely still rely heavily on content information to shape what they download in the online setting (e.g., find worksheets on basic quadratic formula applications for Algebra I). But to help teachers further filter through the overwhelming abundance of resources now available online on any given topic, many websites now add another filtering methods that was not available in the past: *evaluative* metadata. Evaluative metadata provides information about the quality rather than content of the materials, and typically takes the form of user ratings and comments. Commonly found in all kinds of online resource distribution, evaluative metadata enables users' ability to evaluate and filter the vast resources available (Chiao-Fang et al. 2009).

However, these types of rating systems do not always produce outcomes in line with the designer's intent. Prior lab studies of knowledge management systems suggest that the number of user-entered ratings might not affect content searches (Poston and Speier 2005). A survey conducted on Amazon reviews reported that moderate reviews were more helpful than extremely positive or negative comments (Mudambi and Schuff 2010). Consequently, rating systems for teaching resources should not be instituted based solely on expected behaviors without considering the facets of a teacher's decision to select a teaching resource.

Decision heuristics

To theoretically frame the role ratings may play in teachers' selection of resources in online teaching resource exchanges, we began with a more general body of research on human decision-making processes. Research on decision heuristics proposes the existence of a stopping rule (Raab and Gigerenzer 2005). In the case of a teacher looking for a resource, the stopping rule would be the heuristic that triggers the educator to stop looking at other resources, if just momentarily, and read or download the selected resource.

A stopping rule might be simple or complex, perhaps depending on the teacher and their goals when searching for a resource. Todd and Gigerenzer (2000) detail several simple heuristics for stopping rules that have accounted for human decision making in situations ranging from car purchases to college selection to voting. One model of teachers exchanging resources prior to the Internet is an example of a simple heuristic: a co-educator or administration's recommendation alone might be the means for deciding on a resource and operated as the simple decision heuristic.

With the abundance of detailed content and evaluative metadata available in online teacher resource exchanges, there is potential for teachers to utilize more complex decision heuristics for choosing a resource. For example, teachers might follow *Dawes' Rule,* deciding to download a resource based on comparing the number of positives (e.g. high ratings, positive comments) to the number of negatives (e.g. low ratings, negative comments). Or they may follow *Franklin's Rule*, weighting the number of pros and cons differentially, such as weighting comments more heavily than ratings. The *Multiple Linear Regression* heuristic would suggest that the decision to download a resource is a weighting account of pros and cons but also taking into account the strength/extent of the positives and negatives. For example, a resource with one positive rating would be evaluated for download differently than a similar resource that had an equal number of positive and negative ratings.

The multiple linear regression approach might provide more rational decisions (depending on the weights obtained) over the other models, but decision-making researchers have often found that balancing many factors can be cognitively demanding and that even simple heuristics can enable relatively good decisions (Czerlinski et al.

1999). Thus, the availability of many types of metadata does not rule out the potential for teachers to utilize a simple decision heuristic for downloading a resource. Whether teachers use simple decision heuristics that focus narrowly on particular meta-data or use complex decision heuristics that focus broadly on many data sources can have a dramatic effect on the functionalities of a system: how simple the collected information should be, which information should be collected, what are opportunities for one or two less accurate data sources to bias teacher resource selections, etc.

A follow-on research question focuses on the content of these decision heuristics: which data sources tend to be most predictive of choice and is the form of that influence rational? For example, consider findings of a study on the website TPT (Abramovich and Schunn 2012). TPT is an online resource exchange used by millions of teachers that allows educators to buy and sell teaching resources to other users. With more than 530,000 available resources by September 2013, TPT's designers cannot rely on a group of experts to evaluate resources. Instead, TPT provides content search tools and user ratings/comments.

One of the key findings from the study of TPT was that the number rather than content of ratings and comments for teaching resources were highly positively correlated with the main resource use data in TPT, sales. That is, the mere presence of ratings and comments at all, rather than the positive versus negative nature of these comments, appeared to influence user purchases. Further, from multiple regression analyses, the effects appeared to suggest a weighted combination of numbers of ratings and comments drove decision-making, rather than a simple heuristic. However, the data in that study was from aggregate behavior summed to a fixed moment in time, and as a result, it was not possible to determine conclusively whether more ratings and comments led to more sales or whether more sales simply created more opportunities for ratings and comments. In addition, relatively few resources had multiple ratings or multiple comments, and the ratings and comments tended to be highly positive. Thus, the positive/negative nature of the ratings might not have further influenced decision making because there was insufficient discriminating information in them to be useful. Thus, to better understand what type of decision heuristics are used to download resources in online teacher resource exchanges, additional data was required from websites that have more ratings and comments per resource, and ratings with more variation. Further, to better disentangle causal claims, multi-time point data is required in order to determine if existing rating and comments can predict future user activity.

Data source

For our study, we examined data from TFANet TFA website that supports the exchange of teacher resources. TFA is a non-profit organization that seeks to take high-performing, recent college graduates and place them for 2 years in urban and rural schools in the US with underserved students. For the 2011–2012 school year, TFA enlisted 9,000 corps members who taught over 600,000 students.

TFANet was created as a way to support corps members by providing an online network where a variety of resources and services can be accessed. Corps members in need of a teaching resource can go to the resource exchange on the TFANet website and search for a type of resource based on keywords and metadata about each resource. A TFANet user can perform a keyword search for a resource and pre-select by grade, subject, file type, resource, type, author, appropriateness for the school year, or state specificity (Fig. 1).

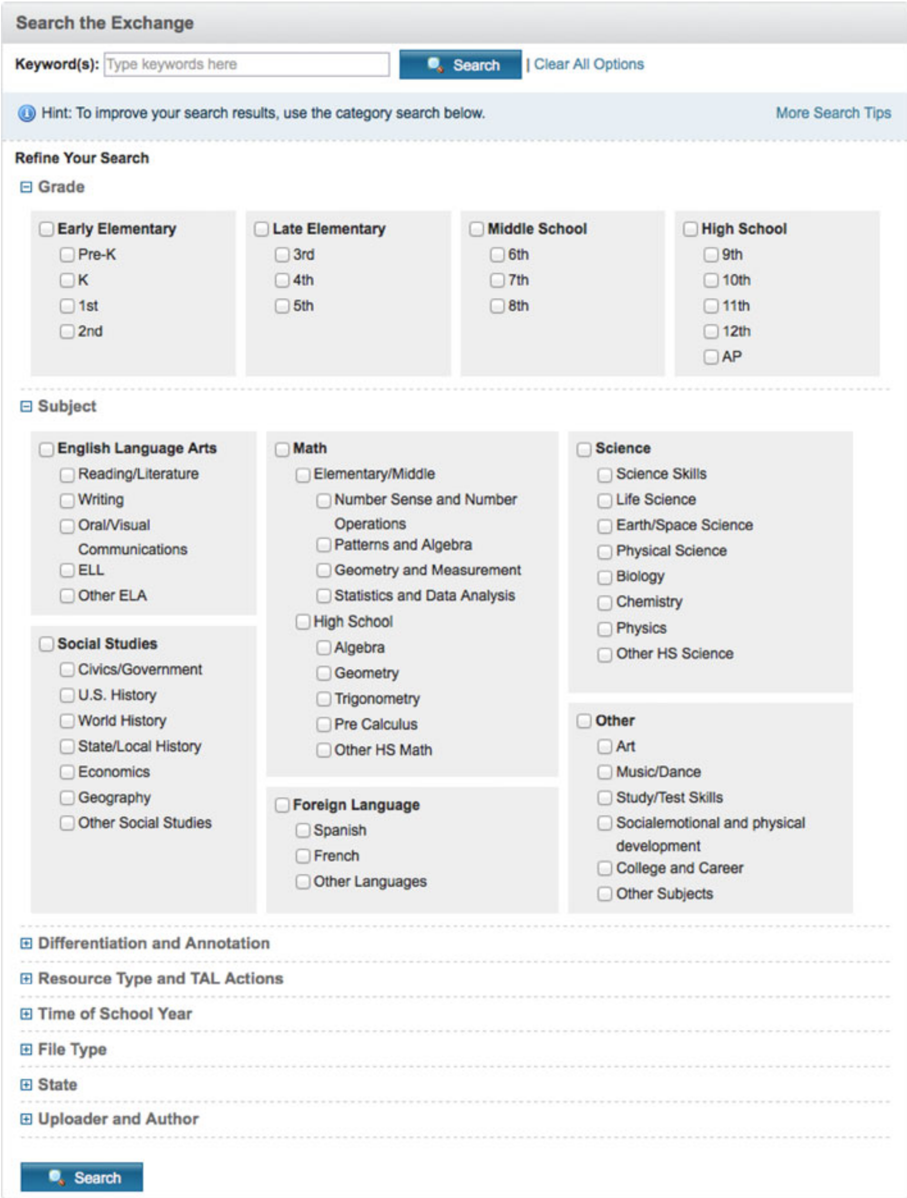

Fig. 1 TFANet search screen

According to TFA, during the 2010 fall semester, 75 % of corps members downloaded resources from TFANet.

Each resource in TFANet also has a web page that supplies additional information about the resource (Fig. 2). This information includes a detailed description provided by the author, an average rating of one to five stars from other TFANet users, the number of ratings, comments from other TFANet users, and a Blue Ribbon icon if TFANet

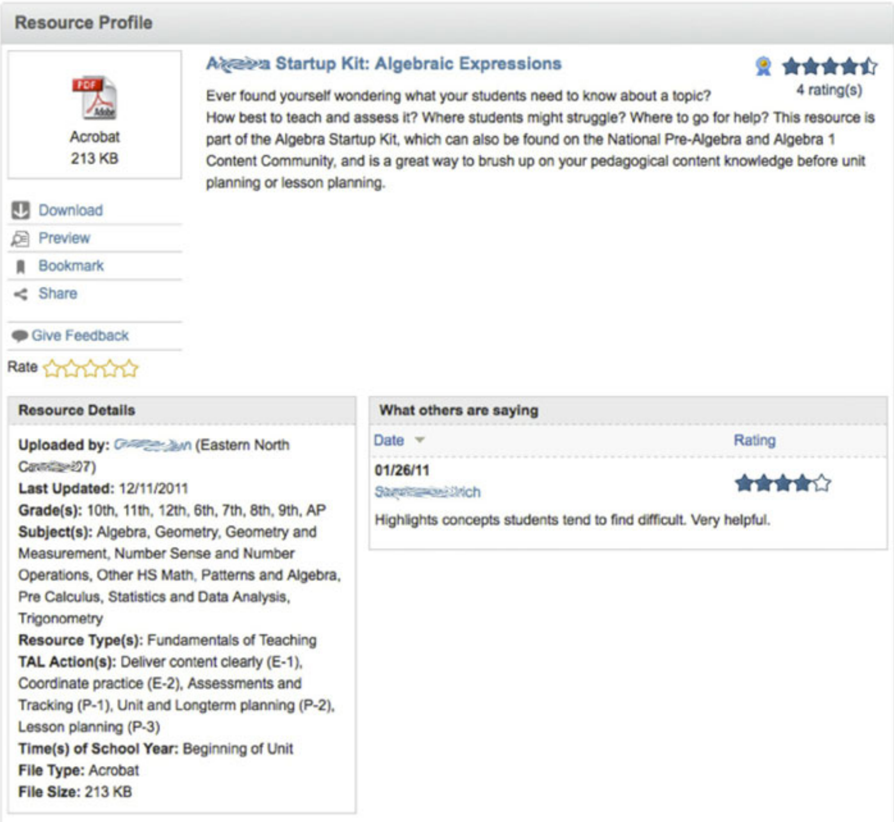

Fig. 2 Example TFANet resource page that provides information about the resources, its applicability, and peer ratings and comments

administrators identified the resource as high-quality. For the rest of our analysis, we refer to this information as evaluative metadata.

TFANet also exhibits the ability for current and former corps members to evaluate resources based on evaluative metadata, including user supplied ratings as well as expert generated ratings (i.e. the blue ribbons). These evaluative metadata are intended to increase system functionality by operating as a quality filter for the large and increasing amount of resources in TFANet.

To serve as the data for the current paper, the administration at TFA graciously provided us a copy of the main TFANet database from multiple time points. Our descriptions of TFANet's user interface are a result of our direct interaction with the system. All of the information provided in this paper has been confirmed for accuracy with TFA administration.

Methods

Our broader research questions are about which evaluative metadata, in the form visible to users, influence a user's choice to download particular resources. We also have a narrower research question about the role of rating and comment frequency and content. Specifically, we examined the following more specific research questions:

RQ1 Which evaluative metadata are most predictive of downloads?

RQ2 Are download decisions best described by a simple or more complex decision heuristics?

RQ3 Given more diversity in rating levels and comment frequency, will resources that are highly rated or rated many times predict the eventual number of times the resource will be downloaded?

Sample and variables

The sample includes only TFA teachers so it is important to describe the context of TFAnet users. The rigor of the TFA admissions process results in a teacher population that is highly motivated. TFA members should also be fairly technically savvy since they are all recent college graduates. Finally, TFA is a national organization, which means that its members represent a wide variety of organizational and policy contexts.

Our analyses examined data from TFANet during a 1-month period between February 10 and March 10 of 2011. By February, most teachers have established a routine and are deep into instruction. Additionally, February occurs before most traditional school testing periods in the US. During this 1-month timeframe, there were 26,959 unique visitors to the resource exchange with 178,626 searches and 79,348 downloads.

Hierarchical linear model

An individual author could influence whether a resource is downloaded by either creating a notably good or bad resource or by attaching unique metadata. To account for this influence on teachers download heuristics, we analyzed the data using a hierarchical linear model (Raudenbush and Bryk 2002). This form of HLM analysis is complementary to educational data mining. Other data mining approaches can offer additional insights, but we believe that HLM offers a particularly appropriate analysis model for this data given its fundamentally nested structure (resources nested inside authors); approaches that do not take into account this structure can artificially decrease standard errors causing an increase in the likelihood of Type I errors. We are also unaware of any other research that utilizes HLM for analyzing online teacher resource exchanges.

Change in downloads as dependent variable

The number of downloads is count data and therefore its distribution is skewed, closely resembling a Poisson distribution. Furthermore, since we examined downloads over a whole month and since each resource we examined was already in the system at the beginning of our time period, we modeled our outcome using a constant-exposure Poisson model, rather than the more typical regression approach built upon normal distributions.

To create a measure of resource downloads, we calculated the increase in total number of downloads for each resource between February 10th and March 10th. Despite some of the resource metadata changing slightly over the course of the month, such as the number of ratings or average ratings, a 1-month time frame effectively balances two opposing challenges: A much shorter time frame would have produced too few downloads to study, and a much longer time frame would have increased the occurrence and severity of changing metadata over the studied download period. The resultant data included 16,863 resources written by 2,149 different authors (mean of 7.85 resources per author, mode of 1, max of 967). 1,300 authors uploaded 2 or more resources.

About 34 % of all resources had no ratings, as expected based on prior work examining resource exchanges (Abramovich and Schunn 2012; Ochoa and Duval 2008). We retained the unrated resources in our analyses because they are naturally occurring and they might change the prominence of other available metadata.

Evaluative metadata used as independent variables

We identified a variety of author and resource factors visible to users to serve as predictors by examining the user interface (see Fig. 2) and considering plausibility of influence (described below). Table 1 provides descriptive statistics on these variables. Although prior research often mentions the importance of copyright information, that information was not found in this site.

The resource-level variables include some surface level characteristics such as the age of the resource in number of years, which was calculated as the difference from the date the first resource appeared in the online exchange (July 11, 2009). Age may influence a sense of timeliness or relevance of a resource.

We also included the length of character descriptions, which was a count of the number of characters used in the description. We theorized that a longer resource description could provide more confidence in the actual contents of a resource.

Table 1 Descriptive statistics of evaluative metadata within TFANet

Variable	Mean	Std. dev.	Min	Max
Person level independent variables (n = 2,149)				
Author number of years since entry into TFA (0 = 1990)	17.17	2.49	0	20
Author Currently in TFA (1 = yes)	0.31	0.46	0	1
Resource level (n = 16,863)				
Dependent variable				
Downloads during 1 month	3.24	3.47	0	51
Independent variables				
Character count of resource descriptions/100	2.79	2.47	0.05	24.77
Age of resource in years (0 = July 11, 2009)	0.91	0.74	0	2.49
Number of ratings (excluding missing ratings)	2.38	2.33	1	38
File format is editable (1 = yes)	0.89	0.31	0	1
Blue ribbon indicator (1 = yes)	0.06	0.24	0	1
Number of comments	0.43	1.09	0	23
Ratings				
Missing average rating (1 = yes)	0.34	0.48	0	1
Average rating between 1 and 2.99 (1 = yes)	0.08	0.27	0	1
Average rating between 3 and 3.99 (1 = yes)	0.20	0.40	0	1
Average rating between 4 and 4.99 (1 = yes)	0.28	0.45	0	1
Perfect 5.0 rating (1 = yes)	0.10	0.30	0	1

A third feature was whether the file format was easily editable (e.g. Microsoft Word, Microsoft PowerPoint) or not (e.g. PDF, JPEG). This variable was a potentially important predictor given the frequent teacher necessity to edit a resource for specific student needs (Brown and Edelson 2003).

We also examined features that were more face-valid indicators of resource quality. For example, experts from TFA had rated some of the resources as being high-quality and had given those resources a "blue ribbon" designation. A second measure of primary interest was the mean rating the resource received from teachers on a scale from 1 (low) to 5 (high). To better enable analysis of interactions of quality and quantity of ratings, we created a set of five binomial variables, one for resources missing ratings and the other four indicating whether the average rating fell within successively higher ranges up to a perfect average rating of "5". By creating a variable for missing ratings we could then directly examine the effects of particular mean rating levels against having no ratings at all (e.g., is a negative rating better or worse than no rating?).

Finally, at the resource level, we also created variables for the number of ratings provided for each resource as well as the number of comments supplied for each resource, given that number of ratings and comments per se were predictive of downloads in past research.

In addition to accounting for effects of particular authors, we also included two variables at the author-level that could also drive download decisions. TFA alumni can continue to participate in TFANet, including uploading and downloading resources. Alumni may be viewed as having greater expertise. Further, this distinction between current corps members and alumni also guides how TFA communicates and supports their current and past members. Note we removed administrator-generated resources because we wished to focus on the exchange of materials among teachers.

The second author-level variable was the number of years since the author had begun working with TFA as a corps member, which may relate to identity effects and preferences for similar cohorts. This was measured starting in 1990 since the oldest corps member contributing a resource began in that year.

Statistical analyses

To answer questions about which factors predicted downloads, we ran three models using HLM 7.0 (Raudenbush et al. 2011). We first examined a null model containing no variables in order to estimate the average number of downloads and to establish the baseline amount of variance existing between authors. The general form of the constant-exposure Poisson model is written as follows:

$$\eta_{ti} = \pi_{0i} \tag{1}$$

$$\pi_{0i} = \beta_{00} + r_{0i} \tag{2}$$

In Eq. 1, η_{ti} represents the expected count of the number of downloads transformed by the log link function ($\eta_{ti} = \log(\lambda_{ti})$, where λ_{ti} is the expected count of the number of downloads); and π_{0i} represents the average log odds of the estimated number of downloads for author *i*. At level 2 (Eq. 2), π_{0i} is a function of the average log odds of the number of downloads for all authors (β_{00}) plus the author specific deviation from the overall average (r_{0i}). Since η_{ti} is the log of the event rate, we can reproduce the overall event rate across all authors by exponentiating the estimate of the coefficient from our null model [$\exp(\beta_{00})$]. Furthermore, we obtain a measure of the variance between authors (r_{0i}) which is assumed to be normally distributed with mean of 0 and standard deviation of unity.

Our second model (described in Eqs. 3, 4 below) simply added predictor variables to the previously described null model. All variables were entered uncentered to ease the interpretation of the intercept. The intercept in these models is the log odds of the number of downloads when all predictor variables equal ‘0’. At the resource level we tested the effect of each predictor variables for random variance at the author level and, where significant variance existed, we retained the random effect. Our second model is as follows:

$$\begin{aligned}\eta_{ti} = {} & \pi_{0i} + \pi_{1i}(Number\,of\,Characters\,in\,Description)_{ti} + \pi_{2i}(Date\,of\,Upload)ti \\ & + \pi_{3i}(File\,Format\,is\,Editable)_{ti} + \pi_{4i}(Blue\,Ribbon\,Indicator)_{ti} \\ & + \pi_{5i}(Number\,of\,Comments)_{ti} + \pi_{6i}(Number\,of\,Ratings)_{ti} \\ & + \pi_{7i}(Average\,Rating\,between\,1\,and\,2.99)_{ti} + \pi_{8i}(Average\,Rating\,between\,3\,and\,3.99)_{ti} \\ & + \pi_{9i}(Average\,Rating\,between\,4\,and\,4.99)_{t} + \pi_{10i}(Perfect\,5.0\,Rating)_{ti}\end{aligned} \tag{3}$$

$$\pi_{0i} = \beta_{00} + \beta_{01}(Current\,TFA\,Member)_i + \beta_{02}(Authors\,Year\,in\,TFA)_j + r_{0i} \tag{4}$$

$$\pi_{1\mathrm{i}\cdots\mathrm{ni}} = \beta_{10\cdots\mathrm{n}0} + \mathrm{r}_{\mathrm{qi;\ where\ n=10\ and\ q=1,3,4,5}}$$

Our third model is an extension of the second model but examines in greater depth the effects of rating number and mean rating level using 16 categories representing combinations of the number of ratings and the average rating (see Table 2). Logically, there should be an interaction between these variables in that more ratings should lead to a larger effect of mean rating level (i.e., quality ratings should be more persuasive when based on more ratings). As shown in Table 2, each subgroup contains at least 100 resources, and each subgroup also differs in mean levels on other predictors, requiring the use of multiple-regression to tease apart their independent effects.

Model comparisons

In order to compare the fit of our data to models 2 and 3 we used the expectation–maximization (EM) algorithm based on a Laplace transformation which produces estimates in HLM that approximate maximum likelihood with reasonable accuracy (Raudenbush et al. 2011). We found that the deviance statistics produced from our third model suggested a better model fit when compared to the deviance from our second model ($\chi^2 = 173.72$, df $= 11$, $p < 0.001$). The Chi square statistic confirms that the reduction in deviance far outweighs the additional degrees of freedom as a result of having more parameters in the model. We detected slight over-dispersion in the empirical Bayes residuals. Therefore, we adjusted for over-dispersion within HLM7.0, but it should be noted that the substantive findings remain the same with either estimation procedure.

Results

Null model: a significant effect of author

We discuss findings from our statistical analyses in order beginning with the null model (see Table 3). Examining the random effects portion of the model, there was significant variation in the estimated number of downloads between authors ($\chi^2 = 4{,}846$, df $= 2{,}148$, $p < 0.001$), highlighting the importance of using a nested model that accounts for author

Table 2 Distribution of resources to subgroups and associated evaluative metadata

	N	Avg. num. comments	% Blue ribbon	% Easy edit	Char. count/100
Missing rating (reference category)	5,797	0.00	04	90	2.27
One rating between 1 and 2.99	645	0.09	01	91	2.15
Two ratings between 1 and 2.99	402	0.18	00	94	2.33
Three ratings between 1 and 2.99	163	0.26	01	99	2.52
Four ratings or more between 1 and 2.99	113	0.68	00	94	2.51
One rating between 3 and 3.99	1,641	0.06	01	91	2.71
Two ratings with an average rating between 3 and 3.99	872	0.20	01	91	2.81
Three ratings with an average rating between 3 and 3.99	425	0.37	04	92	3.09
Four ratings or more with an average rating between 3 and 3.99	393	1.40	07	90	3.37
One rating between 4 and 4.99	1,920	0.08	06	90	2.86
Two ratings with an average rating between 4 and 4.99	1,085	0.24	05	89	3.01
Three ratings with an average rating between 4 and 4.99	634	0.44	09	86	3.39
Four ratings or more with an average rating between 4 and 4.99	1,145	1.96	28	85	4.31
One rating that is a perfect 5.0	1,056	0.15	10	86	3.11
Two ratings with a perfect 5.0 average rating	299	0.48	12	84	3.61
Three ratings with a perfect 5.0 average rating	125	0.76	22	80	3.73
Four ratings or more with a perfect 5.0 average rating	148	1.64	32	80	4.46

effects. As we add variables for models 2 and 3, we will examine the proportion of between author variance that is explained by adding both resource and author-level variables to the model.

Model 2: many independent predictors

Examining the results of model 2 (see Table 3), only three variables were not found to be statistically significant: author's corps year in TFA, having an average rating between 1 and 2.99 (compared with no rating), and the number of comments. However, the number of comments and number of ratings were highly correlated, so in model 3 we examine these variables with an alternate specification.

Many variables were significant in predicting the number of downloads. Given the strength and quantity of these many significant predictors, this model suggests a complex heuristic is involved in teachers' decision making, including a number of relatively superficial factors. For example, the number of characters in the resource description was found to be significant with more characters predicting a higher number of downloads. Two other surface features were significant. First, whether the file was easily editable was a significant variable as was the binomial variable indicating whether the author was a current corps member. The prevalence of these significant surface features suggests that teachers consider many different resource features in order to make their download decisions beyond quality *per se*.

We also found a number of variables more obviously connected to quality to be significant in the model. To illustrate the magnitude of one of our findings, we considered the predicted event rate if all other variables in the model were held to '0': Resources carrying the blue ribbon designation were more frequently downloaded $\sim$0.3 times (model 2, changing the intercept from 2.41 to 2.71) or 0.5 times (model 3).[5] It is interesting that such a strong external quality clue had a relatively small effect on downloads.

Given that the binomial indicator for resources missing a rating was omitted from model 2 (i.e., made the null case), the coefficients for the various mean rating levels should be interpreted relative to the no rating case. In general, most rating levels produced more downloads than having no ratings. Only the lowest mean ratings category was not different in the predicted number of downloads relative to no ratings at all; and contrary to what one might have expected, resources with low mean ratings were not less likely to be downloaded than resources with no ratings. All the higher mean rating categories incrementally predicted significant increases in number of downloads (all $p < 0.001$ except for the difference between 3–3.99 and 4–4.99, which was significant at $p < 0.05$).

A higher number of ratings was also found to be a significant variable predicting a greater number of downloads, replicating previous studies. Number of ratings *per se* is an interesting factor because it is not *prima facie* an indicator of quality since ratings could be either high or low. From the user's perspective, more ratings could reflect greater interest by users downloading that resource (i.e., willingness to rate) or simply social presence (others thought it worth downloading).

[5] These estimates were calculated using the following procedure: we factor in all of a resource's attributes using multiple regression coefficients by first adding and subtracting the log-odds of various predictors before exponentiating to determine the predicted number of downloads.

Table 3 Effects of variables on number of resource downloads

	Null model coeff. (se)	Model 2 coeff. (se)	Model 3 coeff (se)
Author level fixed effects			
Corp year in TFA		0.005 (0.006)	0.005 (0.006)
Current corp member or alumni		0.108 (0.041)***	0.117 (0.035)***
Resource level fixed effects			
Intercept	1.166 (0.015)***	0.879 (0.064)***	0.871 (0.065)***
Character count		0.018 (0.004)***	0.016 (0.004)***
Upload date		0.038 (0.030)***	0.047 (0.017)*
File format is editable		0.093 (0.035)**	0.093 (0.034)*
Blue ribbon indicator		0.117 (0.037)***	0.189 (0.035)***
Number of comments		0.005 (0.014)	0.064 (0.011)***
Number of ratings		0.111 (0.007)***	
Average rating of 1–2.99		0.009 (0.035)	
One rating; range of 1–2.99			0.074 (0.041) ~
Two ratings; range of 1–2.99			0.210 (0.047)***
Three ratings; range of 1–2.99			0.376 (0.066)***
Four ratings or more; range of 1–2.99			0.656 (0.069)***
Average rating range of 3–3.99		0.076 (0.027)**	
One rating; range of 3–3.99			0.094 (0.028)**
Two ratings; range of 3–3.99			0.258 (0.034)***
Three ratings; range of 3–3.99			0.519 (0.041)***
Four ratings or more; range of 3–3.99			0.792 (0.039)***
Average rating range of 4–4.99		0.113 (0.029)***	
One rating; range of 4–4.99			0.131 (0.026)***
Two ratings; range of 4–4.99			0.392 (0.029)***
Three ratings; range of 4–4.99			0.553 (0.035)***
Four ratings or more; range of 4–4.99			0.749 (0.032)***
Perfect 5.0 rating		0.243 (0.034)***	
One rating; perfect 5.0			0.321 (0.029)***
Two ratings; perfect 5.0			0.479 (0.046)***
Three ratings; perfect 5.0			0.641 (0.064)***
Four ratings; perfect 5.0			0.834 (0.056)***
Random effects-between author variance			
Intercept (τ_{00})	0.164	0.101***	0.098***
Character count (τ_{10})		0.127***	0.131***
File format is editable (τ_{30})		0.099***	0.062***
Blue ribbon indicator (τ_{40})		0.013***	0.010
Number of comments (τ_{50})		0.002***	0.002***
% Of variance explained[a]		38	40

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

[a] The percent variance explained was calculated using the following formula $(\tau_{00\mathrm{null}} - \tau_{00\mathrm{model}})/\tau_{00\mathrm{null}}$

Model 3

Findings from model 3 help to further describe the complex influence of both number of ratings and the mean rating from users. Recall that the main difference between model 3 and model 2 was the decision to break down number of ratings from a semi-continuous variable to examine subgroups. Examining the data this way improved model fit. We note several findings when comparing and contrasting model 3 with model 2.

First, similar effects were found for most variables, suggesting stability across the models. Second, one difference between the two models was in the effects of the number of comments. In model 3, resources with more comments were more likely to be downloaded.

Third, findings from the subgroups analyses also demonstrated that a high number of ratings, regardless if the average rating was poor or not, consistently predicted a greater number of downloads (see Fig. 3a representing the interaction of mean rating level and number of ratings). This finding seems to parallel the old adage that there is no such thing as bad publicity. Furthermore, there is also an underlying effect for average ratings since resources with higher ratings are also consistently more likely to be downloaded. What is most striking about this figure though is that the lines for the different mean ratings levels run largely parallel with one another. This is our initial indicator that while a higher average rating predicts a higher number of downloads, a high number of ratings also independently predicts a higher number of downloads. While not seemingly logical, such independent use of these factors are likely a simplification heuristic used to ease this complex decision task.

Using combinations of regression coefficients to predict different event rates

Figure 3a represents the predicted number of downloads for each of the subgroups factoring in only the regression coefficients for the number of ratings and mean ratings (assuming all other variables are "0" in the regression equation). Given how surprising the influence of number of ratings was, we also examined this figure by adjusting for additional variables because each subgroup varied in their overall mean for other significant contributors to whether a resource was downloaded (e.g., blue ribbon status). As there were significant correlations between many of the variables, the multiple regression methodology may have artificially separated underlying effects from the subgroups plotted in Fig. 3a. Figure 3b plots the estimated effects when factoring in the difference in means between subgroups for the other significant variables. Although the lines still largely run parallel to one another, Fig. 3b demonstrates some variance in the slopes of the lines representing an interaction between number of ratings and average rating (i.e., the difference between the lowest ratings and highest ratings is greater when the number of ratings is 4 or more than it is when only 1 rating is supplied).

The individual predictors

Overall, we find that many but not all investigated factors predict download decisions in this online setting: number of ratings, number of comments, mean rating, number of characters in a description, blue ribbon indication, file-format ease of editing, and current status of resource author were all found to be statistically significant independent predictors of downloads.

What was surprising was that our findings confirmed what was only hinted at in the study of TPT—that a high number of low ratings predicts more downloads than a resource

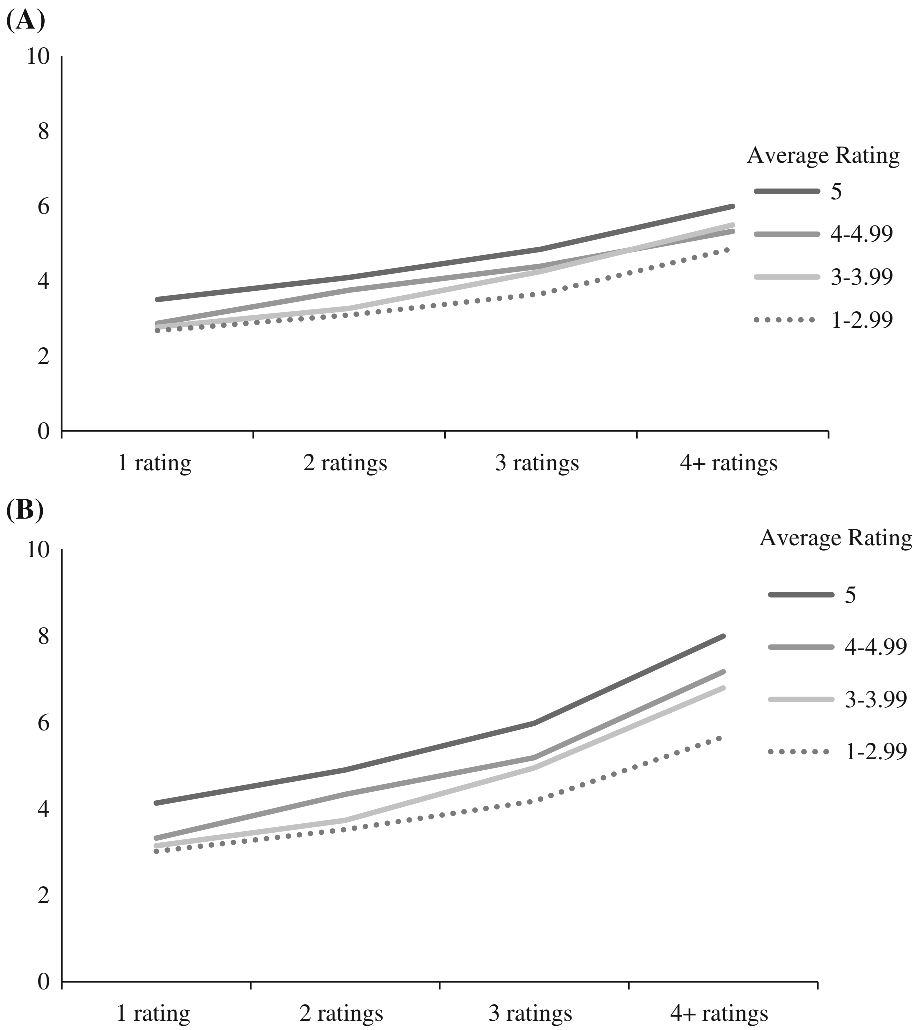

Fig. 3 Estimated number of downloads for subgroups **a** not accounting for subgroup correlations with other significant variables and **b** accounting for subgroup differences on other significant variables

with a low number of high ratings. For example, as illustrated in Fig. 3b, resources with the lowest average rating but with 4 or more ratings will be downloaded more than resources with a single perfect rating. This result seemingly runs counterintuitive to the idea that more low ratings would dissuade others from looking at a resource. This finding seems to indicate that rating contents and comments have a limited influence on a resource's popularity.

We were also surprised to find that the current versus alumni status of a resource's author would predict downloads. This preference for resources from current corps members could be an indication of a current corps member being able to better understand a current corps members resource needs. This preference is probably not an indication of a continual change in teacher resource needs because a resource's upload date was not a predictor.

Our other findings of predictors of downloads were not as surprising. Because blue-ribbon status is an indication of a quality resource it seems logical that its presence would make a resource more attractive for download. Similarly logical is the attractiveness of file resources that are more easily editable since they are easier for educators to alter to their needs. We also believe that the number of characters in a resource description is acting as a proxy measure of the level of a resource's details.

Because number of comments was highly correlated with number of ratings, we were not surprised to see that the number of comments did not predict downloads in model 2 but that it did predict downloads in model 3. The date a resource was uploaded as well as the year that a resource author was in the corps were also not predictors of downloads. A lack of an effect of these variables could be a result of the way this data was presented within TPT. Both require more than a quick glance to process as presented in a resource page and could have been ignored by most users.

Conclusions

Prior research on teacher exchanges has focused on issues of training (Recker et al. 2012), usability (Fitzgerald et al. 2003), standardization of resource descriptions (Vargo et al. 2003), and access rights (Caswell et al. 2008). Here we add to this work by examining the factors that robustly drive download decisions. Based on our findings, there is little support to suggest that a shared simple heuristic based on one or two pieces of metadata drive resource downloads. As an aggregate, teachers were influenced by many factors, following a weighted multiple regression decision making pattern. As we did not examine individual patterns of downloads, we cannot conclude that the majority of users each use complex decision heuristics. However, the data are meaningful to understand the collective behavior of teachers.

We hypothesize that the metadata that we identified as predicting download variance does so because it is used for individuals stopping rule heuristics. This exploratory work suggests there is much to be learned through further research efforts including further quantitative work as well as studying individual teachers via interviews and surveys. For example, additional research could unpack whether teachers could be assuming that their colleagues' ratings are either invalid or not applicable to individual teacher needs. Additionally, the placement of the ratings on the web page for a resource could lessen its impact on teachers' resource download heuristics.

Impacts

It is unlikely that teachers will revert from online back to only using traditional physical network/paper catalog methods for finding teaching resources, and instead this movement towards online methods will likely increase. Even veteran teachers who have amassed a collection of quality teaching resources will need to look for new resources to fit national or local policy changes such as the recent adoption of the Common Core Standards in the US (Gewertz 2012).

We believe the current findings could have several impacts. Designers of teacher resource exchanges, armed with the knowledge of the extent to which ratings predict downloads, can emphasize or de-emphasize resource characteristics in order to achieve desired behaviors. For instance, if certain collections of resources that the designers see as particularly high quality are not being downloaded then the system designers can themselves provide or otherwise solicit reviews for those resources that would then promote more interest and downloads.

In TFANet, many resources have only one rating which still limits the value of ratings as a filter for finding the right resource. A different approach from encouraging more ratings from teachers would be to emphasize ratings from teachers who are better evaluators of resources. For example, if teacher A provides more accurate ratings than teacher B then the online resource exchange can give teacher As ratings more emphasis or weight when reporting reviewers' ratings in aggregate. A resource with only rating might be accurately evaluated if the rating came from an accurate rater.

Teacher professional development based on a specific technology is not unusual to teacher education (Lawless and Pellegrino 2007). However, our findings suggest that teachers would benefit from professional development targeted at how these system integrate into teacher practice (e.g., paying more attention to the interaction of number of reviews and mean rating). At stake is one of a teacher's most precious resources, time. Selection of non-optimal resources will not only result in wasted time but could also still leave the teacher having to generate the resource from scratch.

Online teacher resource exchanges also fit the model of resource-based learning environment for teachers (Hill and Hannafin 2001). The main purpose of an exchange, the creation and review of teaching resources, fulfills some of the elements of good teacher professional development. The resources in an exchange are content specific, actual curricular materials, provide opportunity for reflection through ratings, and are ongoing with new resources consistently being uploaded. With some additional design changes to increase resource ratings, the process of uploading a resource, receiving feedback on the resource from colleagues, and providing ratings on other colleague's resources could provide opportunity for professional development in an activity that teacher would already naturally use as part of their practice.

References

About OER Commons (2013). Retrieved October 9, 2013 from http://www.oercommons.org/about.

Abramovich, S., & Schunn, C. (2012). Studying teacher selection of resources in an ultra-large scale interactive system: Does metadata guide the way? *Computers and Education, 58*(1), 551–559.

Ball, D., & Cohen, D. (1996). Reform by the book: What is: Or might be: The role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 25*(9), 6–14.

Brown, M., & Edelson, D. C. (2003). *Teaching as design: Can we better understand the ways in which teachers use materials so we can better design materials to support their changes in practice LeTUS report series*. Evanston, IL: The Center for Learning Technologies in Urban Schools.

Caswell, T., Henson, S., Jensen, M., & Wiley, D. (2008). Open content and open educational resources: Enabling universal education. *The International Review of Research in Open and Distance Learning, 9*(1).

Chiao-Fang, H., Khabiri, E., & Caverlee, J. (2009). *Ranking comments on the social web*. Paper presented at the International Conference on Computational Science and Engineering.

Clements, K. I., & Pawlowski, J. M. (2012). User-oriented quality for OER: understanding teachers' views on re-use, quality, and trust. *Journal of Computer Assisted Learning, 28*(1), 4–14.

Colton, A. B., & Sparks-Langer, G. M. (1993). A conceptual framework to guide the development of teacher reflection and decision making. *Journal of Teacher Education, 44*(1), 45–54.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & A. R. Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York, NY: Oxford University Press.

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3–14.

Dholakia, U. M., King, W. J., & Baraniuk, R. (2006). *What makes an open education program sustainable? The case of connexions*. Retrieved October 9, 2013 from https://oerknowledgecloud.org/sites/oerknowledgecloud.org/files/36781781.pdf.

Diekema, A. R., & Olsen, M. W. (2011). Personal information management practices of teachers. *Proceedings of the American Society for Information Science and Technology, 48*(1), 1–10.

Doyle, W., & Ponder, G. A. (1977). The practicality ethic in teacher decision-making. *Interchange, 8*(3), 1–12.

El-Hani, C. N., & Greca, I. M. (2013). ComPratica: A virtual community of practice for promoting biology teachers' Professional Development in Brazil. *Research in Science Education, 43*(4), 1327–1359.

Fitzgerald, M. A., Lovin, V., & Branch, R. M. (2003). The gateway to educational materials: An evaluation of an online resource for teachers and an exploration of user behavior. *Journal of Technology and Teacher Education, 11*(1), 21–51.

Gewertz, C. (2012). Educators in Search of Common-Core Resources. Retrieved May 21, 2012 from http://www.edweek.org/ew/articles/2012/02/24/22resources_ep.h31.html.

Gray, L., Thomas, N., & Lewis, L. (2010). *Teachers Use of Educational Technology in U.S. Public Schools (2009): National Center for Education Statistics, Institute of Education Sciences*. Washington, DC: Department of Education.

Gueudet, G., & Trouche, L. (2012). Teachers' work with resources: Documentational geneses and professional geneses. In G. Gueudet, B. Pepin, & L. Trouche (Eds.), *From text to 'lived' resources* (pp. 23–41). Dordrecht: Springer.

Hanson, K., & Carlson, B. (2005). *Effective access: Teachers' use of digital resources in STEM teaching*. Newton, MA: Education Development Center Inc.

Harris, J., & Hofer, M. (2009). *Instructional planning activity types as vehicles for curriculum-based TPACK development*. Paper presented at the Society for Information Technology and Teacher Education International Conference 2009, Charleston, SC, USA.

Hennessy, S., Ruthven, K., & Brindley, S. (2005). Teacher perspectives on integrating ICT into subject teaching: commitment, constraints, caution, and change. *Journal of Curriculum Studies, 37*(2), 155–192.

Hill, J. R., & Hannafin, M. J. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development, 49*(3), 37–52.

Klerkx, J., Vandeputte, B., Parra, G., Santos, J., Assche, F., & Duval, E. (2010). How to share and reuse learning resources: The ARIADNE experience. In M. Wolpers, P. Kirschner, M. Scheffel, S. Lindstaedt, & V. Dimitrova (Eds.), *Sustaining TEL: From innovation to learning and practice* (Vol. 6383, pp. 183–196). Berlin: Springer.

Lau, S.-H., & Woods, P. C. (2009). Understanding learner acceptance of learning objects: The roles of learning object characteristics and individual differences. *British Journal of Educational Technology, 40*(6), 1059–1075.

Lawless, K. A., & Pellegrino, J. W. (2007). Professional development in integrating technology into teaching and learning: Knowns, unknowns, and ways to pursue better questions and answers. *Review of Educational Research, 77*(4), 575–614.

Li, J. Z. (2010). *Quality, evaluation and recommendation for learning object.* Paper presented at the International Conference on Educational and Information Technology, Chongqing, China.

Maloch, B., Flint, A. S., Eldridge, D., Harmon, J., Loven, R., Fine, J. C., et al. (2003). Understandings, beliefs, and reported decision making of first-year teachers from different reading teacher preparation programs. *The Elementary School Journal, 103*(5), 431–457.

Manouselis, N., Vuorikari, R., & Van Assche, F. (2010). Collaborative recommendation of e-learning resources: an experimental investigation. *Journal of Computer Assisted Learning, 26*(4), 227–242.

McCutcheon, G. (1980). How do elementary school teachers plan? The nature of planning and influences on it. *The Elementary School Journal, 81*(1), 4–23.

Morrison, J. (2010). *A research framework to examine the impacts of online peer review on learning outcomes in computer science courses*. Paper presented at the ITS 2010 workshop: Computer-supported peer review in education: Synergies with Intelligent Tutoring Systems Workshop, Pittsburgh, PA.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34*(1), 185–200.

Munby, H. (1982). The place of teachers' beliefs in research on teacher thinking and decision making, and an alternative methodology. *Instructional Science, 11*(3), 201–225.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. doi:10.1037/0033-295X.84.3.231.

Ochoa, X., & Duval, E. (2008). Relevance ranking metrics for learning objects. *IEEE Transactions on Learning Technologies, 1*(1), 34–48.

Ochoa, X., Klerkx, J., Vandeputte, B., & Duval, E. (2011). On the use of learning object metadata: The GLOBE experience. In C. Kloos, D. Gillet, R. Crespo García, F. Wild, & M. Wolpers (Eds.), *Towards ubiquitous learning* (Vol. 6964, pp. 271–284). Berlin: Springer.

Ozogul, G., Olina, Z., & Sullivan, H. (2008). Teacher, self and peer evaluation of lesson plans written by preservice teachers. *Educational Technology Research and Development, 56*(2), 181–201.

Perrault, A. M. (2007). An exploratory study of biology teachers' online information seeking practices. *School Library Media Research, 10*, 1–26.

Poston, R. S., & Speier, C. (2005). Effective use of knowledge management systems: A process model of content ratings and credibility indicators. *MIS Quarterly, 29*, 221–244.

Raab, M., & Gigerenzer, G. (2005). Intelligence as smart heuristics. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence* (pp. 188–207). Cambridge: Cambridge University Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models*. Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, R. T., & du Toit, M. (2011). *HLM 7*. Lincolnwood, IL: Scientific Software International Inc.

Recker, M., Dorward, J., & Nelson, L. (2004). Discovery and use of online learning resources: Case study findings. *Educational Technology and Society, 7*(2), 93–104.

Recker, M., Sellers, L., & Ye, L. (2012). *Investigating impacts of technology-related teacher professional development designs: A comparative case study*. Paper 5: The instructional architect research group. Retrieved October 7, 2013 from http://digitalcommons.usu.edu/iagroup/5/.

Recker, M., Walker, A., Giersch, S., Mao, X., Halioris, S., Palmer, B., et al. (2007). A study of teachers' use of online learning resources to design classroom activities. *New Review of Hypermedia and Multimedia, 13*(2), 117–134.

Remillard, J. T. (2012). Modes of engagement: Understanding teachers' transactions with mathematics curriculum resources. In G. Gueudet, B. Pepin, & L. Trouche (Eds.), *From text to 'lived' resources* (pp. 105–122). Dordrecht: Springer.

Schonfeld, E. (2011). Tablet Education Startup ShowMe Raises $800,000 From Lerer Ventures, SV Angel, And Betaworks. Retrieved September 22, 2012 from http://techcrunch.com/2011/08/16/tablet-education-showme-800000/.

Speer, N. M. (2005). Issues of methods and theory in the study of mathematics teachers' professed and attributed beliefs. *Educational Studies in Mathematics, 58*(3), 361–391.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455–488.

Stern, P., & Shavelson, R. J. (1983). Reading teachers' judgments, plans, and decision making. *The Reading Teacher, 37*(3), 280–286.

Sumner, T., Khoo, M., Recker, M., & Marlino, M. (2003). *Understanding educator perceptions of quality in digital libraries*. Paper presented at the Digital Libraries, 2003. Proceedings. 2003 Joint Conference on.

Tillema, H. H. (2009). Assessment for learning to teach appraisal of practice teaching lessons by mentors, supervisors, and student teachers. *Journal of Teacher Education, 60*(2), 155–167.

Todd, P. M., & Gigerenzer, G. (2000). Précis of simple heuristics that make us smart. *Behavioral and Brain Sciences, 23*(5), 727–741.

Tubin, D., & Edri, S. (2004). Teachers planning and implementing ICT-based practices. *Planning and Changing, 35*(3), 181–191.

Tucker, B. (2011). Teachers swap recipes. *Education Next, 11*(3), 10–17.

Vargo, J., Nesbit, J. C., Belfer, K., & Archambault, A. (2003). Learning object evaluation: Computer-mediated collaboration and inter-rater reliability. *International Journal of Computers and Applications, 25*(3), 198–205.

Weiss, C. H., Cambone, J., & Wyeth, A. (1992). Trouble in paradise: Teacher conflicts in shared decision making. *Educational Administration Quarterly, 28*(3), 350–367.

Samuel Abramovich is an Assistant Professor of Learning and Instruction, and of Library and Information Studies at the University at Buffalo – SUNY.

Christian Schunn is a Professor of Psychology, Intelligent Systems, and Learning Sciences and Policy at the University of Pittsburgh.

Richard Correnti is an Associate Professor of Learning Sciences and Policy at the University of Pittsburgh.