WHERE DID YOU GET THAT LESSON? UNDERSTANDING ONLINE TEACHER RESOURCE EXCHANGES

by

Samuel Abramovich

Bachelor of Arts, Brandeis University, 1998

Master of Teaching, University of Virginia, 2004

Submitted to the Graduate Faculty of

the School of Education in partial fulfillment

of the requirements for the degree of

Educational Doctorate in Learning Sciences and Policy

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Samuel Abramovich

It was defended on

June 27, 2013

and approved by

Richard Correnti, PhD, Associate Professor

Lindsay Clare Matsumura, PhD, Professor

Richard Halverson, PhD, University of Wisconsin – Madison, Professor

Dissertation Advisor: Christian Schunn, PhD, Professor

WHERE DID YOU GET THAT LESSON? UNDERSTANDING ONLINE TEACHER RESOURCE EXCHANGES

Samuel Abramovich, PhD

University of Pittsburgh, 2013

Teachers are increasingly using websites as their primary resource for finding instructional materials. These sites often encourage teachers to both upload and download their own materials while leaving ratings and comments about other resources. Yet despite the large number of participants and amount of data generated by these websites, we know relatively little about how they influence teacher selection of resources. The results of two investigations of large-scale popular sites, TeachersPayTeachers and TFANet, are shared. The examination of the data generated by these sites provided distinct patterns of teacher participation and use. Among the findings is the surprising result that a large number of ratings, positive or negative, is a predictor of a resource's popularity. These findings also lead to suggestions for how teacher educators and instructional designers can improve use and operation of online teacher resource exchanges.

Keywords: Teaching Resources, Metadata, Ratings, Hierarchical Linear Model, Human-Computer Interface

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Over the past 5 years, I have received more support, guidance, and mentorship from my advisor, Chris Schunn than I have any right to receive. Although it is an impossible goal, I hope to emulate his scholarship and generosity for the rest of my career.

I would like to thank my dissertation committee – Richard Correnti, Lindsay Clare Matsumura, and Rich Halverson. Their feedback and support were not only extremely helpful in terms of my scholarship but also in navigating the challenges of the writing process.

I would also like to thank the rest of the LSAP faculty who have been instrumental in my development as a scholar.

Thank you to my fellow LSAP students who help create and balance a community of academic support and rigor.

I would like to thank my in-laws, who supported me and my family in too many ways to count. That is a debt that I can never repay.

I would like to thank my parents for instilling in me the value of education and learning. Their encouragement gave me the fortitude for this endeavor.

I would like to thank my children who are always there to remind me of what is important in life and how balance is necessary.

Most of all, I would like to thank my wife. There are no words that can explain how much she means to me.

1.0 INTRODUCTION

A growing method used by teachers for finding and sharing teacher resources are online exchanges. One teacher, Deana Jump, recently accumulated over $1 million in her total earnings on TeachersPayTeachers, a popular website where teachers can sell resources (Rotherham, 2012). Teach for America's resource exchange, TFANet, has had seventy five percent of its active corps download resources (Tucker, 2011). ShowMe, a site that allows teachers to share pre-recorded lessons, raised $800k in venture capital (Schonfeld, 2011). Educational policy, such as the Common Core State Standards Initiative[1], is pushing teachers to turn to the Internet to find compatible teaching resources (Catherine Gewertz, 2012).

Prior research has revealed the importance of teacher planning on student learning (Peterson, Marx, & Clark, 1978). Specifically, the resources that teachers use will effect their instruction and how their students learn. Online teacher resource exchanges exist to provide teachers a way to share resources with each other, such as for preparation and classroom instruction. However, relative to extensive knowledge about the relationships between teachers and other types of technology (e.g., online teacher forums, multimedia), little is known about how these online exchanges function in practice. This lack of knowledge is surprising, and potentially worrisome, given the established impact of teacher planning and teaching resources on student learning. If online resource exchanges are influencing teaching practices then it is

[1] http://www.corestandards.org

reasonable to also suspect that this technology is impacting student-learning outcomes. Additionally, because of the widespread use of online teacher resource exchanges, this technology is potentially influencing teachers and students on an ultra-large scale.

Consequently, education stakeholders who are concerned with how teachers use technology to prepare for instruction must have a better understanding of online teacher resource exchanges – a technology that is both widely adopted amongst teachers and effects their daily practice. The goal of my dissertation studies was to begin to unpack the potential impact of online teacher resource exchanges in order to better understand how they can inform teacher learning and instruction.

1.1 THE POTENTIAL OF ONLINE TEACHER RESOURCE EXCHANGES

The incorporation of technological innovation is a hallmark of education in the United States. During the industrial revolution, mass production allowed blackboards, chalk, and textbooks to be produced at large enough quantities as to become integral classroom technologies (Mehlinger, 1996). The post World War 2 boom in radio and television allowed educators to explore how to use those technologies to improve instruction (Reiser, 2001). The microcomputer revolution of the 1970s has led to computer technology in schools being commonplace (Niemiec & Walberg, 1992). More recently, the development of Internet based technologies are cited as having the potential to bring positive transformative changes to our education system (A. Collins & Halverson, 2009).

Online teacher resource exchanges are a prime example of how Internet based technologies could have the potential to bring positive transformative changes to our education system. These systems utilize rating systems where teachers assess peers' uploaded resources through leaving comments and numerical evaluations (e.g., 3 out of 5 stars). Other teachers can then use ratings and comments as a recommendation model for what resources they should select when preparing for instruction. Furthermore, the feedback from peers could help a teacher professionally develop their ability to create better teaching resources. Even a brief examination of an online teacher resource exchange indicates how the technology potentially contains elements of peer assessment, professional development, and technology based recommendation models.

The combination of resources and ratings produces a system with the potential for unique, new teaching practices. For example, researcher have established that high quality teacher professional development should be content specific (Borko & Putnam, 1995), use curricular materials (DL Ball & Cohen, 1999), providing opportunity for reflection (Hawley & Valli, 1999), and being ongoing (Guskey & Yoon, 2009). An online teacher resource exchange could provide opportunity for high-quality professional development by focusing teacher reflection, in the form of ratings and comments, on content specific curriculum materials, i.e., the resources available in an exchange. Additionally, this professional development would be ongoing since online resource exchanges allow for new opportunities for reflection through the addition of new teacher-submitted resources.

But while online teacher resource exchanges are potential sources for unique, new teaching practices, there is also the potential that these systems could function as a recreation of established technology supports. Teachers might ignore features of online resource exchanges such as ratings and comments; essentially turning exchanges into online resource libraries with little opportunity for resource recommendation or professional development. Even worse, there is the potential that these systems might harm the practice of teaching. For example, as these online exchanges typically have thousands of resources, a teacher might never find a quality resource given the limited amount of time any teacher has to prepare for instruction.

Further, research findings continue to question the connection between instructional improvement and technology (Hennessy, Ruthven, & Brindley, 2005). The result is that some education researchers believe that the use of technology to improve education is ineffective. Most famously, Larry Cuban (1993) has declared that technological innovations are not important facets of school improvement. In order to avoid becoming quixotic, a researcher investigating online teacher resource exchanges must first understand why there is a dearth of findings that confirm how technological innovations improve education (Kent & McNergney, 1999).

However, before engaging in research, I considered why there is relatively little prior research on online teacher resource exchanges. What makes investigation of online resource exchanges challenging? What discipline should guide the initial research questions for this type of emerging educational technology?

1.2 RESEARCH CHALLENGES OF ONLINE TEACHER RESOURCE EXCHANGES

One challenge is, despite wide belief in some technologies pedagogical potential, research claims that technology is often ineffectual in traditional learning settings. For example, video games are widely believed to be highly effective learning technologies (Jenkins & Squire, 2004). However, studies of their effectiveness in K-12 classrooms often either produce inconclusive results or are seen as impractical (de Freitas & Oliver, 2006). While there is value in hypothesizing about the value of online exchanges, research must also establish how these systems contribute, if at all, to teacher practice. Research on online teacher resource exchanges should focus toward determining the current impact of the technology rather than only focus on online resource exchange's educative potential.

A second challenge to research on online teacher resource exchanges is that the innovation and distribution of educational technology often occurs much faster than the pace of education research. For example, over the past ten years, cellular phones have become a ubiquitous part of our society. Consequently, researchers have explored their potential contribution to learning (Thornton & Houser, 2005). However, the technology of cellular phones has more recently morphed into smart phones and tablet computing, potentially outdating the past research on cellular phones (Banister, 2010). Not all research on technology becomes outdated once the technology changes or evolves, but there are limitations to how research on older technology can be used to support claims of educational effectiveness of emerging technology. Online teacher resource exchanges also have a high rate of innovation and distribution due to their existence on remote servers. Teachers only need an Internet connection and a web browser to access these systems; improvements or changes can be instantaneously distributed without any further technology investment from teachers. Teachers Pay Teachers, one of the exchanges I studied for this dissertation, has seen rapid growth along with major interface changes since its inception. Consequently, online teacher resource exchange researchers must use analyses that can provide findings that keep pace with online exchanges' evolvement.

A third challenge to research on online teacher resource exchanges is reflecting actual teacher practice. For example, researchers have suggested that wikis have strong educational potential and could be easily taken up by educators (Drexler, Baralt, & Dawson, 2008). However, the vast majority of educational wikis have such short lifespans that their educational impact is questionable (Reich, Murnane, & Willett, 2012). In the example of wikis, what education researchers assume would be a popular practice is not. The popularity of some online teacher resource exchanges is evident through number of resources downloaded and uploaded as well as the number of participants. Research on online teacher resource exchanges, if based on systems that have large active populations, will allow for understanding on how teachers use these systems in actual day-to-day practice.

1.3 EDUCATION INFORMATICS

While some education technology systems have become associated with specific theoretical frameworks (e.g., flow with educational games (Chen, 2007), I found no consensus choice of theory for use in research on online teacher resource exchanges. Without an established theoretical paradigm, I instead rely on a range of theory and prior work concerning teachers and educational technology; presented in literature review. However, to guide my research goals, I use the guiding principles of the emerging discipline of education informatics.

Education informatics is an interdisciplinary approach with several definitions. Collins and Weiner (2010) define it as the application of technology to discovering and communicating education information. Srivastava (2012) broadly defines education informatics as the incorporation of information and communication technology (ICT) in education. Levy, et al. (2003) define education informatics as the intersection of information science, education, and computer science. The common theme in these definitions is the dual goals of learning how to use information and learning how to use information to improve learning.

I define the goal of *learning how to use information* as developing understanding that can inform the pedagogy of data use. For example, the ability to be able to find desired information on the Internet is identified as a fundamental 21st century skill (Jenkins, Clinton, Purushotma, Weigel, & Robison, 2007). But compared to other basic skills (e.g., addition, spelling) there is almost a complete absence of understanding of how individuals and communities decide on what information they need and how they seek it. Education informatics researchers strive to understand how information-seeking behavior occurs in order to improve peoples' ability to do so. Applied to research on online teacher resource exchanges, education informatics suggests that researchers should first begin to understand how teachers are using the information in online resource exchanges.

I define the goal of *learning how to use information to improve learning* as synonymous with using data analytics to improve education practice. For example, educational data mining, a type of data analytics, was used to discover patterns in cognitive tutors that indicated student misuse, or 'gaming', of cognitive tutor hint sub-systems (Baker, Corbett, Koedinger, & Wagner, 2004). With the result of this research, cognitive tutor designers and programmers could refine their systems to avoid student misuse of the hint subsystem and improve student learning outcomes. Applied to research on online teacher resource exchanges, education informatics also suggests discovering what information in online exchanges can tell us about the resources that teachers are seeking.

With the research goals suggested by education informatics, my aim is to suggest design changes to online resource exchanges that would improve teacher practice. Knowing what information teachers use in online exchanges would allow for better guidance for teachers searching for resources. Knowing what information can predict the popularity of resources would potentially allow designers to increase the number of high-quality resources.

2.0 LITERATURE REVIEW

2.1 TEACHER RESOURCE EXCHANGES ARE IMPORTANT TO IMPROVING TEACHER PLANNING

2.1.1 Current State of Research on Teacher Planning

For some time, research has revealed the importance of teacher planning on instruction and student achievement (Peterson, et al., 1978). During planning, the instructional goals set and the sources of information used can predict how a teacher attempts to solve instructional problems (Yinger, 1980). Further, the routinization of teacher preparation can provide flexibility and effectiveness during class instruction (Yinger, 1979). Consequently, an established path for improving teacher education is through research on the decisions teachers make when preparing for instruction (Clark, 1988).

Given that Internet based technologies continue to transform so many other aspects of learning and teaching, it is surprising that there is not an abundance of research on how Internet based technologies directly affect teacher planning. The few studies that investigate how Internet technologies influence teacher planning either offer the conclusion that technology does affect teacher planning (Tubin & Edri, 2004) or suggest ways to think about how technology could be used for planning (Harris & Hofer, 2009).

2.1.2 The Importance of Resources for Teacher Planning

In the research on teacher planning, the artifacts, namely lesson plans, provide insight into teacher knowledge. For example, an examination of pre-service teachers' lesson plans revealed their level of understanding of collaborative learning pedagogy (Ruys, Keer, & Aelterman, 2012). Another examination of pre-service teacher lesson plans revealed a lack of understanding of how assessment should be practiced (Campbell & Evans, 2000).

Consequently, if lesson plans reveal teachers' instructional knowledge, one can examine how teachers create their lesson plans to gain insight into how preparation can provide instructional knowledge. For example, the Japanese utilize jugyokenkyuu, lesson study, a collaborative examination of lesson plans, as effective teacher professional development (Lewis, Perry, & Murata, 2006). Further teachers' own creation of new teacher resources or analysis of existing teacher resources provides a learning opportunity for professional development (D. Ball & Cohen, 1996; Leat & Higgins, 2002; Sparks & Loucks-Horsley, 1989).

2.1.3 The Research Imperative: Investigating Current Methods of Teacher Resource Selection

Combining the lack of research on how Internet technologies affect teacher planning with the importance of how teachers select a resource when preparing for instruction creates a clear research imperative. We, as a community concerned with improving teaching and learning, must investigate how Internet technologies, namely online teacher resource exchanges, affect teachers' selection of teaching resources. By unpacking the processes behind the use of online teacher resource exchanges, we can gain current insight into current, large-scale teaching practices and identify ways to engender improvement.

2.2 PRIOR RESEARCH ON TEACHER SEARCH OF ONLINE RESOURCES

2.2.1 Studies of Prevalence of Online Searching Practice

Prior to the advent of teaching materials being available on the Internet, educators were more limited in their ability to acquire teaching resources. Teachers could slowly and infrequently get 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). Once the Internet greatly reduced scarcity and barriers to gaining access to resources, the established models of resource filtering broke (Metzger, Flanagin, & Medders, 2010).

Now many teachers go online to select and immediately obtain resources for instruction just at they also go online as private citizens to purchase other products. The Education Development Center conducted interviews with 26 rural, suburban, and urban STEM teachers. They found that teachers in their study "spend considerable time searching for resources for their classroom, with about half (12) spending between 10 to more than 20 hours a month using the Web for planning and instruction searches." (Hanson & Carlson, 2005). A survey of 68 graduates from a teacher preparation program reported that one of their key uses of a computer in their instruction was locating and gathering teaching materials (Franklin, 2005). A survey of 72 biology teachers revealed how these teachers are increasingly searching the Internet for teaching resources over relying on print materials (Perrault, 2007). Considering that these findings are five to seven years old, with societal increases in Internet access it is reasonable to assume that teachers have either maintained or likely increased the time they spend on the Internet looking for resources.

2.2.2 Studies of Teacher Approaches to Online Search

With teachers spending so much time online looking for resources, 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[2]. Based on just 13 completed surveys, they reported that teachers claim a variety of goals when designing lessons that use online resources, prefer easy to implement resources, and have positive attitudes toward finding teaching resources on the Internet.

Fitzgerald, Lovin, and Branch (2003) conducted a survey of 1,229 users of *The Gateway to Educational Materials*, a website that provides curated links to educational materials (7000 at the time of the survey). When asked, ‘‘what kind of resources or information were you looking for?” the answers included lesson plans (28%), activities (23%), units (10%), and courses (5%). However, when asked, “in what ways will you use this information?” the responses included ambiguous answers such as “classwork” (4%), “homework” (3%), “as is” (3%), “other” (2%), or simply didn’t respond to the question (14%). The last two open-ended questions asked for

[2] NSDL.org

suggested improvements or new features. 105 responses to these questions indicated dissatisfaction with how the site supported users to find resources, including missing resources and inaccurately described content.

Fitzgerald, Lovin, and Branch’s study also included two open-ended questions that asked for suggested improvements or new features. 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 72 New York State biology teachers to determine their online information seeking practices. Citing the large number of resources found online, respondents reported that they 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 (Fitzgerald, et al., 2003).

2.2.3 Key Limitations on Prior Research on Teacher Search: A Consideration of Methods

The previously mentioned studies are left wanting in several regards. Either the number of participants is so low, as in the study by Recker et al. (2007), or that the participants only come from one content area, Biology in Perrault's study (2007), as to question any reported findings. The Fitzgerald, Lovin, and Branch (2003), which managed to avoid the shortcomings of limiting the participants by having 1,229 survey takers, suffered from survey questions that can only suggest how teachers use resource exchanges and not offer any supporting evidence.

Additionally, much of the research on online educational resources is limited in that it 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 that can be acquired online but then be used in more traditional ways during instruction). Recker, Sellers, and Ye (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, 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 & Carlson, 2005). More complex instructional materials often require a robust technology infrastructure in the classroom and alternative pedagogy to that of more traditional materials.

Because, Internet communication technology has a rapid cycle of change, statistics on usage are challenging. Findings from as little at five years ago might have limited relevance to today because the technology or behavior being studied might already be irrelevant since some other technology or behavior has emerged as a replacement. For example, studies from 2000 about teachers who are Internet novices has limited relevance to today's teachers almost ubiquitous Internet access and content (Fitzgerald, et al., 2003). Not only must adjudication be exercised when looking at previous findings but there is also a premium on studies of currently used systems' ability to inform teacher education and practice.

Additionally, most related studies rely on surveys for understanding how teacher use online communities. For example, all of the conclusions from section 2 of this review are entirely based on surveys of teachers or teachers-in-training. However, survey research is limited in its ability to guide understanding of teacher practice (Munby, 1982). Teachers are sometimes unable to voice or explain core elements of their practice (Speer, 2005).

Far fewer studies rely on actual data generated by true resource exchanges. A typical example is a study by El-Hani and Greca (2012) that included usage statistics in 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 online teacher resource exchanges. Other studies, such as the work previous reported by Mimi Recker of Utah State, 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 communities driven efforts.

I suggest that other features or data generated by current resource exchanges can generate a more complete picture of, currently, what types of resources are being selected by teachers.

2.3 DEFINING AN ONLINE TEACHER RESOURCE EXCHANGE

2.3.1 Online Teacher Resource Exchange in Practice

The online teacher resource exchange represents a response to the desire of teachers to find high-quality, ready-to-implement teaching resources. Usually found on the World Wide Web, these systems facilitate the sharing of resources by providing a technical infrastructure for teachers to upload and download a resource.

An exchange is designed to allow teachers of numerous subjects, grade levels, and experience to share lesson plans, unit plans, assessments, or other teacher resources. To facilitate sharing, built into these systems are the opportunity for users to generate metadata. Uploaders categorize their resources according to subjects, grade level appropriateness, and type of resource such as assessment, worksheet, presentation, and review materials. Downloaders can search the available resources by metadata such as subject, grade level, type, and keyword. Searches can also be based on evaluative metadata such as what are the bestselling resources and which resources have the highest rating. Downloaders can further use content descriptions and evaluative metadata to then inform their purchases from the searched items.

Entering a review of a resource is usually accompanied by a rubric of varying complexity. The rating process can be as simple as a five point scale or as complex as a rating system that include categories like accuracy, practicality, and creativity. Users entering a rating also often have the option of writing a comment about the resource that will be displayed along with the average rating of the resource.

A typical use case scenario would begin with a teacher logging into the system and entering some keywords along with a specific grade level and subject area. The results of the search are resources listed according to the relevancy of the search. The teacher can then scroll through the list looking at a short summary, overall rating from other users, number of times the resource had been rated, subject areas, grade levels, type of resource such as assessment or activity, instruction time required, author, and a price. The list can be resorted according to rating, alphabetical according to resource name, price, sales, and date of creation. Once the teacher selects a resource that seems promising they can view the resource's page (see Fig. 1 as an example) where a more detailed description is available along with a preview, individual ratings, individual comments, a question and answer forum for each resource, and a short biography of the resource creator. The teacher can then choose to download the resource or go back to the search results to look for a more promising resource. The teacher can also choose to browse the collection of resources directly looking at a specific grade level, subject area, price category, or bestselling materials.

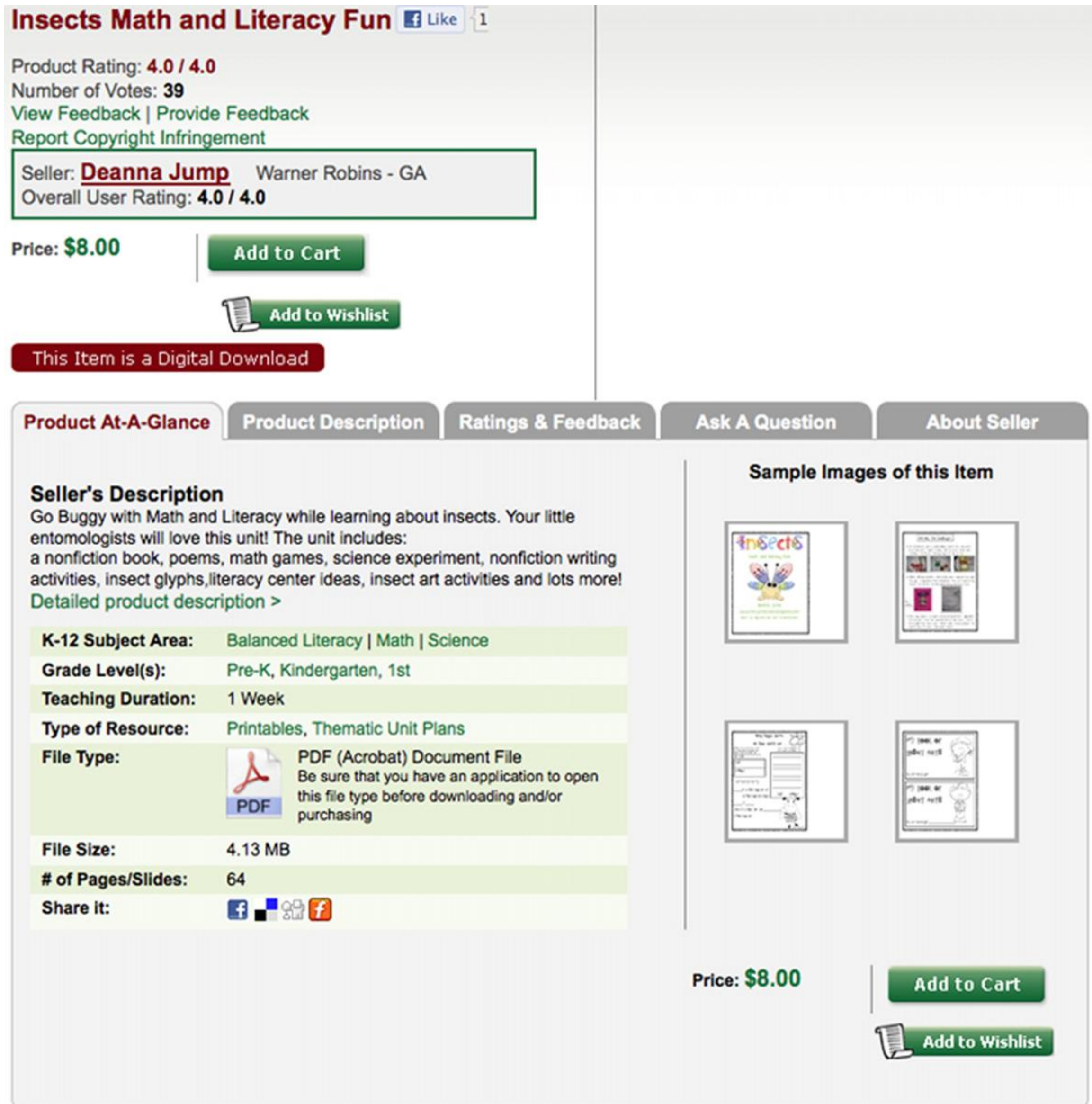


Figure 1: A Typical Online Teacher Resource Exchange Resource Page

2.3.2 The Difference between Online Teacher Resource Exchanges and Digital Learning Resource Exchanges

I wish to make the distinction that online teacher resource exchanges are fundamentally different than websites that focus on distributing or sharing digital learning resources. Digital learning resources can be defined as any digital document that can be used for learning (Ochoa & Duval, 2008). For example, a social studies teacher's standard practice includes using finding digital copies of historical documents to use during instruction (Salinas, Bellows, & Liaw, 2011). Digital learning resources can cover a wide range of types and formats (Barker, 2009).

However, digital learning resources often lack some of the necessary information that a teacher requires before implementing the resources. For example, mathematics teachers have been known to take learning resources that are designed to induce high-level mathematical thinking but proceduralize the mathematics during instruction, lowering the effectiveness of the resource (Stein, Grover, & Henningsen, 1996). A teaching resource might not include the same content as a digital resource or artifact but instead could provide information on effective pedagogies, provide subject matter content for the teacher, or suggest ways that a teacher can relate learning to other parts of a curriculum (Davis & Krajcik, 2005).

Another difference is that learning resource exchanges often have different goals than that of teacher resource exchanges. For example, learning resource systems such as MERLOT (Nesbit, Belfer, & Vargo, 2002) and ARIADNE (Duval, et al., 2001) are designed to be forums where many different types of educators can exchange a wide variety of 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 (Vargo, Nesbit, Belfer, & Archambault, 2003) along with issues of copyright and ownership of resources. This overabundance of metadata can lead to large difficulty in finding appropriate resources (Najjar, Klerkx, Vuorikari, & Duval, 2005).

I note that this distinction is debatable in that any resource used in teacher planning could be handled by teachers in similar fashions. However, the different technical needs of a teacher resource exchange compared to those of a learning resource exchange, as described in this section, merit the distinction in practical research.

2.4 RECOMMENDATION SYSTEMS

Increasingly, recommendation systems have been used to alleviate some of the difficulty in finding teacher resources. These systems utilize computer-based algorithms designed to provide a single or set of resources that match a user's need and preferences. Search engines like Google utilize recommendation systems to provide more accurate search results. Amazon’s product suggestions and Netflix’s movie suggestions are two additional examples of recommendation systems. Recommendation system research in e-learning is increasing through studies on scenario recommendation, adaptive learning, and collaborative filtering (Wu, Xu, & Ge, 2012).

Fully automated recommendation systems have had success in finding indicators of quality for educational resources (Bethard, Wetzer, Butcher, Martin, & Sumner, 2009; M. Recker, et al., 2011) but these indicators often have limited viability. For example, a search engine could recommend a resource based on the website where it originates but this recommendation is only as valid as the quality of the website. Should the website drop in quality, the fully automated recommendation system would be unaware and continue to recommend what would now be lower-quality resources.

To improve results and user satisfaction, developers and researchers have begun to explore collaborative information filtering - including ratings from users to better formulate recommendations (Nikos Manouselis, Drachsler, Vuorikari, Hummel, & Koper, 2011). Ostensibly, introducing a rating into the recommendation algorithm would allow the system to filter out poorly rated resources and rank recommendations according to their rating (M Recker & Wiley, 2001). Additionally, should the recommendation system not provide the wanted resource, a user can browse the available resources by highest rating and see what teacher resources are considered the best.

But the inclusion of ratings into recommendation engines also brings an additional challenge. Walker et al. (2004) found that many educational resources receive few to no ratings and that the ratings that are given often skew positive. They also found that the recommendation system studied provided resources that were of limited usefulness to teachers.

The benefit and challenge of using ratings to filter online educational resources has caused some schizophrenia within the online teacher resource research community. While some researchers have concluded that the use of ratings in recommendation systems has limited use (M Recker, Walker, & Lawless, 2003), others suggest that more investigation into how teachers rely on ratings is needed (Barker, 2009). Extended to online teacher resource exchanges, there is no clear consensus to suggest whether these systems benefit from the inclusion of ratings. It is possible that by including ratings, teachers have the means to filter out resources that would not be suitable. But there is also the possibility that ratings only serve as a distraction, suggesting the wrong resources and increasing the time it takes for teachers to find appropriate resources in online exchanges. I explore this dichotomy in the next section.

2.5 THE ELEPHANT IN THE ONLINE SEARCH ROOM: EVALUATIVE METADATA

2.5.1 Evaluative Metadata in Non-Educational Contexts

In the context of online search, to help users further filter through the abundance of resources now available online on any given topic, many websites add another filtering method 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. Evaluative metadata is commonly found in all kinds of online resource distribution, enabling users' ability to evaluate and filter the vast resources available (Schafer, Konstan, & Riedi, 1999). For example, the online retailer Amazon offers its customers access to evaluative metadata in order to make item selection easier. Amazon customers are encouraged to leave ratings and comments about each of the products available from the online retailer. In turn, Amazon makes the ratings and comments on its products available to its customers along with other metadata such as number of ratings, number of comments, and number of prior purchases. Ostensibly, access to this data adds benefit to users in that they now have simple cues to narrow their selection.

User ratings and comments are also becoming a common feature for websites where a resource or product exchange is between users and does not involve a retailer. By allowing users to rate and comment on videos, YouTube provides users a way to differentiate between large numbers of similar videos. User ratings and comments can be so valuable that some websites, such as Yelp or TripAdvisor, exist primarily as clearinghouses where users post ratings and comments for others.

However, rating systems do not always result in predictable results. Prior lab studies of knowledge management systems suggest that the number of user-entered ratings might not affect content searches (Poston & Speier, 2005). A survey conducted on Amazon reviews reported that moderate reviews were more helpful than extremely positive or negative comments (Mudambi & Schuff, 2010). Further, looking at some of the comments on websites like YouTube, with vulgar and irrelevant comments common, should cause to question the value of comments in helping to find the object of a search.

2.5.2 Traditional Metadata for Teacher Resources

A traditional search online for teaching resources relies on simple metadata about the content of the material such as content type, grade level, difficulty level, duration, etc. This kind of search is similar to what could be done previously through paper catalogs. Teachers will still rely on content information to shape what they download in the online setting (e.g., find worksheets on basic quadratic formula applications for Algebra I). However this metadata is limited in that it makes no distinction on the quality of the resource (M Recker & Wiley, 2001).

When uploading a resource, Teacher Resource Exchanges require the submission of descriptive metadata. This metadata is intended to provide a summary of the content and intended use of the resource, allowing searchers to bypass unwanted results. Examples of this metadata include subject area, resource type, intended grade level, and a short description of the resource. After downloading a resource, users are invited to provide a rating or comment on the resource for future users.

2.5.3 Evaluative Metadata for Teacher Resources

Because of the prior findings on rating systems in general, rating systems for teaching resources should not be instituted based solely on expected behaviors without considering the unique facets of the teacher decision to select a teaching resource. For example, teacher evaluation of a resource can be influenced by unintended consequences (Brown & Edelson, 2003). A teacher's beliefs about subject-specific pedagogies can unfairly influence their evaluation of teaching resources (Remillard, 2012). 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. If a teacher is unfamiliar with particular content knowledge or pedagogies then there exists a potential for important details of a resource to be ignored or improperly evaluated.

2.6 RATINGS AND PEER REVIEW IN TEACHING CONTEXTS

Is it appropriate to apply ratings as an evaluative measure of teaching resources in online systems? Teacher can accurately evaluate their peer's lesson plans and even provide correction to large errors (Ozogul, Olina, & Sullivan, 2008). There are a number of examples of websites for teacher resources that employ ratings as an evaluative filter. The National Education Association[3], the Public Broadcasting System[4], and the Department of Education of Ohio[5] all provide a variety of lesson plans online targeted for teachers and addressing a specific teaching need. Besides using search engines, these systems place more emphasis on evaluative metadata,

[3] http://www.nea.org/tools/BrowseAllLessons.html
[4] http://www.pbs.org/teachers
[5] http://ims.ode.state.oh.us/

allowing users to rate and comment on each lesson plan. 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). 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). Fundamentally, it is difficult to find consensus amongst educators on what constitutes a high quality teaching resource (Sumner, Khoo, Recker, & Marlino, 2003). Even peer review that occurs only online can still be influenced by educational settings (Morrison, 2010). Additionally, pre-service teachers have a preference for more expert feedback over peer feedback (Fu & Hawkes, 2010).

Consequently, the functional behavior of online teacher resource exchanges function is challenging to predict and very little research has examined teacher communities with real-world data (N. Manouselis, Vuorikari, & Van Assche, 2010). Thus it is imperative to discover how these systems work and discover their potential effects on teaching. Understanding what teachers produce ratings, what resources are likely to be rated, and how those ratings are formed could produce significant improvement in the efficiency of recommendation systems.

2.7 RESEARCH QUESTIONS

Carroll, et al. (2003) declare that future research on shared resource systems must address the need for better retrieval of sharable knowledge, improved understanding of shared knowledge, support over time, and management tools and practice. Throughout my literature review, I have raised the following research questions:

1. What metadata in online teacher resource exchanges predicts teacher interest in a resource?

The prior research in section 2.1.2 leaves researchers with a less-than-complete understanding of what types of resources are being selected by teachers. Examining metadata could provide additional insight into giving a complete picture of what teachers use online resource exchanges in their preparation. The answers to this question will allow researchers and implementers of resource exchanges to reasonably predict what resources will be most popular. Future research could then determine if manipulation of the metadata will result in different resource downloads.

2. Are highly rated resources in an online exchange also high quality?

Section 2.1.2 highlighted teachers expressed frustration in finding quality resources. However, section 2.1.3 and 2.2.3 explained how much prior research could not be used to answer if the resources being downloaded are also high quality. Because it is important for teachers to have high quality teaching resources, answering this question can contribute to understanding how a teacher resource exchange can promote downloading of high quality resources.

3. Is evaluative metadata used as the primary factor in selection of resources?

The commonality of evaluative metadata, detailed in section 2.5.1, would imply that it is used by teachers when selecting online teacher resources. Verifying this claim is important since the problems with evaluative metadata of teacher resources detailed in section 2.5.3 could lead to unintended consequences if teachers rely on the data. Relying on evaluative metadata might lead to selection of an inappropriate resource where reliance on other metadata might improve the selection process.

The following two papers, combined as initial research efforts on online teacher resource exchanges, answer these research questions, and create a foundation of finding for further research.

3.0 PAPER 1 - STUDYING TEACHER SELECTION OF RESOURCES IN AN ULTRA-LARGE SCALE INTERACTIVE SYSTEM: DOES METADATA GUIDE THE WAY?[6]

3.1 INTRODUCTION

The Internet and Internet-based technologies allow for the creation of what has been dubbed ultra-large-scale interactive systems; systems that leverage millions of users and huge databases to create unique and complex tools (Gabriel, Northrop, Schmidt, & Sullivan, 2006). Such systems, such as YouTube, Ebay, and Wikipedia, have begun to impact many facets of modern society and, because of the newness of these technologies, science has just begun to unravel how people use them.

Ultra-large-scale interactive systems have also begun to change how teachers prepare for instruction (Franklin, 2007; Jaen, Bohigas, & Novell, 2007; Russell, Bebell, O'Dwyer, & O'Connor, 2003). Currently, a teacher with access to the Internet now has at their disposal an almost unlimited number of resources provided by large interactive online systems. With this access, the previous problem of how to give teachers access to the resources (Greenhow, Robelia, & Hughes, 2009; Watson, 2006) has been supplanted with the new challenge of how to support teachers' effort in finding the appropriate resources (Hill & Hannafin, 2001; Jaen et al.,

[6] Published January, 2012 in *Computers & Education, 58*(1), 551-559.

2007). This challenge is particularly important given that the creation and analysis of teacher resources provides a learning opportunity for teachers (Ball & Cohen, 1996; Leat & Higgins, 2002; Sparks & Loucks-Horsley, 1989) that can result in improved student learning.

Fortunately, ultra-large-scale interactive systems have various features designed to help users navigate the large amount of resources made available. The most common method is the inclusion of search engines (Fallows, Rainie, & Mudd, 2004), such as is found in Amazon. For example, users can search Amazon for a book based on content or metadata like genre or date of publication. However, search engines using only content and genre/date metadata are possibly limited in their ability to return optimal results. Specifically, a lack of completeness, accuracy, or uniqueness of a resource's content and metadata could generate search results that are skewed or incorrect.

Another common method for assisting user search is to allow for and take advantage of user ratings or comments on a specific resource (Schafer, Konstan, & Riedi, 1999); these ratings and comments are what we call "evaluative" metadata. Users looking for a book at Amazon can narrow their search results by looking at what ratings or comments that other users, often described as the crowd, have left about the resources. These ratings and comments allow for lists of resources to be ordered by ratings, essentially creating recommendations from the input of other users. Evaluative metadata is commonly found in systems that wish to leverage their users' ability to evaluate in order to provide additional filtering of their vast resources. Other examples include Youtube, which provides ratings and comments of submitted movies, and Ebay, which provides ratings and comments of buyers and sellers. Of course, ratings can be biased and comments can be noisy, limiting the benefit of evaluative metadata.

If we are to better understand how ultra-large-scale systems impact teachers then it is important to look at how teachers are currently selecting resources beyond content or keyword search. How are teachers using evaluative metadata when selecting resources in ultra-large- scale interactive systems? Based on the preponderance of crowd-sourced “evaluative” metadata in highly successful commercial websites, we selected what we observed to be the most prevalent types: comments, ratings, and popularity. Based on their association with commercial success, we hypothesized that these three specific “evaluative” metadata would drive teacher selection of resources in large- scale online systems of teacher materials. To test our theory, we conducted an observational study of an existing popular ultra-large- scale online system called TeachersPayTeachers.

3.1.1 Theoretical Background

A recurring theme in research on how teachers use technology and the Internet involves ways in which metadata could contribute to more optimal resource selection. Recker et al. (2005) conducted educator workshops and found gaps in metadata that otherwise would have helped educators discover appropriate resources. Wang and Hsu (2006) designed a demonstration system for how different types of searches of metadata can result in good resource selection. Others have explored the design of systems in which accurate metadata could also help teachers find resources that minimize the need for adaption for instruction (Recker, Dorward, & Nelson, 2004; Suthers, 2001). Many computer scientists have focused on the creation, testing, and implementation of metadata standards for online learning resources (Anido et al., 2002; Bohl, Schellhase, Sengler, & Winand, 2002; Gonzalez-Barbone & Anido-Rifon, 2008; Plodzien, Stemposz, & Stasiecka, 2006).

Additionally, research on measuring usability in digital libraries has created heuristic models that can be applied to a variety of repositories of online resources (Jeng, 2005). Often the models for high quality digital libraries include objective measures of metadata quality (Goncalves, Moreira, Fox, & Watson, 2007) or suggest methods for improving metadata accuracy (Nesbit, Belfer, & Vargo, 2002).

Seemingly absent is research on how metadata impacts teacher selection of resources in the type of large-scale interactive environments becoming more common on the Internet. This absence is indicative of the social-technical gap that often occurs in this type of research (Ackerman, 2000). The social approach examines the social needs of a particular group, in this case teachers, given existing technologies and suggests how new technology could meet those needs. The technology approach usually designs new technologies to benefit a social group, such as designing innovative technology to benefit teachers. The missing approach involves examining how new technology is currently being used by these social groups, such as how teachers are currently using innovative technology.

To help close this gap, a good strategy would involve examining teacher use of metadata in ultra-large-scale interactive systems. Large- scale interactive environments present unique opportunities for system designers and teachers to access large quantities of materials. Furthermore, metadata can provide a window into how teachers are finding resources in these systems.

We limited our investigation by focusing on the use of evaluative metadata. There are studies that have looked at how best to incorporate evaluation of resources into searchable metadata (Nesbit & Li, 2004; Nesbit, Li, & Leacock, 2006) and examined how teacher communities can collaborate to provide more accurate evaluative metadata (Recker, Walker, &

Lawless, 2003). Given that Internet search engines are becoming more powerful and that non-evaluative metadata is static, we believe that finding content-relevant teaching resources based on non-evaluative metadata will eventually become a relatively straightforward process. We suggest that evaluative metadata, in terms of its potential ability to distinguish quality levels among relevant resources, is where the power of ultra-large-scale interactive systems lies. However, this expectation does not imply that all methods of collecting and using evaluative metadata will be successful there likely will be some socio-technical challenges to effectively manage collection and use.

Conducting any laboratory analysis of ultra-large scale interactive systems is unrealistic simply because of the number of subjects involved. Finding thousands of users to participate in an experiment is impractical. The only practical option is to analyze a system that currently operates with both a very large user-base and a very large amount of resources. These requirements necessitate finding a system that is reliable and sustainable (Schlager, Fusco, & Schank, 2002) with a critical mass of users and a robust incentive system for attracting new users, adding new resources, and evaluating existing resources (Ackerman, 2000). Analyzing a currently operating system also has the added benefit of placing our findings within the reality of current teacher behavior.

3.1.2 Data Source

The online system we examined is a website called TeachersPayTeachers (TPT) that allows teachers to sell resources they have designed and created directly to other teachers. The operators of TPT graciously provided us with a complete copy of the main database with user name and credit card information removed. Included in the database were resource names, descriptions, overall rating of the resource, number of times a resources had been rated, number of times a resources had been sold, price of the resource, comments left by users for each resource, a resource ID number, and the ID number of the author. Also included was metadata generated by the author of the resource and displayed on each resource's webpage in TPT. The metadata consisted of grade level appropriateness, subject area, and file type. Standard web metrics such as web page hits were not available.

The following description of TPT is a result of direct interaction with the system along with discussion with TPT's owner. All of the information provided has been confirmed for accuracy with the owner.

Operating since February of 2006 to March of 2010 at the time of the database dump, TPT had over 200,000 registered users, 62,781 resources, and over $900,000 in sales. TPT allows teachers of numerous subjects, grade levels, and experience to put lesson plans, unit plans, assessments, or other teacher resources up for sale. Built into the system is the opportunity for buyers of materials to generate metadata in the form of several ratings and comments. Sellers categorize their resource according to subjects, grade level appropriateness, and type of resource such as assessment, worksheet, presentation, and review materials. Buyers can search the available resources by the non- evaluative metadata such as subject, grade level, type, and keyword. Searches can also be based on evaluative metadata such as what are the bestselling resources and which resources have the highest rating. Buyers can further use content descriptions and evaluative metadata to then inform their purchases from the searched items.

Writing reviews of resources that cost money in TPT is only allowed for registered users who have purchased and downloaded the resource. The rubric is a rating scale with 8 rankings in the form of grade letters (F, D, D+, C, C+, B, B+, A) across six categories: Overall Quality, Accuracy, Practicality, Thoroughness, Creativity, and Clarity. Ratings for each category are transformed by TPT to a five-point scale (0–4 presented to the tenth decimal point), averaged, and displayed prominently for each resource. In addition, alongside the author's name is the average rating of all of their submitted resources. Users entering a rating also have the option of writing a comment about the resource that will be displayed along with the average rating of the resource.

A typical use case scenario would begin with a teacher logging into the system and entering some keywords along with a specific grade level and subject area. The results of the search are resources listed according to the relevancy of the search. The teacher can then scroll through the list looking at a short summary, overall rating from other users, number of times the resource had been rated, subject areas, grade levels, type of resource such as assessment or activity, instruction time required, author, and a price. The list can be resorted according to rating, alphabetical according to resource name, price, sales, and date of creation. Once the teacher selects a resource that seems promising they can view the resource's page (see Figure 2) where a more detailed description is available along with a preview, individual ratings, individual comments, a question and answer forum for each resource, and a short biography of the resource creator. The teacher can then choose to add the resource to their shopping cart or go back to the search results to look for a more promising resource. The teacher can also choose to browse the collection of resources directly looking at a specific grade level, subject area, price category, or bestselling materials. Once a resource has been purchased, teachers are encouraged to use the purchased resource in instruction before returning to rate it (see Figure 3). Ratings are not required but encouraged (Figure 4).

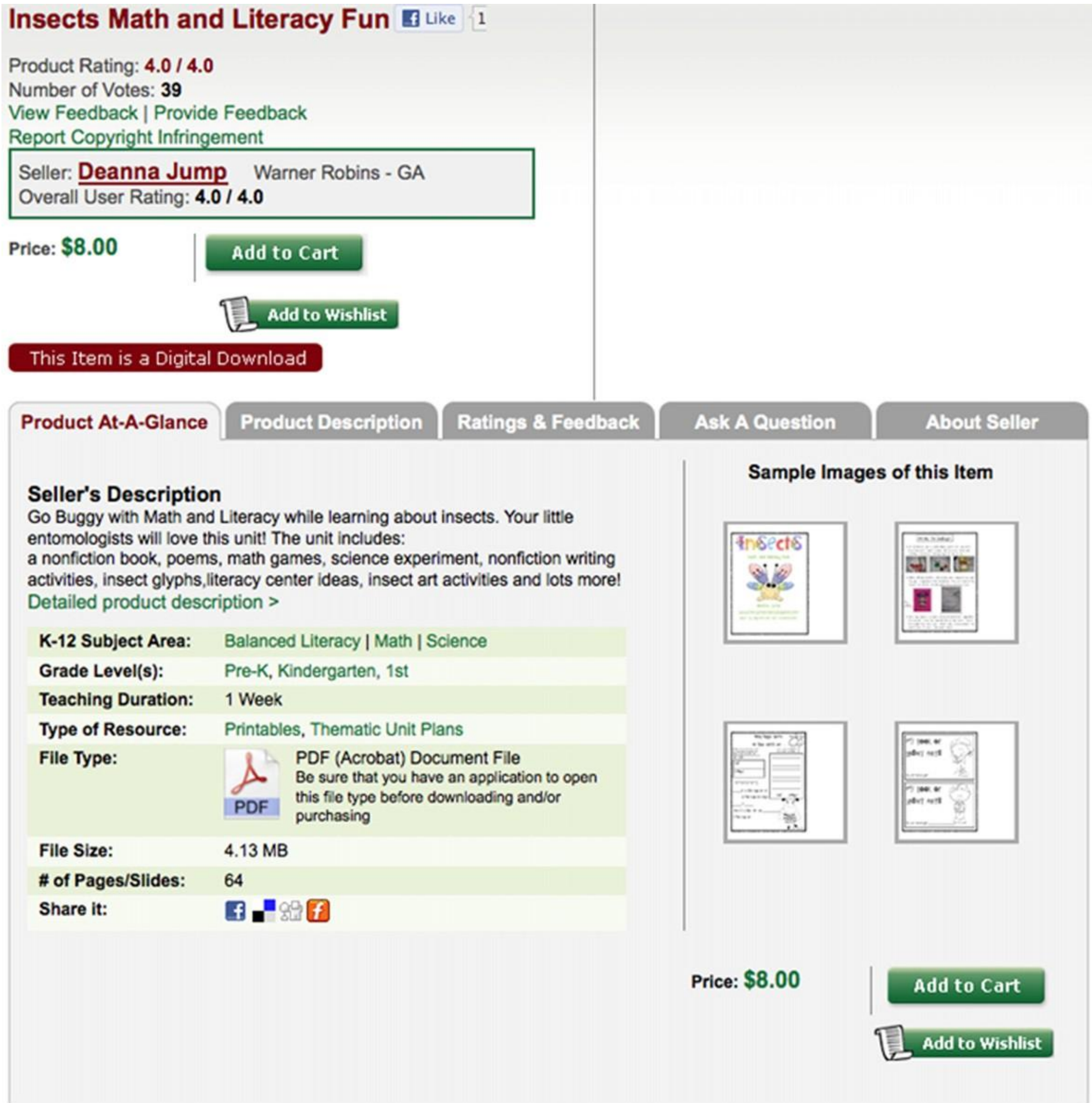


Figure 2: Resource Page

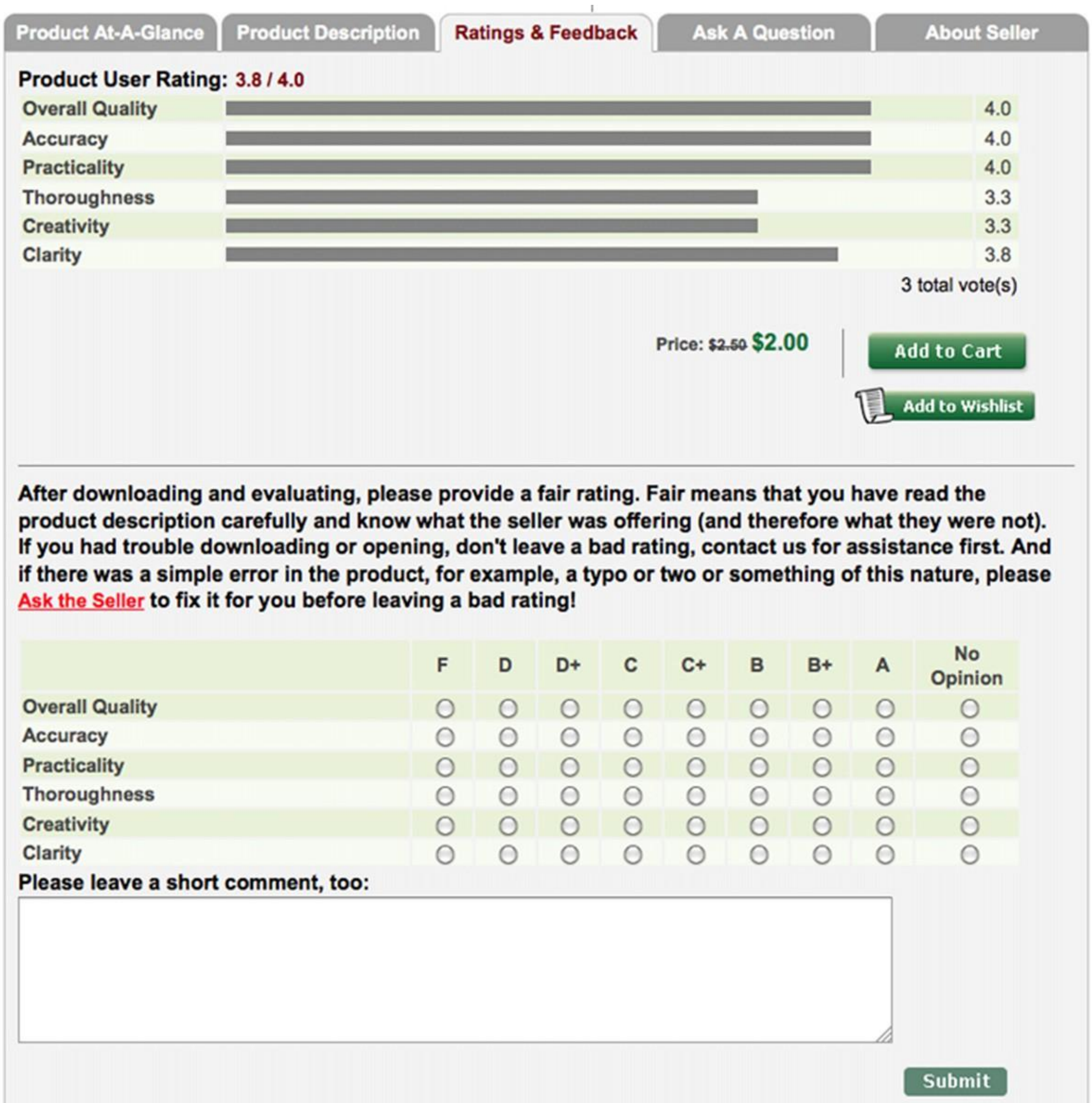


Figure 3: Rating Rubric for Resources

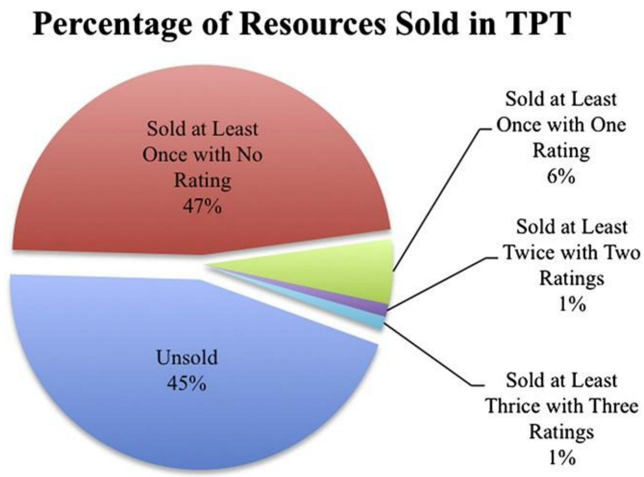


Figure 4: Percentage of Resources Sold in TPT with Number of Ratings

TPT meets the previously outlined requirements for sustainability and incentivized participation. The large number of participants insures a critical mass of resource submissions and purchases. The fact that resources are sold incentivizes sellers to provide resources and the overall low cost of the resources, typically 1 to 5 dollars per resource, incentivizes buyers to purchase resources.

As an ultra-large-scale interactive system, the issue of search and selection is critical. What evaluative metadata are teachers using to select resources?

3.2 STUDY 1: EVALUATIVE METADATA CONSISTENCY

3.2.1 Methods

The measure by which resource metadata is useful to teachers is to the degree that teachers can find, identify, select, and acquire the resource that they are seeking (Bruce & Hillmann, 2004).

To abstractly determine metadata quality, we could have examined completeness, accuracy, provenance, consistency, conformance, or accessibility (Bruce & Hillmann, 2004). Each of these measures has drawbacks for the purpose of determining a practical use to teachers. For example, it is hard to attach an analysis to a rating that could be made for subjective reasons.

TPT introduces another quantitative measure of curriculum value, namely sales. Teachers spend an average of $623 dollars a year of their own money on classroom preparation (National Teaching Realities Survey, 2010), a non-trivial expenditure for a teacher in the current global economy. As a result, the number of sales of a resource indicates a certain valuation by the purchasing teachers. Evaluative metadata that correlates positively with sales might indicate teacher criteria for selection.

Because evaluative metadata is optional in TPT, it is not necessarily the case that users would rate products or leave comments. The absence of such metadata would limit its use. Further, large skews in the metadata, such as all positive reviews, would limit the value as well. So to determine what evaluative metadata guides teachers to select resources in an ultra-large-scale interactive system we first looked for some overall patterns of the evaluative metadata. Our research questions were:

1. What percentages of items that are sold receive evaluative metadata (i.e., have a rating)?
2. What is the distribution of evaluative metadata (i.e., ratings), positive and negative, amongst items that are sold?

To examine the content of the thousands of comments in the database, we used Pennebaker, Booth, and Francis's Linguistic Inquiry and Word Count (LIWC) (2007). LIWC is a computer program that can identify positive and negative affect words with high reliability and has been successfully used in a variety of research contexts (e.g., Paletz & Schunn, 2009).

Having established the basic descriptive data, we then turned to correlational analyses of evaluative metadata and sales:

3. Does the evaluative metadata relate to how the resources are valued? Is the evaluative metadata and average sales related?

3.2.2 Results & Discussion

Mining the TPT database for answers to these questions revealed some surprising findings. The vast majority of sold items never receive any additional evaluative metadata. Of the 34,602 items that were sold at least once, approximately 86% of the sold items never receiving any type of rating, 10% rated once, 2% rated twice, and 2% rated three times or more. Similarly, 88% of all items sold at least once receive no comments, 7% received one comment, 3% received two comments, and 2% received three comments or more. This level of activity is analogous to the 90-9-1 rule of participation (Nielsen, 2006).

Turning to the content of the ratings, of those items that receive at least one rating, the vast majority of the ratings are very positive: 76% received a mean rating of three or higher and 40% have a mean rating of four, the highest possible rating. A similar pattern was observed with resource comments left by users. According to our LIWC analysis, positive comments outweighed negative comments 13 to 1. Thus the evaluative metadata generated in the form of ratings or comments is usually very positive and generally not very discriminating. The large amount of positive ratings and comments could be the result of teachers not perceiving quality differences between different resources and finding them all sufficient for instructional purposes.

Another possibility could be that teachers are only choosing to provide ratings and comments in the positive cases. Additional research is necessary to better understand the nature of the positive ratings.

To examine the relationship among variables, we present analyses using Pearson correlations and linear regression - the R2 captures the percent of variance accounted for by the analyses. The large skew in all the variables raises a question about whether raw or transformed data, or whether Pearson or Spearman correlations should be used. There are a large absolute number of data points in the tails. Further, those tails are pragmatically important to viability of TPT and describe the most common monetary transactions. Therefore, we prefer to report the Pearson correlations that proportionally weight the contributions of these data points in the tails. However, analyses using log- transformed data or Spearman correlations produce similar conclusions.

Despite the relatively small percentage of items that are rated, the very large pool of sold items (N = 34,602) meant that we could still determine with high precision how closely sales are related to evaluative metadata. The correlations are presented in Fig. 4.

Three of the four key relationships were highly statistically significant and two of the four were at least moderate in strength. Positive emotion words were significantly associated with sales, but at a very weak level; further the relationship with negative emotion words was not statistically significant ($r = -0.02$, $p > 0.19$). Ratings correlate with sales at a weak level, which could indicate that sales are weakly related to material quality, as according to TPT users, or that displayed ratings of quality is weakly related to sales.

Another salient finding is that sales are more strongly related to number of ratings and comments than to average rating. Combining number of ratings and comments into a linear regression reveals that they account for just under half of the variance in sales ($R2 = 0.47$, $F (2,34,602) = 15,415$, $p < 0.0001$), with independent statistically significant contributions from both factors. Here there is likely a bidirectional influence: items that have higher sales also have more opportunities for generating evaluative metadata while evaluative metadata may influence others to buy.

The small correlation between sales and mean rating does not have that possible mediation through number of opportunities to provide ratings so the evidence of a direct but weaker relationship is more evident. However, it is important to acknowledge possible third variable confounds, such as some topics being more popular with higher quality ratings and higher sales. Including topics, grade, type of resources in the multiple regression did not reduce the relationships between mean quality ratings and sales, and thus a topic-based third variable confound is unlikely to be the driving factor.

But the initial aggregate analyses for number of ratings and number of comments do not specify a causal directionality because we did not have time-stamped information about when the evaluative metadata was generated. We cannot definitively determine what were the number of sales for a resource prior to generation of the evaluative metadata and what were the number of sales after the evaluative metadata was visible to potential buyers. To partially address this issue, we ran the same statistical analyses for items that sold at least fifteen times. The premise for this analysis is that buyers will eventually be influenced by a critical number of ratings and comments no matter what other reasons for purchasing a resource (Standifird, 2001). Consequently we hypothesize that by the time a resource has been purchased a minimum of fifteen times then the critical number of ratings and comment necessary to influence sales will have been reached. The minimum of fifteen sales is a simple threshold chosen for being well above the average number of sales per resource while still providing a large enough number of resources to merit statistical analysis.

The same statistical patterns that existed for all items that sold at least once also existed for those items that sold 15 times or more as well, including the variance of sales predicted by number of ratings and comments (R2 ¼ 0.48, F (2,5080) ¼ 2332, p < 0.01). If the effect had been purely from sales to opportunity to rate and comment, then we would have expected the correlation between sales and number of ratings and comments to be at least partially attenuated in this higher sales subset.

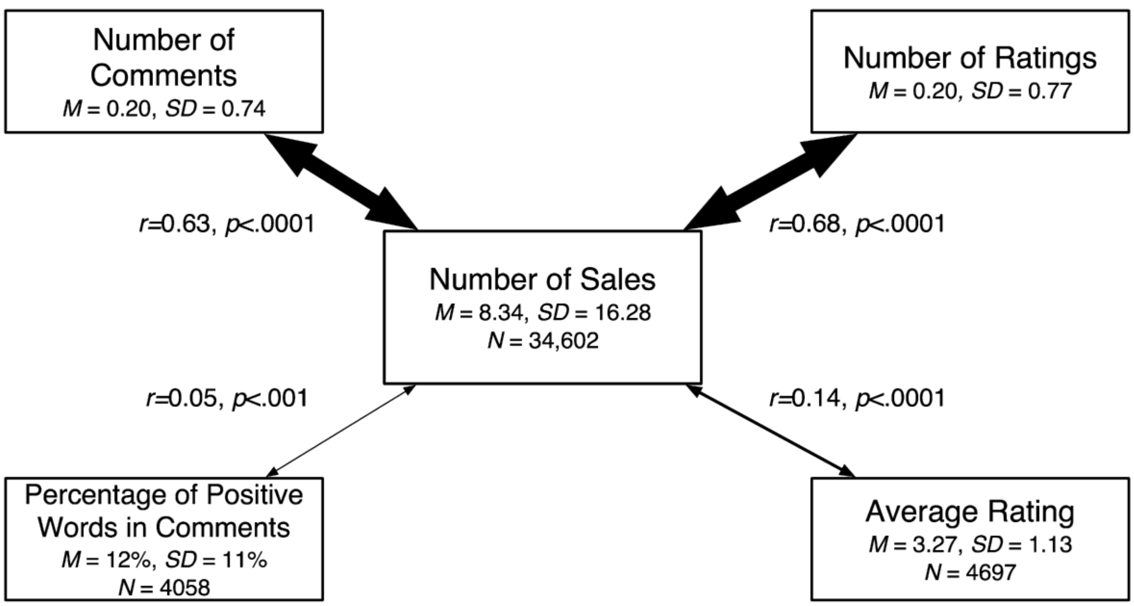


Figure 5: The strength of correlations between Evaluative Metadata and Sales, with Evaluative Metadata broken into comments vs. ratings by their number vs. content. Thicker arrow bars represent a stronger correlation with sales.

3.3 STUDY 2: EVALUATIVE METADATA VALIDITY

Although the initial analyses suggest a weak relationship between sales and user quality ratings and comments, we do not yet know whether the evaluative metadata could indicate a quality resource because we do not yet know the validity of user ratings and comments. Further, it is interesting to examine whether number of comments and ratings is perhaps also indirectly correlated with resource quality. With such a skew in user ratings, perhaps the presence of evaluative metadata could be a useful clue to teachers regarding material quality. Finally, to evaluate the overall success of the search tools in providing access to higher quality materials, it is interesting to see how sales are associated with material quality. Therefore, our next step was to determine the relationship of TPT's evaluative metadata (sales, mean rating, comment content, number of ratings, and number of comments) to actual material quality by comparing them to ratings generated by curriculum experts.

3.3.1 Methods

TPT includes content from a very wide range of domains and therefore it was necessary to focus on a particular content domain in order to narrow the scope of our study to meet our available capitals. We chose to analyze math resources because of our access to math resource experts in addition to the widely recognized importance of providing math teachers with good curriculum materials (Stein, Smith, Henningsen, & Silver, 2000). The percentage of resources in TPT varied somewhat by subject area, with math resources accounting for 11% of the total. The patterns that we established in Study 1 for all TPT resources also existed for the subset of math resources, further legitimizing our selection of the domain for further study.

Table 1: Number of math resources selected for analysis by experts by removing resources no longer for sale, removing cases too rare to produce informative results, and balancing cell size across cells that could be examined.

Number of Items per Cell		Number of Ratings		
		Hi = 3+	Med = 2	Low = 1
Average Rating	Hi = 3–4	18 (of 19 possible)	21 (of 26 possible)	20 (of 137 possible)
	Med = 2–2.9	0 (of 1)	0 (of 2)	20 (of 21 possible)
	Low = <2	0 (of 0)	0 (of 2)	24 (of 28 possible)

In the TPT database, there were 4,060 math resources that had at least one sale in its database. Because we wanted to compare user ratings with experts, we eliminated those resources that did not have a rating. It is possible that some purchases of resources are made simply for curiosity or an impulse purchase, especially when offered for free or very cheaply, so we also restricted the pool of math resources to those that cost $2 or more. After further eliminating resources that were no longer for sale (and thus we could not acquire them), we randomly selected a stratified sample of resources that represented the different combination of number of ratings and average rating for the resources of TPT from a possible pool of 236 math resources (see Table 1). Because of how ratings were distributed, not all combinations of rating number and average rating occurred often enough to be studied. Resources that were rated frequently and given a poor or average rating were extremely rare. Consequently our sampling was done such that the correlations of expert ratings with each of quality and number of ratings or average rating could be examined while holding the other dimension constant. We eliminated certain cells from our analysis since they would not have provided enough resources to allow for statistical findings. This filter resulted in our elimination of the rare cases of any resource that had 2 or more ratings in addition to an average rating of less than 3. For the resources in the remaining cells, we selected all available resources (that were still for sale). The one exception to this complete inclusion was for resources that were rated highly only once, which had a pool of 137 possible resources. Of those 137, we selected 20 at random in order to prevent them from having an overwhelming affect on our analysis based on sheer sample size and to make the rating task for experts manageable. This sampling choice was also consistent with our overall research approach of examining the naturally occurring behavior within the system. Overall, the resources had a mean price of $3.08 with a standard deviation in price of $1.29.

We selected five math resource experts based on the criteria that they had previously taught math in a K-12 school setting and were either pursuing their doctorate or had already received it in a field directly related to mathematics education. Three of the experts are currently studying advanced math education topics and one rater, with a doctoral degree in education, is an assistant professor in a math department.

The experts were given the same rubric and description of the resource available in TPT. The resources were randomly presented in their original format in an online structure similar to how the resources are made available in TPT.

For 12 of the resources, some change to the format was required. These resources were originally formatted for a Smartboard and had to be converted to a more compatible web format. In these cases, the experts, all familiar with how a Smartboard works, were told which resources were converted for their review and were asked to judge those resources more favorably on the basis that some formatting had been altered likely to the resources' detriment.

Judging the quality of instructional resources is a difficult task, even for experts. Ideally it requires knowing the intended learning goals and the way in which the materials are to be used (Kesidou & Roseman, 2002). Experts were told to use the information provided by the seller regarding intended use and intended goals. However much information was provided. Resources could be classified by their author for any grade level and up to three different categories. Although the experts reported the task difficult, reliability of the expert judgments at the level of average quality rating was moderate but acceptable given the difficulty of the task and the unknown quality of the rubric used by TPT, Cronbach's Alpha = 0.64.

Because there were a more manageable set of comments, the comment content for all 103 resources were classified by hand as either positive, negative, or simply resource related (e.g., "I used this resource in my classroom."). By way of validation, LIWC ratings were significantly correlated with human ratings: positive rated comments with percentage of positive emotion words ($r = 0.61$, $p < 0.01$), and negative rated comments with percentage of negative emotion words ($r = 0.28$, $p < 0.01$). The lower correlation of negative ratings likely stems from their rare occurrence (see below).

3.3.2 Results

Unsurprising to us, expert ratings of the resources were much lower than the ratings given by users of TPT. The average expert rating was 1.9, with no expert giving a mean rating for a resource higher than 2.9, in contrast to a mean user rating of 3.3 for these items. Experts in general tended to be harsher in their evaluation than novices. We fully expected this result given that experts have the ability to be more critical because of their familiarity with the many potential flaws found in typical resources for supporting student learning.

It is interesting that expert ratings were significantly correlated with sales ($r = 0.19$, $n = 103$, $p = 0.04$). Thus, there is some evidence at the TPT structure does enable some discrimination according to expert-determined material quality, although the relationship is weak. Further, this finding suggests that users could use prior sales information as a valid but weak clue regarding material quality.

Turning to the more putative system measures of material quality, the correlation between the expert ratings and the mean rating from TPT was not statistically significant ($r = 0.15$, $p = 0.13$). Thus, the ratings found in the system are not in themselves useful predictors of quality according to experts. Because we selected items with varying quality ratings, the explanation cannot lie in the skew in the ratings. However, the very small number of ratings used to produce these means may be an underlying cause. It is likely that there is a lot of variability across teachers just as there is across experts and only one rating is not sufficient data to produce a reliable quality estimate.

An examination of the content of comments in TPT also revealed no significant correlations. The number of overall comments ($r = 0.01$, $p = 0.45$), the number of positive comments ($r = 0.01$, $p = 0.54$), the number of negative comments ($r = 0.01$, $p = 0.26$), or the number of resource comments ($r = 0.01$, $p = 0.91$) were not correlated with expert ratings.

We could discern no other statistically significant correlation between the experts' rating of resources and the other metrics from TPT other than sales, including the correlation between experts' rating and the number of ratings ($r = 0.14$, $p = 0.14$).

The correlation between expert ratings and number of ratings was the closest to statistically significant, but when ran as a multiple linear regression with sales and number of ratings as the independent variable and expert ratings as the dependent variable, the independent contribution of number of ratings was not statistically significant ($p = 0.17$).

Our conclusion is that the positive ratings and comments described in Study 1 are not the equivalent of expert analysis. Additionally, the other available metadata for each resource is also not the equivalent of expert analysis. The positive correlation between sales and expert ratings does indicate to a degree that teachers are purchasing quality resources but the criteria for selecting the resources is probably not directly based on the evaluative metadata available in TPT.

3.4 GENERAL DISCUSSION

So what evaluative metadata drives teachers to select resources in ultra-large-scale interactive systems? Sales did correlate weakly with TPT ratings and, to an even lesser degree, expert ratings. These statistically significant relationships could indicate that the bestselling items in TPT are due in part to the evaluative metadata listing their rated quality. However, the correlations between user ratings and actual quality are low enough that it would be hard to conclude that an expert-determined level of quality was anything more than a minor influence on sales. A safer inference is that mean rating metadata does have a small influence on sales but not to a degree that a buyer can be assured of quality based on the rating alone or that a seller can be assured that a highly selling item is because of higher ratings than a poorly selling item. These findings are important in that they serve as a benchmark for comparison to other teacher resource systems that rely on peer- based ratings.

A stronger correlation occurred between sales and the number of ratings and comments of a TPT resource. One explanation of this relationship is that the relative scarcity of evaluative metadata accentuates those items that do receive evaluation. Because TPT has so many resources, using the system without whittling down choices might be overwhelming (Schwartz et al., 2002). Users might seek some initial way to differentiate the vast number of items in the database in order to choose from a smaller pool. Items that have evaluative metadata are rarer and, by setting a minimum rated threshold, allow a user to have a more comfortable item pool to choose from, even if those ratings are negative (Smith, Menon, & Sivakumar, 2005).

Turning to the issue of the accuracy of the evaluative metadata, it is not surprising to find that experts rated the resources in TPT much lower than TPT users. Curriculum experts should have a much higher expectation of quality for teacher resources or then they wouldn’t be experts. In addition, teachers may have rated the materials on a different basis than the ones sellers intended or experts used, resulting in a higher rating. The reliability of the rating rubric must also be examined to investigate consistency of the ratings across users.

Another reason why number of ratings might have a large effect on sales is that some of the TPT default views list the most reviewed items in descending order. With companies such as Google and Microsoft producing more optimized search engines it is possible that users have increasing amounts of faith in how search engines list results.

3.4.1 Future Implications

Despite the findings presented, additional research is necessary to better understand how teachers interpret and use evaluative metadata in resource exchange systems. We only conclude that the evaluative metadata in TPT is not the equivalent of expert reviews. We hope to further unpack how teachers use metadata when selecting resources with additional research such as surveys and case studies.

Our findings do allow us to theorize about possible motivations that drive participation in TPT. The most obvious potential motivators are profit for sellers and the ease of finding a specific resource for buyers. However, ultra-large-scale interactive systems must also rely on a community of users who identify with the goals of the system and provide the data necessary for operation. These communities can be based on social-bonds or identity-bonds. Social-bond based communities are where the participants have a social connection with each other that exist independent of the online community. An example would be a community of teachers at a singular school. Identity-bond based communities occur where the participants desire to be in a community is based on a perceived identification with the community independent of any social-bonds. An example would be a music teacher who seeks an online community of other music teachers despite being part of a school's teacher community (Ren, Kraut, & Kiesler, 2007).

TPT could also rely on an identity-based community to insure its success, namely the contribution of resources and associated metadata. The resource classification and ratings system are designed to be similar to what teachers use for their students, utilizing commonly accepted vocabulary and performance indicators. Essentially TPT, capitalizing on that teachers have a strong identity with their professional community, encourages an identity-based community (Goodson & Cole, 1994). While we were not focused on understanding how motivation impacts teacher resource selection, this topic is an important subject to explore if we are to form better conclusions about teacher behavior online. For instance, altering the incentives for evaluating resources to better leverage the benefits of an identity-based community could encourage more generation of evaluative metadata (Cheng & Vassileva, 2006) and increase TPT's ability to more accurately evaluate resources.

An ultra-large-scale interactive system could fit a model of a knowledge building community (Scardamalia & Bereiter, 1994) or of a resource-based learning environment (Hill & Hannafin, 2001). By being focused on the creation and review of teaching materials such as problems, lessons, or assessment the system would fulfill some of the elements of good teacher professional development include being content specific (Borko & Putnam, 1995), use of actual curricular materials (Ball & Cohen, 1999; Hawley & Valli, 1999), providing opportunity for reflection (Hawley & Valli, 1999), and being ongoing (Hawley & Valli, 1999). Findings from research on ultra-large-scale interactive systems could benefit teachers who do not use online resources by uncovering improvements that could be applied to all teacher professional development. Another benefit could be an improved machine learning approach to identifying resource quality (Bethard, Wetzer, Butcher, Martin, & Sumner, 2009) or more efficiency in the usability design of the systems (Sumner, Khoo, Recker, & Marlino, 2003).

As we develop a better understanding of how teachers currently use these systems we can close the social-technical gap of our understanding and begin to mold better learning opportunities for our teachers. Hopefully, in the near future, we can directly measure the effect between ultra-large-scale interactive systems on student learning.

4.0 PAPER 2 - THE ROLE OF EVALUATIVE METADATA IN AN ONLINE TEACHER RESOURCE EXCHANGE[7]

Abstract

This paper examines the online teacher resource exchange for Teach for America to examine how metadata (including ratings) influences how teachers select resources online. Using downloads as the dependent variable, a hierarchical linear model was used to account for resource author's influence. The findings include evidence that larger numbers of ratings predict more downloads than average rating. Utilizing decision heuristics theory as the base of a theoretical framework, the authors conclude that there is no evidence that teachers rely on a singular simple heuristic when determining whether to download a resource.

4.1 INTRODUCTION

Education researchers have previously established the impact of teacher decision-making (Munby, 1982). Decisions a teacher makes both prior, during, and post instruction can impact student learning (Colton & Sparks-Langer, 1993). Many researchers have examined the decisions that teachers make during planning prior to instruction since those decisions can be

---

[7] Under Review in *Educational Technology Research & Development*.

indicative of and influential on teacher behaviors, which in turn influence what students learn (Doyle & Ponder, 1977; Maloch, et al., 2003; Stern & Shavelson, 1983; Weiss, Cambone, & Wyeth, 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 & Trouche, 2012).

As more educators choose to go online to gather resources (Gray, Thomas, & Lewis, 2010; M Recker, Dorward, & Nelson, 2004), new questions emerge for how information and communication technologies on the Internet impact or influence a teacher's decision to select a particular resource. For example, online teacher resource exchanges have emerged as a response to the desire of teachers to find high-quality, ready-to-implement teaching resources. Usually found on the World Wide Web, these systems facilitate the sharing of resources by providing a technical infrastructure for teachers to upload and download a resource.

However, given that Internet based technologies continue to transform so many other aspects of learning and teaching, it is surprising that there is little research investigating how Internet based technologies could influence teacher decision making on resource selection. The few studies that investigate how Internet technologies influence teacher planning either offer the unlikely conclusion that technology does affect teacher planning (Tubin & Edri, 2004) or are limited to suggesting ways to think about how technology could be used for planning (Harris & Hofer, 2009).

We, as a community concerned with improving teaching and learning, must investigate how Internet technologies such as online teacher resource exchanges effect teachers' selection of teaching resources. The evaluation of educational resources can have a direct effect on teacher practice and teacher learning (DL Ball & Cohen, 1999; Leat & Higgins, 2002; Sparks & Loucks-

Horsley, 1989), and consequently student performance (J. Hill & M. Hannafin, 2001). By unpacking the processes of online teacher resource exchanges, we can gain current insight into current teaching practices and identify ways to improve the process.

4.2 THEORETICAL BACKGROUND

Prior to the advent of teaching material being available on the Internet, educators were more limited in their ability to acquire teaching resources. Teachers could slowly and infrequently get 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. Once the Internet greatly reduced scarcity and barriers to gaining access to resources, the established models of resource filtering broke (Metzger, et al., 2010). Now many teachers go online to select and immediately obtain resources for instruction just at they also go online as private citizens to purchase other products.

4.2.1 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 still rely 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, Khabiri, & Caverlee, 2009). Ostensibly, access to this data adds benefit to users in that they now have simple cues to narrow their selection.

However, these types of rating systems do not always result in predictable results. Prior lab studies of knowledge management systems suggest that the number of user-entered ratings might not affect content searches (Poston & Speier, 2005). A survey conducted on Amazon reviews reported that moderate reviews were more helpful than extremely positive or negative comments (Mudambi & Schuff, 2010). Consequently, rating systems for teaching resources should not be instituted based solely on expected behaviors without considering the unique facets of the teacher decision to select a teaching resource. For example, teacher evaluation of a resource can be influenced by unintended consequences (Brown & Edelson, 2003). A teacher's beliefs about subject-specific pedagogies can unfairly influence their evaluation of teaching resources (Remillard, 2012; Stein, et al., 1996). 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. If a teacher is unfamiliar with particular content knowledge or pedagogies then there exists a potential for important details of a resource to be ignored or improperly evaluated.

4.2.2 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. Learning resources are any digital document that can be used for learning (Lau & Woods, 2009; Ochoa & Duval, 2008) Teaching resources might include some of the same information as a learning resource but should also include additional information. For example, math teachers have been known to take learning resources that are designed to induce high-level mathematical thinking and during instruction proceduralize the mathematics, lowering the effectiveness of the resource (Stein, et al., 1996). A properly designed teaching resource might not include a digital resource or artifact but instead could provide information on effective pedagogies, provide subject matter content for the teacher, or suggest ways that a teacher can relate learning to other parts of a curriculum (Davis & 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) and ARIADNE (Klerkx, et al., 2010) are designed to be forums where many different types of educators can exchange a wide variety of 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, Klerkx, Vandeputte, & Duval, 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 & Olsen, 2011), and teachers may prefer to use other methods, like evaluative metadata, to guide their search for materials.

Teacher resource exchanges are often designed for either a specific subset of teachers or a specific teaching need. The National Education Association[8], the Public Broadcasting System[9], and the Department of Education of Ohio[10] all provide a variety of lesson plans online targeted for teachers and addressing a specific teaching need, ignoring metadata standards. Instead of being guided primarily by the metadata standards, these systems employ evaluative metadata, allowing users to rate and comment on each lesson plan. Theoretically, the ratings and comments on these teaching resources allow users to better select the resource that match teacher and student need. However, relatively little is known about how teachers make use of these sites. Most saliently, questions emerge about how teachers use evaluative metadata to decide on a resource.

4.2.3 Decision Heuristics

In studying how teachers using evaluative metadata to guide resource selection, it is of course important to examine which kinds of metadata guide teacher selections: number of ratings, mean rating level, number of comments, length of comments, etc. But it is just as important to study whether and how teachers integrate these factors together to be able to inform policy makers, researchers, and website designers.

To develop a theory for how teachers select 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 &

[8] http://www.nea.org/tools/BrowseAllLessons.html
[9] http://www.pbs.org/teachers
[10] http://ims.ode.state.oh.us/

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 resource selected.

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 (see Table 2). 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 the strength/extent of the positives and negatives. For example, a resource with one positive rating would be evaluated for download differently that a similar resource that had an equal number of positive and negative ratings.

Table 2: Stopping Rule Heuristics

Simple Heuristics	**Explanation**
Recognition	Based on a resemblance to prior chosen resources
Take the Best	Using the one best factor or test that had previously worked for selecting a resource
Take the Last	Using the one factor or test that had last worked for selecting a resource
Complex Heuristics	
Dawes' Rule	Counting the number of Pros vs. Cons
Franklin's Rule	A weighted counting of Pros vs. Cons
Multiple Linear Regression	A weighted sum of the extent of Pros vs. Cons

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, Gigerenzer, & Goldstein, 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.

Recent research examined the website TeachersPayTeachers (TPT) (Authors, 2012). TPT is an online resource exchange that allows educators to buy and sell teaching resources to other users. With more than 62,781 resources by March 2010, TPT's designers could not rely on a group of experts to effectively evaluate and filter resources for the resource purchasers without amassing huge costs. Instead, TPT provides content search tools and user ratings/comments;

each resource has its own page where buyers can provide a rating and comment about their purchase, potentially helping a future buyer choose a resource to purchase.

One of the key findings from the study of TPT was that the number of ratings and comments for teaching resources were highly positively correlated with the main resource evaluation criteria in TPT, sales. That is, the mere presence of ratings and comments at all appeared to influence user purchases. Further, from multiple regression analyses, the effects appeared to suggest a weighted combination of these two factors drove decision-making, suggesting some correspondence to Franklin's rule or Multiple Regression models of decision-making and not 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 level of the ratings might not have further influenced decision making because there was insufficient information in them. 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.

4.3 DATA SOURCE

For our study, we examined the data metrics from TFANet - Teach for America's (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 two years in urban and rural schools in the US with underserved students. For the 2011-2012 school year, TFA will enlist 9,000 corps members who will teach 600,000 students and is projected to continue to grow with the goal of placing 15,000 teachers in 60 different regions by 2015. In addition, TFA continues to receive both private and public funding, including a 2010 $50 million dollar grant from the US Federal Department of Education.

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 (Figure 6). According to TFA, during the 2010 Fall semester, 75 percent of corps members downloaded resources from TFANet.

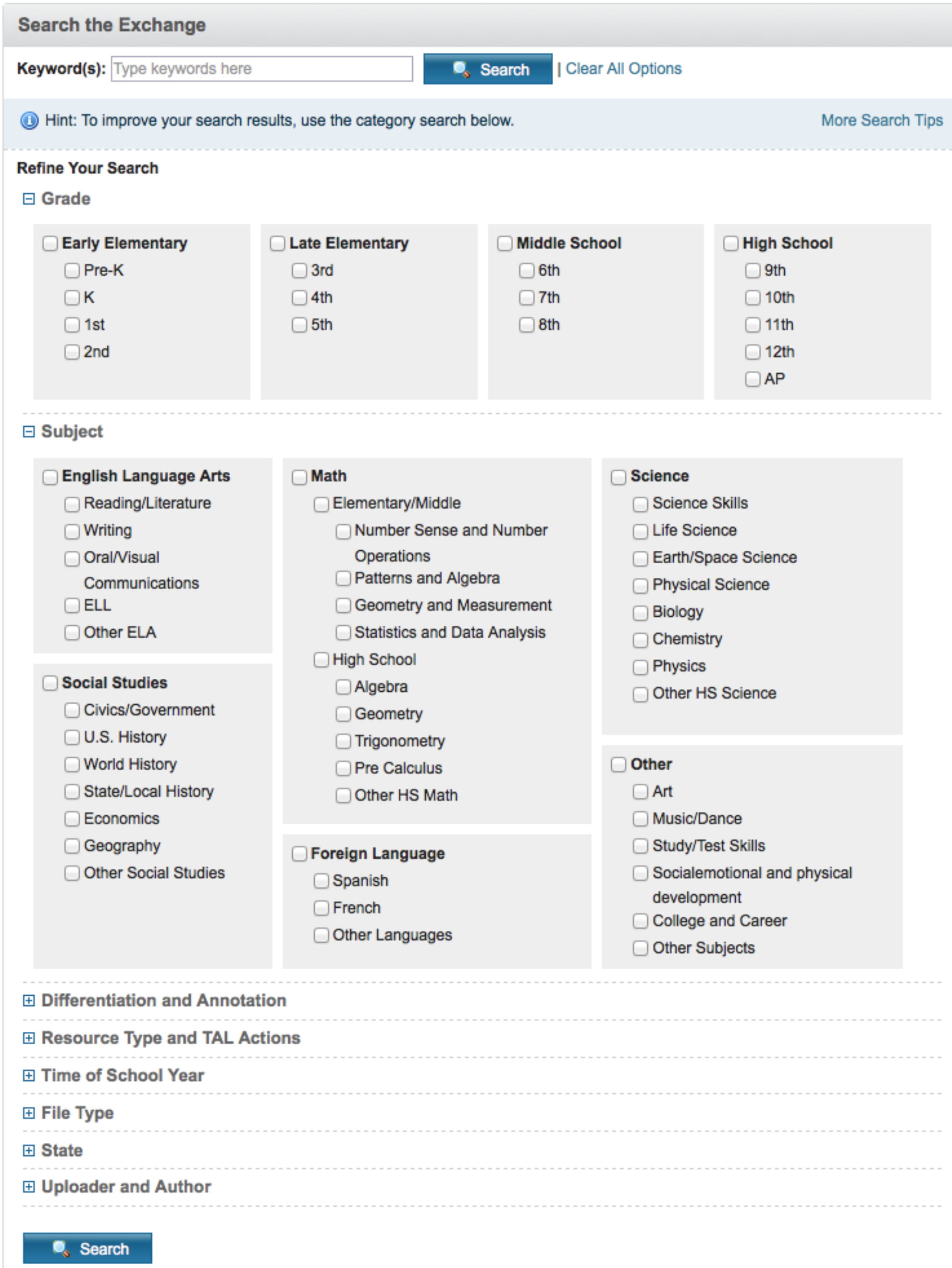


Figure 6: TFANet Search Screen

Each resource in TFANet also has a web page that supplies additional information about the resource (Figure 7). 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 administrators identified the resource as high-quality. For the rest of our analysis, we refer to this information as evaluative metadata.

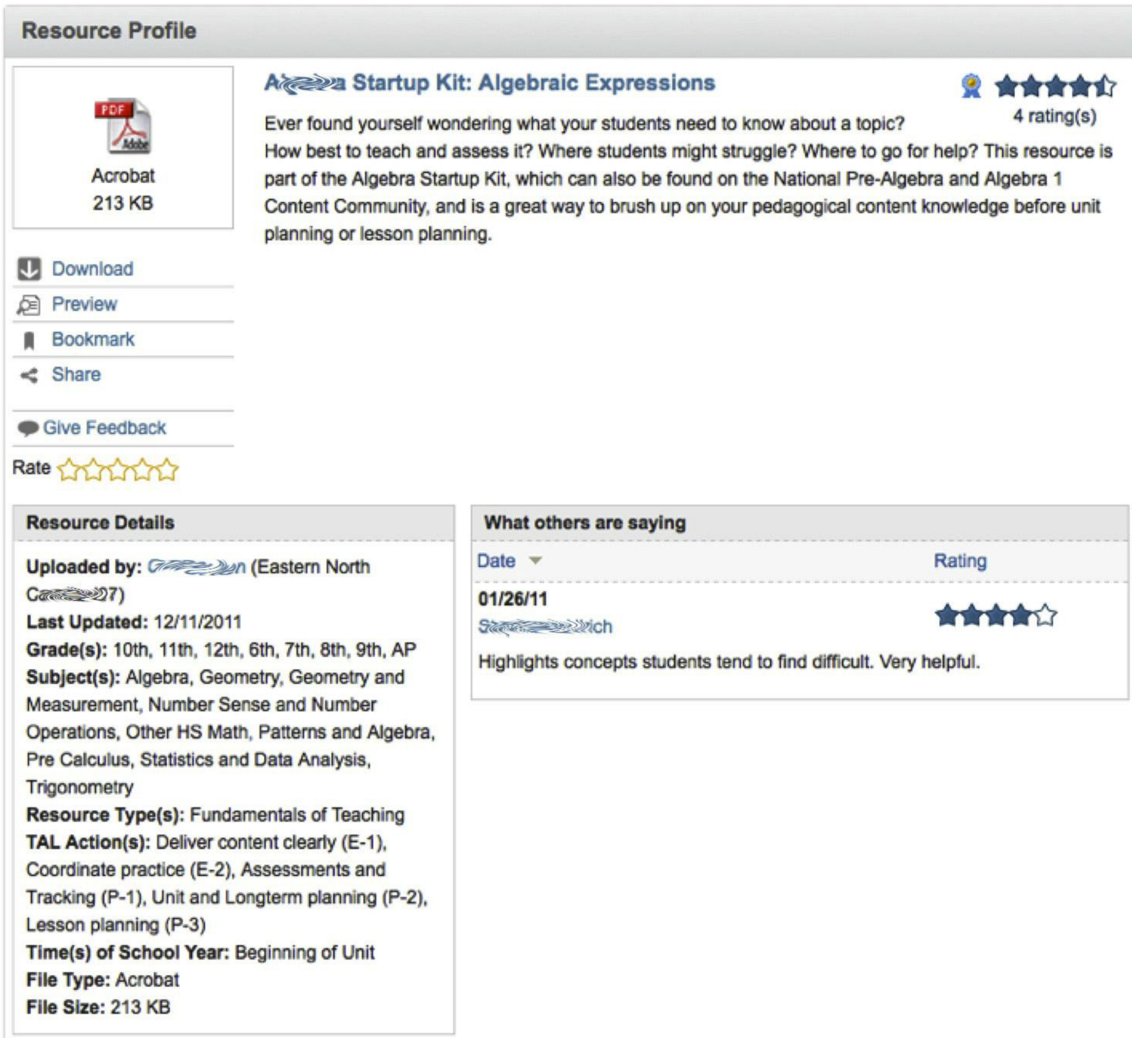


Figure 7: Example TFANet resource page that provides information about the resources, its applicability, and peer ratings and comments.

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), is 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. Included in the database are individual resource names, descriptions, dates of upload, ratings, number of comments, number of ratings, number of downloads, and author's name and region. Descriptions of TFANet's user interface are a result of direct interaction with the system. All of the information provided in this paper has been confirmed for accuracy with TFA administration.

4.4 METHODS

Our underlying research question when studying online teacher resource exchanges is whether the evaluative metadata, in the form visible to users, will influence a user's choice to download particular resources. Would a resource that is highly rated or rated many times predict the eventual number of times the resource will be downloaded?

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 decision heuristic like 'Take the Best' or more complex decision heuristics like 'Franklin's Rule' or 'Linear Regression'?

Here we are seeking to characterize general heuristics of teachers using an online resource exchange to inform overall outcomes from their download decisions. Thus, we are not going to focus in this paper on identifying patterns that are individually true of members, which may not match patterns of the aggregate (Siegler, 1988).

4.4.1 Sample and Variables

The sample includes only *Teach For America* 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 also attracts individuals who are comfortable with self-starting and bootstrapping their own teaching education since TFA's mission is to place teachers in high-needs schools with less training than traditional teacher education programs. TFA members should also be fairly technically savvy since they are all recent college graduates and experienced with the current state of learning technologies used at higher-education institutions. Finally, TFA is a national organization, which means that its members represent a wide variety of organizational and policy contexts. By focusing on teacher-generated resources, our findings can provide insight to what many would consider the prototypical type of teacher who would benefit from participating in a national online resource exchange.

Our analyses examined data from TFANet during a one-month period between Feb. 10 and Mar. 10 of 2011. The decision to examine these data in February was intentional since it was likely to be fairly representative of ‘regular’ teachers’ practice. 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 one-month timeframe, there were 26,959 unique visitors to the resource exchange with 178,626 searches for resources and 79,348 resource downloads.

4.4.1.1 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 needed to select a research method that could account for the variance in downloads attributed to authors.

Consequently, we examined the data using a hierarchical linear model (Raudenbush & Bryk, 2002). Since many authors created more than one resource, we chose a two level model (i.e., the resource data was nested within authors). By choosing a hierarchical linear model (HLM) for our analysis, we could account for variance in downloads attributed to individual authors and produce findings on the resource metadata independent of major author influence (i.e., appropriate standard errors).

4.4.1.2 Change in downloads as dependent variable

The number of downloads is count data and, like most count data, the distribution (i.e., modal count of 0 followed by a highly skewed and tapering distribution) fits a Poisson curve. Furthermore, since we chose to examine downloads over a single month and since each resource we examined was already in the system at the beginning of our time period, each resource had an equal chance of being downloaded. Therefore, we modeled our outcome using a constant-exposure Poisson model, which utilizes a log-link function to account for the distribution of our count data. Furthermore,

In order to examine the effects of evaluative resource metadata on the incidence of resource downloads, we first calculated the increase, if any, of the total number of downloads for each resource between Feb. 10th and Mar. 10th. Despite some of the resource metadata changing over the course of the month, such as the number of ratings or average ratings, a one-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 of changing metadata over the studied download period. The resultant data included 16,863 resources written by 2,149 different authors. The average number of resources for each author was 7.85 with a mode number of resources of 1. The minimum number of resources uploaded by an author was 1 and the maximum was 967. 1300 authors uploaded 2 or more resources.

Although our analysis was performed on 16,863 resources, not all of these resources were rated by individuals using the system. About 34% of all resources were missing a rating, but that was to be expected based on prior work examining resource exchanges (Authors, 2012; Ochoa & Duval, 2008). However, we decided to retain all resources in our analyses since unrated resources are important to include in an analysis of a teacher resource exchange for two reasons. First, they are naturally occurring and as a result must be considered when looking at overall system behaviors. Second, a missing rating might increase the influence of other metadata. For instance, if a user is interested in a resource that is unrated, then this might change the influence of other available metadata. These unrated resources are kept in the regression model that includes number of ratings and levels of ratings through a categorical treatment of rating levels (described below).

4.4.1.3 Evaluative metadata used as independent variables

To determine additional predictor variables besides author, we assumed the role of the downloader and examined the user interface of the TFANet resource exchange (see Figure 7). From this we were able to identify 9 different variables visible to users: description of a resource, date a resource was uploaded, rating (if any), number of ratings, file type, a blue ribbon indicator, and the number of comments. This list of metadata variables is not exhaustive but we believe we have selected the metadata that was the most influential based on its availability to users and ease of understanding (e.g., copyright information, a variable often discussed and prominently displayed in some resource exchange sites, was not presented metadata in this site). We list these variables along with some descriptive statistics in Table 3.

Table 3: Descriptive Statistics of Evaluative Metadata within TFANet

Variable	Mean	St. 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)	.31	.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)	.89	.31	0	1
Blue Ribbon Indicator (1=Yes)	.06	.24	0	1
Number of Comments	0.43	1.09	0	23
Missing Average Rating (1=Yes)	.34	.48	0	1
Average Rating between 1 and 2.99 (1=Yes)	.08	.27	0	1
Average Rating Between 3 and 3.99 (1=Yes)	.20	.40	0	1
Average Rating Between 4 and 4.99 (1=Yes)	.28	.45	0	1
Perfect 5.0 Rating (1=Yes)	.10	.30	0	1

Our independent variables include both resource-level and author-level attributes. The resource-level variables include some surface level characteristics such as the age of the resource in number of years, which was calculated from the date the first resource appeared in the online exchange (July 11, 2009). We included this variable to test if the age of a resource influenced teachers' downloads heuristics. We also included the length of character descriptions, which was a count of the number of characters used in the description. Teachers, when viewing a resource's page on TFANet, are presented with the author's description of the resource. We theorized that a longer resource description could influence a teacher's choice to download the resource.

A third surface level feature of resources included in our models as a binomial variable was whether the file format was easily editable (e.g., Microsoft Word, Microsoft PowerPoint) or not (e.g., PDF, JPEG). Most of the resources (about 89%) were easily editable. This variable was a potentially important predictor given the frequent teacher necessity to edit a resource for specific student needs (Brown & Edelson, 2003). A teacher could prefer to download a resource knowing that they will have to make significant edits to it.

We also examined more substantive features of the resources. For example, experts from TFA had rated some of the resources as being high-quality and had given those resources a "blue ribbon" designation. Thus, we included as a binomial variable whether the resource was considered blue ribbon. A second measure of primary interest to us was the average rating the resource received from teachers on a scale from 1 (low) to 5 (high). 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" (approximately 10 percent of all resources had a perfect rating). By creating a variable for missing ratings we could then in our HLM compare the variance of certain ratings against having 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. These two variables were highly related (r = .749), but we maintained both in analysis since we were interested in exploring the unique contribution of each of these variables since both are visible to teachers when deciding to download a resource. This association between these two variables led us to examine and compare two different statistical models below.

We also included variables at the author-level. Although new resources can be supplied by current core members, alumni, and other various TFANet administrators, for our purposes we removed administrator-generated resources because we wished to focus on the exchange of materials among teachers. We examined two author variables. TFA alumni can continue to participate in TFANet, including uploading and downloading resources. Because an author's years of participation in TFA are part of the displayed metadata for a resource, we created a variable indicating whether the author was a current corps member or TFA alumni. This distinction between current corps members and alumni is also guides how TFA communicates and supports their current and past members. The second author-level variable was the number of years since the author had begun working with TFA as a corps member. This was measured starting in 1990 since the oldest corps member contributing a resource began in that year. About 31% of authors were current corps members and most resources were contributed by current and recent members.

4.4.2 Statistical Analyses

As previously discussed, we structured the TFANet data for analysis so that resources and their associated metadata were nested within individual authors. We employed a series of hierarchical linear models (HLM) to examine our nested data. HLM models were created using the statistical software HLM 7.0 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2011).

Recall from our research questions that we sought to understand whether evaluative metadata were predictive of the number of downloads and whether we could identify decision heuristics teachers used to download resources. To answer these questions we ran three models. 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:

$$\eta_{ti} = \pi_{0i} \qquad [1]$$

$$\pi_{0i} = \beta_{00} + r_{0i} \qquad [2]$$

In equation 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)[11]; and π_{0i} represents the average log odds of the estimated number of downloads for author *i*. At level 2 (equation 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

[11] This transformation occurs automatically within HLM as a result of running a Poisson model.

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 equations 3 and 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'. We added variables at both the resource level and at the author level (see Table 3 for means and standard deviations of these predictors) to understand the effects of each of these variables on the number of downloads. 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:

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

$$\pi_{0i} = \beta_{00} + \beta_{01}(\text{Current TFA Member})_{i} + \beta_{02}(\text{Author's Year in TFA})_{j} + r_{0i} \qquad [4]$$

$$\pi_{1i\ldots ni} = \beta_{10\ldots n0} + r_{qi;\ \text{where } n=10 \text{ and } q = 1,3,4,5}$$

Our third model is an extension of the second model which examines a similar set of predictors. The only difference is that we examined the estimated number of downloads for resources based on 16 categories determined both by the number of ratings and by the average rating (see Table 4 for specification of variables). In this model we retained number of comments as a semi-continuous variable but used the number of ratings to create subgroups with the categorical indicators of average ratings. There were two reasons why we thought this would improve model fit. First, the distribution of number of ratings is skewed since many resources received zero ratings, suggesting categorizing the variable may be preferable. Second, the number of comments and number of ratings were highly correlated (r=.749) making multicollinearity a potential problem when both variables were included simultaneously in our analyses. Additionally, our decision was based on our emerging theory that a large number of low ratings predicted more downloads than a small number of high ratings, as well as our theory about how teachers would respond when viewing evaluative metadata. We note that there is still potential for collinearity after this analysis.

Table 4: Distribution of Resources to Subgroups and Associated Evaluative Metadata

	N	Avg. Num. Comments	% Blue Ribbon	%Easy Edit	Char. Count /100
Missing Rating (Reference Category)	5797	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	1641	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	1920	0.08	06	90	2.86
Two Ratings with an Average Rating Between 4 and 4.99	1085	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	1145	1.96	28	85	4.31
One Rating that is a Perfect 5.0	1056	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

As shown in Table 4, each subgroup contains at least 100 resources. In addition, the subgroups also differ in their averages for select variables that were also included in both models 2 and 3. For example, consider the distribution of the blue ribbon resources. While resources rated 2.99 or below have only about a 1% chance of being designated as blue ribbon, resources with an average perfect rating of 5 have about a 32% chance of being blue ribbon. This is important to consider since estimates of the β's from this model are not independent and can be added to find the predicted log event rate for any resource (given its particular attributes across all of the predictor variables). As we demonstrate, this log event rate can be converted to a predicted event rate through exponentiation.

4.4.2.1 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, Bryk, Cheong, and Congdon, 2004). These models produced a deviance statistic and therefore allowed for model comparisons. 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 (χ^2 = 173.72, df=11, p<.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 examined the empirical Bayes residuals from these models and detected slight over-dispersion in these models. Therefore, we re-examined model 1 and model 2 using penalized quasi-likelihood (PQL) estimation and adjusting for over-dispersion within HLM7.0. In our tables we display the findings obtained from our models using this constant-exposure Poisson model accounting for overdispersion, but it should be noted that the substantive findings remain the same with either estimation procedure.

4.5 RESULTS

4.5.1 Null Model

We discuss findings from our statistical analyses in order beginning with the null model (see Table 5). We obtained two pieces of information from the null model which did not include any predictor variables. First, the estimate of the overall log odds of number of downloads across all resources and authors (β_{00} = 1.166) was converted to an event rate of 3.21 (λ_{ti} = exp(β_{00})) which is very close to the raw mean of downloads (μ = 3.24) presented in Table 3. Second, examining the random effects portion of the model, there was significant variation in the estimated number of downloads between authors (χ^2 = 4,846, df=2,148, p<.001), highlighting the importance of using a nested model that accounts for author 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.

4.5.2 Model 2

Findings from model 2 including all of the variables listed in Table 3 produced a number of significant findings, which challenges the notion that teachers using TFAnet rely on a simple decision heuristic in order to make their decisions about resource downloads. Only three variables were not found to be statistically significant. Both the author's corps year in TFA and having an average rating between 1 and 2.99 (compared with no rating) were not significant, suggesting they may not need to be factored into a design heuristic. Additionally, the number of comments was not significant in model 1. As we described previously, however, the number of comments and number of ratings were highly correlated, so in model 3 we examine these variables with an alternate specification.

Table 5 : 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 Teach for America		.005*** (.006)	.005*** (.006)
Current Corp Member or Alumni		.108*** (.041)	.117*** (.035)
Resource Level Fixed Effects			
Intercept	1.166*** (.015)	.879*** (.064)	.871*** (.065)
Character Count		.018*** (.004)	.016*** (.004)
Upload Date		.038*** (.030)***	.047*** (.017)*
File Format is Editable		.093**** (.035)	.093*** (.034)
Blue Ribbon Indicator		.117*** (.037)	.189*** (.035)
Number of Comments		.005*** (.014)	.064*** (.011)
Number of Ratings		.111*** (.007)	
Average Rating of 1 to 2.99		.009*** (.035)	
One Rating; Range of 1 to 2.99			.074*** (.041)~
Two Ratings; Range of 1 to 2.99			.210*** (.047)
Three Ratings; Range of 1 to 2.99			.376*** (.066)
Four Ratings or More; Range of 1 to 2.99			.656*** (.069)
Average Rating Range of 3 to 3.99		.076** (.027)	
One Rating; Range of 3 to 3.99			.094*** (.028)
Two Ratings; Range of 3 to 3.99			.258*** (.034)
Three Ratings; Range of 3 to 3.99			.519*** (.041)
Four Ratings or More; Range of 3 to 3.99			.792*** (.039)
Average Rating Range of 4 to 4.99		.113*** (.029)	
One Rating; Range of 4 to 4.99			.131*** (.026)
Two Ratings; Range of 4 to 4.99			.392*** (.029)
Three Ratings; Range of 4 to 4.99			.553*** (.035)
Four Ratings or More; Range of 4 to 4.99			.749*** (.032)
Perfect 5.0 Rating		.243*** (.034)	
One Rating; Perfect 5.0			.321*** (.029)
Two Ratings; Perfect 5.0			.479*** (.046)
Three Ratings; Perfect 5.0			.641*** (.064)
Four Ratings; Perfect 5.0			.834*** (.056)
Random Effects-Between Author Variance			
Intercept (τ_{00})	.164	.101***	.098***
Character Count (τ_{10})		.127***	.131***
File Format is Editable (τ_{30})		.099***	.062***
Blue Ribbon Indicator (τ_{40})		.013***	.010~
Number of Comments (τ_{50})		.002***	.002***
% of variance explained‡		38	40

p<.05, ** *p*<.01, ****p*<.001

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

Many variables were significant in predicting the number of downloads. Given the strength and quantity of these many predictors, this model suggests a complex heuristic is involved in teachers' decision making. 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.

We also found a number of other variables 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. In model 2, the estimated number of downloads due to a resource being given blue ribbon status changes from the overall intercept of 2.41 to 2.71[12]. In model 3 the effect of blue ribbon designation increases such that the predicted number of downloads changes from 2.39 to 2.89. However, as described earlier, it is somewhat unrealistic to assume all other resource attributes are '0' since previously we observed that blue ribbon resources were unequally distributed across our subgroups (see Table 4). This further underscores the point that complex decision heuristics are likely applied by teachers when downloading resources since each of these variables does not operate alone in the absence of other resource attributes.

---

[12] These estimates were calculated using the following procedure: It is possible to 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.

Given that the binomial indicator for resources missing a rating was omitted from model 1, the coefficients for the various average 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 average 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 average ratings were not less likely to be downloaded than resources with no ratings. All the higher average rating categories incrementally predicted significant increases in number of downloads. Chi-square statistics revealed that successively higher average ratings were significantly different from the previous level (all of these tests produced significant findings at a level $p<.001$ except for the difference between average rating of 3-3.99 and 4-4.99 which was significant at $p<.05$).

In addition to the influence of the average rating score from users a higher number of ratings was also found to be a significant variable predicting a greater number of downloads. 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). Finally, while the random effects portion of the model indicates significant differences remain between authors, the variables added to the model explained 38% of the variance between authors in the intercept when compared to the null model.

Furthermore, the random effects portion of the model also suggests that teachers' decisions to download resources do not follow a simple heuristic. The presence of random variation among authors for the effects of many of the resource-level predictors suggests there are probably unmeasured facets to teachers' decision-making heuristic. In other words, the fact that the effect of variables such as the count of characters in the description, whether the file is editable, whether it merits a blue ribbon distinction, and the number of comments varies among authors suggests that there are other characteristics (e.g., topic) beyond those captured in model 2 that might also influence teachers' decisions.

4.5.3 Model 3

Findings from model 3 help to further describe the complex influence of both number of ratings and the average 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, suggesting a non-linear prediction for categories describing simultaneously number of ratings and average rating fit the data better than treating number of ratings as semi-continuous. 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 for most surface level and substantive attributes of resources and authors. Additionally, both models explained a similar proportion of the variance between authors (40% for model 3 compared with 38% for model 2). Second, one difference between the two models was in the effects of the number of comments. In model 3, when subgroups were examined, the number of comments was shown to have a positive relationship to the number of downloads. All other things being equal, 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. These findings are demonstrated graphically in Figure 8. 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 indicating the average ratings of resources 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.

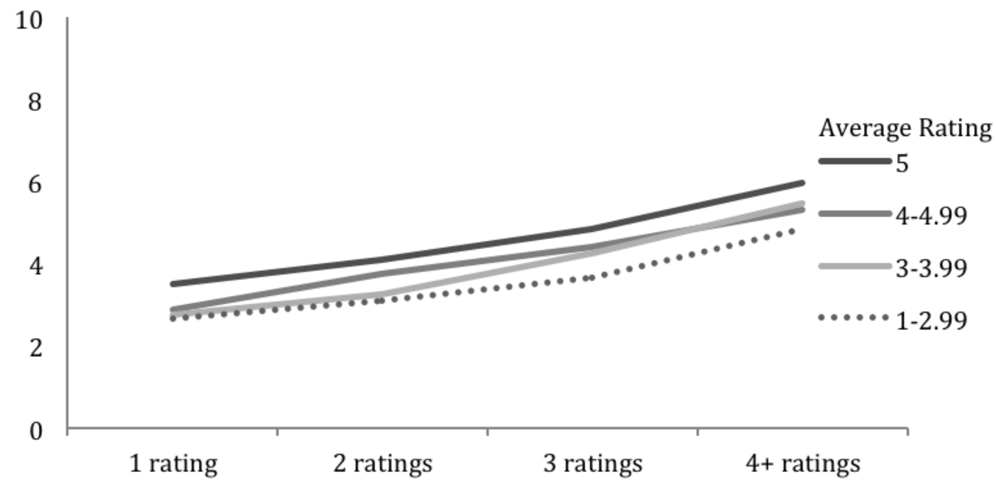
Finally, in the random effects portion of the model, the variability across authors diminishes for the effect of being designated as a blue ribbon resource. Thus, by categorizing resources based on number of ratings and average rating, we significantly reduced the variability between authors for the effect of blue ribbon, suggesting these predictors likely do not operate in isolation from one another and, perhaps, the importance of considering them simultaneously in a complex decision heuristic.

4.5.4 Using Combinations of Regression Coefficients to Predict Different Event Rates

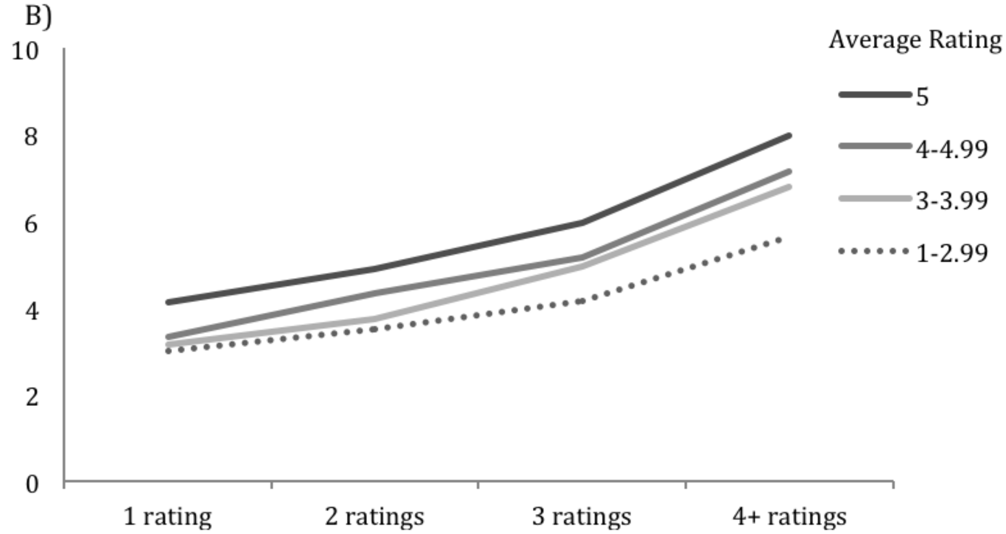
Although the findings presented thus far are logical in how the subgroups are ordered from fewest number of downloads to largest number of downloads, the difference in magnitude is not as great as we might have expected. Counter-factually, the analysis underlying Figure 8a assumes all extraneous variables were constant at ‘0’ when we calculated the expected number of downloads for each subgroup. However, we know from Table 4 that subgroups varied on many variables that we know also predicted number of downloads. In order to account for the effects of these other variables, we recalculated the expected number of downloads using both the subgroup means from Table 3 and the estimated coefficients from Model 3. We plotted the resultant downloads for subgroups in Figure 8b. Although Figure 8b demonstrates a slight 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 lines still largely run parallel to one another.

Figure 8: Estimated number of downloads for subgroups A) with additional variables and B) without additional variables.

A)



B)



4.5.5 The Individual Predictors

Overall, we find that many but not all investigated factors predict download decisions in this online setting: number of ratings, average 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 independent predictors of downloads. Further, we found that date of upload, number of comments, and the year that a resource author was in the corps were not significant independent predictors of downloads.

We were not surprised to find that number of ratings and number of comments are separate predictors of resource downloads since this confirmed the findings of the previous study of TeachersPayTeachers. What was surprising was that our findings confirmed what was only hinted at in that previous study – that a high number of low ratings predicts more downloads than a resource with a low number of high ratings. For example, as illustrated in Figure 8b, resources with the lowest average rating but with 4 or more ratings will be downloaded more than resources with a single perfect rating. In other words, resources with a small amount of high ratings were downloaded less than resources with a large amount of low ratings. This result seemingly runs counterintuitive to the idea that more low ratings would dissuade others from looking at a resource. While we wish to avoid inferences based on causality, this finding seems to indicate that ratings and comments are a significant but limited influence on a resource's popularity.

We were also surprised to find that the current vs. alumni status of a resource's author would predict downloads. Specifically, our findings indicate that current TFA corps members' authored resources are more likely to be downloaded than those from alumni members. This preference for resources from current corps members could be an indication of a current corps member being able to better understand or current corps members' resource needs. This preference could also be an indication of a continual change in teacher resource needs. For example, if a school administration changes curriculum then current teachers could have different resource needs than past teachers. However, if this were true then we would have expected the date of a resource's upload to have a negative prediction on downloads. Instead, 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 and thus it is also logical that it too would predict downloads.

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.

4.6 CONCLUSIONS

Based on our findings, there is little support to suggest that the majority of users rely on a shared simple heuristic based on one piece of evaluative metadata to choose a resource to download. If this simple heuristic were used, we would have expected to find a limited number of influential variables that predicted a high proportion of the variance in downloads.

As an aggregate, teachers were influenced by many factors, following a weighted multiple regression decision making pattern. We cannot conclude that the majority of users each use complex decision heuristics since we did not account for how the relative uniqueness of a resource (e.g., a specific type of resource with a specific subject for a specific grade) impacts a teacher's decision to download. Calculating how the unique qualities of a resource would impact a teacher's desire to download it is beyond the scope of this paper. However, in terms of predicting and changing actions, the data are meaningful to understand the collective behavior of teachers. Just the strong predictive quality of the number of ratings and average ratings alone does cause us to hypothesize that TFANet users are likely relying on a heuristic combining at least two dimensions to select a resource.

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.

4.6.1 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 policy changes such as the acceptance of the Common Core Standards (C. Gewertz, 2012). The number of teacher resource exchanges and the amount of resources available are a testament to this trend. Just in our data alone, the TFANet users downloaded 57,945 resources in a one month period. For the first quarter of 2012, TeachersPayTeachers had over $2,000,000 of teacher resources. For these reasons alone, it is important to continue to understand how teachers select resources online.

We believe the findings in this paper 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.

Our finding of a lack of normative interaction of review positivity and number of reviews (e.g., the difference in downloads for four ratings of a perfect 5 against four ratings with mean of less than 3 was the same as the difference in downloads of only one perfect 5 against 1 rating less than 3) was very surprising and could be used as a means to revisit teacher decision-making. Why would a teacher be more interested in downloading a resource that had many low ratings? One possible reason for this is that the user interfaces of teacher exchanges influences teacher behavior toward this result. Future research can investigate to what degree does Internet communication technologies influence teacher decisions. Another possible reason for the lack of a normative interaction between positivity and reviews is that teachers are not using a consistent heuristic for choosing resources. Developers of knowledge building communities (Scardamalia & Bereiter, 1994) or of resource-based learning environments (J. Hill & M. Hannafin, 2001) can use the findings from this paper as the basis to generate professional development that improves how teachers understand the role of evaluative metadata.

4.6.2 Future Directions

There are many additional analyses that could further unpack how teachers choose resources to download in online teacher resource exchanges. For example, while teacher's could have a very specific topic, subject, or other need when looking for a teaching resource, it is unknown how close of a match to the desired topic, subject, or need would trump the overall quality of the resource for any teacher. Surveys of teacher resource preference would be the optimal way of determining the impact of resource 'uniqueness'.

Another unknown is the influence of existing ratings on how a teacher will rate a resource. Will low ratings influence a teacher to give another low rating? A study to understand this influence would involve having teachers determine if they want to download a resource independent of rating and give a rating to a resource without knowing existing ratings.

It is also unknown if more accurate resource descriptions would predict more or less downloads. A third study would involve comparing more and less accurate resource descriptions for similar resources in order to determine any changes in predicting downloads.

Finally, further analysis is necessary to better understand the influence of comments associated with ratings. There are a wide number of analytic techniques that could be used to explore open-response fields, including latent semantic analysis and linguistic word counts. These methods could reveal how certain types of comments could independently or in conjunction with certain types of ratings predict teachers' interest.

5.0 CONCLUSIONS

In the introduction to this dissertation, I highlighted three challenges to research on online teacher resource exchanges: research should initially focus on determining the current impact of the technology rather than focus on online resource exchange's educative potential, researchers must use a theoretical framework that can provide findings that keep pace with online exchanges' evolution, and research should gain understanding on how teachers use these systems in actual day-to-day practice.

By using data from two different in-use, large-scale, and popular resource exchanges (i.e., 'real-world' systems), my findings are thereby applicable to current use. Findings produced from a laboratory setting or contrived technology would suffer from doubt about their generalizability since no contrived setting can match the highly varied and surprising behavior of large-scale use. Further, my findings are based on current teacher practice rather than a population who is either mimicking teachers (i.e., graduate students) or has minimal teaching experience (i.e., pre-service teachers). Finally, my analyses relied on metrics of metadata that provide specific measures of teacher behavior rather than the potential ambiguity of self-reporting in survey and interviews between behavior and self-reported reported behavior.

5.1 SUMMARY OF FINDINGS

5.1.1 Evaluative Metadata can Predict Measures of Teacher Interest

The initial research question that arose from my literature review was, "What metadata in online teacher resource exchanges predicts teacher interest in a resource?" The findings from my TPT research focused specifically on evaluative metadata and provided strong indication that evaluative metadata as a whole (i.e., comments, ratings, and popularity measures) predicted teacher interest in a resource (i.e., sales). However, while I assumed there was some bidirectional relationship between TPT's evaluative metadata and the sales of a resource, I also uncovered the curious finding that number of ratings and comments were more strongly correlated to sales that the average rating of a resource. This hinted at the possibility that numbers of ratings and comments for a resource were independent predictors of teacher interest in a resource, separate from the interest predicted by a resource's average rating.

To more fully answer what metadata predicts teacher interest, I conducted my second study using a different data set, TFANet, and using different methodology, hierarchical linear models. The TFANet study confirmed and expanded my prior findings; again there was strong indication that evaluative metadata as a whole predicted teacher interest in a resource (i.e., downloads). The conclusion I draw is not that all metadata predicts teacher interest in a resource but that *metadata, in general, is predictive of metrics representing teacher interest*.

The TFANet study also confirmed what was only hinted at in the TPT study; the number of ratings of a resource predicts future downloads independent of the average rating. While the findings do seem to indicate a significant but small interaction effect between the number of ratings and the average rating, there was clear evidence for the surprising outcome that a large number of low ratings predicted more downloads than a small number of high ratings.

While I cannot conclude that evaluative metadata can predict levels of teacher interest for all online teacher resource exchanges, having found these patterns in two different datasets using two different methods of analysis does provide me with confidence to conclude that designers of *online resource exchanges can use evaluative metadata, as a whole, to predict which resources will be popular with teachers*.

5.1.2 Ratings and Quality

A second question from my literature review was, "Are highly rated resources in an online exchange also high quality?" The comparison between expert reviews and teacher ratings from the TPT study led me to conclude that *highly rated resources are not necessarily high-quality resources*. In other words, there is evidence that a curriculum experts' determination of the quality of a resource is different than teachers' determinations.

I was not surprised that expert ratings of resources were much lower than the ratings within TPT considering that curriculum experts have the ability and experience to produce severer evaluations. Flaws in resources that might be unnoticed by teachers could be uncovered by experts because they are more familiar with the latest pedagogical methods. What was surprising was that there was no correlation between expert ratings and teacher ratings. Further research is necessary to determine if this non-correlation is typical for all resource exchanges. The combination of commerce with teacher resources in TPT could be a factor in the non-correlation. If a teacher is spending money on a resource then they might have an entirely different rubric to determine dollar value. Similarly, the novice teachers in TFANet (i.e., first year core members) might also contribute to the non-correlation. The specific needs of a new teacher could result in a heuristic that differs from an expert, who would be evaluating based on a general teacher need (Westerman, 1991). However, I believe the non-correlation does potentially mean that *teachers, when evaluating resources generated by each other, are using different criteria than what an expert would use in determining a resource's level of quality*.

5.1.3 No Single, Simple Heuristic Based on Evaluative Metadata

The final question from my literature review was, "Is evaluative metadata used as the primary factor in selection of resources?" Analysis from my TPT study revealed that even among resources that were purchased, a relative scarcity of resources received evaluative metadata. This alone is an indicator that evaluative metadata could not be the only factor for a teacher selecting a resource in TPT since this type of metadata often does not exist. But even amongst the resources that were rated, the number of ratings was more strongly correlated to sales than the average rating.

The finding that the number of ratings is an independent predictor of teacher interest was confirmed and expanded upon in my TFANet study. For example, a high number of low ratings in TFANet predicts more downloads than a resource with a low number of high ratings. Resources with the lowest average rating but with 4 or more ratings were downloaded more than resources with a single perfect rating. Not only does this finding seem to indicate that ratings and comments are a limited influence on a resource's popularity, but that teachers, as a whole, are not using a single simple heuristic based on evaluative metadata to select a resource. If teachers were using a simple heuristic based on evaluative metadata then I would have expected to find one type of evaluative metadata to be a much stronger predictor than other types of metadata.

Like my prior conclusions, I do not suggest that no teacher uses a single, simple download heuristic based on evaluative metadata, but that having found these patterns in two different datasets using two different methods of analysis does provide me with confidence to conclude that a large plurality of *teachers in online teacher resource exchanges use a download heuristic that includes evaluative metadata*.

5.2 IMPLICATIONS OF FINDINGS

5.2.1 Correcting the Dearth of Evaluative Metadata

If designers of online teacher resource exchanges wish to use evaluative metadata to help teachers select resources then they must address the paucity of ratings. In TPT, the absence of ratings for a large number of resources likely results in limiting the effectiveness of using evaluative metadata as a guide for finding the optimal resource (Walker, et al., 2004). Similarly, in TFANet, even thought the majority of downloaded resources have a rating, those ratings are small in number. This could be a problem if a teacher who relies on ratings or comments would ignore a resource that is optimal but has not been vetted by a number of colleagues. A simple solution would be for operators to incentivize downloaders to rate more resources. For example, in order to download a specific resource, teachers might be required to rate a number of additional resources that have yet to be evaluated. However, given that this is a rather obvious suggestion and that my findings are based on 'real world' data, I expect that many operators find encouraging teachers to leave ratings to be challenging. Even TFANet, a system that caters to a highly motivated population, finds the mode number of ratings for resources at 1.

Even if a teacher is comfortable with only one rating before making a download decision, that solitary rating could have a strong influence on the decision heuristic. However, there are circumstances where society has decided that a single opinion has value in informing heuristics (e.g., movie review blogs, restaurant review websites) (Smith, Menon, & Sivakumar, 2005). 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 A's 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.

5.2.2 Improving the Accuracy of Evaluative Metadata

Regardless of the number of ratings in an online teacher resource exchange, if the designers are hoping to create resource ratings that are the equivalent of expert opinion then they potentially need to alter or change the rubric used by their raters. A high quality rubric can improve inter-rater reliability on assessments (Jonsson & Svingby, 2007). The lack of correlation between the existing ratings in TPT and expert opinions could be the result of a rubric that allows raters to determine their own criteria of quality. Although I have not yet compared TFANet ratings and expert opinion, I expect to find no correlation partially because the rating rubric (a simple choice of 1 to 5 stars) is very simplistic and subject to interpretation.

But creating a single reliable rubric for all types of teacher resources in different content areas could be a task that is extremely difficult (Jonsson & Svingby, 2007). Beyond just creating an improved general rubric for evaluating teacher resources, operators of teacher resource exchanges might consider using different rubrics for the wide variety of different types of resources often available in online exchanges. For example, an essay is assessed differently than a multiple-choice test yet both are often assigned a letter grade. An online teacher resource exchange could provide raters with different rubrics for lesson plans, worksheets, or multimedia rather than attempting to create a single rubric that might not fairly evaluate the strengths and weaknesses of the different types of resources. For example, an English resource designed for grammar instruction might deserve a high rating only if it's content is grammatically accurate. However, a math resource might deserve a high rating if it provides teachers an opportunity to engage students with forming mathematical procedures that connect to fundamental mathematical concepts, even if it is grammatically flawed (Remillard, 2012).

5.2.3 Creating Teacher Professional Development for Online Exchanges

Teacher professional development based on a specific technology is not unusual to teacher education (Lawless & Pellegrino, 2007). However, the findings from my dissertation suggest that teachers could benefit from professional development targeted at how these system integrate into teacher practice. Not only are student learning outcomes potentially a result of teacher selection of an optimal resource, selection of non-optimal resources will result in wasted time and require the teacher to generate the resource from scratch. The determination that there is little evidence of teachers using a single download heuristic suggests that teachers are using a variety of heuristics, which means there is a possibility that bad or non-optimal heuristics are used. Professional development on this topic could build a teacher's knowledge on balancing what the metadata says about a resource with whether the teacher should examine the resource and strengthen skills for selecting optimal resources.

For example, a teacher that is introducing a topic to her students for the first time might be looking for a specific type of resource best suited for that learning scenario. Searching an online teacher resource exchange might reveal dozens of resources that are approximate matches for the teacher's resource needs. With the ability to develop a reliable heuristic (Todd & Gigerenzer, 2000), a teacher could create a threshold that eliminates many inappropriate resources and still preserve an appropriate selection of potential resources. If the topic is common (e.g., the American revolution) then a teacher might need to only examine the highest rated resources while a less common topic (e.g., the Continental Army's invasion of Canada) might require expanding a rating threshold.

Online teacher resource exchanges also fit the model of resource-based learning environment for teachers (J. R. Hill & M. J. 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: resources that 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 tweaks, the process of uploading a resource, receiving ratings on authored resources 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.

5.2.4 Design Principles

For designers of online teacher resource exchanges, there are specific principles that emerge as a result of my dissertation findings. Achieving a desired number of ratings per resource cannot be assumed to be a natural function of including teacher ratings. Consequently, a key design principle is to *Monitor and Promote Ratings*. Designers must monitor the number of ratings for resources and likely adjust the technical infrastructure to promote (i.e., increase) ratings for resources that have teacher interest (i.e., downloads or sales) but a dearth of ratings.

Additionally, if designers hope to have high correlation between the ratings generated by teachers and what an expert would determine, then careful consideration and measurement are necessary to promoting this outcome. A second design principle is *Design Rating Rubrics for Accuracy*. The rubric used in the rating process can have due influence on the rating generated

(Jonsson & Svingby, 2007). Consequently, if ratings do not correlate with expert opinion, a designer could exert influence through the rubric (e.g., specifying categories for analysis, requiring justification) in order to influence the rater toward the desired accuracy. To increase accuracy, the rubric would need to support increases of reliability (i.e., similar ratings from different raters) and validity (i.e., increased correlation with experts), while not hampering related behaviors (e.g., the number of ratings per rater).

5.3 FUTURE QUESTIONS AND LIMITATIONS

It is unlikely that teachers will revert from using the Internet to find resources back to using more traditional methods; instead it is more likely that the movement towards using online teacher resource exchanges will increase. Online teacher resource exchanges provide an unparalleled number of resources to anyone with an Internet connection and metrics measuring use indicate a growing popularity.

The findings presented in my dissertation are a first step in understanding online teacher resource exchanges. Additional research is necessary to better understand how these systems are impacting teaching.

5.3.1 Generalizability

A notable limitation of my dissertation is that my conclusions have limited generalizability. The two online exchanges studied in my dissertation were created for specific purposes and populations. TPT is a for-profit enterprise that encourages teachers to earn money from selling products they have created for other teachers. TFANet was designed to support TFA's members during their two years of service. Other resource exchanges will have different purposes which could result in different patterns of use (Diekema & Olsen, 2011).

Although both systems are designed for teachers, the participants in either system are potentially unique and not accurate representatives of other teachers for the purposes of understanding resource exchanges. Participants in TPT are teachers are likely comfortable with monetization of teaching but capitalizing teacher work might be morally objectionable to a significant number of teachers. Participation in TPT could also be prohibited in a teacher's contract, with the employer owning the rights to the teacher's resources. For TFANet, the selection process of TFA means that a significant number of participants received only minimal teacher training and have limited teaching experience.

Fortunately, there are many currently operating online resource exchanges designed for other groups of teachers. As previously mentioned, the National Education Association, the Public Broadcasting System, and the Department of Education of Ohio all provide a variety of lesson plans online targeted for teachers and addressing a specific teaching need. Because these systems automatically generate relevant data in their normal operation, confirmation of my dissertation findings is possible by conducting similar investigations with additional datasets.

5.3.2 Is There No Such Thing as Bad Publicity?

The lack of normative interaction of review positivity and number of reviews was very surprising. Why would a teacher be more interested in downloading a resource that had many low ratings? One possible cause could be that user interfaces influences teacher behavior (Li, 2010). For example, the TPT default view lists the most reviewed items in descending order.

Many Internet search engines rank results starting with the most accurate or likely result. A teacher used to similar ranking of search results might assume the same for TPT and choose their purchases accordingly, resulting in an artificial inflation of number of ratings predicting sales. Additionally, the placement of evaluative metadata within the visual design of an exchange could affect teachers' download heuristics. Metadata that is prominent could potentially have a larger impact than metadata this is graphically minimized. Qualitative interviews of teacher along with experiment on user interfaces might be able to determine this influence. Additionally, experimentation with user interfaces might allow for direct comparison of optimal technological designs.

Another cause for a teacher to be more interested in downloading a resource that had many low ratings could be based on the paucity of evaluative metadata for some resources. A teacher might seek some initial way to parse a large set of resources into a smaller pool (Smith, et al., 2005). By setting a minimum rated threshold, a teacher could have a more comfortable resource pool to choose from, even if those ratings are negative. Additionally, different teachers would have different thresholds of ratings and average ratings. Qualitative interviews and surveys of teacher participants will provide further insight into teachers' resource heuristics and potentially determine if there is no such thing as bad publicity in online teacher resource exchanges.

5.3.3 Teacher Communities

The success of an online resource exchange is dependent on a community of users who identify with the goals of the system, provide the resources and ratings necessary for operation, and download the resource for their teaching needs. Without participation, an exchange would become a static online repository and risks becoming irrelevant (Mimi Recker, et al., 2005). Future research should determine what type of community, if any, are prevalent in online exchanges in order to make more informed design recommendations. For example, the incentives for evaluating resources might differ based on the type of community bonds manifested. An identity-based community might be more incentivized by increasing specific teacher visibility within the system (e.g., teacher of the week) while a social-based community might be more incentivized by prominent displays of contribution (e.g., clear display of number of ratings and their usefulness).

5.3.4 Additional Analyses on TPT and TFANet

Teachers could have a very specific resource need that would trump the overall quality of the resource for any teacher (Diekema & Olsen, 2011).Future statistical models of TPT or TFANet could explain more variation in interest by including a variable that measures the uniqueness of a resource. For example, a resource that is badly constructed could still be highly downloaded only because it is the only resource of its kind.

Additional analysis is also necessary to better understand the influence of comments associated with ratings. There are a wide number of analytic techniques that could be used to explore open-response fields, including latent semantic analysis and linguistic word counts (Paletz & Schunn, 2009). These methods could reveal how certain types of comments could independently or in conjunction with certain types of ratings predict teachers' interest.

Another analysis should be the examination of different users in online resource exchanges. My analyses have been focused on unpacking understanding of online exchanges systemically. However, there is a limitation in this approach because of the variation in the teachers who participate in these systems. Teachers with different backgrounds and experience could have different approaches to using online exchanges (Westerman, 1991). For example, a novice science teacher might be interested in downloading many different types of resources where an experienced math teacher might focus on providing a small number of accurate ratings to favorite resources. To better understand how different users engage in online exchanges, analysis of novice users, expert users, high downloaders, high uploaders, and high raters is necessary.

Implicit in the distinction between novice and expert users is whether online exchanges experience significant change over time. This could be examined from the user perspective; after how much participation in an exchange does a novice user become an expert? Additionally, there is also the possibility of systemic change over time. An online teacher resource exchange might engender different levels of participation expending on the system's maturity. For example, a new exchange that has a relatively small number of resources might motivate participants to upload large quantities of their own work while a mature system with many resources might discourage uploading. To uncover whether change over time exists in exchanges I must engage in further analyses that focuses on growth patterns of resource and teacher participation.

BIBLIOGRAPHY

Ackerman, M. S. (2000). The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human-Computer Interaction, 15*(2), 179-203.

Anido, L. E., Fernandez, M. J., Caeiro, M., Santos, J. M., Rodriguez, J. S., & Llamas, M. (2002). Educational metadata and brokerage for learning resources. *Computers & Education, 38*(4), 351-374.

Authors, 2012.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students "game the system." Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Vienna, Austria.

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.

Ball, D., & Cohen, D. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco, CA: Jossey-Bass Inc.

Banister, S. (2010). Integrating the iPod Touch in K–12 education: Visions and vices. *Computers in the Schools, 27*(2), 121-131.

Barker, L. J. (2009, June). Science teachers' use of online resources and the Digital Library for Earth System Education. In *Proceedings of the 9th ACM/IEEE-CS joint conference on digital libraries* (pp. 1-10). Austin, TX: ACM

Bethard, S., Wetzer, P., Butcher, K., Martin, J. H., & Sumner, T. (2009, June). Automatically characterizing resource quality for educational digital libraries. In *Proceedings of the 9th ACM/IEEE-CS joint conference on digital libraries* (pp. 221-230). Austin, TX: ACM.

Bohl, O., Schellhase, J., Sengler, R., & Winand, U. (2002, December). The Sharable Content Object Reference Model (SCORM) - a critical review. In *Proceedings of the International Conference on Computers in Education* (pp. 950-951). IEEE Computer Society.

Borko, H., & Putnam, R. (1995). Expanding a teacher's knowledge base: A cognitive psychological perspective on professional development. In T. Guskey & M. Huberman (Eds.), *Professional development in education: New paradigms and practices* (pp. 35-65). New York, NY: Teachers College Press.

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?* (Design Brief, LeTUS Report Series). Evanston, IL: The Center for Learning Technologies in Urban Schools.

Bruce, T. R., & Hillmann, D. (2004). The continuum of metadata quality: Defining, expressing, exploiting. In D. Hillmann, & E. Westbrooks (Eds.), *Metadata in practice* (pp. 238-256). Chicago, IL: ALA Editions.

Campbell, C., & Evans, J. A. (2000). Investigation of preservice teachers' classroom assessment practices during student teaching. *The Journal of Educational Research, 93*(6), 350-355. doi: 10.1080/00220670009598729

Carroll, J., Choo, C., Dunlap, D., Isenhour, P., Kerr, S., MacLean, A., & Rossen, M. B. (2003). Knowledge management support for teachers. *Educational Technology Research and Development, 51*(4), 42-64. doi: 10.1007/bf02504543

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM, 50*(4), 31-34.

Cheng, R., & Vassileva, J. (2006). Design and evaluation of an adaptive incentive mechanism for sustained educational online communities. *User Modeling and User-Adapted Interaction, 16*(3), 321-348.

Chiao-Fang, H., Khabiri, E., & Caverlee, J. (2009, August 29-31). Ranking comments on the social web. In *International Conference on Computational Science and Engineering* (Vol. 4; pp. 90-97). IEEE. doi: 10.1109/CSE.2009.109

Clark, C. M. (1988). Asking the right questions about teacher preparation: Contributions of research on teacher thinking. *Educational Researcher, 17*(2), 5-12. doi: 10.3102/0013189x017002005

Collins, A., & Halverson, R. (2009). *Rethinking education in the age of technology*. New York, NY: Teachers College Press.

Collins, J., & Weiner, S. (2010). Proposal for the creation of a subdiscipline: Education informatics. *Teachers College Record, 112*(10), 2523-2536.

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.

Cuban, L. (1993). Computers meet classroom: Classroom wins. *The Teachers College Record, 95*(2), 185-210.

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.

de Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education, 46*(3), 249-264.

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.

Drexler, W., Baralt, A., & Dawson, K. (2008). The Teach Web 2.0 Consortium: A tool to promote educational social networking and Web 2.0 use among educators. *Educational Media International, 45*(4), 271-283.

Duval, E., Forte, E., Cardinaels, K., Verhoeven, B., Durm, R. V., Hendrikx, K., Wentland Forte, M. Ebel, N. Macowicz, M., Warkentyne, K., & Haenni, F. (2001). The Ariadne knowledge pool system. *Communications of the ACM, 44*(5), 72-78.

El-Hani, C., & Greca, I. (2012). ComPratica: A virtual community of practice for promoting biology teachers' professional development in Brazil. *Research in Science Education*, 1-33. doi: 10.1007/s11165-012-9306-1

Fallows, D., Rainie, H., & Mudd, G. (2004). *The popularity and importance of search engines* (Tech. Rep.). The Pew Internet & American Life Project. Retrieved July 15, 2013 from http://www.pewinternet.org/Reports/2004/Data-Memo-on-Search-Engines/Data-Memo.aspx?view=all

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.

Franklin, C. A. (2007). Factors that influence elementary teachers use of computers. *Journal of Technology and Teacher Education, 15*(2), 267-293.

Franklin, C. A. (2005). Factors that influence elementary teachers' use of computers. Paper presented at the annual American Educational Research Association meeting, Montréal, Québec.

Fu, H., & Hawkes, M. (2010). Technology-supported peer assessment as a means for pre-service teachers learning. Paper presented at the Workshop on Computer Supported Peer Review in Education at the Tenth International Conference on Intelligent Tutoring Systems, Pittsburgh, PA.

Gabriel, R. P., Northrop, L., Schmidt, D. C., & Sullivan, K. (2006). Ultra-large-scale systems. In *Companion to the 21st ACM SIGPLAN symposium on object-oriented programming systems, languages, and applications* (pp. 632-634). Portland, OR: ACM.

Gewertz, C. (2012, Feb. 29). Educators in search of common-core resources. *Education Week, 31*(22), 1-12.

Goncalves, M. A., Moreira, B. L., Fox, E. A., & Watson, L. T. (2007). “What is a good digital library?" - A quality model for digital libraries. *Information Processing & Management, 43*(5), 1416-1437.

Gonzalez-Barbone, V., & Anido-Rifon, L. (2008). Creating the first SCORM object. *Computers & Education, 51*(4), 1634-1647.

Goodson, I., & Cole, A. (1994). Exploring the teacher. *Teacher Education Quarterly, 21*(1), 85-105.

Gray, L., Thomas, N., & Lewis, L. (2010). *Teachers' use of educational technology in U.S. public schools: 2009*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Greenhow, C., Robelia, B., & Hughes, J. (2009). Learning, teaching, and scholarship in a digital age - Web 2.0 and classroom research: What path should we take now? *Educational Researcher, 38*(4), 246-259.

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, Germany: Springer.

Guskey, T., & Yoon, K. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495-500.

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. In *Proceedings of the Society for Information Technology & Teacher Education International Conference 2009* (pp. 4087-4095). Chesapeake, VA: AACE.

Hawley, W., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 127-150). San Francisco, CA: Jossey-Bass Inc.

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., & Hannafin, M. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development, 49*(3), 37-52. doi: 10.1007/bf02504914

Jaen, X., Bohigas, X., & Novell, M. (2007). The need for virtual information managers in education. *Computers & Education, 49*(2), 254-268.

Jeng, J. (2005). What is usability in the context of the digital library and how can it be measured? *Information Technology and Libraries, 24*(2), 46-56.

Jenkins, H., & Squire, K. (2004). Harnessing the power of games in education. *Insight, 3*(1), 5-33.

Jenkins, H., Clinton, K., Purushotma, R., Weigel, M., & Robison, A. J. (2007). Confronting the challenges of participatory culture: Media education for the 21st century. MacArthur Foundation Publication. Retrieved July 15, 2013 from http://fall2010compositions.pbworks.com/f/JENKINS_WHITE_PAPER.pdf

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144.

Kent, T. W., & McNergney, R. F. (1999). *Will technology really change education? From blackboard to web.* Thousand Oaks, CA: Corwin Press, Inc.

Kesidou, S., & Roseman, J. E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching, 39*(6), 522-549.

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). Springer Berlin Heidelberg.

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.

Leat, D., & Higgins, S. (2002). The role of powerful pedagogical strategies in curriculum development. *Curriculum Journal, 13*(1), 71-85.

Levy, P., Ford, N., Foster, J., Madden, A., Miller, D., Nunes, M. B., McPherson, M., & Webber, S. (2003). Educational informatics: An emerging research agenda. *Journal of Information Science, 29*(4), 298-310.

Lewis, C., Perry, R., & Murata, A. (2006). How should research contribute to instructional improvement? The case of lesson study. *Educational Researcher, 35*(3), 3-14. doi: 10.3102/0013189x035003003

Li, J. Z. (2010, September 17-19). Quality, evaluation and recommendation for learning object. In *Proceedings of the International Conference on Educational and Information Technology (ICEIT)*, Chongqing, China. doi: 10.1109/ICEIT.2010.5607654

Maloch, B., Flint, A. S., Eldridge, D., Harmon, J., Loven, R., Fine, J. C., Bryant-Shanklin, M., & Martinez, M. (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., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011). Recommender systems in technology enhanced learning. In *Recommender systems handbook* (pp. 387-415). Springer US.

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. doi: 10.1111/j.1365-2729.2010.00362.x

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.

Mehlinger, H. D. (1996). School reform in the information age. *Phi Delta Kappan, 77*(6), 400-407.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication, 60*(3), 413-439.

Morrison, J. (2010, June). 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.

Najjar, J., Klerkx, J., Vuorikari, R., & Duval, E. (2005). Finding appropriate learning objects: An empirical evaluation. In A. Rauber, S. Christodoulakis, & A. Tjoa (Eds.), *Research and advanced technology for digital libraries* (Vol. 3652, pp. 323-335): Springer Berlin Heidelberg.

National Teaching Realities Survey. (2010). Culver City, CA: Kelton Research. Retrieved July 15, 2013 from //multivu.prnewswire.com/mnr/officemax/43900/

Nesbit, J., & Li, J. (2004). Web-based tools for learning object evaluation. In *Proceedings of international conference on Education and Information Systems: Technologies and Applications* (pp. 21-25). Orlando, FL.

Nesbit, J., Belfer, K., & Vargo, J. (2002). A convergent participation model for evaluation of learning objects. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie, 28*(3), 105-120.

Nesbit, J., Li, J., & Leacock, T. (2006). Web-based tools for collaborative evaluation of learning resources. *Journal on Systemics, Cybernetics and Informatics, 3*(5), 102-112.

Nielsen, J. (2006). Participation inequality: encouraging more users to contribute. Retrieved July 15, 2013 from http://www.useit.com/alertbox/participation_inequality.html

Niemiec, R. P., & Walberg, H. J. (1992). The effects of computers on learning. *International Journal of Educational Research, 17*(1), 99-108.

Ochoa, X., & Duval, E. (2008). Relevance ranking metrics for learning objects. In *IEEE Transactions on Learning Technologies, 1*(1), 34-48. doi: 10.1109/TLT.2008.1

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): Springer Berlin Heidelberg.

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. doi: 10.1007/s11423-006-9012-7

Paletz, S. B. F., & Schunn, C. D. (2009, October 2-3). A new metric for assessing group level participation in fluid teams. Paper presented at the Conference on Science and Innovation Policy conference, Atlanta, GA.

Pennebaker, J., Francis, M., & Booth, R. (2007). Linguistic Inquiry and Word Count (LIWC2007) [Computer software]. Austin, TX: University of Texas.

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

Peterson, P. L., Marx, R. W., & Clark, C. M. (1978). Teacher planning, teacher behavior, and student achievement. *American Educational Research Journal, 15*(3), 417-432.

Plodzien, J., Stemposz, E., & Stasiecka, A. (2006). An approach to the quality and reusability of metadata specifications for e-learning objects. *Online Information Review, 30*(3), 238-251.

Poston, R. S., & Speier, C. (2005). Effective use of knowledge management systems: A process model of content ratings and credibility indicators. *MIS Quarterly, 29*(2), 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, UK: Cambridge University Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications, Inc.

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

Recker, M., & Wiley, D. (2001). A non-authoritative educational metadata ontology for filtering and recommending learning objects. *Interactive Learning Environments, 9*(3), 255-271. doi: 10.1076/ilee.9.3.255.3568

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

Recker, M., Dorward, J., Dawson, D., Halioris, S., Liu, Y., Mao, X., Palmer, B., & Park, J. (2005). You can lead a horse to water: teacher development and use of digital library resources. In *Proceedings of the 5th ACM/IEEE-CS joint conference on digital libraries* (pp. 1-8). Denver, CO: ACM.

Recker, M., Leary, H., Walker, A., Diekema, A., Wetzler, P., Sumner, T., & Martin, J. (2011, April). Modeling teacher ratings of online resources: A human-machine approach to quality. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, LA.

Recker, M., Sellers, L., & Ye, L. (2012). Investigating impacts of technology-related teacher professional development designs: A comparative case study. The Instructional Architect Research Group. Retrieved July 15, 2013 from http://digitalcommons.usu.edu/iagroup/5

Recker, M., Walker, A., & Lawless, K. (2003). What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. *Instructional Science, 31*(4), 299-316. doi: 10.1023/a:1024686010318

Recker, M., Walker, A., Giersch, S., Mao, X., Halioris, S., Palmer, B., Johnson, D., Leary, H., & Robertshaw, M. B. (2007). A study of teachers' use of online learning resources to design classroom activities. *New Review of Hypermedia and Multimedia, 13*(2), 117-134.

Reich, J., Murnane, R., & Willett, J. (2012). The state of wiki usage in U.S. K–12 schools: Leveraging web 2.0 data warehouses to assess quality and equity in online learning environments. *Educational Researcher, 41*(1), 7-15.

Reiser, R. A. (2001). A history of instructional design and technology: Part I: A history of instructional media. *Educational Technology Research and Development, 49*(1), 53-64.

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, Germany: Springer.

Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies, 28*(3), 377-408.

Rotherham, A. J. (2012, September 20). Should teachers be allowed to sell their lesson plans? Time. Retrieved on July 15, 2013 from http://ideas.time.com/2012/09/20/should-teachers-be-allowed-to-sell-their-lesson-plans/

Russell, M., Bebell, D., O'Dwyer, L., & O'Connor, K. (2003). Examining teacher technology use: Implications for preservice and inservice teacher preparation. *Journal of Teacher Education, 54*(4), 297-310.

Ruys, I., Keer, H. V., & Aelterman, A. (2012). Examining pre-service teacher competence in lesson planning pertaining to collaborative learning. *Journal of Curriculum Studies, 44*(3), 349-379. doi: 10.1080/00220272.2012.675355

Salinas, C., Bellows, M. E., & Liaw, H. L. (2011). Preservice social studies teachers' historical thinking and digitized primary sources: What they use and why. *Contemporary Issues in Technology and Teacher Education, 11*(2), 184-204.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences, 3*(3), 265-283.

Schafer, J. B., Konstan, J., & Riedi, J. (1999). Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on electronic commerce* (pp. 158-166). Denver, CO: ACM.

Schlager, M., Fusco, J., & Schank, P. (2002). Evolution of an on-line education community of practice. In K. A. Renninger & W. Shumar (Eds.), *Building virtual communities: Learning and change in cyberspace* (pp. 129-158). Cambridge, UK: Cambridge University Press.

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/

Schwartz, B., Ward, A., Monterosso, J., Lyubomirsky, S., White, K., & Lehman, D. R. (2002). Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology, 83*(5), 1178-1197.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology*: General, *117*(3), 258-275. doi: 10.1037/0096-3445.117.3.258

Smith, D., Menon, S., & Sivakumar, K. (2005). Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing, 19*(3), 15-37.

Sparks, D., & Loucks-Horsley, S. (1989). Five models of staff development for teachers. *Journal of Staff Development, 10*(4), 40-57.

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.

Srivastava, P. (2012, Jan. 3-5). Educational informatics: An era in education. Paper presented at the IEEE International Conference on Technology Enhanced Education, Amritapuri, India.

Standifird, S. S. (2001). Reputation and e-commerce: eBay auctions and the asymmetrical impact of positive and negative ratings. *Journal of Management, 27*(3), 279-295.

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. In *Proceedings of the 3rd ACM/IEEE-CS joint conference on digital libraries* (pp. 269–279). Houston, TX: IEEE Computer Society.

Suthers, D. D. (2001). Evaluating the learning object metadata for K-12 educational resources. In *Proceedings of the IEEE international conference on advanced learning technologies* (pp. 371-374), Madison, WI: IEEE.

Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning, 21*(3), 217-228.

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: Educators use web sites and social networks to share lesson plans. *Education Next, 11*(3), 78-79.

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.

Walker, A., Recker, M., Lawless, K., & Wiley, D. (2004). Collaborative information filtering: A review and an educational application. *International Journal of Artificial Intelligence in Education, 14*(1), 3-28.

Wang, H.-C., & Hsu, C.-W. (2006). Teaching-material design center: An ontology-based system for customizing reusable e-materials. *Computers & Education, 46*(4), 458-470.

Watson, G. (2006). Technology professional development: long-term effects on teacher self-efficacy. *Journal of Technology and Teacher Education, 14*(1), 151-166.

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

Westerman, D. A. (1991). Expert and novice teacher decision making. *Journal of Teacher Education, 42*(4), 292-305.

Wu, B., Xu, W. X., & Ge, J. (2012). Recommendation research in e-learning. *Applied Mechanics and Materials, 121*, 3067-3071. doi: 10.4028/www.scientific.net/AMM.121-126.3067

Yinger, R. (1979). Routines in teacher planning. *Theory Into Practice, 18*(3), 163-169.

Yinger, R. (1980). A study of teacher planning. *The Elementary School Journal, 80*(3), 107-127.