2012-12-14

Measuring the Technical Difficulty in Reusing Open Educational Resources with the ALMS Analysis Framework

Seth Michael Gurell
Brigham Young University - Provo

Measuring Technical Difficulty in Reusing Open Educational

Resources with the ALMS Analysis Framework

Seth M. Gurell

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

David Wiley, Chair
Rick West
Randall Davies
Charles Graham
Andrew Gibbons

Department of Instructional Psychology and Technology

Brigham Young University

December 2012

ABSTRACT

Measuring Technical Difficulty in Reusing Open Educational
Resources with the ALMS Analysis Framework

Seth M. Gurell
Department of Instructional Psychology and Technology
Doctor of Philosophy

The Open Educational Resources (OER) movement was started roughly ten years old (Wiley & Gurell, 2009). Since that time thousands of resources have been produced. Though these resources have been used both for classroom development and for the autodidact, the development of OER was not without problems. Incompatibility between Creative Commons licenses has made revising and remixing two resources difficult, if not impossible (Linksvayer, 2006). Tools to help educators find appropriate educational resources have been necessary but are still nascent. Educators' perceived quality issues have also hampered adoption (Wiley & Gurell, 2009). The result is that resources were only being minimally reused (Wiley, 2009). One possible reason observed for the limited reuse was the barrier of technology. Some resources were easier to view, revise and remix from a technical perspective than others.

Hilton, Wiley, Stein, and Johnson (2010) created the ALMS analysis framework to assess the technical openness of an OER. Although the ALMS framework allowed for an assessment of OER, no pilot instrument was reported in the Hilton et al. (2010) article. The framework has not been tested because there is no known rubric with which measurement can occur. Consequently, Hilton et al.'s framework needed to be further developed and tested against a range of open educational resources.

This dissertation examined the ALMS analysis, which was previously only a concept, in order to create a concrete framework with sufficient detail and documentation for comparisons to be made among OERs. The rubric was further refined through a Delphi study consisting of experts in the field of OER (n=5). A sample of OERs (n=27) rated by a small group (4) was conducted to determine inter-rater reliability. Intra-class correlation indicated moderate agreement (ICC(2,1) =.655, df=376, 95% CI [.609, .699]). Findings suggested that the degree of technical difficulty in reusing OERs can be measured in somewhat reliable manner. These findings may be insightful in developing policies and practices regarding OER development.

Keywords: OER, Open Educational Resources, ALMS, Delphi study

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Background

In the United States, as elsewhere, it has been illegal to copy or share content, including educational material, or any other kind of creative works without express permission of the copyright holder (United States Copyright Office, 2008). Consequently, individuals or organizations that wish to provide free educational materials for teachers and students to use in a variety of ways must include copyright statements on their websites or methods of granting these permissions. Copyright has varied significantly from country to country (Grossman & Lei, 2002; Marron, 2000; Correa, 2000), though international laws and treaties are minimizing these differences (Ginsburg, 2000). Copyright law has influenced a number of fields including education. One way it affected education is by creating barriers to reusing educational materials.

The concept of revising educational materials was started at least forty years ago (e.g, Gerard, 1969). Wayne Hodgins was believed to have coined the term *learning objects* (LO) in 1994 to describe an approach to designing educational materials that would maximize their use and reuse in different contexts (Wiley, 2002). Unfortunately, LO were the subject of both great hyperbole and significant confusion and debate regarding definitions (Friesen, 2004; Wiley, 2008). Professional standards and organizations such as ARIADNE (2000), IMS (2000), and SCORM (2008) created their own framework for LO development, maintenance, and interoperability. However, researchers focused on design and technical issues largely to the exclusion of legal and copyright issues. Wiley (2009) argued that the same copyright concerns and permission ambiguities that have been a concern for teachers have hindered LO from living up to the promises made by the advocates (Parrish, 2004; Polsani, 2006; Friesen, 2004).

A major step towards ameliorating the copyright issue was the creation of Creative Commons licenses. Creative Commons licenses were developed to help resolve concerns and ambiguities over when and how educational resources might be used, modified, and reused (Elkin-Koren, 2006; Caroll, 2006; Creative Commons Corporation, n.d.a). For example, the Creative Commons Attribution-Noncommercial license specified that anyone may use, copy, share, and alter a set of materials without further need to contact the copyright holder provided they credit them appropriately and that the material was not used for commercial purposes. A licensing web application on the Creative Commons website made the process of applying a Creative Commons license to an educational (or other cultural) work a reasonable process (Atkins, Brown and Hammond, 2007). When a more generous copyright license like Creative Commons was applied to educational materials, those materials were identified as open educational resources.

Reuse of Open Educational Resources

The term *open educational resource*, or OER, was coined at a UNESCO workshop in 2002. OER have been defined as educational resources (e.g. lessons, plans, quizzes, syllabi, instructional modules, encyclopedia entries, and simulations) that were freely available for use, reuse, adaptation, and sharing (Gurell, 2008a). An OER may be open or closed to differing degrees depending on how it is licensed. For example, an OER might be licensed to prevent derivative works. In that situation the OER was open with respect to use and reuse but not adaptation and remixing.

Although the number of resources available has grown, the rate of reuse of those open educational materials has been minimal. Duncan (2009) conducted a study of reuse in the OER repository Connexions. The author found that a third of all of the modules in that repository

were never reused and only fifteen Connexions modules, out of 4,713, were reused more than five times. A similar study on the Connexions repository found that most changes to content were made by the original authors and no other users (Petrides, Nguyen, & Jimes, 2008). A limitation of these two studies was that they only measured use and reuse within the Connexions repository, which was likely chosen because of its technical capabilities for supporting, and consequently measuring, reuse. Other researchers similarly concluded that reuse is difficult. Geser (2007) and Wiley (1999) stated that the paradigm of educational resources being easily interoperable with one another has largely failed.

One possible reason for the limited reuse observed may have been the barrier of technology. Some resources have been easier to view, revise and remix than others. For example, a PDF cannot be easily changed. However, the software to view the file was widely available. In contrast, the XCF file format of the GIMP photo editor was a fully open image format, but cannot be viewed in any other popular software programs. Geser (2007) created a set of criteria for defining what makes an open resource.

- Access to open content (including metadata) is provided free of charge for educational institutions, content services, and the end-users such as teachers, students, and lifelong learners.

- The content is liberally licensed for re-use in educational activities, favorably free from restrictions to modify, combine, and repurpose the content. Consequently, the content should ideally be designed for easy re-use as open content standards and formats are being employed.

- Educational systems/tools software is used where the source code is available (i.e. Open Source software) and there are open Application Programming Interfaces (open APIs)

and authorizations to re-use Web-based services as well as resources (e.g. for educational

content RSS feeds). (p. 20)

Geser's framework supported only a binary open/closed perspective towards resources. In

other words, a resource was considered either fully open or fully closed. The framework

included financial, legal, and technical criteria. Hilton, Wiley, Stein and Johnson (2010)

described the 4Rs framework considering legal dimensions of openness only. The 4Rs of

openness were reuse, revise, remix and redistribute. *Reuse* was using a resource as-is without

any modification. *Revision* was any kind of alteration to a resource prior to use. *Remix* was the

combining of any two resources together. *Redistribute* was the sharing of reuse with others. The

resource may or may not have been modified prior to redistribution.

ALMS Analysis

With that delineation of types of openness, Hilton, Wiley, Stein and Johnson (2010)

introduced the ALMS analysis for assessing the technical aspects of revising and remixing an

OER. ALMS was the acronym for Access to Editing Tools, Level of Expertise required to revise

or remix, Meaningfully editable, and Source Files. The ALMS analysis was based on openness

with respect to reuse or redistribution. Unlike Geser, Hilton et al. (2010) suggested that

openness was a continuum rather than a binary distinction.

The ALMS analysis framework consisted of the following elements:

- Access to Editing Tools**.** The tools necessary to edit a particular resource may vary. For
  example, the best way to alter a PDF has been through Adobe Acrobat Pro, which costs
  several hundred dollars. In contrast, editing HTML can be done with a free text editor.

- Level of Expertise required to revise or remix**.** The technical ability necessary to make a
  change affects its reusability. Altering an FLA file may mean understanding how to use

the Flash development editor and how to program in ActionScript. Editing a DOC file

merely requires the use of a word processor.

- Meaningfully Editable. Some resources can be edited, but not in a meaningful way.

  Hilton et al. (2010) used the example of a PDF of scanned handwriting. Although the

  PDF could be altered, the core content of the file, the handwritten text, cannot be

  changed.

- Source Files. Source files are defined as file(s) of origin. As an example, Flash files are

  distributed to end users in SWF format, but are edited by developers in FLA. Video

  editing projects often have multiple source files that are compiled and compressed into

  one file that is added to a DVD, or distributed on the Internet.

Applicability of ALMS Analysis

Although the ALMS framework allowed for a conceptual assessment of open educational

resources no actual measurement was reported in the Hilton et al. (2010) paper. The framework

had not been used extensively because there was no rubric corresponding to it. Consequently,

there has been no known systematic scale measuring the technical difficulty of reuse of open

educational resources. A study of technological difficulty was thought to help determine,

quantitatively, the degree to which technology is a barrier to reuse. A greater understanding of

the role of technology in preventing reuse has several future applications. Individual educators

may use it to gauge what kinds of resources are easiest for reuse. Those who create and manage

repositories could use an ALMS rubric to determine what kinds of OER best meet the

revise/remix goals of the repository. Funding agencies could also set policy based on the ALMS

rubric in terms of sponsoring OER creation.

Research Question

The benefits of an ALMS rubric cannot be realized unless it was sound from a statistical standpoint. Therefore, the purpose of this research was to explore whether an ALMS rubric, once created, can be validated as an instrument. The study answered the following question: To what degree can the technical difficulty of reusing OER be measured using the ALMS framework in a way that is valid and reliable? The first step in answering this question was to create an ALMS rubric and the second step was to validate that instrument by determining the degree of inter-rater reliability.

Chapter 2

Literature Review

The theoretical underpinnings of OER are critical to an understanding of their reusability. Consequently, several of the foundations are reviewed in this chapter. The concept of *object-oriented programming* (OOP) is addressed first. OOP is important to OER because the concept of reuse for learning objects, and OER had its origins in OOP. Next, *learning objects* (LO) are discussed as an important intermediate stage between OOP and OER. Finally, OER is addressed specifically. As these three concepts are reviewed, it will be argued that each concept was driven by certain promises some of which failed to materialize. This review will show that problems with OOP continued to be present in LO, which in turn influenced OER.

Before reviewing the literature related to OOP, LO, and OER are addressed, it is important to discuss how the sources utilized in this review are perhaps different from other dissertations. Each of these topic areas were discussed in scholarly publications, journals, and conferences. However, some topics covered by this review are limited by their nature. The technical nature of these topics seemed to have lent itself to being presented in conference proceedings, white papers, and textbooks rather than in peer-reviewed journals. This preference in the literature was particularly true when exploring OOP and reuse since that was an issue that seemed to more directly affect practicing programmers. LO did have a history of notable scholarly publications though the LO community also had vigorous discussion through conference proceedings, white papers, and blog posts.

Literature regarding OER also faced challenges. The field of OER itself is perhaps a little over ten years old (Wiley & Gurell, 2009). Journal publications regarding OER often took the form of special issues or single papers within a journal. The scholars in the OER community,

as well as others, have expressed concern about the cost of journals and availability of access (Monbiot, 2011). Stephen Downes summarized his feelings stating "from my perspective life is too short to have to deal with arbitrary reviewers and edits well past the point of diminishing returns" (Downes, 2012). As another example, Danah Boyd, a prominent scholar on teen use of technology, even pledged to avoid publishing in journals that are not available openly (Boyd, 2008). Some scholars have also pledged to support open access to journal articles exclusively (Suber, n.d.; Schechter, n.d). Though some scholars choose to publish in openly available journals, some simply preferred to publish on the Internet and bypass the journal altogether. These factors may have contributed to a lack of publications in high-impact, peer-reviewed journals. In this literature review, citations from sources that are considered unorthodox or unscholarly are done with the understanding that they may be such. It was believed that these sources are necessary to provide the most accurate discussion of how these concepts have evolved by both scholars and practitioners.

Object-Oriented Programming (OOP)

The origins of Object-Oriented Programming were traced back to 1967-1969. The very beginning started with the FLEX machine, which used references to objects (Kay, 1993). The references pointed to a particular part within the FLEX machine's memory that stored numerical values. The FLEX machine also had the ability to stop and start interaction on objects. These features were later iteratively developed into Simula, which was considered to be the first OOP language (Nierstrasz, 1986). Since that time there have been many programming languages that have enhanced and extended the OOP paradigm (Booch et al., 2007), including Pascal, C++, Java, and Ruby.

Characteristics. Some have argued that OOP has become the paradigm of choice among programmers (e.g. Hayes, 2003) though it has also been argued that OOP was never truly prominent (Ben-Ari, 2010). Nonetheless, OOP had become such a transformative force in computing that Stroustroup (1987) noted "'Object-oriented' has in many circles become a high-tech synonym for 'good'" (p. 51). Given this past aggrandizement of OOP, closer examinations of its features were warranted. Although each OOP language has its own unique features, all OOP languages share some commonalities including abstraction, encapsulation, inheritance, and instantiation.

Abstraction. Abstraction has been a part of OOP since its beginning. The name of the first OOP language, Simula, exemplified this relationship. The name was chosen because the object-oriented features of the language were meant to mirror or "simulate" the reality of real-life objects (Booch et al., 2007). To understand how Simula mirrored reality, it is necessary to define objects and abstraction. Stefik and Bobrow (1985) defined an object as "entities that combine the properties of procedures and data since they perform computations and save local state" (p. 41). By "local state" Stefik and Bobrow referred to the values of variables at a given point in time. Armstrong (2006) in a survey of articles relating to OOP defined objects as: "an individual, identifiable item, either real or abstract, which contains data about itself and descriptions of its manipulations of the data" (p. 125). Both of these definitions described an object as something that has properties, and those properties can be altered through interaction. Armstrong's definition explicitly allowed for abstract objects though Stefik and Bobrow allowed for it implicitly. Bian (2007) described three categories of abstraction. The first category was object-oriented analysis, which was generating a conceptual model. The second category was object-oriented design in which the conceptual model is given a systems or causal layer of

reasoning. More simply stated the second category was an object that took in or possessed some

kind of input(s) that is then processed to generate some kind of output(s). The third category was

OOP where abstraction was actually implemented within the code.

Abstraction, then, was the process of examining phenomena and "abstracting" or

determining an object's essence. This process was how Simula mirrors or relates to real-life

objects. The context determined which attributes of the object were chosen to represent it.

Booch et al. (2007), who gave importance to uniqueness in defining abstraction, stated "the

essential characteristics of an object that distinguish it from all other kinds of objects and thus

provide crisply defined conceptual boundaries, relative to the perspective of the viewer" (p. 38).

Madsen and Møller-Pedersen (1988) recommended abstracting properties of an object that

cannot be measured. Snyder (1986) placed greater emphasis on the methods or means by which

attributes of an object are accessed or modified. Ultimately, the attributes selected depended on

the aims of the program.

For a simple example of abstraction, one might imagine a circle. There are several

attributes that might be important in abstracting it. A circle has a radius, a circumference, and an

area. If the circle is depicted graphically, there would be a line or border representing the circle.

This line might have thickness or a color attribute. These attributes must be represented as a

particular data type. For example, the radius might bean integer and the area a decimal point

number. The border color might be stored as a hexadecimal number (e.g. #f356de). There is not

necessarily any single data type that best reflects a given attribute. As an example, a radius

might be best as an integer in some situations, because decimal approximation is not needed, but

in other cases a decimal point number might be required because higher precision is needed.

How the representation of these attributes is changed or adjusted depends on encapsulation.

Related to abstraction was the concept of a class. A class is a construct composed of common properties or patterns (Madsen & Møller-Pedersen, 1988). The word "class" can be thought of as similar to "type" (Booch et al., 2007). In that sense, a class was what organizes objects (Capretz, 2003). In a sense, a class was a group of abstractions to describe an object.

Encapsulation. Abstraction and encapsulation were complimentary concepts (Booch et al., 2007). Snyder (1986) defined encapsulation as "a technique for minimizing interdependencies among separately-written modules by defining strict external interfaces" (p. 39). To better understand encapsulation, consider the example of the circle used in the previous section. A circle object, again, might have the attributes of radius circumference, border thickness and border color. These attributes would then be accessed through certain functions. For example, the circle object might have the function *getRadius()* in which the numeric value of the radius is returned. Another function, *setRadius(),* might be the only function permitted to change the radius attribute for the circle object.

Encapsulation, in a sense, closes off an attribute so that only the designated functions interact with an object's attributes. This design paradigm was considered advantageous, since it prevented undesired or unintended interaction between an object and the rest of the code (Snyder, 1986). Code maintenance is also easier since problems with an object could be traced to a limited number of functions that interact with an object.

Modularization was another benefit of encapsulation. The attributes of an object and potentially none of the code outside of the object needed to change. As an example, if the radius attribute of a circle was changed from an integer to a decimal point number, only functions within the object need to be changed (Booch et al., 2007). This aspect of OOP would be later adopted by learning objects. It should be noted that individual circumstances within OOP code

may necessitate changes outside the function, but ideally they would be minimal. Changing a

class directly is one way to make changes to an object, but another method was through

inheritance.

   Inheritance. Inheritance is a concept in OOP in which one object is used to create another

object (Stefik, 1985). These relationships often have an "is-a" relationship (Booch et al., 2007).

As an example, a townhouse is a kind of home. This relationship might also be thought of as

parent-child (Pokkunuri, 1989). The number of attributes inherited can vary. Some classes that

utilize inheritance might extend the features of an object, or it may subtract from them

(Stroustrup, 1987).

   Inheritance has several advantages. If two distinct abstractions share similarities, the

functions and descriptions of attributes be inherited and do not need to be duplicated.

Additionally, if a change is made to the original object, other objects with inherited properties

are changed. This inheritance mechanism could save programmers time since they did not have

to make changes to every object that inherits attributes from the altered object (Pokkunuri, 1989).

   A more concrete example of inheritance may be useful. If a programmer created a

Sphere class for a program, they might have the class inherit properties from the circle class. For

example, the *Sphere* class might inherit the radius and circumference attributes and add a volume

attribute. The *Sphere* class might also have a fill color or some other attribute that identified

what is contained within the sphere. The programmer saved time in not having to develop radius

and circumference attributes for the *Sphere* class and it could be extended to cover new

attributes.

   This type of inheritance illustrated is known as single inheritance. Another type is

multiple inheritance. Multiple inheritance is a situation in which attributes from two or more

classes are inherited to form an entirely new class (Booch et al., 2007). As an example, imagine a *Citrus* class that had attributes that might account for flavor or level of citric acid. This *Citrus* class was merged with a *Sphere* class to form an *Orange* class. Multiple inheritance allowed for increasingly sophisticated objects. As with single inheritance, multiple inheritance allowed for changes to the original classes without having to rewrite the classes that inherit these attributes (Nierstrasz, 1989). The exact rules that guided this process vary from language to language (Stefik & Bobrow, 1985).

    Instantiation. Instantiation is the act of creating multiple copies of an object for interaction with the program as a whole (Stefik & Bobrow, 1985). It stands to reason that OOP would allow for multiple copies of the same object, especially if the goal of OOP was to model real-world interaction. However, it is important to understand that the benefit of multiple objects in OOP is not simply a matter of having multiple copies, but that each copy had the same attributes and methods. Because each object was an exact copy, it "permits multiple distinct uses of identical parts" (Stefik & Bobrow, 1985, p. 52). As an example, a program may have an *Employee* object that stores information relating to an employee. Because of instantiation, the program could have multiple *Employee* objects that could have added or removed data to each independently of one another.

    An Example. The following example demonstrated abstraction, encapsulation, inheritance, and instantiation. A video game programmer created a simulation in which the player shoots flying enemies. To create the class, the programmer might reflect on what attributes were important to model. The programmer might select attributes such as name, speed, a maneuverability factor, maximum flying height, or ability to fly backwards. The programmer decided how to quantify these attributes through data types such as strings (e.g. a

phrase), integers, decimal point numbers, or Boolean values. These attributes were then

encapsulated so that only specified methods could access these attributes. One particular

example might be a *setSpeed()* method that set a new speed for the object. A *getSpeed()* method

might provide the current speed when that information is needed by other objects or the program

itself. Once satisfied with the design of the base class, the programmer began extending the

class. For example, a *Dragon* class inherited attributes from the *Flying Object* class with the

added attributes of flame breath, tail attack, scale defense or others. The programmer might also

create a *Robot* class that inherited attributes from the *Flying Object* class with the added

attributes of laser cannons, armor defense, or others. Through multiple inheritance, the attributes

of the *Dragon* and *Robot* class could be merged to create a *Robot Dragon* class. Because of the

OOP concept of instantiation, the programmer may create as many dragons, robots, and robot

dragons as the computer's resources will allow.

Criticisms. No programming paradigm is perfect for all situations. OOP is no exception,

and there are several limitations to this paradigm. Criticisms of OOP varied. Some criticisms

have more to do with the way it was marketed rather than any objection with the paradigm itself

(Kölling, 1999). One criticism of OOP related to abstraction. Some natural phenomena did not

translate well to OOP. For example, geographic features were difficult to break down into

discrete objects (Bian, 2007). More generally, Bian (2007) noted that "the contradiction between

a discrete definition of objects and the perceived continuous world may affect the applicability of

the object-oriented approach in spatial representation" (p. 269). Distinguishing contexts in

which discrete definitions are acceptable may be difficult.

In addition to the difficulty of dividing real world phenomena into discrete units, there is

the problem of capturing relationships between objects. Kester (1993) provided several

examples of real world phenomena that arise, not from the interaction of one or two objects, but from many objects interacting with one another. Kester provided one example with electrical circuits. Electrical circuit components can generate heat and electromagnetic fields, and disrupt airflow within a circuit. A few individual components may not generate much heat. However, when several of these items are together on the same circuit it may cause a problem. Competent engineers often plan around these issues and the sizes of these effects are minimal. Nonetheless, they are examples of real world phenomena that occur that an OOP programmer might not consider or discover.

According to Cardelli (1996) there were other practical programming limitations to OOP. In terms of scale, OOP was advantageous for small-scale projects but cumbersome in larger projects. Cardelli (1996) claimed that OOP was inherently less efficient because of the effort needed for encapsulation. Cardelli also noted that the time taken to compile code that utilizes OOP was often longer than the compile times of other methods.

Cardelli (1996) noted that while reuse occurs successfully in some OOP projects, there were challenges to the promise of reuse. One obstacle was that as a project's complexity grows, there is an increasing need for programmers to understand how attributes and methods are implemented even if they were not making direct changes to the code. A related obstacle was the tendency to override or alter object methods in such a way that is counterproductive to, or otherwise undercuts, the intended functionality of the program.

Another criticism of OOP related to its emphasis on reuse. Lewis, Henry, Kafura, and Schulman (1992) conducted a study to measure actual object reuse. Computer Science students (n=42) were given the task to build a program. The students were randomly assigned into two groups. The first group was required to use Pascal. Participants were presumably not using one

of the variations of Pascal that allowed for OOP development. The second group was required to

use C++, a programming language that is often associated with OOP. These two groups were

each further subdivided into three groups that received none, moderate or strong encouragement

to reuse code. The resulting code was measured across several different metrics such as number

of errors, edits and test runs. Although the article did not say directly, it implied that the actual

code produced by study participants was examined to determine whether or not they actually

used OOP.

Between OOP and non-OOP the overall main effect favored OOP, meaning it had fewer

errors, edits, and test runs to develop a working project. However, in the group in which no

reuse was encouraged, there was no significant difference in the number of errors between the

groups using OOP and the one that did not. For groups that received moderate to strong

encouragement to reuse, OOP scored better across some metrics. Lewis, Henry, Kafura and

Schulman (1992) found that those who received strong encouragement to reuse actually

increased the number of errors when using an OOP approach. The authors hypothesized that

subjects who were strongly motivated to reuse used code did it inappropriately because of that

motivation. Therefore, according to the study, one of the disadvantages of OOP could be the

tendency to reuse in situations where it was not appropriate, ultimately hindering the

performance of the program as a whole.

These criticisms and limitations have not stopped OOP from becoming a significant

paradigm in computer science (Hayes, 2003). Learning objects and open educational resources

have inherited many of these limitations. These situations are discussed in the sections that

follow.

Learning Objects (LO)

Learning objects (LO), as discussed in this section, derive some of their conceptual grounding from OOP. LO also forms much of the basis for OER as a concept. Learning objects gained notoriety during the late 1990s and early 2000s though academic research is still ongoing. Examples of ongoing research include Santacruz-Valencia, Navarro, and Kloos (2009), Wiley (2010), and Geselbracht and Reisner (2010). The origin of LO could be disputed, though Wiley (2002) placed it with David Merrill's Component Display Theory as part of the development of the concept of "knowledge objects" (p. 5). The actual term "learning object" was used first by Wayne Hodgins when he named a working group "Learning Architectures, APIs, and Learning Objects" (Wiley, 2002, p. 3). The LO concepts precipitated large expenditures by governments, corporations, and universities (Friesen, 2004) but the results fell short of expectations. Before discussing the limitations of LO, a definition needed to be established.

Definition. Like OOP, the intent of this literature review was to define, show usage, and then explain shortcomings or limitations of LO. Establishing a definition of LO was difficult because of the widely different understandings of what constitutes LO.

It should be noted that there is even disagreement as to whether LO should be the preferred term. McGreal (2004) noted terms for the same idea including *educational object*, *content object*, *asset*, *learning resource*, *unit of learning*, and several others. McGreal separated the definitions into three categories: *anything*, *anything digital*, *anything for learning and a specific learning environment*. The wide range of categories for definitions demonstrate just how disparate the notion of LO has been. Using the categorization found in McGreal (2004), each of the three categories will be discussed.

The first category, that a LO can qualify as anything, was best typified by Downes (2003). Downes argued in a blog post, his preferred method of scholarly discourse, that "there is

no reason to restrict *a priori* what counts as a learning object." (The argument was further

extended that there is no tangible benefit to limiting a definition of open educational resources).

In the comments section of the post, McGreal argued that there are benefits to *a priori*

definitions such as standardization and opportunity for improvement. Like McGreal, some

authors found this perspective frustrating since a concept that can include everything seems self-

defeating. Sosteric and Hesemier (2002) argued "a definition that includes 'everything' is not a

definition at all" (p. 3).

The second category of definitions of LO given by McGreal (2004) was anything digital.

Wiley was typically the author cited to represent this category of LO definition. More

specifically, Wiley (2002) defined LO as "any digital resource that can be reused to support

learning" (p. 6). McGreal's categorization may seem misplaced since Wiley's definition seems

more restrictive. Perhaps the reason for McGreal's categorization was that Wiley would not

necessarily determine a digital object's value in an education context *a priori*. In this respect,

anything digital could potentially be reused to support learning. The digital nature of LO is

important as it allows for potentially unlimited duplication of a given object and that ability is

integral to LO. Wiley argued that this definition was narrow enough to be useful, but broad

enough to take into account objects that may not have had an explicit instructional purpose but

were nonetheless useful for learning. In particular, this definition did not exclude much of the

content found on the Internet (Wiley, n.d.). Sosteric and Hesemier (2002) suggested that this

kind of definition was limited as LO are commonly thought to be more than just the individual

object itself. Attributes such as discreteness, reusability, and searchability were typically

considered part of LO, but this is not necessarily true with every digital object.

The third category was anything, digital or non-digital, that has the purpose of supporting learning. This category is best exemplified by the IEEE Learning Technology Standards Committee that defined a LO as "as any entity, digital or non-digital, that may be used for learning, education or training" (IEEE, 2002, p. 5). This definition was understood to include "learning objectives, persons, organizations, or events" (Hamel & Ryan-Jones, 2002, para. 6). An objection to this definition might be the difficulty of determining whether or not an object has pedagogical intent. Pedagogical intentness is the result of the context, both the context of creation and its intended use. For example, a Batman comic book was created with the intent to entertain. However, in the context of an art classroom it may take on instructional value.

McGreal (2004) concluded by suggesting that the best LO are ones that closely mirror what educators actually use in practice. Therefore, the definition McGreal decided upon is "any reusable digital resource that is encapsulated in a lesson or assemblage of lessons grouped in units, modules, courses, and even programmes" (p. 28). This definition clearly excluded non-digital materials and even digital materials that do not have an explicitly educational purpose, making it one of the more restrictive definitions. It also excluded digital resources stored in decontextualized repositories.

The dispute over definitions continues. Polsani (2006), frustrated by the lack of consensus around a definition, stated "the terms Learning Objects (LOs) and Reusable Learning Objects are frequently employed in uncritical ways, thereby reducing them to mere slogans" (p. 1). Greaves, Roller, and Bradley (2010) termed LO as Reusable Learning Objects (RLO) and stated that RLO must be reusable across institutions and contexts to accommodate different users, but give little else in the way of insight. Over time authors tend to choose a definition, with only minimal effort to address the differences in definitions (Jones, 2006; Ternier et al.,

2008; Ochoa & Duval, 2008; Gonzalez-Barbone & Anidorifon, 2008). Whatever the definition used, in the late 1990s researchers, corporations, and educators began to create objects that they termed as LO.

Examples. If one were to use the most expansive definition of LO that includes virtually anything, then LO have been present since the beginning of time. Using stricter definitions, LOs emerged around the time the term was first used, about 1994 (Wiley, 2002). Using 1994 as a starting date, the history of LO can be thought of in three stages. The three stages are Early (1994 – 1998), Middle (1999 – 2004) and Late (2005 – present). There was a certain amount of arbitrariness in the demarcation of these periods. In the case of LO there were not necessarily definitive milestones to represent shifts in the practice or research of LO. Additionally, these divisions were as much shaped by what happened as what did not happen during these time periods. Each period will be discussed in order.

Early-LO (1994 -1998). Early LO were characterized by two types of usage in the literature. The first were conceptual papers that introduced the concept of LO to interested parties. Stahlke and Nyce (1996) provided an introduction to LO and suggested that LO was a paradigm that distance education courses should use entirely. Clark (1998b) gave an overview of LO and suggested that LO would greatly help for-profit entities in the future because of the cost savings in reusing LO.

The second usage of LOs in the literature during this time period was specific examples of projects done by researchers. Examples are wide-ranging and eclectic. Clark (1998a) related efforts from several institutions to create objects in Java, QuickTime, and VRML. The author stated that LO have been "successfully" integrated into introductory science courses at the University of Michigan (para. 34). No further details about the use of LO in that particular case

were given in the article. Clark called for a "shared development mode" to allow for

standardization and interoperability (para. 38). Ingargiola et al. (1994) discussed the creation of

a LO repository for teaching artificial intelligence. This particular repository utilized graphical

features in order for students to see the relationships between individual LO within a system.

The article provided only a description of the repository and not any learning outcomes or

experimental results. Bassi, Cheney, and Lewis (1998), like Clark (1998a), gave an example of a

LO project outside of academia. In the article the authors used Motorola's University, which

launched in 1996, as a case study of how LO changed corporate learning. Motorola's University

stored LO relating to a wide range of topics from "technical training to management education"

(p. 33). The rise of LO repositories outside academic settings, like the Motorola University,

presaged the Mid-LO period.

Middle-LO (1999 - 2004). The Middle-LO period was characterized by increased interest

and activity in LO. Not coincidently many of the peer-reviewed publications relating to LO were

published in this period. For example, Wiley's (2000) definition fell into this period, as does the

IEEE (2002) definition.

Neven and Duval (2002) provided a survey of LO of that time. In their article ten

repositories were compared. These repositories ranged from 48 LO in the Lydia repository to

MERLOT, which had 7,408 LO. The objects were situated in various fields, but many objects

were designed towards STEM disciplines. Most repositories were hosted by a university or non-

profit organization. Whether that was reflective of repositories as a whole, or if it was simply

something in common among those selected in the survey, is unknown. Neven and Duval were

vague as to whether commercial LO repositories were considered. Although most of the article

discussed various technical features of LO repositories, the authors did note a few items relating

to the design and use of LO. First, they noted that peer review, while useful, was only a part of

about half of the repositories, likely because peer review is time-consuming and needs oversight.

Second, they noted that some LO repositories gave users a personal workspace allowing for the

collection and sorting of LO. Features such as these could be potentially useful in aiding reuse.

In addition to features within repositories, formal technical standards and specification emerged

during this period. One of the most notable was SCORM.

SCORM, an acronym for *Shareable Object Content Reference Model*, was technically

created in 1997. However, it was during this middle-LO period that SCORM gained

prominence. SCORM was a collection of technical specifications intended primarily for

increasing the interoperability of LO with learning management systems. SCORM was a

product of Advanced Distributed Learning (ADL), an initiative of the Department of Defense

(Neumann & Geys, 2004; Lehman, 2007). The Department of Defense was interested in

interoperability as each branch of the U.S. military was purchasing a tremendous amount of e-

learning material. Unfortunately, prior to SCORM, this material was difficult to reuse between

branches. Therefore, the hope behind SCORM was to create a standard that was context and

content-neutral. That is to say, SCORM materials could be anything from antenna repair

instruction to sexual harassment training. The SCORM materials could be used as part of a

course, or self-contained training unit (Jones, 2002).

Over time SCORM began to be incorporated by higher education learning management

systems and interest in higher education grew. By adopting SCORM, gains in efficiency and

interoperability were hoped for in education (Bohl et al., 2002). The intended purpose of

SCORM in higher education was that a course could be moved to different Learning

Management Systems (LMS) with descriptive information included. Without SCORM or a

common standard, moving courses or LO between systems can be time and resource-intensive.

Kazi (2004) discussed the use of SCORM to create an intelligent tutoring system. Yang et al.

(2003) suggested an LMS that is was designed specifically with SCORM usage in mind. This

proposed LMS would have an algorithm that deconstructed objects to gather information about

an object and organize it accordingly. Objects could then be assembled through the LMS, or

downloaded for educators to use offline (Yang et al., 2003). Simões and Horta (2004) suggested

extending the SCORM specification for bibliography and evaluation as they are items that are

useful to higher education. Papers such as these characterized the interest in taking standardized

LO and expanding its uses to be more encompassing of the learning experience.

      Nonetheless, there were critiques of SCORM as well. Bohl, Schellhase, Sengler, and

Winand (2002) pointed out some disadvantages in SCORM. One deficit was that the ability to

use a SCORM object was highly dependent on the technical capabilities of the LMS. Another

deficit was the inability to fully separate an object from its context. As a result, SCORM objects

arranged together appeared to be "a hotchpotch of ill-matched content" (Bohl et al., 2002, pg. 2).

Godwin-Jones (2004) pointed out that although the SCORM standard allowed for prerequisites

to be specified, that part of the standard was rarely used. The issue of inconsistent appearance

and structure was also raised by the author who expresses a desire for more specifications in this

area. Abdullah, Bailey, and Davis (2004) noted that a common criticism of SCORM was that it

encouraged simplistic teaching and was ill-suited towards education. Abdullah, Bailey, and

Davis mentioned this criticism when referring to IMS Simple Sequencing, a standard that

mandates a linear sequence of content with compulsory assessments to prevent accessing content

out of order. They pointed to this development as something that lends itself to simplistic

learning.   Many of the arguments presented against SCORM could also apply to LO in general.

These limitations of LO are discussed below. While the limitations stated are noteworthy, SCORM continued to be used into the Late-LO period.

    Late-LO (2005 - present). The Late-LO period was not characterized so much by what it contains, but what was not present in the period. After 2004, there appeared to be a decided decline in attempts to define LO. The reason for this decline was unknown, and would presumably be debated just as vigorously as the debate over the definition itself. As stated previously, authors seemed to choose either Wiley (2000) or IEEE (2002) definitions and declined to discuss the rest of the debate. This period was characterized by an increase in the number of peer-reviewed papers addressing LO as the concept moved further into mainstream education technology.

    Topics of papers and usage of LO within the Late-LO period are diverse. Some authors choose to address best practices in developing LO. Polsani (2006) issued prescriptive ideas for LO development, namely that the key to LO was to abstract them as much as possible and eliminate any prevailing instructional design methodology. More of Polsani's critiques are discussed later in this chapter. Ochoa and Duval (2008) suggested metrics for ranking LO. These metrics are based on Duval's "Quality in Context." This concept of quality was divided into three categories: topical relevance, personal relevance and situational relevance. Topical relevance was defined as how well the domain of the LO matched the search query. Personal relevance was the degree that a LO matched previous preferences. Situational relevance was the degree that a LO matched the situation that triggered the search query. Ochoa & Duval suggested several metrics for measuring each of these categories. For example, a LO might be more highly ranked in terms of personal relevance if it contained a particular metadata value that a user has searched for in the past. Other topics discussed in the Late-LO period included papers

addressing subsets or specializations of LO. Gonzalez-Barbone and Anidorifon (2008) gave

step-by-step instructions for creating a SCORM object. Cebeci and Tekdal (2006) gave a

framework for designing audio LO. Santacruz-Valencia et al. (2009) also provided a framework

for the assembly of LO.

Other papers in this LO period were closer to case studies. Greaves, Roller, and Bradley

(2010) presented an example of LO being used in a Business Studies curricula. Qualitative

comments found that students were receptive to using LO because of their flexibility and their

use in supplementing the curriculum. Balatsoukas, Morris, and O'Brien (2008) outlined content

models and methods of aggregating LO at Cisco Systems. Krauss and Ally (2005) conducted an

evaluation on a single pharmacology LO to examine reactions from students and instructors.

The researchers used a questionnaire to measure student approval (n=10). The questionnaire was

a 5-point scale (i.e. Strongly Disagree, Disagree, Neutral, Agree, Strongly Agree). In many of

these categories over 70% of respondents agreed or strongly agreed with the value and usability

of the LO. The Krauss and Ally (2005) article was noteworthy in that it addressed a single LO in

contrast to other articles that examined the state of LO repositories. The issues surrounding a

single LO such as colors and animation were different than the issues facing repositories as a

whole (e.g. discoverability, storage, metadata).

Lehman (2007) divided LO repositories into three categories: general, discipline-specific,

and commercial/hybrid. General repositories included ARIADNE (http://ariadneeu.org) and

MERLOT (http://www.merlot.org). These repositories might have included any number of

objects across different disciplines. As the name implies, discipline-specific repositories focused

on a single discipline. Examples of discipline-specific repositories included Health Education

Assets Library (HEAL) (http://www.healcentral.org) and Math Forum (http://mathforum.org).

Lehman provided only one example of a commercial/hybrid repository, which was XanEdu (http://xanedu.com). These repositories were geared towards faculty and instructional designers as opposed to learners directly.

McGreal (2008) divided LO repositories into three categories: content repositories, linking repositories, and hybrid repositories. Content repositories host the LO on the site itself. Curriki was one example (http://www.curriki.org/)as a wiki dedicated to K-12 LO. The second kind of repository, the linking repository (sometimes referred to as a referatory because it referred a user to a different location), did not actually host any material but instead linked to the LO. The LO might be hosted on any number of repositories or web sites. Searching LO across several sources was made possible through "federated search" that gathered the results from multiple search queries and combined them into to one list of search results for the user (McGreal, 2008). MERLOT was an example of a linking repository. The third category given by McGreal, the hybrid repository, both hosted and linked to LO. ARIADNE (http://www.ariadne-eu.org/) was an example of a hybrid repository. A search query such as "rocks" returned results from OER Commons, Globe.org and the Smithsonian.

With these various types of repositories, the focus of the LO community became less about the composition and presentation of an individual LO and more about how these items were presented in the aggregate. What future directions LO production and the community surrounding it will take is uncertain. However, it was clear that LO has received, and continues to receive, criticism. These criticisms extend beyond the definitions already discussed and extend to usage and design.

Criticisms. Earlier in this chapter, the debate and discussion surrounding LO was addressed. Further, it was stated that a formal definition, or even vague consensus, was never

achieved. Although the lack of agreement on a definition was significant, issues with LO extend

beyond the definition to several different areas. These problems included ideology, granularity,

separation of visual design, and metadata.

   Difficulty in determining size. According to Wiley, Gibbons, and Recker (2000) granularity

was "the 'size' of a LO, while combination refers to the manner in which objects are assembled

into larger structures to facilitate instruction" (p. 2). Like the issue of defining LO, the

granularity of a LO was a source of ongoing discussion and debate with little resolution. The

discussion of granularity starts early within the LO literature, and largely proceeds by metaphor.

   The most popular analogy for LO in its early days was LEGO blocks. LEGO blocks

were interchangeable and simple to use. In some ways this analogy represented the ideal hoped

for at the time. Each individual LO was intended to be completely compatible almost with every

other LO, just as most LEGO blocks were compatible with almost every other block. That

compatibility was intended to provide maximum freedom to for potentially unlimited

combinations of LO. Like LEGO blocks, LO was supposed to be easy to use, so that non-

instructional designers could implement them. The LEGO analogy also had the advantage of

being widely recognized (Wiley, 1999).

   Unfortunately, this analogy appeared to have significant limitations. First, LO were not

completely interchangeable or compatible. One of the biggest sources of incompatibility was the

inherent limitations of instructional design. Not every LO combination made sense

instructionally. Moreover, the idea to make LO simple prevented LO from becoming

instructionally useful. Because of these limitations, Wiley (1999) saw the need for a different

analogy. Parish (2004) also noted that the ease of interoperability was misstated.

Wiley (1999) suggested that LO had properties like an atom. An atom could be combined with other elements though these combinations must abide by governing laws of chemistry and physics. Combining atoms together required some skill and knowledge. Likewise, the combinations of LO should abide by instructional design principles and should be done by those who have knowledge of the principles. Parrish (2004) asserted that the atom analogy was limited as it used the analogy of a physical object when a communication analogy was much more appropriate. Parrish chose the communication metaphor of creating a film. Film sequences cannot be rearranged in any order, but must be arranged according to filmmaking principles. Despite these alternate analogues the LEGO analogy remained and many papers continued to use it as an analogy for LO (Hodgins, 2002; Nash, 2005; Ravotto & Fulantelli, 2007).

Debate regarding the granularity extended beyond analogy to the specifics of LO size. Some researchers tended to think of granularity in terms of the number of ideas expressed. Sicilia and Garcia (2003) reported that several authors stated a LO should be confined to an idea or set of closely related ideas, or a single learning objective. Polsani (2006) also stated that a LO should consist of a "few related ideas" that "forms the basic building block of rational knowledge." Metros (2005) argued that any LO must have "1) a learning objective, 2) a practice activity, and 3) and assessment" (p. 6). Wiley, Waters and Dawson (2004) asked the question "can you ever imagine wanting to teach some portion of this topic without teaching the others" (p. 7)? Although this suggestion seems to be straightforward, breaking down any complex task or domain into completely discrete portions seemed difficult, if not impossible.

Other ideas regarding LO granularity were more focused on the LOs relation to the greater context and reusability. Longmire (2000), in addition to restricting LO to a single learning objective, stated that a LO should be "free-standing" and "modular." For Wiley, the LO

granularity was summarized by the question "What degree of granularity of learning objects best promotes the instructional use of the learning objects" (Wiley, 1999)? Wiley's question remained unanswered though the literature seemed to suggest the size should be smaller rather than larger (Quinn & Hobbs, 2000; Polsani, 2006; Hamel Ryan-Jones, 2002). Nash (2005) pointed out that size should not be thought of just in terms of length of duration or content, but small in terms of technical file size as well.

     Separation of and differences in visual design. Another critique regarding LO was how it was visually presented. The debate centered on the question of whether the visual design should be separated from the content itself. Wiley (1999) compared LO to granite blocks that can be painted any number of colors, however the underlying structure remained the same. Bannan-Ritland, Dabbagh, and Murphy (2002) gave the analogy of layers to an onion. The objectives might be one layer, while the visual interface would be another. Polsani (2006) stated that a "successful" LO is dependent on the separation of content and usage, which presumably includes visual content. In terms of implementing this separation, the use of XML was noted or suggested (Polsani, 2006; Mohan & Brooks, 2003; Neven & Duval, 2002; Downes, 2001).

     Although the idea of separating visual presentation from content seems logical, and can be done through means of XML, there were criticisms of this approach. Parrish (2004) asserted that if visual consistency in general was important, then it follows that the visual presentation of LO mattered as well. Wiley, Waters, and Dawson (2004) stated that the placement of LO next to each other "color and absorb each other's meanings" (Mason, 2006). For example, a poorly designed LO may increase cognitive load and therefore alter the effectiveness of the next LO in the instructional sequence, even if the second LO was well designed. In summary, the

sequencing and appearance of each LO cumulatively affected the learner's instructional experience.

     Challenges in determining appropriate metadata. Another unresolved issue was that of metadata. Metadata can have any number of definitions. *Metadata* was commonly defined as "data about data" (Boyle & Cook, 2001; McGreal, 2004). Friesen (2003) took a more functional position, defining metadata as "systematic description to facilitate searching and administration (p. 59)." Friesen's definition perhaps best explained why metadata exists.

     Finding and managing LO has occupied significant attention within LO research (Qin & Hernandez, 2004). Some of this importance has been as part of repository management (Metros, 2005). However, the ultimate goal of metadata was not simply for repository management but for creating instruction itself. It was hoped that an automated custom curriculum could be generated with the selection and order of LO being determined by a computer with the help of metadata (Mohan & Brooks, 2004; Sicilia & Garcia, 2003).

     Metadata has been a part of formal LO since the beginning. As noted earlier, the U.S. Department of Defense Advanced Distributed Learning Initiative began work on SCORM in 1997 (Jones, 2002). That work contained a metadata specification for describing learning materials (Lehman, 2007). The IEEE's Learning Object Metadata standard contained over 70 metadata fields (Gonzalez-Barbone & Anidorifon, 2008) and was based on earlier metadata work by IMS in the US as well as ARIADNE in Europe.

     There is another common set of metadata known as Dublin Core. The Dublin Core specification contained thirteen categories: Creator, Subject, Description, Publisher, Contributor, Date, Format, Identifier, Source, Language, Relations, Coverage, and Rights (Dublin Metadata Initiative, 2010). Though these specifications and standards were created and set early in the

field of LO, debate and discussion surrounding metadata continued. One of the issues debated

was to what extent the time-intensive task of creating metadata could be automated. Wiley

(2001) suggested that large objects were best-suited for automation because of the difficulty of

determining the context of small objects. Farrell, Libund, and Thomas (2004) asserted that basic

automated metadata was enough for end-users. Nesbit, Belfer and Vargo (2008) recommended

using automated metadata that would then be extended manually by experts.

Beyond the issue of automation was the question of what kind of metadata should be

attached to a LO. More specifically, there was an issue of how to best represent context to

facilitate LO selection and reuse. Hodgins (2002) made the distinction between "objective" and

"subjective" metadata. Objective metadata were attributes such as author, date, and

identification number. Subjective metadata included quality and reviews. Hodgins (2002)

placed greater value on subjective metadata because it was critical in generating a personalized

learning experience. Wiley (2001) made the same distinction in metadata and noted that

subjective metadata must vary between LO, while objective metadata could largely be held

consistent. Though subjective metadata was valued, determining what information to include

was problematic. Boyle and Cook (2001) understood the utility of being able to include

information about the instructional use of a LO, but believed that it was difficult to provide that

kind information without the backing of a theory. Jonassen and Churchill (2004) shared

concerns about the desire for standardization and need to take the learner's context into account.

Parrish (2004) acknowledged the possibility that metadata will remain simplistic because of the

inherent limitations of metadata definitions.

Ideological concerns. A final group of critiques that will eventually apply to open

education as a whole as well is loosely be grouped under ideology. This term is not used

disparagingly, but rather refers to concerns that reflect underlying values and not the logistical or practical issues used in creating, using, and reusing LO.

Parrish (2004) raised ideological concerns about LO. According to Parrish, LO "assumes a broad willingness to share and to borrow that may not exist. It also assumes that the processes of composing, distributing, and combining learning objects does not create substantial new costs" (p. 51). The reusability of an individual LO was also called into question. According to Parrish the more generally reusable an object was, the less likely it was to be effective in the local context. Wiley (2004) termed this dilemma as the "the reusability paradox" (p. 3).

Gur and Wiley (2008) criticized LO on the grounds that they objectify teaching. The authors considered teaching objectified when the LO focus was only on things that could be measured and standardized. These things often excluded elements of human interaction such as debate and dialogue. A second concern of Gur and Wiley was that LO "deskills" instructors because it largely excluded them from process of designing instructional material (para. 2). Through that design process teachers become marginalized and alienated from the teaching itself. Gur and Wiley (2008) also argued that part of this objectification extends to people through the process of reification. The result is that people are viewed as something to be manipulated and controlled. Gur and Wiley summarized stating, "Instructional designers need to question the trend of packaging and delivering that have been taken from other industrial enterprises" (para. 57). However, the article was unclear as to whether it was advocating instructor-only development of course materials or a process that allows the instructor and instructional designer to share power.

None of the issues facing LO across these time periods were ever satisfactorily addressed. These ambiguities remained perhaps because the community at large ceased to engage one

another in the Late-LO period. Consequently, it is not surprising that LO has not accomplished

the envisioned goal of learning that was automatically generated and customized for each

learner.

The issues surrounding LO were illustrative of the difficulties of reuse. Metadata could

be a significant barrier to reuse. If educators could find or properly identify a LO, then the

likelihood of reuse was minimal. Even if the "proper" LO was found, there were other barriers

to reuse. Mismatches in the granularity of two LO made bringing the two together difficult.

Different paradigms regarding the separation of visual design and the content itself caused

further problems. One instructor might have wanted to use a particular LO, but lack the skills,

time, or resources to generate a visual layer. In other cases, an undesired visual content layer

may be so intertwined with the content that trying to separate the two for reuse would be

prohibitive. Even if these barriers are addressed, some may question whether the attempt to

reuse was worth it at all, as it may lead to the objectification of education.

Open Educational Resources (OER)

Though the barriers of appropriate size, visual design, metadata, and ideology remained

the LO movement gave rise to the Open Educational Resources (OER) movement. Although the

origins of the OER movement began during the middle-LO period, when interest in LO remained

high, the OER movement increased in growth and in some ways overtook LO as a topic of

interest for practitioners and scholars alike.

In the early 2000s, researchers began to note the effect of copyright on LO. Muzio and

Heins (2002) developed "personal" and "shared" LO libraries, to ensure proper attribution and

control by the copyright holder. Neven and Duval (2002) proposed a similar distinction in

partitioning reuse. Both Hamel and Ryan-Jones (2002) and Muzio and Heins (2002) discussed

copyright concerns within the context of LO but gave no indication what, if any, resolution

might be found. Other authors such as Bohl, Scheuhase, Sengler, and Winand (2002) and

Mohan and Brooks (2003) acknowledged that copyright is an issue, but presented it as being

outside the consideration of their respective papers. Wiley, Waters, and Dawson (2004)

suggested that the unsuccessful attempts of the music industry to enforce copyright laws should

be illustrative for anyone who hoped to make money exchanging and distributing LO.

The acknowledgement of the problematic relationship between LO and copyright set the

stage for the OER movement. As Wiley, Waters, and Dawson (2004) pointed out the availability

of content was not necessary enough to meet an institution's needs. Instead, institutions must be

able to make changes to meet their needs of their learners.

Copyright issues. However, many countries including the U.S. have laws that restrict

usage without explicit consent from the copyright owner. The OER movement intended on

granting permission beyond traditional copyright (United States Copyright Office, 2008) so that

explicit permission was not required for each instance of reuse. Openness beyond the default

permission can be a question of degree.

The extent of the permissions could be framed around the "4R's" (Hilton, Wiley, Stein &

Johnson, 2010). The 4R's were "Reuse," "Redistribute," "Revise" and "Remix." Each of these

"R's" represented a degree of openness.

- Reuse. A resource can be used without the difficulty of obtaining explicit permission for

  the usage. The use may be whole or in part.

- Redistribute. Redistribution is sharing of the work with others. Typically this reuse

  occurs with attribution to the original copyright holder.

- Revise. Revision is the ability to change a work for different purposes. The revision might be small or substantial.

- Remix. Remixing is the combining of two or more resources. The resulting work constitutes a new work.

Creative Commons licensing provided assistance in granting each of these "R's." All Creative Commons licenses allowed for "Reuse" and "Redistribution." "Revision" and "Remix" are allowed unless the No-Derivatives clause is invoked, in which case the resource must remain intact. Any of these degrees of openness can be ameliorated through the Non-Commercial clause that prevented commercial usage of a resource.

Resource and project growth. Since 2002, the number of OERs had increased significantly. Tens of thousands of OERs are now available on the Internet with more being added each day. One of the best-known OER repositories had been the online encyclopedia Wikipedia. Along with it were other Wikimedia projects such as Wikibooks that provided collaboratively-developed books, and Wikiversity, dedicated to various kinds of OER (WikiMedia Foundation, 2009a; WikiMedia Foundation, 2009b). Other OER projects included Curriki, which has been focused on the K-12 materials for K-12 educators, and WikiEducator, a site dedicated to OER development for commonwealth nations (Curriki, 2009; Commonwealth of Learning, 2009). Each of these projects allowed anyone to edit or alter content, with the only requirement being a valid e-mail address to register. However, many teachers have preferred to use educational materials with a clearer origin and stronger affiliation with an institutional brand recognized for high quality.

The most prominent example of these kinds of expert-centric OER has been MIT OpenCourseWare (OCW). In this case, OCW referred to several OERs presented together as a single course or courses. Some OCW were relatively simple, such as the syllabus of a course made available with Creative Commons licensing. Other OCW were quite elaborate. For example, those courses might contain videos of each lecture, supplemental text, quizzes, lecture notes, or rubrics with scoring samples. Each of these OCW courses had both the institution's name and a professor's name attached to it. The material represented professor's time, effort, and intellectual contribution. With those contributions come expectations and assumptions about credibility and expertise. MIT was the first to start an OCW initiative in 2002 (MIT OCW, n.d.). At the time of this writing, there are dozens of OCW initiatives at institutions worldwide (OpenCourseWare Consortium, n.d.). For an educator concerned about wiki pages being potentially edited by anyone, OCW offered a valuable way to obtain OER approved by experts in their respective fields.

Both forms of OER continue to grow. The Open Education movement has spent millions of dollars on development and infrastructure (Atkins, Brown, & Hammond, 2007). As the number of OER continued to grow, an ecosystem of freely available software tools has grown around the content. Some of these tools focused on the development, management, and dissemination of OER and OCW. For example, the open source eduCommons platform was developed as a content management system specifically designed for publishing OCW.

The Folksemantic project was developed as a search engine and recommender system for OER and OCW that helped teachers and others find openly licensed materials. Creative Commons also created its own search engine for OER/OCW (known as DiscoverEd; http://discovered.creativecommons.org/search/), as has the OCW Consortium (Far, 2011).

Another helpful tool, previously mentioned, is the Creative Commons license web application (http://creativecommons.org/choose/). This web page asked users to fill out the form, and as a result, created a Creative Commons license for the work specified. This tool has been important to individual OER contributions as most individuals cannot afford to have legal counsel create a custom license for each work. The license web application was simple enough that complex legal issues can be overcome by educators. These tools demonstrated a community that has been continuing to experiment with ways to ease the use of OERs.

   Technical barriers to reuse. Despite the proliferation of open education projects, little reuse seemed to take place. Browne et al. (2010) interviewed academics to find out what some of the obstacles were to OER adoption, use, and reuse. Those interviewed cited numerous barriers such as perceived quality/authority of OER and a lack of incentives to participate. Faculty also expressed concerns about support and improper use of the materials. Windle, Warrad, and McCormick (2010) also stated that lack of cultural support was a detriment. They also repeat some of problems found with LO, like the contextual specificity, as barriers to OER sharing.

   Duncan (2009) provided an examination of reuse within the Connexions repository (http://cnx.org). Connexions was an OER repository that consists of modules that can be authored by anyone. These modules could be gathered together in a collection (Connexions, n.d.). The modules can also be reused as the basis of another module, or can be reused across a potentially infinite number of collections. Each module in Connexions contained a metadata record, though some metadata records are more complete than others. Fields in the metadata record included name, creation date, author, and version history. Duncan used the metadata records to determine the modules that were original, and those that were derivatives. Approximately one-third of the modules used were never used in any collection. One-fourth of

the modules were reused. The reused modules made up one-fourth of total module use. The findings showed that while some reuse did occur, by and large most modules were not reused more than once. This study was limited to reuse within the Connexions repository itself and does not include any reuse that might have occurred outside of Connexions.

Petrides, Nguyen, and Jimes (2008) also examined reuse within Connexions, but the methodology was grounded in a qualitative approach. Reuse in the context of this study was "defined as the remixing or adaptation of OER for new/local purposes." The authors conducted phone interviews with 11 Connexions contributors to find out their motivations for contribution. They asked questions relating to their activities in Connexions and OER. Additionally, each contribution to Connexions could have included a log file with notes about the change. Participants were also asked clarifying questions regarding these log files. They found that updating a module was the most common reason for reuse. Typically the update would take the form of fixing errors or adding visual elements to existing content. Some reuse occurred as part of collaboration. From analysis of the Connexions repository, the researchers also concluded that authors were hesitant to reuse others' work.

In discussions with Connexions contributors who ceased adding content to the repository, a number of barriers previously discussed were raised. In addition, a lack of technical skills was cited as one of the reasons they ceased to contribute. Contributors who continued adding content took technical barriers as a challenge and expressed satisfaction once the content was uploaded (Petrides, Nguyen, and Jimes (2008).

Beggan (2010) also noted technical barriers to reuse. The author was part of a team developing an OpenCourseWare project at the University of Nottingham. Focus groups were conducted to determine issues that might affect the project's development. As a result of their

findings they concluded that "Technological barriers can be a very real issue to open publishing and additional resources dedicated to content conversion may be required" (p. 6). More specifically, they noted that anything beyond creating plain text proved problematic.

Though this issue of technical experience was acknowledged in the literature, there has been no attempt to systematically measure the degree to which technology is a barrier in OER reuse. As mentioned in Chapter 1, reuse of OER has been minimal. A lack of reuse was a significant problem for the OER community, since one of the primary reasons for licensing content openly is to give users the ability to revise or remix. If researchers had a better sense of the degree that technical problems are a barrier to reuse, they might better understand the problem of OER reuse itself. Even if the technical barriers to reuse are found to be insignificant, the finding will be valuable because it could suggest other significant areas to research.

Connections between OOP, LO, and OERs

Before beginning research it was necessary to have a conceptual framework to clarify what was needed for reuse from a technical standpoint. The ALMS framework provided a way to consider reuse with enough detail to be operationalized. To review, the ALMS analysis consists of the following questions:

- Access to editing tools?

- Level of expertise required to revise or remix?

- Meaningfully editable?

- Source-file access?

These criteria focused on practical considerations for technical reuse and avoided many of the dogmatic stances that have been present in open education and the OER movement.

Though the ALMS analysis was unique in the literature, there were limitations. First, the ALMS analysis was only a conceptual framework. There was no known systematic way to conduct an ALMS analysis. The second limitation, related to the first, was the unknown relationship between the ALMS analysis and the body of OER existent. If a collection of OERs from various repositories were collected and methodically measured for technical barriers to reuse, conclusions might be made about the state of OER reusability in general. The purpose of this dissertation was to address this minimally explored area of reuse and increase understanding of OER reuse as a concept and in practice.

The narrative of OOP to LO to OER is one of an interdisciplinary field attempting to find a guiding metaphor. The field of educational technology has ties to computer science and it is not surprising to use a popular development paradigm from computer science like OOP. Unfortunately, the limitations of OOP were never fully addressed by the LO community. Each of the OOP concepts, abstraction, encapsulation, inheritance and instantiation, were never fully incorporated into LO.

Though LO did realize the promise of instantiation as LO could be reproduced infinitely, LO failed to successfully incorporate other OOP concepts. Abstraction, a key element in OOP, seemed to be counter-productive for LO. The "reusability paradox" demonstrated that the more a particular LO was abstracted, the less valuable the LO become. The context of a LO, both internal and external, proved problematic for encapsulation and inheritance. The design of a LO faced two difficult decisions. If a LO was designed to be completely encapsulated, then its size become large and unwieldy. If the LO was designed too small, then the designer or educator must add context, assuming it has permission, or as part of a surrounding context.

Wiley (2009b) explicitly claimed that OER has overcome many of the problems with LO.

OER grants legal permission, most often through Creative Commons licensing, for educators to change a resource to address these issues. Nonetheless, any attempts to modify a LO to better meet the principles incorporated by OOP can be time and resource-intensive. Even if OER was modified to take advantage of OOP principles, the benefits of automated instruction and interoperability in a larger context largely may be unrealized. Additionally, the concerns of Gur and Wiley (2008) and Parrish (2004) regarding the objectification of education are frequently raised but never resolved.

Summary

Although the history of OOP, LO, and OER showed disagreement about the very fundamentals, each concept has been adopted to varying amounts by educators. The literature indicated that given the complexity of issues that surround reuse, it would be naïve to assume that these issues could be easily solved. Therefore, it was necessary to focus on a single aspect of reuse.

The literature reviewed on the topic of reuse was limited, what literature was available indicated that technical barriers were one obstacle. Therefore, this dissertation centered on technical limitations of OER reuse. The purpose of this research was to better understand what was acting as a barrier to reuse and to what extent it was a problem. Ultimately the intent of the research was to address to what degree these technical barriers can be measured in a reliable and valid manner.

Chapter 3

Methodology

The purpose of this dissertation was to address the following question: Can the technical difficulty of reusing OER be measured using the ALMS framework in a way that is valid and reliable?

There were any number of ways to address this question depending on epistemological views, methodology preferences, and personal biases. For this dissertation, the methodology selected was intended to provide an expert-driven, quantitative approach to the analysis of OER reuse from a technical perspective. To this end, the methodology consisted of three parts. The first part was the drafting of a rubric using the ALMS framework as a guide. The second part utilized a Delphi study analysis to provide evidence for construct validity and help refine the items used in the rubric. The final step in the validation process pilot tested the instrument against a sample of OERs to obtain an estimate of reliability for the measures produced.

Initial draft of the ALMS rubric

The first part of the study was concerned with developing the rubric for rating OER reuse. Rubrics have been used in many different education and psychology contexts to evaluate learners or programs (Bresciani et al., 2009; Penny, Johnson & Gordon, 2000; Roblyer & Wiencke, 2003). Rubrics have also been used in higher education mostly in performance assessment (Redy & Andrade, 2010).   Generally, there are two types of rubrics: holistic and analytic. Holistic rubrics typically measure overall performance or attributes. Analytic rubrics are used to measure individual components or attributes of the construct being measured. A holistic rubric might be used when what is being measured requires that all the components of

the performance must be present to some degree. The overall rating was only understood by considering the entire performance including all aspects of the skill. In contrast, analytic rubrics were better suited for tasks that have defined parts that may or may not be required elements when determining the quality or value of the performance or ability being measured (Mertler, 2001). For this study an analytic rubric was used. An analytic rubric was better suited for this study in comparison to a holistic rubric because an analytic rubric allowed for measurement of each component of the ALMS framework. Some components would be present in varying degrees for each OER being considered.

The ALMS analysis, as discussed in Chapters 1 and 2, formed the theoretical basis for the rubric. It provided evidence of content and construct validity. I constructed the rubric with Dr. David Wiley. By following the prescribed framework we attended to the important aspects of the constructs being measured. Each category of the ALMS analysis was anticipated to have some kind of representation within the rubric. As an expert in both LO and OER, and as the creator of the ALMS analysis framework, Wiley had expertise to assist in the construction of the rubric. I was also qualified to construct the rubric based on experience with OER for over four years including creating a guide to developing OER (Gurell, 2008).

During the development of the draft rubric, emphasis was placed on what a hypothetical educator might do to revise or remix an OER. A series of possible criteria was developed. The practicality for educators to measure each criterion was considered. These steps were then considered with respect to the ALMS framework and what factors applied to each part of the rubric. Extensive discussion took place to determine whether elements from one part of the rubric were too similar to other parts and how the two might be distinct conceptually. For example, there was significant discussion relating to the differences between Meaningfully

Editable and Source Files, since the two components of the framework were closely related, but still separate.

Two drafts of the rubric were created and discussed during these discussions. When a third, more stable draft of the rubric was created, three sample OER were rated. The OER selected were expected to score high, low and mid-range on the ALMS rubric. The intent of using different types of OER was to gain a sense of how well the rubric discriminated among OER and investigate the face validity of scores. After this small sample rating exercise, the resulting scores were discussed and rubric criteria were adjusted accordingly (see Appendix A).

Delphi Study

Once an initial version of the rubric was created, there was a need to establish its usefulness and validity. This was done through a Delphi study. This section will first discuss validity issues and then how a Delphi study was used to address these issues.

Validity issues. Validity is the "extent to which a measure reflects only the desired construct without contamination from other systematically varying constructs" (Judd, Smith, & Kidder, 1991, p. 51). There are four potential evidences of validity: face, content, criterion, and construct (Moskal & Leydens, 2000). Messick (1994) also discussed consequential validity as an evidence of validity based on the social values conveyed. Each of these types of evidences were distinct and in order for a rubric to be valid it should demonstrate multiple evidences. Face validity was "the degree to which a measure appears to be related to a specific construct" as determined by persons who are not experts (Furr & Bacharach, 2008, p. 173). Criterion validity is the degree to which scores predict certain future performance or correlates with a known measure of the construct (Furr & Bacharach, 2008). In this study, face validity and criterion

validity were not directly considered. Face validity was not considered because it was a

problematic concept with some questioning whether it belongs in scientific research (Downing,

2006; Nevo, 1985; Bornstein, Hill, & Stepanian, 1994). Criterion reference validity, sometimes

referred to as predictive validity (Babbie, 2000), was also not considered in this study. The

reason it was not considered was because there were no known measures of the construct to

correlate results. In addition, it does not take the conceptual meaning of the test, the focus of the

study, into account (Furr & Bacharach, 2008). Likewise, consequential validity, whether the

results are interpreted and used correctly and appropriately, was not considered. The

consequences of the ALMS rubric scores were discussed as part of the Delphi study.

The two measures of validity that were considered in this study are construct and content

validity. Construct validity is the "degree to which an assessment instrument measures the

targeted construct" (Haynes, Richard, & Kubany, 1995, p. 241). Content validity is the degree

that the assessment is "relevant to and representative of the targeted construct" (Haynes, Richard,

& Kubany, 1995, p.241).   Initial evidence of validity for the creation of this instrument was

provided by using the ALMS framework.

Background of the Delphi study. In order to verify evidence of validity, a Delphi study

was used. The Delphi study is a process where experts are solicited for their opinion regarding a

particular topic. It has been used in several fields such as health care (Powell, 2003) and

education (Murray & Hammons, 1995). The Delphi study methodology was originally

developed by the Rand Corporation to "obtain the most reliable consensus of opinion of a group

of experts" (Dalkey & Helmer, 1963, p. 468). There were three key features of Delphi studies:

anonymity, controlled response, and statistical treatment of group opinion (Dalkey, 1969).

Given these principles, a Delphi study was different than a group discussion. The intent behind

specifying anonymity and controlled response was to prevent any individual from overpowering or inappropriately influencing the rest of the group (Dalkey & Helmer, 1963). The statistical treatment of group opinion was intended to provide an objective measurement.

The Delphi method has been applied with many variations. Nworie (2011) noted that many Delphi studies used a questionnaire though there are some that utilized an open-ended question format. A questionnaire provided greater focus for study participants, but risked missing information or perspectives that might be captured from a more open-ended approach. The method by which consensus was determined also varies among Delphi studies. Some studies used a percentage of agreement as a measure of consensus and other studies considered consensus achieved when participants no longer shifted their responses (Nworie, 2011). Sample size in Delphi studies have ranged from four to 171 participants. The Delphi method has been applied in several dissertations and used for over 400 journal articles (Skulmoski, 2007).

Structure of Delphi Study. The Delphi structure had several components. In this section, the selection of Delphi participants was discussed. Next, the method used by participants to rate the rubric was fully explained. After the first round, the process for subsequent rounds was outlined.

Delphi participants. The five participants in the Delphi study were Steve Carson, Scott Leslie, Ahrash Bissell, Cable Green, and Karen Fasimpaur. The first step in the Delphi study was to select participants who have expertise regarding OER. This study was closer to the original Delphi study sizes with five participants selected based on their expertise in OER.

Steve Carson was External Relations Director for MIT OCW. He was also President of the OpenCourseWare Consortium. Carson was one of the leaders in developing MIT OCW, and

has undertaken several research initiatives to examine OER usage. Carson has an MFA from

Emerson College (Carson, n.d.).

Scott Leslie was a Manager of Client Services specializing in Open Education at the

BCcampus in Canada. Leslie has developed OER repositories for BCcampus and the website

http://freelearning.ca.   He was recently an OLNet Fellow at the Open University of the UK

where he studied the reuse of OER, making him a well-suited candidate for this study.

Dr. Ahrash Bissell is a former director of ccLearn, which was a division of Creative

Commons dedicated to open education efforts. Bissell received a doctorate in Biology from the

University of Oregon (Alexandria Archive Institute, n.d.). At the time of this study he worked

for the Monterey Institute for Technology and Education (MITE), an organization that was active

in the OER community through its National Repository of Online Courses (NROC) and

HippoCampus collections.

Cable Green was the Director of Global Learning at Creative Commons. Prior to joining

Creative Commons, Green developed a project to share college courses as open educational

resources (Creative Commons, n.d.).

The final participant was Karen Fasimpaur. Fasimpaur was an active participant in

developing, presenting, and blogging about OER in K12 context. Fasimpaur was the Founder of

the Kids Open Dictionary, an open educational resource.

Procedure. In the first round the rubric was introduced to each participant. Participants

were instructed that the purpose was to review a rubric developed by David Wiley and myself

that measured technical barriers to reusing OER as opposed to other areas such as overall quality.

The ALMS rubric was divided into sections that covered each of the components of ALMS (i.e.

a section for Access to Editing Tools, Level of Expertise, Meaningfully Editable, and Source

Files). Each of the sections of the rubric consisted of multiple items (e.g. Level of expertise

needed to rate the OER). Associated with each item was a response scale that indicated the

degree that the item applied to the OER in question. The response scale was fully anchored,

which meant each score on the response scale had its own unique label. Delphi participants were

asked to rate each label of the rubric according to three categories: Reliability, Desirability, and

Feasibility.

　　　　　Reliability in the context was meant to be the same as the statistical sense (i.e., the degree

to which a respondent might consistently or accurately rate the object in question on that

particular item). However, no actual statistics were calculated by raters to determine their score.

Instead, reliability in this context was intended to be a predictive reliability. With respect to the

Reliability score, raters were being asked whether they believed the label would allow raters to

accurately and consistently indicate their response.

　　　　　Desirability was a measure of how valuable each label was with respect to measuring the

item it was associated with it. A label may have been considered less desirable if it doesn't

accurately measure the item itself. Additionally, a label may have been rated less desirable

because the item itself was less important, or because that particular score on the response scale

wasn't considered important in comparison to others.

　　　　　Feasibility meant the degree to which raters believed each label to be practical for raters

to assess. As an example, a label that required raters to search for suitable software program for

reuse by searching 30 search engine pages may have been desirable, but was probably not

feasible.   These categories were taken from a Delphi study of tobacco subsidies policy in

Kentucky (Rayens, 2000).

Participants were asked to review the anchors or labels of each item in the rubric according to Reliability, Desirability, and Feasibility on a scale of 1-4. A 4 rating was considered most reusable, while a 1 rating would be considered least reusable. Participants were given an opportunity to provide free-form responses and asked to recommend critical areas that the draft rubric may have overlooked or the elimination of unnecessary items.

The Delphi study was designed to be four rounds rather than have a point of convergence for the scores. In the first round of the Delphi study, participants were asked to rate all four sections of the rubric. After that first round, means were compared using a series of ANOVAs along with Tukey's HSD as a *post-hoc* test (Gravetter & Wallnau, 2008). These statistics were calculated to best determine which questions needed the most revision. In addition to the ANOVA calculations, the means for every category in each section of the rubric was calculated. For example, the mean for the Reliability category across all Access to Editing Tools was calculated. That mean was combined with Desirability and Feasibility means to form one overall mean for that portion of the rubric. If that overall mean exceeded 3.25, then that section of the rubric was considered complete and was not rated in future rounds of the rubric. If a section of the rubric did not meet 3.25, then it would be rated in future Delphi rounds.

In subsequent rounds, two parts of the rubric were rated in each round. Study rounds after the first were limited to two sections because of concerns regarding rater fatigue. Sections of the rubric rated after the second round were presented to raters in ascending order (i.e. lowest score first). Similar to the first round, participants were asked to rate each label across the three categories on a scale of 1-4. The rubric in the first round was distributed by e-mail. After the first round, responses to the rubric were collected using Qualtrics survey software because it was easier for users to enter scores.

Ratings of Sample OERs

Once the instrument was created, checked, and revised through a Delphi study, the ability of raters using the instrument to provide consistent estimates of reusability was tested. Reliability can be defined as "essentially a synonym for consistency and replicability over time, over instruments and over groups of respondents." (Cohen, Mannion, & Morrison, 2000). In the context of this inter-rater exercise, reliability is the degree to which multiple raters can use the rubric and come up with consistent results. Note that this definition of reliability was different than how it was used in the Delphi study. In the Delphi study reliability scores were a prediction based on expert opinion. The method of asserting reliability of scores in this part of the study was produced by the rubric to rate a sample of OER. To best evaluate the reliability of scores, it was necessary to select a varied and wide range of OER samples for raters to consider. That meant samples from a few different repositories and from several different disciplines were each rated by a variety of trained raters. In order to achieve this diversity in resources, samples were taken from three broad categories: Humanities, Social Sciences, and Sciences. These three categories are not meant to be representative of every academic discipline. Instead the intent of these categories was to measure whether there would be any variation in score based on the domain and the reusability of the OER. The language was kept broad because a specific domain might be available in one sample repository but not in another.

Table 1

Sample distribution to determine inter-rater reliability

Type	NROC	MIT OCW	WikiEducator
Humanities (e.g. Art, Literature)	3 samples	3 samples	3 samples
Social Sciences (e.g. Education, Psychology)	3 samples	3 samples	3 samples
Sciences (e.g. Biology, Math)	3 samples	3 samples	3 samples

A total of 27 samples were rated. The sample consisted of audio, video, animation/slides, PDF, webpages or text, and images. These OER were chosen to represent a broad range of OER from a technical perspective. Table 2 presents the quantity of each media type sampled.

These particular samples were selected to create discrimination between items. The samples were chosen by myself prior to rubric construction. A concern was that knowing which repositories were being rated prior to developing the rubric could influence the development of the rubric in such a way that the criteria selected would be more applicable to those repositories than others. It was not anticipated that this limitation in the design would interfere with the conclusion reached.

The repositories selected were representative of well-known OER repositories. Each of these repositories differs in their approach to OER. MIT OCW organizes its content by course and subdivided it by week or individual lecture. WikiEducator is organized in the loose manner that often typifies wikis. NROC presents its information as courses with different multimedia components. Efforts were made to find OER from different parts of the repository. In the case of the NROC repository, only one college Humanities course, Religions of the World, was

available. During the selection process, efforts were also made to choose OER that were

reflective of a variety of media types.

Table 2

Quantity of media types in sample OER

Type	NROC	MIT OCW	WikiEducator
Animation/Slides	3 samples	0 samples	1 samples
Audio	0 samples	1 sample	0 samples
Image	0 samples	0 samples	3 samples
PDF	3 samples	4 samples	0 samples
Webpage/text	2 samples	2 samples	5 samples
Video	1 sample	2 samples	0 samples

One important note related to the choice of repositories and the Delphi participants.

Many of the Delphi participants had direct relationships with the repositories selected. Delphi

participants were not notified which repositories the samples would be taken from. Since Delphi

participants did not know which repositories were to be being sampled the threat of bias was

somewhat mitigated. Naturally, some Delphi participants still may have chosen to bias their

feedback to look favorably on repositories with which they are involved regardless of their

knowledge of the sample.

Participants. Ideally those who would rate the samples would closely resemble educators

who reuse, revise and remix OER in order to ensure that the scores generated by the raters

closely matched other educators who would most likely benefit from the rubric.   Unfortunately,

there was very little information on these educators (Duncan, 2009; Harley, 2008). Since the rubric developed in this study focused on the technical aspects of reuse, the qualifications for a rater included computer skills. For this study, participants needed the skills typically taught in a pre-service teacher program.

Participants chosen to rate the OER sample were drawn from a teaching college. These students were graduate students in an instructional technology program. Although there may be differences in computer skills between participants, the minimum requirements on completed coursework helped to establish a basic level of competence. Because little was known about the average technical ability of educators interested in OER reuse, it was not inconsistent to set the standard at a very basic level. Future research could explore how different levels of technical knowledge influenced scores.

Some of the participants already had a relationship with researchers. Since I did not have a particular bias in the scores, raters were assured during training that high scores were not necessarily desirable.

Rater training and rating process. Four raters participated in this part of the study. Training was given to participants prior to rating. Training was held in a graduate student lab at a pre-specified time. Each rater had their own computer. Raters were trained as a group. Additionally, raters were instructed to resolve any questions with me rather than consult with other raters.

The training consisted of 15 minutes of instruction that contained an overview of what open educational resources are and Wiley's 4R's of openness, as discussed in Chapter 1. The raters were also introduced to each component of the ALMS rubric as revised after the Delphi

study. A sample OER from the MIT OCW collection was presented and rated as a group, facilitated by the researcher.

Prior to the training, I generated a list of possible concerns the participants might have during rating. These possibilities were addressed as part of the instruction. Many of the possible concerns related to media types. The raters were told to assume that text from a web page or PDF file could be copied via the cut and paste feature of most operating systems, rather than editing the file itself. Videos that could not be saved directly to the hard drive were considered non-editable. Interactive slideshows were specified as Flash objects that were considered editable presuming the source file was available, though the expertise needed to edit a resource of this type would be high. The final instructions to raters related more to act of rating itself. The raters were told that the rating was not being timed, but that the rating process would take about an hour and a half, though they were free to take breaks as needed.

Links to the sample OER were listed on a webpage provided to each of the raters. Each sample OER had a single sentence reuse case associated with it. The reuse case was meant to help focus raters if there was uncertainty regarding how to rate a particular resource. The reuse case was also meant to address concerns that Delphi study participants had regarding what the rubric was rating.   The ALMS rubric was converted to an online survey using the Qualtrics survey software. Each sample OER had its own unique URL to associate responses with the OER being rated. After the training was complete, it took about two hours for all of the raters to complete scoring of the samples.

Measurement of inter-rater reliability. Once the training was complete, the scores generated were analyzed. The first, and most important, analysis was inter-rater reliability. An intra-class correlation was calculated based on rater scores.   If the intra-class correlation

exceeded 0.61 further data analysis continued. The 0.61 level was chosen based on Shrout

(1998) which considered 0.61-0.80 to be "*moderate*" agreement. Use of the word "moderate" is

based on a revision of the Landis and Koch (1977) taxonomy of adjectives to describe levels of

agreement (pg. 309). Shrout (1998) stated the following regarding the moderate level:

> "Moderately reliable measures would still lead to bias as explanatory variables in
>
> regression analyses, and would be somewhat inefficient as outcome measures, but these
>
> untoward effects are limited. Substantial reliability in the only category where the
>
> adjective conveys the sense that measures are generally adequate" (pg. 309).

Although the ALMS rubric was developed with the intent of achieving the highest level

of agreement, a moderate level was deemed sufficient for this study. It was deemed sufficient

for two reasons. The first reason was that this ALMS rubric represented the first attempt to

systematically measure the obstacles in reusing OER. For that reason it was not likely that the

rubric would reach agreement above 0.81. Furthermore, because there are no known equivalent

rubrics, it was considered more beneficial to the OER community to release a rubric with

moderate agreement rather than waiting for a rubric with substantial agreement.   The second

reason for the 0.61 level was time taken to develop the rubric. Delphi studies are time-intensive

revising the ALMS rubric to reach a 0.81 could take any number of rating and revision exercises

if the level of agreement could be reached at all. Therefore, it stood to reason that a modest level

of agreement would be deemed acceptable for this study with the understanding that lower

agreement levels equaled lower confidence.

Once inter-rater reliability was established, the ratings underwent several analyses. The

means among raters, disciplines and repositories were compared. Since the samples of OER were

rated in sets of three, the mean scores of the sample based on their order will also be compared.

The three parts to this study (i.e. the rubric development, Delphi study, and inter-rater reliability) represents one of several different ways to systematically examine the obstacles of OER reuse from a technical perspective. The technical reuse rubric was an implementation of the ALMS analysis concept. This rubric was further refined by an expert-driven Delphi study that attempted to incorporate much of the existing understanding of OER reuse. Lastly, a representative sample of OER was rated and these ratings were then analyzed for patterns.

The underlying purpose of this study is to work towards better understanding the possibilities revision and remixing, which are central to purpose of OER. If scholars and practitioners better understand revision and remixing, they can develop and implement best practices. However, if over time it is found that anything more complex than basic resources can be practically shared, then the goals of the OER community would be brought into question. Therefore, this study was considered valuable in establishing the efficacy of OER.

Chapter 4

Results

This study centered on the creation and validation of the ALMS rubric. Development of

the ALMS rubric took place in several stages. The initial draft rubric development was

completed as part of establishing the methodology of this dissertation. This chapter discusses the

process, some of the logistical challenges, and decisions made developing the ALMS rubric.

Refinements were made to further improve the rubric. Results of an exercise to determine the

degree of inter-rater reliability were reported.

Refinement of the ALMS rubric (Delphi Study Findings)

The rubric went through four revision rounds based on four cycles of a Delphi study. Not

all portions of the ALMS rubric were reviewed during each round of the study to limit the time

commitment on Delphi study participants. Several papers noted that time constraints are a

common problem with the Delphi study methodology (Woudenberg, 2006; Pollard et al., 2009;

William & Webb, 1964; Gordon & Helmer, 1964).

In the first round of the Delphi study, the entire rubric was presented. Presenting the

entire rubric to participants gave them an opportunity to review the rubric in its entirety before

giving specific feedback. This procedure gave me as the researcher an opportunity to make

major changes to the rubric before subsequent rounds. After the first round, two sections of the

rubric (e.g. Access to Editing Tools, Level of Expertise) were considered for the second and third

rounds. Rater fatigue after the first round resulted in establishing of a 3.25 as a threshold for

sufficient agreement. The fourth round only considered one section of the rubric because the

other sections of the rubric met the 3.25 mean across all three categories. Each label of the

rubric was rated on a scale of 1 to 4 across the three categories (Reliability, Desirability, and

Feasibility). The definitions for each of these terms was explained Chapter 3. At the end of each round, the mean score for each category was calculated along with a total mean across all categories.

Murray and Hammons (1995) suggested that a Delphi study should end when stability is achieved. Powell (2003) examined Delphi studies to determine how consensus was considered to be achieved. The review concluded there was no universally accepted way to determine consensus. Powell noted that one method was determining a particular level of agreement ranging from 55% to 100%. I decided that a mean score of 3.25 across Reliability, Desirability and Feasibility metrics constituted sufficient agreement. Because it took longer for some sections of the rubric to achieve this level of score, some sections of the rubric received more feedback than others.

In addition to calculating the mean across categories for the rater, the results of the first round were further analyzed to better understand the relation between raters and scores. The first analysis was a comparison of means among raters across raters for the 72 labels rated combining Reliability, Desirability, and Feasibility scores into a single overall score.

The mean of the raters was above 3.00, with the exception of Rater 2. The lower mean for Rater 2 may be attributable to conceptual concerns about the rubric itself. These concerns are discussed below. The standard deviation of the raters suggests that although the means were relatively high, scores varied to a fair degree.

The Reliability category had the lowest mean for the majority of the raters. This low rating suggested some concerns about the ability to score that label reliably. The standard deviation of some categories is 0.00 because they gave the same rating for that category over the

entire rubric. The uniformity of scores was concerning and may indicate raters did not closely review some parts of the rubric.

Table 3

Overall mean and standard deviation from first round Delphi study

Rater	Mean	N	Std. Deviation
Rater 1	3.17	72	.504
Rater 2	2.69	72	.464
Rater 3	3.81	72	.597
Rater 4	3.04	72	.777
Rater 5	3.22	72	.537
Total	3.19	360	.685

In addition to analyzing means, a series of ANOVAs were done to look at the differences between questions and raters. Raters were compared using a Tukey's HSD *post-hoc* test.

The *post hoc* tests showed that Raters 2 and 3 were significantly different from the other raters. With respect to the mean, Rater 2 was notably lower than the average while Rater 3 was notably higher than average. Rater 2 may have scored lower on average because of free-form comments made about the rubric that questions how accurate the rubric might be given the variety of contexts that reuse occurs. Rater 3 may have been higher because that particular rater likely had the most practical experience rating OER.

These analyses show that although the majority of ratings and raters were positive with respect to the rubric, additional iterations were needed to enhance the instrument's construct

validity. However, the Access to Editing Tools section of the rubric needed the most

improvement.

      Access to Editing Tools. The idea of access to editing tools was difficult to represent in a

rubric. Any number of factors could contribute to the ability to reuse. Time, language,

bandwidth, and type of operating system could all influence whether a particular tool is

accessible or not. I considered hypothetical situations to think through issues regarding access.

As a result, nine questions were developed for the Access to Editing Tools section of the rubric

(see Appendix A).

Table 4

Mean and standard deviation by rater and category for first round Delphi study

Rater	Category	Mean	Std. Deviation
Rater 1	Reliability	2.83	.504
	Desirability	3.20	.408
	Feasibility	3.48	.511
Rater 2	Reliability	2.83	.381
	Desirability	2.44	.507
	Feasibility	2.83	.388
Rater 3	Reliability	4.00	.000
	Desirability	3.44	.917
	Feasibility	4.00	.000
Rater 4	Reliability	2.83	.917
	Desirability	3.04	.611
	Feasibility	3.26	.752
Rater 5	Reliability	2.83	.381
	Desirability	3.80	.408
	Feasibility	3.00	.000
Total		**3.19**	**.685**

Table 5

Mean and significance by rater and category for first round Delphi study

(I) Rater	(J) Rater	Mean	Sig.
Rater 1	Rater 2	.472	.000
	Rater 3	-.639	.000
	Rater 4	1.25	.704
	Rater 5	-0.56	.979
Rater 2	Rater 1	-.472	.000
	Rater 3	-1.111	.000
	Rater 4	.764	.004
	Rater 5	-.528	.000
Rater 3	Rater 1	.639	.000
	Rater 2	1.111	.000
	Rater 4	.764	.000
	Rater 5	.583	.000
Rater 4	Rater 1	-.125	.704
	Rater 2	.347	.004
	Rater 3	-.764	.000
	Rater 5	-.181	.347
Rater 5	Rater 1	.056	.979
	Rater 2	.528	.000
	Rater 3	-.583	.000
	Rater 4	.181	.347

Note: Standard Error was .098 across all comparisons

The first question was whether the appropriate software was pre-installed with the operating system.   From the perspective of an educator who is looking to reuse a resource, they would naturally look for a program that is already available to them. Programs such as TextEdit and Wordpad, which are available on Mac OS X and Windows respectively, could open and edit several different file formats. Because the ability to open files was a binary distinction, the 2 and 3 for that criterion in the rubric were not available, which left the only option a yes/no question.

The second question related to whether the application is downloadable from the Internet. If a hypothetical educator was unable to find a program preinstalled on their computer they might search the Internet for a suitable program. The criterion for this second question was challenging to construct. Some file formats are easier to find than others. For example, finding a program that edits PDFs would be fairly straightforward (though there are other obstacles to reuse). In contrast, finding an appropriate program to edit Real Media files could be more difficult. Initially the criteria specified a length of time associated with the search (e.g. found through 15 minutes or more of search), but it was decided that this was too time-consuming to accurately measure. Additionally, most users of the rubric would not actually test those criteria. Instead, the initial draft included broad language to indicate the amount of effort it took to find the appropriate software.

The three questions that followed are related to the software application downloaded from the Internet. If no software was available, then the response would be "N/A". The first in this series of questions was how much the application costs. Since educators have varying degrees of financial ability it was deemed appropriate to consider this obstacle. The language in the criteria was broad enough to allow for interpretation depending on context. The second

question related to the size of the program. In many international settings, bandwidth availability

can be highly variable and expensive (Singh, 2000). As with the other questions relating to

programs downloaded from the Internet, the broad language was meant to enhance flexibility.

The third question was whether the program was dependent on other software or programming

libraries. The question mostly considered Linux operating systems where a program might be

available for free, but require several libraries to install. This task can make the process

complicated, time-consuming and an obstacle to reuse.

Once a hypothetical educator has looked for a program pre-installed on the computer or

searched for a program to download, a next step might be to search for a web service to assist in

reuse. The first question was whether it is available with or without restrictions. "Restrictions"

refers to web services that allow for some remix with some kind of caveat. For example, a video

conversion website might allow for file conversion, but would place a watermark on the image,

or only permit videos of a certain duration. The next question related to cost. Some web

services have paid options but the cost may be prohibitively high. The final question asked was

whether it was necessary to establish an account in order to use the web service. In a sense, the

act of creating an account was considered a minor, but notable, barrier to reuse.

The final question of the Access to Editing Tools portion of the ALMS rubric was

whether text could be cut-and-pasted into a word processor. There was some concern about

adding this question, because it was specific to a type of media whereas the previous questions

could apply whether an educator was revising or remixing an image, video, or text. Nonetheless,

it was deemed important to ask whether cut-and-paste was applicable as an educator might use

that as one possible way to revise or remix. This question remained in the rubric.

Although these questions were intended to best capture the technical barriers to reuse, the questions posed by the rubric demonstrate the difficulties of assessing the technical context of an educator. Responses from the first round of the Delphi study indicated that the Access to Editing Tools needed the most revision.

First round of Delphi. Response from the first round of the Delphi study was a mixture of positive and negative reactions. The mean of ratings for the Access to Editing Tools in all three categories ($M = 2.87$) was notably lower than the mean for the entire rubric ($M = 3.19$). The standard deviation for access scores was high ($SD = .84$). One reason for this lower mean score can be attributed to Rater 1 ($M = 1.84$). This low mean may be attributable to concerns given in the free-form comments:

> Especially for the first two sections, but really for all of the sections, I am having trouble considering each rubric item in the absence of a specific revise/remix context. So, for my purposes, I have adopted the following scenario as a default. My responses in the first two sections are based, at some level, on this scenario; my responses for the last two sections were less dependent on this specific scenario. A teacher would like to assemble a digital "handout" for her 9$^\text{th}$ grade class. She intends to re-use parts of the resource for a class presentation, but the resource will also be uploaded to the class LMS so that the students can work from and refer to it later. She will need to find and then incorporate text, images, and video (with sound) into the resource. She will need to alter some of the text to fit the new context, and she will need to trim and embed the video pieces. She would also like to provide either a dubbed or captioned version of the video for her ESL students. The new, remixed resource will be available as a single file (or at a single URL) with embedded links, as appropriate, to the source materials…

> This issue is comparable to the endless discussions around "quality." What is a "high quality" resource for one audience might be totally inappropriate (and therefore of "low quality") for

another audience – while some "quality" attributes are objective elements of an OER, many attributes are not, especially those dealing with questions of "appropriateness" and "utility."

Rater 1 suggested in another section of the free-form comments that this concern relates more to the Access to Editing Tools and Level of Expertise, since the ability to meaningfully edit and the availability of source files are more objective measures and less dependent on context. Although this comment validated concerns about context with the Access to Editing Tools portion of the rubric, the comments left concerns on how to revise the rubric as a whole.

Revising the rubric to take hypothetical use cases into account would have been problematic. Although use cases were considered when constructing the rubric, they were only done to find conceptual flaws in the criteria considered. The intent behind the rubric was to create criteria that might account for the majority of potential uses. If the rubric was built around several use cases it would potentially have been limited to those use cases and reduce the overall utility of the rubric. Nonetheless, context did seem crucial to understanding how to accurately measure the difficulty of accessing tools for remixing.

Free-form comments from other raters also shared concerns about context. Many of the raters were concerned about the broad language for several of the criteria. Raters wanted specific numbers stated for cost and file size because of the varying degree that these issues are a problem for educators. As stated earlier, the intent was to keep the language broad so the rubric could be interpreted in local contexts. Although I considered that idea to be worthwhile, it also became clear that the broad language may cause problems for achieving inter-rater reliability.

The combination of these concerns necessitated a drastic look at how the concept of access was addressed in the rubric. The derived solution would need to consider context, at least

conceptually, while simultaneously ensuring enough stability in scores to achieve the necessary degree of reliability.

Although there were concerns about the Access to Editing Tools portion of the rubric, I remained convinced that the ideas placed there were appropriate. The proposed combination of pre-installed software, programs downloadable from the Internet, and web service was not at issue. Rater 5 suggested that the question about programming libraries should be merged with the question of whether the software can be downloaded. However, none of the other questions were deemed inappropriate by any of the raters.

If questions were considered reasonable by participants, and the goal was to make the rubric as widely applicable as possible, it stood to reason that the scores might be better suited to represent a general accessibility to editing tools, rather than the editing tools in a specific context. Assuming a score could be given that best represented the general accessibility to editing tools, it was considered possible to give a pre-determined score to represent access. That pre-determined score would be then averaged with the scores from other parts of the rubric.

It was decided that the easiest way to determine the general accessibility to editing tools was through the file type (e.g. PDF, DOC, and MP3). A fair amount of information can be determined from the file type. For example, if the OER being measured was in the MP3 format, then it is understood to be an audio file with several programs available for revising and remixing. A PDF file might have more possibilities, since it most commonly contains text and images, but it still has several programs or web services that could be used for editing. Each of these file types could be given a score based on the responses I gave to the questions in the rubric.

A second consideration also needed to be given to the operating system. Rater 4 noted that the software pre-installed on a computer varies by operating system. Therefore, the pre-determined scores would take the operating system into account. The intent was to create a table of scores based on common file types associated with OER and typical operating systems used. This table could anticipate the majority of likely combinations that an educator might encounter to provide the best possible utility for the rubric. Furthermore, the table could conceivably be automated into a form that the user is prompted for the file type and operating system and receives the Access score in return.

The first step in creating the Access table was to divide OER into the major media categories: image, video, audio and text. Then, based on my experience, generate a list of the most common file types for that particular type of media. These file types are listed as part of Table 6.

The next step was to rate them across each of the three major operating systems: Windows, Mac OS X and Linux. Note that Linux is referred to as an operating system, even though it might be better described a collection of closely related operating systems sharing the same kernel. Much of the scoring was done based on my expertise, though some research was necessary to verify file compatibility with certain programs. The scores to each question, like the first round of the study, were on a scale of 1-4, with 1 suggesting an answer unfavorable to revision or remix. The score to each question was averaged for all of questions in Access to Editing Tools portion of the rubric to provide a final score. In addition, a list of 2-3 possible tools was given to demonstrate how an OER of that file type might be remixed.

Table 6

List of file types for Access to Editing Tools table (second round of Delphi study)

Image	Video	Audio	Text
JPG	WMV	MP3	DOC
PNG	AVI	WAV	TXT
GIF	MPG	OGG	RTF
TIFF	SWF	AIFF	PDF
PSD	FLV	WMA	DOCX
	MP4	RM	
	MOV		

Second round of Delphi study. The construction of the Access table resulted in 522

individual scores for a total of 58 file type/operating system combinations. Consequently it was

decided that it would be too cumbersome to have participants rate each Access table score.

Instead, raters were provided free-form comments about the access table. Comments were

aggregated at the end of the round and were distributed to participants via e-mail.

Images. The first question raised by Delphi participants with regards to the Windows

scores was whether Photoshop files (PSD) should be included. The PSD format was included

because it was one of the most common formats that permitted multiple layers in an image.

Rater 3 expressed concern because the PSD file format was intended to be used with one

program, Photoshop. Other file formats such as PNG or JPG are not associated with a particular

program. Although, the PSD file was different than the others, I felt that its inclusion was still

warranted because of its popularity.

Another PSD issue related to the Linux operating system. Photoshop does not have an official version of the program that works for Linux. However, Photoshop can be used on Linux through use of an emulator. One of the most common emulators for Linux is WINE. WINE allowed users to run Windows software with one of the Linux distributions. One of the raters felt that using WINE was equivalent to using another operating system and did not reflect actual conditions when using a Linux distribution. I agreed and this item was adjusted in subsequent rounds.

Another rater asked whether SVG should be listed. SVG is an open format without licensing or intellectual property constraints. SVG was not originally selected because the format has not gained widespread adoption. Despite low adoption, it was decided that inclusion in the Access table would take very little time and made sense given the openness of the file format.

Video. There were fewer comments regarding video editing tools. However, those comments illustrate some of the challenges in measuring the technical context of video files. These challenges may have also applied to other media types.

The first comment with respect to video related to file conversion. Video files, as with other files, can be converted to different formats. Since the Access table was based on the file format, there was concern that the ability to change formats was not taken into account. This comment did raise a valid point, because there were many free tools that allowed for file conversion. Although there was the possibility that an educator might use video conversion software, I believed that a typical educator would not do so. It was believed that many educators would prefer to abandon a resource rather than go through an additional step. Nonetheless, it should be acknowledged that this assertion should be researched further with actual educators.

The second comment related to a concern about the nature of editing. More specifically, video tasks can potentially involve several skills or only a specific set. One rater phrased their concern as follows:

> I can't help but feel that these rankings need to be specified to reflect the specific nature of the edits in question. The challenge with video is that there are so many more things that someone could do, assuming proper expertise and tools. The ratings here seem to presume that "remix" will only consist of cutting, pasting, and stitching together different video segments. There might some additional voiceover or other sounds, and there might be some cute graphical effects. But it is also theoretically possible to literally edit the videos - use different actors in certain scenes, replace background imagery, etc. I agree that this sort of editing is currently outside the technical abilities (and interests) of the vast majority of people, but perhaps not for long. A simple statement that codifies the scope of the edits that we believe to be of interest, and therefore relevant to these ratings, would be handy.

This comment illustrates the wide range of potential reuse cases that can be conceived for OER. Although this rater's point had merit, the purpose of this rubric was to account for the typical educator. Because this issue was not one they might commonly face, the criterion was not changed.

Audio. There was only one detailed comment regarding audio files. The reason for the lack of comments may be that participants already raised similar concerns with respect to video. In the audio portion of the Access table one of the possible editing programs listed for Linux was Adobe Audition. As with Adobe Photoshop, using Adobe Audition on Linux would require WINE or a similar emulator. One of the raters suggested that using an emulator to run a program

on Linux meant they are not really using Linux. I agreed that this concern was valid and edited the Access table to remove Adobe Audition from the list of editing programs for Linux.

Overall. The feedback from the second round was helpful in clarifying acceptable reuse scenarios and challenges in presenting context. As with the first round, the participants did not have any concerns about the questions themselves. One of the raters did note that it was confusing to list questions such as "Are you able to cut-and-paste the text into a word processor or text editor?" in each part of the rubric, even though the cells listed a not applicable or "N/A" score. The series of questions relating to the web services were also duplicated across all operating systems, and this was confusing to at least one rater. The questions were kept in the rubric, at least for subsequent rounds, to maintain a parallel structure. Keeping the questions in the rubric did not influence scoring. However, I acknowledged that the future versions of the rubric may need to change to better clarify which questions are actually applicable to specific media types.

Another comment about the rubric overall had more significant implications. A participant suggested that the variance of scores among operating systems was not particularly notable. If true, that would suggest that the file format was far more important than the operating system. There are potentially three reasons why the importance of the operating system may be minimized in instances of revising and remixing. First, the very nature of web services means that they can be run on any operating systems that have a web browser installed. If there is a difference in ability of web services it is usually because the web browser does not, or cannot, fully produce the necessary functionality.

The second reason why an operating system may have less bearing on the scores was because some of the editing programs were available on two, and possibly three, operating

systems. For example, photo-editing software was available for Windows, Mac OS X and Linux

distributions. The text editing program Libre Office (formerly Open Office) was available for all

three major operating systems as well.

The third reason that operating systems may have less influence than previously thought

was the number of files that could be edited by a given program. Since the file types chosen

were some of the more popular, it stood to reason that there would be a least one program

available for each operating system that allowed for revising and remixing. The minimal

difference forced me to reconsider creating separate Access tables for each operating system.

The feedback from the second round of the Delphi study demonstrated that the concerns

were largely that of formatting and changes to individual scores. The largest conceptual

challenge was the lack of distinction among scores based on the operating system. Alterations to

the rubric based on this challenge, as well as others raised in that round, were made in

preparation for the third round of the Delphi study.

Third round of Delphi study. The first step in revising the Access table for the third round

of the Delphi study was to adjust scores based on feedback from participants. The PSD file

format score was the most significant change. Because the use of emulators such as WINE to

simulate a Windows environment on Linux was deemed inappropriate, the score was lowered

significantly. Other changes to scores were mostly minor adjustments, often less than a .5

difference to the total score.

The next step in revising the rubric was to create a single score that represented the

access to editing tools across Windows, Mac OS X and Linux operating systems. To do so, the

means of scores across all three operating systems was calculated. The rationale for this

calculation was that a mean score would represent access, or lack thereof, in the broadest and

most inclusive sense. A weakness to using a simple mean was that it may not have properly represented what a specific educator might face. Although Linux has been adopted in educational settings, the majority of educators in the US were likely to use Windows or Mac OS X for their daily work tasks. That meant a unified score might underestimate or overestimate the difficulty of reusing OER if the score for that particular operating system was notably different than the others. Because the scores were similar enough in many instances, it was determined that a simple mean for a given media type across all operating systems was appropriate.

The response from the third round of the Delphi study was not as detailed as the previous two rounds of the study. There might be any number of explanations for less participation, but two seemed to be most likely. The first possible reason was that participants were becoming increasingly fatigued with the response process. One participant did decline to provide feedback in the fourth round of the study for that reason. The fatigue likely came from rating each individual label of the ALMS rubric. The overall length of time taken for multiple rounds of the Delphi study was likely a contributing factor as well. The fifth chapter of this dissertation will discuss Delphi study participation in greater detail.

The second possible reason was that the Access table was achieving agreement among raters. There were fewer conceptual concerns in this round. Free-form responses from the round such as "seems fine" added some credibility to this conclusion.

Since the responses from the third round were minimal, the feedback will be discussed broadly, rather than separating each response by media type. One of the first items of feedback was an adjustment to the image scores. This concern related to the assertion within the third round rubric that image editing software was pre-installed with the Mac OS X operating system. The feedback highlights one of the difficulties in rating editing tools on Mac OS X. The OS X

operating system, with some exceptions, was only available on Macintosh (Apple) computers. When someone, including a typical educator, purchases a Macintosh computer it typically comes with the iLife software suite. This software suite typically has programs that can be used for revising and remixing. For example, iPhoto allows users to edit and revise photos, including commonly used image files.

One of the raters pointed out that technically the operating system itself does not have those programs pre-installed. Therefore, the score should not be a 4 to indicate high reusability. Although this argument was technically accurate the score was ultimately not changed. The rationale for keeping the score the same was, for the typical educator, the distinction between pre-installed programs and programs that are part of the operating system was not important. The score was kept for the next round of the rubric, but that decision could be reversed in future iterations.

Summary of findings regarding Access to Editing Tools. The significant changes that the Access to Editing Tools portion of the rubric went through as part of the Delphi study were unanticipated. During the development of the rubric, the Access to Editing Tools was the most debated and discussed. An all-encompassing rubric appeared impossible, because of the diversity of editing situations. Nonetheless, it is worth noting that the study participants seemed satisfied with criteria selected.

Use of an Access automated decision table, rather than a rubric, both solved and created problems. An Access table provided consistent scores based on the file type, which was an objective measure. Because the Access table has scores for the most commonly used file types, it could represent much, though not all, of the hypothetical revise/remix use cases. The table presented the rationale for the score in a transparent manner. The simplicity of the Access

formula allows for granular changes necessary as the technological context that educators work in evolves.

The Access table did have disadvantages. As stated earlier, scores on the Access table are entirely dependent on the file type used. Although the file type can play a significant role in the access table, it does not account for all aspects of access. For example, consider an application that costs $20 for one license. Individually this cost may seem reasonable, but to scale that cost across several dozen computers, as a department at a higher education institution might, may be prohibitive. For the purposes of rating, the reuse/remix was presumed to be an individual effort. A related concern, raised by one of the study participants, was that some file formats can be converted into different formats. If a technologically adept educator can change the file format, then that complicates scoring for possible revising/remixing scenarios. As stated previously, the ability to edit and change was not judged to be overly compromising to the Access table conceptually because the file conversion process involves an additional barrier to reuse beyond the act of reuse.

The continually shifting technological landscape presented another limitation of the Access table. For example, the WebM video format was an emerging format supported by Google (WebMproject.org, n.d.), but not included in the Access table. This format was being adopted by some web browsers, but the long-term future of the format remained uncertain. Although the longevity of a format could have been factored into the rubric, it might have changed the scores so they are not reflective of the challenges faced by educators currently remixing and revising. Given that, it was conceivable that a file format might score well with respect to Access to Editing Tools at a given point in time, but score poorly a year or two later.

Therefore, the Access table, by necessity, would need to be revisited periodically to update the scores reflecting the current technology environment.

Though these limitations are notable, the Access table did seem to be the best way to ensure consistent scores that are maximally representative of the difficulty experienced by educators. Because the number of scores in the Access table were more than what the raters could score practically, this section was considered complete after changes made following the third round of Delphi study. Future research regarding the Access table will be discussed in Chapter 5.

Level of Expertise. As with the Access to Editing tools, it could have been be argued that the Level of Expertise portion of the rubric was a function of the context of the revise/remix situation. However, work went forward developing an approach by which the level of expertise could be calculated by the nature of the resource. The challenge was to delineate revising/remixing expertise into smaller, more measurable constructs.

 The most logical way to separate the various components appeared to be by the different media types that make up a resource. For this rubric, media types were divided into the following categories: text, image, audio, video, programming, and specialty. The programming aspect of expertise was meant to address any kind of conditional logic employed in the resource. For example, an interactive slideshow might have some kind of scenario-based resource where different users have a unique experience based on their responses. In that situation, the resource would be considered to require programming expertise, even if it was simply a matter of planning links within PowerPoint.

The last category, specialty, was meant to be a miscellaneous category that would anticipate any other specialized knowledge necessary to revise or remix the resource. At the

time of the rubric's construction, the specialty category was intended for animation, which did

not fit any of the other categories satisfactorily. It was anticipated that many resources would be

simple with respect to the number of media types used. In some cases, a resource would only

consist of one media type. This diversity in resources posed challenges to the accuracy of the

rubric. As shown later in the chapter, some media types like animation were so difficult to

account for that they were ultimately not considered by the rubric.

Similar to the Access to Editing Tools portion of the rubric, the first draft of Level of

Expertise section consisted of four levels of criteria. A "1" indicated extensive expertise would

be required, a "2" significant expertise, a "3" some expertise and "4" indicated no expertise was

needed. After applying the draft rubric to a sample of OER it became clear that the difference

between a level 2 and 3 was not significant. Therefore, it was proposed that level 2 and 3 be

merged to become 2.5. Merging the two levels reduced the overall structure for that part of the

rubric to three levels (see Appendix A). These levels correspond roughly to the common

understanding of beginner, intermediate and advanced. The disadvantage to these levels was that

it would no longer run precisely parallel with other parts of the rubric.

Another part of the rubric that was revised as part of this initial development phase was

the description of the criterion. To address the difficulty of expressing an abstract level of

expertise, one example was given for each criterion of the rubric. For example, extensive

expertise with respect to text editing uses the example of formatting a multi-column layout.

Some video expertise was defined as adding transitions or filters to a video clip. These examples

were intended to help raters calibrate their answers.   Because only one example was provided,

and no single example could be all encompassing, the ability for these examples to assist in

rating extensively was limited.

First round of Delphi study. Next to the Access to Editing Tools portion of the rubric, the

Level of Expertise scored second to lowest (Access to Editing Tools Mean=2.87, Level of

Expertise, Mean=3). Some of the reasons for the low score are evident in the free-form

comments discussed in this section. Descriptive statistics are also illustrative of both the lower

mean and disagreement among raters.

Rater 4 scored this portion of the rubric lowest. The fairly high standard deviation for

Rater 4 (SD=.492) was largely the result of the slightly higher scores for the Feasibility aspect of

the cells (M=2.65, SD=.493). Rater 2 had the highest standard deviation (SD=.634). That rater

seemed most concerned with the Desirability aspect of the cells (SD=.562) while giving the same

score, 3, across all Reliability and Feasibility ratings. Rater 1 did not distinguish between any of

the aspects of the cell and gave the entire portion of the rubric a rating of 3. Rater 3 was close to

rating the entire portion a 4, with only some deviation (SD=.231).

Table 7

Mean scores and standard deviation for Level of Expertise (first round)

Rater	Mean	Std. Dev.
1	3	.0
2	2.78	.634
3	3.94	.231
4	2.39	.492
5	2.89	.572
Mean	3	.386

The results from the rubric show the diversity of thought among raters. Some, like Raters 3 and 4, seem to make a broad judgment as to the Reliability, Desirability, and Feasibility of measuring the level of expertise needed. Rater 4 clearly viewed the Reliability and Desirability of this part of the rubric as problematic, perhaps because the Reliability and Desirability were considered more context-dependent.

Most of the free-form comments made during the first round of the Delphi study addressed the rubric as a whole or Access to Editing Tools. As mentioned in the Access to Editing Tools section of this chapter, one of the raters had concerns about the desire for context when considering the particular revising/remixing situation. In that comment, the rater compared rating the reusability of object as having the same subjectivity as rating the "quality" of an OER. The comment referenced several debates within the OER community regarding how to determine the quality of OER. Implicit in the comment was the understanding that those debates have generally been viewed as fruitless.

Adequately addressing this concern, whether it was for the Access to Editing Tools or Level of Expertise portion of the rubric, was out of scope for the study. The very nature of this rubric presupposed that the variation in context was sufficiently minimal such that a rubric could account for the difficulty in revising/remixing in a sufficiently consistent and reliable manner. Additionally, during the first two rounds of the Delphi study, the rubric was meant to measure the potential of the resource to be revised or remixed. As an example, an interactive slideshow might include audio, text, and video. In order to have the maximum flexibility in revising or remixing an educator would need to have expertise in audio, text and video. The necessity of these skills for the greatest potential in revising or remixing means that there would be greater difficulty even if the particular context only calls for skills in one media type. This concept of

maximum reusability was something that was revisited with respect to ALMS rubric rater training. During rater training the raters were told to consider the maximum reusability of an object.

There were two other free-form comments that were narrower. Rater 5 commented on the difficulty of having someone rate programming expertise, even if they had some technology expertise. This same rater was also concerned about the specialty expertise rating and considered it "too vague." These two items were revised in the next round of the study.

Second round of Delphi study. The changes to the Level of Expertise portion of the rubric were not as drastic as the changes to the Access to Editing Tools. The comments and scores did not suggest extensive revisions were needed, other than what was raised in the first around about the conceptual objectives of the rubric.

The first step in revising the rubric was to address the concerns about the programming and specialty expertise. After discussion, I decided the categories were too vague for raters to use accurately. Consequently, the specialty expertise question was removed from the rubric. The programming expertise item was collapsed to become a binary question asking whether any programming expertise was needed. In other words, the 2.5 rating was eliminated for the programming criterion, leaving only a "N/A", 1, or 4 rating possible.

The second step in revising the rubric was to add detail to the examples in some cells in this portion of the rubric. For example, in the first round the cell for extensive expertise gave the example "multi-column layout." This example was expanded in the second round to "Setting up a multi-column layout; extensive use of graphics and text. Example: newsletter layout." These examples provided greater clarity to raters regarding the intended differences between scores.

One participant dropped out of the study between the first and second rounds. Second round scores indicated a higher degree of satisfaction with the Level of Expertise section.

Table 8

Mean scores and standard deviation for Level of Expertise (second round)

Rater	Mean	Std. Dev.
1	3.81	.587
2	3.90	.293
3	3.33	.713
4	2.79	.558
Mean	3.46	.538

The mean increased by almost .5 from the first round of the study (M=3.46). Interestingly, the standard deviation increased as well (first round SD=.386). The increased standard deviation suggested that some raters, particularly Rater 1, viewed measuring the Level of Expertise with increased scrutiny. The questions relating to text and images scored the highest (3.61 and 3.63 respectively). The question relating to programming expertise, despite the changes made from the first round, scored the lowest (M=2.99). The expertise of the raters may be indicated by the means of the criteria. Text and image editing are more likely to be prevalent in comparison to programming expertise. Alternately, the low mean may represent a lack of confidence in potential raters to determine the necessity of programming given a particular resource. Regardless of the reason, the lower rating was concerning because the raters seemed to have difficulty in the concept of measuring the necessity of programming for reusability.

There was one free-form comment with respect to the Level of Expertise that was worth noting. The rater expressed concerns regarding what exactly was meant by Level of Expertise:

It is still quite confusing, but I appreciate that you are trying to simplify a complex space. It is hard to keep straight what exactly it is that is being rated here. Is it the person, the activity, the resource in hand, the resource we desire, etc?

This ambiguity was understandable and something that was encountered when constructing the rubric. As mentioned when discussing comments from the first round of the study regarding Level of Expertise, the rubric was constructed with the intent to look at what sort of expertise was needed to exercise maximum reusability. This explanation was apparently not made sufficiently clear to Delphi participants. Still, even with an explanation, the potential confusion poses a problem for raters. During inter-rater reliability, which will be discussed later in the chapter, raters were given a sample reuse scenario to help guide scoring decisions.

At the beginning of the study it was decided that an overall mean for a section of the rubric that exceeded 3.25 was sufficient to consider the rubric complete. Because the Level of Expertise portion of the rubric exceeded 3.25 (M=3.46) this portion of the rubric was considered completed. Although the changes made to the Level of Expertise were not as dramatic to those made Access to Editing Tools portion of the rubric, the revisions are suggestive of the challenges in estimating expertise. The difficulty in separating the intended use of a resource and its attributes seemed to be difficult. Accounting for properties that fit outside the clearer lines of audio, video, and text also seemed to be problematic. The ability of this portion of the rubric to accurately measure expertise was reduced in nuance to enhance reliability. Nonetheless, through their scores and comments, the study participants seemed to find the Level of Expertise portion of the rubric conceptually consistent.

Meaningfully Editable

Conceptually, the Meaningfully Editable portion of the rubric was designed to account for incongruity within editing programs to revise or remix a file. Some programs offer an ability to edit a file, but the kinds of edits allowed are not always meaningful in the revise / remix context. For example, a PDF that consisted of scanned images may allow for notes or annotations to be added, but none of the underlying content would be editable. For that reason the word "meaningful" was important in describing what this section of the ALMS rubric was attempting to accomplish.

One of the first challenges of creating the Meaningfully Editable portion of the rubric was distinguishing it from the Source Files portion. In order to make meaningful edits it is necessary to have the source files. Therefore, the two concepts were considered conceptually correlated. However, the two concepts are different enough to constitute two distinct portions of the rubric.

For an educator attempting to revise/remix a resource, having the source file is not the same as being able to meaningfully edit it. Using the aforementioned example regarding a PDF with scanned images as an example, having the source image files would not necessarily enhance the revisability or remixability of the object. The source files are no more alterable than the resulting resource. Contrast this example with that of a JPG image file that was generated from a PSD, or Photoshop file. The JPG might be a composite of several images together. If the educator had access to the PSD file, they might have the ability to edit each of individual images that made up the composite.

Like the Level of Expertise portion of the rubric, this portion of the rubric was divided by media type in order to delineate more clearly in what ways the resource was meaningfully

editable. Instead of designating a programming category, the Meaningfully Editable portion of the rubric only contained a question asking whether other, or miscellaneous, components of the OER were editable. The difference in scoring was the amount of the OER that was meaningfully editable.

The scoring was dependent on the amount that could be meaningfully edited, therefore an example in each cell was not necessary. Instead, each criterion provided an example of how one part of a resource could be embedded in another (see Appendix A). For example, when assessing the amount of the video that can be meaningfully edited, the example given was "video stored independently, not in Flash object." Although these examples specify a particular file type, it was believed that raters would be able to successfully generalize the example to a given resource, regardless of its media type.

First round of Delphi study. Scores from the first round of the study were more positive than the Access to Editing Tools and Level of Expertise portion of the rubric. Still, raters were not uniformly positive in their ratings.

The means for Raters 2 and 3 were notably different (M=2.70, M=3.73). Rater 2 scored other portions of the rubric lower in comparison to other raters, so the relationship in this first round was not surprising. The scores for Rater 2 suggested that he or she lacked some confidence in the rubric conceptually. By contrast, Rater 3 had a high mean (M=3.73), though there was also a high standard deviation (SD=.686). In examining the scores, it was clear that high standard deviation for Rater 3 resulted from lower Desirability scores. Further analysis revealed that the lower scores were concentrated in the cells for the 2 and 3 scores for each of the criterion. These lower scores may indicate that the rater did not see the purpose of making distinctions between degrees of editability within the resource. The rater may have felt that what

was most desirable was binary distinction. In other words, a resource can either be meaningfully edited or it cannot.

Table 9

Mean scores and standard deviation for Level of Expertise (first round)

Rater	Mean	Std. Dev.
1	3.53	.503
2	2.70	.462
3	3.73	.686
4	3.25	.795
5	3.33	.475
Mean	3.31	.690

The mean for Rater 4 (M=3.25) was close to overall mean for raters (M=3.33). However, the standard deviation for Rater 4 was very high (SD=.795). The high standard deviation seemed to be due to lower Reliability scores. However, Rater 4 did give higher scores to Desirability and Feasibility. These scores seem to indicate that Rater 4 had concerns about being able to reliably measure the editability of a resource, but believed that it was desirable and feasible. The possibility of this concern was furthered by Rater 4. The rater stated "Audio and video can sometimes be edited through non-obvious means." The exact meaning of the comment was not clear. The comment may have been in reference to programming or settings

within an interface. For example, the audio in a video could be removed via settings in Flash, or specifications in an HTML tag.

The variety of editing methods was a concern when developing the rubric. Still, I believed that the rubric was comprehensive enough to encompass the majority of the revising or remixing scenarios. Nonetheless, it was worth noting that future raters using the rubric may be unaware of all the possible methods that can be used to meaningfully edit a resource.

Summary of Meaningfully Editable. Because the ratings from the first round of the Delphi study (M=3.31) met the threshold to be considered sufficient agreement (3.25), the Meaningfully Editable rubric was not included in subsequent Delphi rounds. Although the raters appeared to have some concerns, particularly scores of 2 or 3, raters acknowledged that this concept was important with respect to measuring the technical difficulty of revising or remixing OER.

The higher means in comparison to Access to Editing Tools and Level of Expertise may be related to the nature of the concept of editability. Raters expressed concerns about the first two parts of the rubric with respect to the generalizability of the criteria and their ability to remain relevant in different contexts. These concerns often came because of the relevant properties of an object change from context to context. In the case of the Meaningfully Editable concept, it may have been easier for raters to consider the editability of an object to be strictly a property of the object itself. Consequently, the scores from raters were confident that the Meaningfully Editable portion of the rubric could be reliable.

Source Files. The Source Files portion of the rubric was reflective of the nature of multimedia development. Within multimedia development there are two general types of files: the files that the end user interacts with and source files. The files that the end user interacts with are not typically designed for editing. In contrast, source files are designed specifically for

editing. For example, when producing a video, someone might have several video clips, audio clips, and a project file that are recognized by the editing program. The result of the editing would be exported into one file intended for consumption by end users. This distinction between distribution files and source files was not always present. A Microsoft Word document would be both a source and distribution file because all of the content intended for distribution would be available. Given the distinction it was necessary to assess whether the source files for a given resource were available.

As with the Meaningfully Editable and Level of Expertise portions of the rubric, the Source Files were divided by media type. The scoring between each question was determined by the amount of source files present for that media type. As with the Level of Expertise portion of the rubric in the first round, the Source Files section had two criteria assessing whether programming files or "other" components were available. The "other" category was meant to be a miscellaneous category that captured source files not considered by the other media types similar to the specialty category in the Level of Expertise section of the rubric.

Similar to the Meaningfully Editable section, the Source Files part of the rubric gave one example per criterion, rather than give one for each cell, assuming that this would be sufficient for raters. Each example was relevant to the media type in question. For example, the image source file gave the example of "a Photoshop image with all layers available." These examples share the same disadvantage with the Meaningfully Editable portion of the rubric in that the examples are not all-encompassing and some raters could be uncertain how to rate Source Files given a previously encountered type of OER.

First round of Delphi study. Response from the first round of the Delphi study was

positive, but not as strong as the Meaningfully Editable portion of the rubric. Though the scores

were not as high as Meaningfully Editable, the scoring by raters was similar.

Table 10

Mean scores and standard deviation for Level of Expertise (first round)

Rater	Mean	Std. Dev.
1	3.17	.504
2	2.69	.464
3	3.81	.597
4	3.04	.777
5	3.22	.537
Mean	3.19	.685

Rater 2 scored this portion of the rubric lowest (M=2.69). As mentioned previously,

Rater 2 consistently scored sections of the rubric lowest, likely because of the concerns about the

rubric overall. Rater 3 scored the highest of all raters (M=3.81). These raters were both the

highest and lowest respectively in relation to the Meaningfully Editable portion of the rubric.

The standard deviations for Rater 3 were slightly smaller in Source Files (SD=.597) in

comparison to Meaningfully Editable (SD=.686), which may indicate slightly more confidence

in the Source Files portion of the rubric. The scores for Rater 4 had the highest standard

deviation, as they did for Meaningfully Editable (SD=.777). Rater 4 did not make any comments

regarding the source file so no particular insight was gained, except the scores for Reliability are

lower than Desirability and Feasibility. These scores, similar to those for Meaningfully Editable, suggested that Rater 4 did not believe the Source Files could be reliably rated, though it was desirable and feasible.

Rater 5 explained their low Reliability scores given by stating "low [scores] because I expect most reviewers won't know enough to look around for the files." This comment presented a significant concern when rating. Depending on how the resource was presented, the source may be located in any number of places. It was easy to imagine that some future raters would have difficulty giving ratings in the Source Files portion of the rubric.

A solution was not immediately apparent. By necessity, at least with respect to the ALMS framework, the Source Files needed to be considered. I determined that the best solution would be to address this as part of rater training. The overall mean for the rubric (3.19) was close to the threshold that it would be considered sufficient, but did not reach it. Therefore, this portion of the rubric needed to be reviewed by study participants again.

One curious phenomenon with the criteria scores was that the audio criteria scored notably lower than other questions. The scores seemed to be lower because of the ratings given to the 2 or 3 labels for that criterion. The cause for this lower score was unknown.

The desired edits for this round had two purposes. First, the scores needed to be raised to meet the threshold score of 3.25. Second, changes made to other parts of the rubric necessitated revision to ensure a more parallel structure with the rest of the rubric.

The first change was to merge the 2 and 3 scores to form 2.5. The merging of these two scores was done with the Level of Expertise portion of the rubric prior to being distributed to Delphi study participants. The change was not made to the Source Files portion of the rubric at the time because it was not seen as necessary.

The second change was to eliminate the "other components" criteria from the rubric. This miscellaneous category was eliminated from the Level of Expertise portion of the rubric because it had scored poorly. The programming criterion was also eliminated from Source Files to match the Level of Expertise portion of the rubric after the first round.

The final change was to clarify the example used for the video criteria. Rater 5 found the one given in the first round to be too vague. Therefore, it was revised to clarify exactly what constituted a source file for video.

Fourth round of Delphi study. Though one of the aims of the Delphi study was to receive as much feedback as possible on each portion of the rubric, inevitably some limitations needed to be imposed. The first round of the Delphi study presented all portions of the rubric to study participants. As stated earlier, results of the first round created concern that the number of ratings required by study participants would discourage participation. Therefore, subsequent rounds of the Delphi study only presented two sections of the rubric at a time. The Source Files portion of the rubric was prioritized last because the mean of scores (3.19) was only .06 less than the threshold of 3.25.

The disadvantage of this approach was that participation was low by the final round. Even with reducing the number of sections reviewed per round of the study, the amount of rating needed was apparently too much. One rater contacted me to state explicitly that they did not see a purpose of rating the rubric in such detail. In reviewing the Delphi study process I believed that the rater was correct and only criterion as a whole should have been rated.

Consequently, only one participant responded to the final round of the Delphi study. The ratings given by Rater 1 suggested that concerns about the Source Files portion of the rubric remained.

Rater 1 seemed to have confidence that the source files for text components of a resource could be rated. However, Rater 1 scored the images media type lower than other any section. The reason for the low score was not communicated. It may be that the rater was unsure whether someone using the rubric could correctly identify whether a given image was in fact a source file. The other two media types, audio and video, received identical scores across all labels. Those ratings suggest that the rater believed that someone using the rubric was equally likely to identify audio and video source files. The rater did give lower scores for the labels with a score of 2.5 for images, audio and video. There may have been some concern with the ability to rate partial access to source files of that media type. No free-form comments were made on the rubric to provide additional clarification.

Table 11

Mean scores by question from fourth round of Delphi study (Rater 1)

Media type	Mean	Std. Dev.
Text	4	.000
Images	2.78	.786
Audio	3.11	.737
Video	3.11	.737
Mean	**3.25**	**.565**

The resulting overall mean (3.25) from the only participant in the fourth round of the Delphi study met the threshold score to consider this portion of the rubric complete. The lack of participation in the fourth round was concerning and meant uncertainty as to whether the scores were representative of how other study participants might have perceived the Source Files

portion of the rubric. Nonetheless, the score was sufficient to consider the rubric ready for inter-

rater reliability.

Calculation of Inter-Rater Reliability

Once each part of the rubric reached the specified point of convergence through the

Delphi study, the next step was to determine the degree of inter-rater reliability for people using

the final rubric. In parallel with measuring inter-rater reliability the intent was to measure a

sample of OER to determine scores typical for the sample's repositories. These scores would

give an indication of the amount of technical difficulty necessary to revise or remix the OER.

Initial analysis of the ratings consisted of basic statistics. These basic statistics were

calculated using SPSS and Microsoft Excel. The first analysis was to explore whether there were

any differences between repositories.

Table 12

Mean scores and standard deviation by repository

Repository	Mean	Std. Dev.
NROC	2.36	.33
MIT OCW	2.99	.2
WikiEducator	3.16	.15
Mean	2.84	.23

Mean scores indicate that WikiEducator was seen as the easiest to revise/remix, with

resources from MIT OCW the next easiest. Resources from the NROC repository were scored

the lowest, though the standard deviation indicates the greatest variability (SD=.33). The lower

standard deviation for WikiEducator and MIT OCW indicate less variability among ratings for resources in those repositories.

The next step was to analyze the ALMS scores with respect to the discipline of resources rated (i.e. humanities, social sciences and sciences). The intent of this analysis was to conduct preliminary examinations into the differences among OER from different disciplines.

Means between Humanities and Social Sciences were 0.19 apart (M=3.02 and M=2.83). The differences between Social Sciences and Sciences means was similar. The reason for the differences was unknown. The standard deviation for ratings of sciences resources was 0.10 higher than for the other two disciplines. The reason for this discrepancy was also unknown. It may be that at least some of the Science resources had a greater degree of multimedia than resources from other disciplines. Likewise, the Social Science resources could have more multimedia than Humanities resources. Because the number of OER rated was limited and not randomized these results cannot be generalized.

Table 13

Overall Mean of scores and standard deviation by type

Repository	Mean	Std. Dev.
Humanities	3.02	.16
Social Sciences	2.83	.19
Sciences	2.67	.32
Mean	2.84	.22

The next step was to analyze the scores according to the order that they were rated. Resources were rated in order of the repository and discipline. For example, all the NROC

Humanities resources were rated sequentially, followed by the NROC Social Science resources.

The assumption prior to the analysis was that the order would not affect the scores.

Table 14

Mean scores and standard deviation by order

Order	Mean	Std. Dev.
1	2.91	.17
2	2.81	.25
3	2.80	.27
Mean	2.84	.23

Mean scores indicate that there were slight differences among OER rated in a series. The

mean for resources that were rated first tended to be higher, while it decreased for those rated

second and third. The cause for this variability was unknown, but does not necessarily indicate

an ordering effect. The difference may be attributable to the resources themselves. The higher

standard deviation that resulted when a user rated the second or third OER in the series may be

because they become increasingly uncertain about scoring as they are exposed to more OER of

the same type. Because each rater scored the resources in the same order and the repositories

were rated sequentially and was not randomized, there was no way to determine whether order

effects did or did not occur.

The final analysis with respect to ratings was the differences among sections of the

rubric. The Access table was not part of the analysis, since those scores were automatically

generated based on the file type.

The mean for Source Files was the lowest, which suggests that some resources had no source files available. Meaningfully Editable was rated the highest because many of the resources had some degree of editability. High standard deviations among all sections of the rubric suggested that resources possessed the qualities measured to varying levels.

Table 15

Mean scores and standard deviation by section of ALMS rubric

Section	Mean	Std. Dev.
Level of Expertise	3.40	.58
Meaningfully Editable	3.53	.72
Source Files	3.18	.77
Mean	3.37	.69

Although examining the means of the ratings provided interesting information, what was most important was the degree of inter-rater agreement. An intra-class correlation coefficient was selected because it was applicable for studies where more than two raters were rating the same sample. The intra-class correlation coefficient for these scores indicated moderate agreement (ICC(2,1) =.655, df=376, 95% CI [.609, .699]).

In summary, an ICC of .655 met the minimum threshold of .60 that was pre-determined as being sufficient agreement before the beginning of the study. However, the degree of agreement did not indicate a high degree of agreement. The implications of the ICC will be discussed in chapter 5.

Chapter 5

Discussion

In this discussion the results of both the Delphi study and inter-rater reliability will be addressed. The practical application of these results will be proposed as well. Conclusions and recommendations for further research based on the study will be given.

Summary of findings

The purpose of this study was to address this question: Can the technical difficulty of reusing OER be measured using the ALMS framework in a way that is valid and reliable? The research required developing a rubric. Wiley and I, both experts in the field of OER, formulated a draft rubric for the ALMS framework. The process of developing the rubric was more challenging than anticipated. Some elements of the ALMS framework were more difficult to separate than others.   For example, the ability to access source files and whether a particular OER was editable were closely connected. Other parts of the rubric were challenging to operationalize, such as the Access to Editing Tools. This portion of the rubric was difficult because there are several ways that an educator might obtain editing tools to revise or remix a particular OER. The initial plan to maintain absolutely parallel scoring criteria (e.g. 1-4) proved problematic for some portions of the rubric such as Level of Expertise.

The Delphi study helped refine the construct validity of the rubric. Participants in the Delphi study had the greatest amount of concern for the first two portions of the rubric relating to Access to Editing Tools and Level of Expertise. These two constructs were perceived to be highly context-dependent and participants questioned whether they could be reliably measured. In general, participants also had concerns with criteria that asserted a degree of

revisability/remixability. These criteria, which had possible of scores of 2 and 3, rated lower with respect to Reliability, Desirability, and Feasibility in comparison to scores of 1 and 4. These slightly lower ratings suggest that the raters were more comfortable with a binary scoring in determining whether a particular media type (e.g. text, video) was meaningfully editable or whether the source files are available.

     Delphi study participants seemed to agree that the components of the ALMS rubric were consistent with the technical issues that educators face in revising/remixing OER overall. Many of the suggested revisions included changes to examples provided for individual criterion. Some criteria were eliminated from the rubric based on participant feedback. These criteria include addressing programming and animation ability. Although losing these criteria could be problematic in measuring a minority proportion of OER, doing so made the rubric more reliable for the vast majority of OER.

     The greatest change to the ALMS rubric was the conversion of Access to Editing Tools portion of the rubric to an Access table. In a sense, the Access portion is taken out of the rubric because the scores are pre-determined. By necessity, the Access table was limited to the most popular file types and there are some OER not accounted for by the table. Though the scores are likely to be more reliable, since they are based on a single objective measure (i.e. file type), the Access table will need to be reviewed periodically by experts in the field to ensure the appropriate file types are included.

     The third part of the study, in which a sample of OER was rated, indicated that at least some measure of agreement could be made with the current form of the rubric. This assertion was supported by the Intra-Class Correlation. The amount of agreement was high enough to reach the predetermined level of .60, but was not as impressive as originally hoped. This finding

suggested that further training of the raters may increase agreement though further research would be needed to verify.

The scores resulting from the sample rating were informative. The scores indicated that there were differences among repositories with respect to the remixability/revisability of their resources. However, inferences about the overall quality of the resources should not be made from these scores. The NROC repository resources tended to be the most sophisticated with respect to interactivity and that may have adversely affected the ability to revise or remix. Conversely, resources in WikiEducator could be revised or remixed to a greater degree, but may lack the refinement of the other two repositories. This finding harkens back to Wiley's (2001) decade old reusability paradox, wherein educational effectiveness and reusability are inversely related.

The difference in scores among disciplines was interesting. Further research should be conducted as to whether there is a systemic difference among disciplines with respect to technical barriers in revising or remixing. More research should also be conducted to determine whether order affects the scores given to a resource.

Practical Application

The practical application of the results of this study most directly affects two populations: educators and those who produce open educational resources. For educators, the practical application of this study is in understanding how difficult a given OER might be to revise or remix. The ALMS rubric in its current state is usable for educators with the understanding that it could be further refined and improved. Educators should also understand what factors are considered in the process of revising or remixing, but need to exercise caution in generalizing results. They should also be careful in comparing two resources with small variations in scores,

as the differences may be a result of rater training or disagreement. Depending on their

experience, they may be ready to use the ALMS rubric with minimal training. More extensive

training may be needed for other educators.

For those involved in creating OER, the question is how these findings might influence

the types and designs of OER they produce. Claiming that OER can be revised or remixed

because of their license, without consideration of technical barriers, is not tenable. That

assertion is not to suggest that an OER isn't beneficial unless its contents can be revised or

remixed to high degree. As one rater noted in their feedback within the first round of the Delphi

study, concerns about technical barriers are only part of the considerations in disseminating

OER. However, the full benefits of OER are not possible if the technical barriers are too high for

educators.

Limitations

As with any research, this study had limitations. First, the Delphi study methodology

had several known limitations. These limitations included the decision of who is an expert and

whether they really possessed the expertise needed. In this study the experts were selected by

Wiley and me. Woudenberg (1991) suggested that outliers may be more likely to drop out of a

Delphi study as opposed to those closer to consensus. Thus, the results of the study may be

misleading with respect to agreement due to the experts selected.

Another limitation of the Delphi study methodology was the reduced interaction provided

by somewhat constrained method of feedback (Landeta, 2006). This particular implementation

of the Delphi study methodology also had limitations. The number of participants was five and

with attrition the number of participants who provided feedback, particularly in later rounds, was

low. The low participation was concerning when determining how the scores represent the understandings within the OER community.

The determination of inter-rater reliability also had limitations. The number of participants was below 5 and other sources of disagreement may have appeared if the number of raters was higher. The selection of raters may also have been problematic. All raters were selected from an education technology program, though the experience of the individual raters varied. Nonetheless, the lack of representation across disciplines may have influenced the scores. Conversely, variation in scoring may be accounted for by the raters' lack of experience with OER or reusing resources from the Internet, and not actual disagreement. Lastly, the selection of OER that made up the sample may not be representative of the field as a whole. Although efforts were made to find a variety of OER, it is possible that other collections may have ALMS rubric scores that vary significantly from those rated in this dissertation.

These limitations, while real, do not severely reduce the applicability of the study findings. The results represent important advances in our understanding of the reusability of OER from a technical perspective. However, understanding the limitations of the study helps to indicate future areas of research.

Recommendations and future research

The first series of recommendations relates to the Delphi study. Prior research indicates that the Delphi study is a time-intensive research methodology, particular in comparison to other methods (Woudenberg, 2006; Gordon & Helmer, 1964). This study is not an exception. Therefore, it is recommended that future research using the Delphi methodology allocate a significant amount of time to gathering response feedback. Another recommendation is to keep the data collection mechanism as simple as possible. In the first round of the study, participants

were asked to record scores through a word processor document that was then submitted to me.

Feedback from participants indicated that this method was unnecessarily time-consuming. In

subsequent rounds of the study Qualtrics survey software was used to collect scores. Use of

survey software is recommended for other researchers wishing to use the Delphi study

methodology.

Asking Delphi study participants to rate each label in the rubric was a mistake. Instead,

each criterion should have been rated. In doing so, raters would have been better directed

towards the goal of determining the construct validity of the criteria. It would have also greatly

reduced rater fatigue.

There are several possibilities for future research. One possibility is to expand the

sample of OER rated through the ALMS analysis to gain a better sense of the technical difficulty

of reusing OER in other repositories. While that research could be potentially useful, the

continually evolving nature of OER may make maintaining the currency of findings difficult. A

possible solution may be automation of rating. For example, a web browser extension might

scan a web page and detect the OER on that page. Scores would be automatically given based

on the file type. The web browser extension could infer from the file type what media types

were used. Automating rating offers interesting potential to compare thousands of OER quickly.

Despite the potential, it is unknown whether the scores could be accurately automated.

Another possibility for future research is the exploration of what the ratings mean to

educators. Although ALMS rubric scores may be useful in quantitative sense, tying those scores

to the experiences of educators who are attempting to revise/remix OER for their classroom may

provide greater insight into what resources are useful for them. It may be that resources rated

below a certain score, in a practical sense, cannot be revised or remixed from an average

educator's perspective. Interviews with educators may raise issues that need to be addressed in future iterations of the rubric. More research is needed with respect to the processes that educators revise or remix to better understand the subtleties of the process. By examining the social implications of how the rubric is used the consequential validity of the rubric would be better understood.

A third area of research is further refining the rubric. As stated earlier, the scores only resulted in moderate agreement. In order for the rubric to gain additional validity and utility the inter-rater agreement needs to be increased. Future research could experiment with incremental iterations of the rubric to determine whether the scores can be raised. Decisions made during the course of this study most notably that of the Access table, could also be examined in greater detail to enhance validity.

Conclusion

This study is part of the ongoing discussion in the field of OER, focusing on the topic of revising and remixing. Although this is one of the first attempts to develop a method of systematically measuring the technical difficulty of reusing, this research is dependent on work that has been done by others in the field. Prior research, particularly with learning objects, has demonstrated how conflicted understandings are with respect towards reuse. The findings of this study may advance the body of knowledge in the field, but the validity of these findings is dependent on the premises that have been made before. As stated in Chapter 2, there is a sense that some of the issues revolving around the concept of reuse are still unresolved.

The utility of these findings are also dependent on the advancement of the field of OER itself. The challenges of sustainability, intellectual property concerns and some trends in computing in general, such as the preference of app stores and computers that provide

increasingly less customization, are a few of the problems faced by the field currently. These problems are not trivial and will require significant innovation to maintain momentum. The rise of Massive Open Online Courses (MOOC) may indicate a change in the role of OER in education.

Regardless of how these issues are resolved, the fundamental issues remain. Developing rules for an ever changing technological landscape makes stable research difficult. The challenge of addressing context systematically is not easily solved. In conducting this research, I believe that many of these issues may never be truly solved, but only addressed to fulfill the most pragmatic needs of educators.

References

Abdullah, N. A., Bailey, C., & Davis, H. (2004). Augmenting SCORM manifests with adaptive

links. *Proceedings of the fifteenth ACM conference on Hypertext & hypermedia -

HYPERTEXT '04*, 183. New York, NY: ACM Press. doi: 10.1145/1012807.1012857.

American Educational Research Association, American Psychological Association, and National

Council on Measurement in Education. (1999). Standards for educational and

psychological testing. Washington, DC: American Educational Research Association. As

cited in Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction*.

Thousand Oaks, CA: Sage Publications.

Alexandria Archive Institute. (n.d.). *People: Ahrash Bissell*. Retrieved from

http://www.alexandriaarchive.org/people_ab.php.

ARIADNE (2006) ARIADNE Foundation for European Knowledge Pool. Retrieved from

http://www.ariadne-eu.org/

Atkins, D.E., Brown, J.S. and Hammond, A.L. (2007). A Review of the Open Educational

Resources (OER) Movement: Achievements, Challenges, and New Opportunities.

Retrieved from http://www.oerderves.org/wp-content/uploads/2007/03/a-review-of-the-

open-educational-resources-oer-movement_final.pdf

Armstrong, D. J. (2006). The quarks of object-oriented development. *Commun. ACM*, *49*(2),

123–128. doi:10.1145/1113034.1113040

Babbie, E. R. (2000). *The Practice of Social Research (9th ed.),* Belmont, CA: Wadsworth Pub

Co.

Bassi, L., Cheney, S., & Lewis, E. (1998). Trends in workplace learning: Supply and demand in

interesting times. *Training and Development*, *52*(11), 51-54, 56-60, 62-64, 66-75.

Balatsoukas, P., Morris, A., & O'Brien, A. (2008). Learning objects update: Review and critical approach to content aggregation. *Educational Technology & Society*, *11*(2), 119-130.

Beggan, A. (2010). Opening up: Staff attitudes to open learning. *Presented at the OCW Consortium Global 2010, Hanoi, Vietnam: OCWC*. Retrieved from http://unow.nottingham.ac.uk/focus_groups.pdf

Ben-Ari, M. (2010). Objects never?: Well, hardly ever! *Communications of the ACM*, *53*(9), 32–35. New York, NY: ACM. Retrieved from http://portal.acm.org/citation.cfm?id=1810905.

Bian, L. (2007). Object-oriented representation of environmental phenomena: Is everything best represented as an object? *Annals of the Association of American Geographers*, *97*(2), 267-281. doi: 10.1111/j.1467-8306.2007.00535.x.

Bohl, O., Scheuhase, J., Sengler, R., & Winand, U. (2002). The sharable content object reference model (SCORM) - a critical review. *International Conference on Computers in Education, Proceedings, IEEE Computer Society*. 950-951. doi: 10.1109/CIE.2002.1186122

Booch, G., Maksimchuk, R., Engle, M., Young, B., Conallen, J., & Houston, K. (2007). *Object-oriented analysis and design with applications* (3rd ed.). Reading, MA: Addison-Wesley Professional. Retrieved from http://portal.acm.org.erl.lib.byu.edu/citation.cfm?id=1407387.

Boyd, D. (2008). Open-access is the future: Boycott locked-down journals. *Apophenia*. Retrieved from http://www.zephoria.org/thoughts/archives/2008/02/06/openaccess_is_t.html

Boyle, T., & Cook, J. (2001, December). Towards a pedagogically sound basis for learning object portability and re-use. In *Meeting at the Crossroads. Proceedings of the 18th*

Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education. The University of Melbourne (Vol. 101, p. 109).

Browne, T., Holding, R., Howell, A., & Rodway-Dyer, S. (2010). The challenges of OER to academic practice. *Journal of Interactive Media in Education*, *2010*(01), 1-15. Retrieved from http://jime.open.ac.uk/jime/article/viewArticle/2010-3/html.

Capretz, L. F. (2003). A brief history of the object-oriented approach. Luiz Fernando Capretz Department of Electrical & Computer Engineering. *Software Engineering Notes*, *28*(2), 1-10.

Cardelli, L. (1996). Bad engineering properties of object-orient languages. *ACM Computing Surveys (CSUR)*, *28*(4es), 150. New York, NY: ACM. doi: 10.1145/242224.242415.

Carroll, M. W. (2006). Creative Commons and the new intermediaries. *Michigan State Law Review*, *2006*(1), 45–65. Retrieved from http://works.bepress.com/michael_carroll/1/

Carson, S. (n.d.). About me. Retrieved from http://tofp.wordpress.com/about/.

Cebeci, Z., & Tekdal, M. (2006). Using podcasts as audio learning objects. *Learning*, *2*.

Clark, D. J. M. (1998). Developing, integrating, and sharing web-based resources for materials education. *JOM*, *50*(5). Retrieved from http://www.tms.org/pubs/journals/JOM/9805/Clark/Clark-9805.html

Clark, R. (1998). Recycling knowledge with learning objects. *Training and Development - Alexandria*, (October), 60-61. Retrieved from http://clarktraining.com/content/articles/LearningObjects.pdf.

Cohen, L, Manion, L., and Morrison, K. (2000). *Research methods in education*. London: New York, NY: Routledge Falmer.

Commonwealth of Learning. (2009). *WikiEducator: About*. Retrieved from

http://www.wikieducator.org/WikiEducator:About

Connexions. (n.d.). *Connexions - about us*. Retrieved from http://cnx.org/aboutus/

Cooper, B. B. S., & Gargan, A. (2009). Rubrics in education. *Phi Delta Kappan*, *91*(September),

54-56.

Correa, C. M. (2000). *Intellectual property rights, the WTO, and developing countries: The

TRIPS agreement and policy options*. London: Zed Books.

Creative Commons Corporation. (n.d. b). *About licenses*. Retrieved from

http://creativecommons.org/about/licenses/

Creative Commons Corporation. (n.d.a). *What is CC?* Retrieved from

http://creativecommons.org/about/what-is-cc

Curriki (n.d.). *Frequently asked questions*. Retrieved from

http://www.curriki.org/xwiki/bin/view/Main/FAQ

Dalkey, N. C., Corporation, R., Scientific, C. F., & (US), T. I. (1969). *The Delphi method: An

experimental study of group opinion*. Santa Monica, CA: Rand Corp. Retrieved from

http://www.rand.org/content/dam/rand/pubs/research_memoranda/2005/RM5888.pdf

Dalkey, N.C., Helmer, O. (1963). An experimental application of the Delphi method to the use of

experts. *Management Science*, *9*(3), 458-467. Retrieved from

http://www.jstor.org/stable/2627117

Downes, S. (2003). *Copying with digital rights management*. Retrieved from

http://www.downes.ca/cgi-bin/page.cgi?post=5300

Downes, S. (2012). IRRODL – A new edition has been published. *OLDaily*. Retrieved from

http://www.downes.ca/post/57223

Downes, S. (2001). Learning objects: Resources for distance education worldwide. *The International Review of Research in Open and Distance Learning*, *2*(1), 1-31. Retrieved from http://www.irrodl.org/index.php/irrodl/article/viewArticle/32/378inga

Downing, S. M. (2006). Face validity of assessments: Faith-based interpretations or evidence-based science? *Medical Education*, *40*(1), 7-8. Retrieved from EBSCO*host*

Dublin Core Metadata Initiative. (2010, October 11). *Dublin core metadata element set, version 1.1.* Retrieved from http://dublincore.org/documents/dces/

Duncan, S. (2009). *Patterns of learning object reuse in the Connexions repository*. Dissertation Abstracts International (Vol. 70, p. 73). Ann Arbor, MI: University Microfilms International. Retrieved from http://digitalcommons.usu.edu/cgi/viewcontent.cgi?article=1398&context=etd

eduCommons. (n.d.). *Welcome to eduCommons*. Retrieved from http://educommons.com/

Elkin-Koren, N. (2006). Creative Commons: A skeptical view of a worthy pursuit. *The Future of the Public Domain.* In P. Bernt Hugenholtz & Lucie Guibault (Eds.), London: Kluwer Law International. Retrieved from http://ssrn.com/abstract=885466

Farr, P. (2011). A new chapter for OCW search. *OCW Search blog*. Retrieved from http://www.ocwsearch.com/blog/news/a-new-chapter-for-ocw-search

IEEE. (2002). Draft standard for learning object metadata. *Citeseer*. Retrieved from http://www.webs1.uidaho.edu/fritz/edui/LOM_1484_12_1_v1_Final_Draft.pdf.

Farrell, R. G., Liburd, S. D., & Thomas, J. C. (2004). Dynamic assembly of learning objects. *Proceedings of the 13th International World Wide Web conference on Alternate track papers & posters* (Vol. 1, p. 162–169). New York, NY: ACM. Retrieved from http://portal.acm.org/citation.cfm?id=1013367.1013394

Friesen, N. (2004). Online education using learning objects. In R. McGreal, (pp. 59-70). London: Routledge. Retrieved from http://learningspaces.org/n/papers/objections.html

Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage Publications.

Gravetter, F. J., & Wallnau, L. B. (2008). *Essentials of statistics for the behavioral sciences*. Belmont, CA: Wadsworth Publishing Company.

Geser, G. (2007). *Open educational practices and resources: OLCOS roadmap 2012*. Open eLearning Content Observatory Services. Retrieved from http://www.olcos.org/cms/upload/docs/olcos_roadmap.pdf

Ghaemi, H. (2011). Is Rasch model without drawback? A reanalysis of Rasch model limitations. *Modern Journal of Language Teaching Methods*, *1*(2), 31-39.

Geselbracht, M. J., & Reisner, B. A. (2010). Inorganic chemistry learning objects for use in the general chemistry curriculum. *Journal of Chemical Education*, *87*(7), 756-757. doi:10.1021/ed100366m

Ginsburg, J. C. (2000). International copyright: From a "bundle" of national copyright laws to a supranational code? *SSRN Electronic Journal*, (10). doi: 10.2139/ssrn.222508

Godwin-Jones, R. (2004). Learning objects: Scorn or SCORM? *Language Learning & Technology*, *8*(2), 7-12.

Gonzalez-Barbone, V., & Anidorifon, L. (2008). Creating the first SCORM object. *Computers & Education*, *51*(4), 1634-1647. doi: 10.1016/j.compedu.2008.04.004

Gordon, T.J. & Helmer, O. (1964). Report on a long-range forecasting study. Retrieved from http://www.rand.org/pubs/papers/2005/P2982.pdf

Greaves, L., Roller, S., & Bradley, C. (2010). Repurposing with a purpose: A story with a happy ending. *Journal of Interactive Media in Education*, *2010*(01).

Gronlund, N. E. (2002). *Assessment of Student Achievement* (7th ed.). Boston, MA: Allyn & Bacon.

Gur, B. S., & Wiley, D. A. (2008). Instructional technology and objectification. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, *33*(3). Retrieved from http://www.cjlt.ca/index.php/cjlt/article/view/159/151

Gurell, S. (2008). (Ed.) *The OER handbook 1.0.* Retrieved from http://www.lulu.com/content/paperback-book/open-educational-resources-handbook-for-educators-version-10/3597933

Hamel, C. J., & Ryan-Jones, D. (2002). Designing instruction with learning objects. *International Journal of Educational Technology*, *3*(1), 111–124. Retrieved from http://www.ascilite.org.au/ajet/ijet/v3n1/hamel/

Harley, D. (2008). Why understanding the use and users of open education matters. In T. Iiyoshi & M. S. V. Kumar (Eds.), *Opening up education* (pp. 197-211). Hong Kong: MIT Press. Retrieved from http://api.ning.com/files/FXR*bxQZxzwFlhNUygPKDUiKS*sSZaAyqjOMpf1uvHKES3 ZsAcIEsUZW5sXumBHEgEA53JbP53MS0I8mwgucDAfvdIxM*9do/elibrary_Carnegie OpeningUpEducationpdf_720.pdf

Hayes, B. (2003). The post-OOP paradigm. *American Scientist*, *91*(2), 106. doi: 10.1511/2003.2.106

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological

    assessment: A functional approach to concepts and methods introduction to content

    validity. *Psychological Assessment*, *7*(3), 238-247.

Hodgins, H. W. (2004). The future of learning objects. In J. R. Lohmann & M. L. Corradini

    (Eds.), *2002 ECI Conference on e-Technologies in Engineering Education: Learning

    Outcomes Providing Future Possibilities* (Vol. 46, pp. 76-82). Davos, Switzerland:

    Education Technology Publications. Retrieved from

    http://services.bepress.com/cgi/viewcontent.cgi?article=1012&context=eci/etechnol

    ogies

Hilton III, J., Wiley, D., Stein, J., & Johnson, A. (2010). The four 'R's of openness and ALMS -

    analysis: Frameworks for open educational resources. *Open Learning: The Journal of

    Open and Distance Learning*, *25*(1), 37. doi:10.1080/02680510903482132

Ingargiola, G., Hoskin, N., Aiken, R., Dubey, R., Wilson, J., Papalaskari, M.-A., Christensen,

    M., et al. (1994). A repository that supports teaching and cooperation in the introductory

    AI course. *ACM SIGCSE Bulletin*, SIGCSE '94, 36–40. doi:10.1145/191029.191048

IMS (2010). *IMS Global Learning Consortium*. Retrieved from http://imsproject.org/

Jonassen, D. H., & Churchill, D. (2004). Is there a learning orientation in learning objects?

    International Journal on e-Learning, (June), 32-41.

Jones, E. R., & D, P. (2002). Implications of SCORM and emerging e-learning standards on

    engineering education the need for e-learning standards. *ASEE Gulf-Southwest Annual

    Conference, Lafayette, LA*. American Society for Engineering Education.

Jones, R. (2006). Designing adaptable learning resources with learning object patterns. *Journal of Digital Information 6*(1). Retrieved from

http://journals.tdl.org/jodi/article/viewArticle/60/62

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social science*. Fort Worth: Holt, Rinehart, and Wanston.

Kay, A. C. (1993). The early history of Smalltalk. *ACM SigPlan Notices*, *28*(3), 69–69. Association for Computing Machinery Special Interest Group On. Retrieved from http://thedavincipursuit.com/page24/page25/files/The_Early_History_of_Smallt.pdf

Kester, J. E. (1993). Some limitations of object-oriented design. *IEEE Aerospace and Electronic Systems Magazine* (Vol. 8, pp. 14-16). IEEE. doi: 10.1109/62.257112

Klebl, M., Krämer, B. J., & Zobel, A. (2010). From content to practice: Sharing educational practice in edu-sharing. *British Journal of Educational Technology*, *41*(6), 936-951. doi: 10.1111/j.1467-8535.2010.01134.x

Kölling, M. (1999). The problem of teaching object-oriented programming. *Journal of Object Oriented Programming*, *11*(8), 8-15. Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.8294&rep=rep1& type=pdf

Krauss, F., & Ally, M. (2005). A study of the design and evaluation of a learning object and implications for content development. *Learning*, *1,* 1-22.

Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change*, *73*(5), 467–482. doi:10.1016/j.techfore.2005.09.002

Lehman, R. (2007). Learning object repositories. *New Directions for Adult and Continuing Education*, *113*, 57-66. doi: 10.1002/ace

Lewis, J. A., Henry, S. M., Kafura, D. G., & Schulman, R. S. (1992). *On the relationship between the object-oriented paradigm and software reuse: An empirical investigation* (p. 23). Retrieved from http://eprints.cs.vt.edu/archive/00000295/

Longmire, W. (2000, March). A primer on learning objects. *Learning Circuits*, 1–4. Retrieved from http://www.astd.org/LC/2000/0300_longmire.htm

Mackintosh, W. (Last Updated 20 Jan 2011). *Formal Biography: Wayne Mackintosh*. Retrieved from http://wikieducator.org/User:Mackiwg

Madsen, O., & Møller-Pedersen, B. (1988). What object-oriented programming may be-and what it does not have to be. *ECOOP'88 European Conference on Object-Oriented Programming* (p. 1–20). New York, NY: Springer. Retrieved from http://www.springerlink.com/index/12320674m5429123.pdf

Messick, S. (1994). The Interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher. 23*(2), 13-23.

McGreal, R. (2004). *Online education using learning objects*. New York, NY: Psychology Press

McGreal, R. (2008). A typology of learning object repositories. *Handbook on Information Technologies for Education and Training*, 5–28. New York, NY: Springer. Retrieved from http://www.springerlink.com/index/K801214426K36LJN.pdf

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25), 1–10. http://www. ericae. net/pare/getvn. asp? v-7&n25. Retrieved from http://www.jcu.edu/academic/planassess/pdf/Assessment Resources/Rubrics/Other Rubric Development Resources/Designing Scoring Rubrics.pdf

Metros, S.E. (2005). Learning Objects: A Rose by Any Other Name. *EDUCAUSE Review. 40*(4), 12-13. Retrieved from

http://www.educause.edu/EDUCAUSE+Review/EDUCAUSEReviewMagazineVolume4

0/LearningObjectsARosebyAnyOther/157986

MIT OCW. (n.d.). *Our history*. Retrieved from

http://ocw.mit.edu/OcwWeb/web/about/history/index.htm

Mohan, P., & Brooks, C. (2003). Learning objects on the semantic web. *Learning*. IEEE

Computer Society. Retrieved from

http://www.computer.org/portal/web/csdl/doi/10.1109/ICALT.2003.1215055

Monbiot, G. (2011, September 29). Academic publishers make Murdoch look like a socialist.

The Guardian. Retrieved from

http://www.guardian.co.uk/commentisfree/2011/aug/29/academic-publishers-murdoch-

socialist

Mosier, C. I. (1947). A critical examination of the concepts of face validity. *Educational and

Psychological Measurement*, *7*(2), 191-205. doi: 10.1177/001316444700700201

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability.

Practical Assessment, Research & Evaluation, *7*(10). Retrieved from

http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ638498

Murry Jr, J. W., & Hammons, J. O. (1995). Delphi: A versatile methodology for conducting

qualitative research. *Review of Higher Education*, *18*(4), 423–36. Retrieved from

http://eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ508592

Nash, S. S. (2005). Learning objects, learning object repositories, and learning theory:

Preliminary best practices for online courses location of learning objects for use in e-

learning. *Learning*, *1*.

Neumann, F., & Geys, R. (2004). SCORM and the learning grid. *4th International LEGE-WG*

   Workshop Towards a European Learning Grid Infrastructure Progressing with a

   European Learning Grid (pp. 1-7). Stuttgart, Germany.

Neven, F., & Duval, E. (2002). Reusable learning objects: a survey of LOM-based repositories.

   Proceedings of the tenth ACM international conference on Multimedia (p. 291–294).

   New York, NY: ACM. Retrieved from

   http://portal.acm.org/citation.cfm?id=641007.641067

Nierstrasz, O. (1989). A survey of object-oriented concepts. *Object-Oriented Concepts,*

   Databases and Applications, 3–22. Citeseer. Retrieved from

   http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.5247&rep=rep1&

   type=pdf

Nworie, J. (2011). Using the Delphi technique in educational technology research. *TechTrends*,

   55(5), 24-30. doi:10.1007/s11528-011-0524-6

Ochoa, X., & Duval, E. (2008). Relevance ranking metrics for learning objects. *IEEE*

   Transactions on Learning Technologies, *1*(1), 34-48. doi: 10.1109/TLT.2008.1

OpenCourseWare Consortium. (n.d.). *Consortium members*. Retrieved

   from http://www.ocwconsortium.org/members/consortium-members.html

Parrish, P. E. (2004). The trouble with learning objects. *Educational Technology Research and*

   Development, *52*(1), 49–67. New York, NY: Springer. doi: 10.1007/BF02504772

Penny, J., Johnson, R. L., & Gordon, B. (2000). The effect of rating augmentation on inter-rater

   reliability: An empirical study of a holistic rubric. *Assessing Writing*, *7*.

Petrides, L., Nguyen, L., & Jimes, C. (2008). Open educational resources: Inquiring into author

   use and reuse. *International Journal of Technology Enhanced Learning*. Retrieved from

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=Abstr

actPlus&list_uids=161964317217753078related:9ocNOrBpPwIJ

Petrides, L., & Jimes, C. (2006). Open Educational Resource: Toward a new educational

paradigm. *iJournal*, (14). Retrieved from

http://www.ijournal.us/issue_14/ij_14_04_articleframe_Petrides_Jimes.html

Pokkunuri, B. P. (1989). Object oriented programming. *ACM SIGPLAN Notices*, *24*(11), 105-

149. Addison-Wesley, Reading, MA. doi: 10.1145/71605.71612

Pollard, R. Q., Dean, R. K., O'Hearn, A., & Haynes, S. L. (2009). Adapting health education

material for deaf audiences. *Rehabilitation Psychology*, *54*(2), 232-8. doi:

10.1037/a0015772

Polsani, P. (2006). Use and abuse of reusable learning objects. *Journal of Digital Information*,

4(3). Retrieved from http://journals.tdl.org/jodi/rt/printerFriendly/89/88

Powell, C. (2003). The Delphi technique: myths and realities. *Journal of Advanced Nursing*,

41(4), 376-82. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12581103

Ravotto, P., & Fulantelli, G., (2007). The Sloop idea: sharing free/open learning objects. In *Sloop

– Sharing Learning Objects in an Open Perspective*. Retrieved from

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.108.7649&rep=rep1&type=pdf

#page=9

Rayens, M. K., & Hahn, E. J. (2000). Building consensus using the policy Delphi method.

Policy,Politics, & Nursing Practice, *1*(4), 308. Thousand Oaks, CA: SAGE Publications.

doi: 10.1177/152715440000100409

Roblyer, M. D., & Wiencke, W. R. (2003). Design and use of a rubric to assess and encourage

interactive qualities in distance courses. *American Journal of Distance Education*, *17*(2),

77–98. Retrieved from http://www.informaworld.com/index/783721333.pdf

Santacruz-Valencia, L. P., Navarro, A., Aedo, I., & Delgado Kloos, C. (2009). Comparison of

knowledge during the assembly process of learning objects. *Journal of Intelligent*

Information Systems, *35*(1), 51-74. doi:10.1007/s10844-009-0088-5

Schechter, S.E. (n.d.). About (FAQ). *Research Without Walls*. Retrieved from

http://www.researchwithoutwalls.org/

Shaw, S., & Sniderman, S. (2002). Reusable learning objects: Critique & future directions.

World Conference on E-Learning in Corporate, Government, Healthcare, and Higher

Education 2002, *2002* (1), 2154-2157.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in*

Medical Research, *7*(3), 301–17. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/9803527

Sing, S.H. (2000). *Ways and means of bridging the gap between developed and developing*

countries. New York, NY: Information Technology and Public Administration at United

Nations. Retrieved from

http://unpan1.un.org/intradoc/groups/public/documents/APCITY/UNPAN014648.pdf

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007.). The Delphi method for graduate research.

Journal of Information Technology Education, *6*, 1-21. Retrieved from

http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.151.8144

Sosteric, M. & Hesemier, S. (2004). A first step towards a theory of learning objects. *Education*

Using Learning Objects, *3*(2). Retrieved from

http://books.google.com/books?hl=en&lr=&id=v3vTOljVoG0C&oi=fnd&pg=PA30&dq=A+first+step+towards+a+theory+of+learning+objects&ots=lej6o2X7wS&sig=hOb5qhSXjFGPKQTOmyv3KKNiAco

Slife, B. D., & Williams, R. N. (1995). *What's behind the research? Discovering hidden assumptions in the behavioral sciences*. Thousand Oaks, CA: Sage Publications.

Stahlke, H. F. W., & Nyce, J. M. (1996). Reengineering higher education: Reinventing teaching and learning. *Cause/Effect*. *19*(4), 44-51.

Stefik, M., & Bobrow, D. (1985). Object-oriented programming: Themes and variations. *AI magazine*, *6*(4). Retrieved from http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/508

Stroustrup, B. (1987). What is "object-oriented programming?" *ECOOP'87 European Conference on Object-Oriented Programming* (p. 51–70). New York, NY: Springer. Retrieved from http://www.springerlink.com/index/j26h7136tl638602.pdf

Snyder, A. (1986). Encapsulation and inheritance in object-oriented programming languages. *ACM SIGPLAN Notices*, *21*(11), 38-45. doi: 10.1145/960112.28702

Suber, P. (n.d.). *Stronger pledges*. Retrieved from http://www.openaccesspledge.com/?page_id=21

Ternier, S., Massart, D., Campi, A., Guinea, S., Ceri, S., & Duval, E. (2008). Interoperability for searching learning object repositories. *D-Lib Magazine*, *14*(1/2). doi:10.1045/january2008-ceri

Udas, K., Merkel, C. (Ed.). (2009). *Summary – content is infrastructure*. Retrieved from http://cnx.org/content/m19810/latest/

United State Copyright Office. (2008). *Copyright basics*. Retrieved from

http://www.copyright.gov/circs/circ1.pdf

WebMproject.org (n.d.). *Frequently asked questions*. Retrieved from

http://www.webmproject.org/about/faq/

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve*

student performance (1st ed.). Hoboken, NJ: Jossey-Bass.

WikiMedia Foundation. (2009a). *Wikibooks: What is Wikibooks?* Retrieved from

http://en.wikibooks.org/wiki/Wikibooks:What_is_WikibooksWikiMedia Foundation.

(2009b). Wikiversity: Welcome, newcomers. Retrieved from

http://en.wikiversity.org/wiki/Wikiversity:Welcome,_newcomers

Wiley, D. A. (2002). Connecting learning objects to instructional design theory: A definition, a

metaphor, and a taxonomy. In D. A. Wiley (Ed.), *The Instructional Use of Learning*

Objects: Online Version. Retrieved from http://reusability.org/read/chapters/wiley.doc

Wiley, D. (2005). *On the Lack of Reuse of OER*. Retrieved from

http://opencontent.org/blog/archives/900

Wiley, D. (2009). Impediments to learning object reuse and openness as a potential solution.

Revista Brasileira de Informática na Educação (Brazilian Journal of Computing in

Education). 17(3). Retrieved from http://www.br-

ie.org/pub/index.php/rbie/article/view/1022

Wiley, D. (2002). Learning objects need instructional design theory. *The ASTD e-Learning*

handbook, 115–126. Retrieved from

http://books.google.com/books?hl=en&lr=&id=uXA-

xCItrc4C&oi=fnd&pg=PA115&dq=Learning+objects+need+instructional

+design+theory&ots=1dCl8N2BiV&sig=H0RYpDOHr9v9pE7c4ahhx2k4Kq4

Wiley, D., Waters, S., Dawson, D., Lambert, B., Barclay, M., Wade, D., et al. (2004).

Overcoming the limitations of learning objects. *Journal of Educational Multimedia and*

Hypermedia 13, 507-521. Chesapeake, VA: Association for the Advancement of

Computing in Education. Retrieved from http://davidwiley.org/docs/overcoming.pdf

Wiley, D. (2010). Learning objects, content management and e-learning. *Content Management*

for E-Learning. New York, NY: Springer.

Wiley, D.(n.d.). The learning objects literature. http://opencontent.

org/blog/wp-content/uploads/2007/07/wiley-lo-review-final.pdf

Williams P. & Webb C. (1994) The Delphi technique: a methodological discussion. *Journal of*

Advanced Nursing 19, 180–186.

Woudenberg, F. (1991). An evaluation of Delphi. *Technological Forecasting and Social Change*

40(2), 131-150, ISSN 0040-1625.

Appendix A

ALMS Rubric (First Round)

Title of OER: _____

URL of OER: _____

Operating System: _____

Access to Editing Tools

	1	2	3	4	
Is an appropriate software application pre-installed with the operating system?	No	-	-	Yes	N/A
Is an appropriate software application downloadable from the Internet?	Not found through Internet search	Found through significant research	Found through some research	Found through Internet search easily	N/A
How much does the application cost?	Program cost is expensive	Program cost is significant	Some program cost	No program cost	N/A
How large is the program to be downloaded?	File to be downloaded is prohibitively large	File to be downloaded is a significant size	File to be downloaded is a non-trivial size	File to be downloaded is small	N/A
Does the program depend on other software or programming libraries?	Program depends on a large amount of libraries or other programs	Program depends on a significant amount of libraries or other programs	Program depends on some libraries or other programs	Program does not depend on any libraries or files	N/A
Is appropriate functionality available from a web-based	No	Access with significant restrictions	Access with some restrictions	Access with no restrictions or sign-up only	N/A

service?					
How much does the service cost?	Service cost is expensive	Service cost is significant	Some service cost	No service cost	N/A
Does the service require the user to have an account?	Service requires extensive account with major privacy concerns	Service requires an account with a significant amount of information	Service requires an account with some or no information	Service does not require an account	N/A
Are you able to cut-and-paste the text into a word processor or text editor?	No	-	-	Yes	N/A

Subtotal for Access: _____ (Total / No. of non-N/A answers)

Level of expertise required to edit

	1	**2.5**	**4**	
What degree of text editing expertise is required?	Extensive text editing expertise required *(Example: multi-column layout)*	Some text editing expertise required *(Example: footnotes, headers)*	Minimal text editing expertise required *(Example: bolding, headings)*	N/A
What degree of image editing expertise is required?	Extensive image editing expertise required *(Example: masking, clone stamp)*	Some image editing expertise required *(Example: multiple layers, filters)*	Minimal image editing expertise required *(Example: cropping, brightness/contrast)*	N/A
What degree of audio editing expertise is required?	Extensive audio editing expertise required *(Example: remixing/revising multiple sources)*	Some audio editing expertise required *(Example: audio filters, transitions)*	Minimal audio editing expertise required *(Example: trimming and saving in*	N/A

			appropriate format)	
What degree of video editing expertise is required?	Extensive video editing expertise required *(Example: remixing/revising multiple sources)*	Some video editing expertise required *(Example: transitions, filters)*	Minimal video editing expertise required *(Example: trimming and saving in appropriate format)*	N/A
What degree of programming expertise is required?	Extensive programming expertise required *(Example: data structures such as Object-Oriented Programming)*	Some programming expertise required *(Example: functions, if/then statements)*	Minimal programming expertise required *(Example: variation manipulation, basic arithmetic)*	N/A
What degree of other specialty expertise is required?	Extensive specialty expertise required *(Example: Advanced animation or 3D modeling)*	Some specialty expertise required *(Example: intermediate animation or 3D modeling)*	Minimal specialty expertise required *(Example: basic animation or 3D modeling)*	N/A

Subtotal for Expertise: _____ (Total / No. of non-N/A answers)

Meaningfully Editable

	1	2	3	4	
What portion of the text in the OER is editable? *(Example: Text stored independently, not embedded in an image)*	No text can be edited	Some text can be edited	A significant amount of text can be edited	All of the text can be edited.	N/A

What portion of the images in the OER is editable? (Example: Images stored independently, not embedded in a PDF)	No images can be edited.	Some images can be edited	A significant amount of the images can be edited	All of the images can be edited	N/A
What portion of the audio in the OER is editable? (*Example: Audio stored independently, not embedded in a video*)	No audio can be edited	Some audio can be edited	A significant amount of the audio can be edited	All of the audio can be edited	N/A
What portion of the video in the OER is editable? *(Example: Video stored independently, not embedded in a Flash object)*	No video can be edited	Some video can be edited	A significant amount of the video can be edited	All of the video can be edited	N/A
What portion of the other components in the OER is editable? *(Example: 3D-Model)*	None of the other components of the OER can be edited	Some of the other components of the OER can be edited	A significant amount of the other components of the OER can be edited	All of the other components of the OER can be edited	N/A

Subtotal for Editable: _____ (Total / No. of non-N/A answers)

Source files available

	1	2	3	4	
What portion of the source files for the text in the OER	Text source files are not available	Some text source files are	A significant amount of	All text source files are available	N/A

is available? *(Example: a Word document available for a PDF)*		available	the text source files are available		
What portion of the source files for the images in the OER is available? *(Example: a Photoshop file with all layers available)*	Image source files are not available	Some image source files are available	A significant amount of the image source files are available	All image source files are available	N/A
What portion of the source files for the video in the OER is available? *(Example: raw, unedited audio files available)*	Audio source files are not available	Some audio source files are available	A significant amount of the audio source files are available	All audio source files are available	N/A
What portion of the source files for the programming code in the OER is available? *(Example: source code downloadable from a website)*	Source code files are not available	Some source code files are available	A significant amount of the source code files are available	All source code files are available	N/A
What portion of the source files for the other components of the OER is available? *(Example: FLA files for SWF flash objects)*	Other component source files are not available	Some source files are available	A significant of source for other components are available	All source files for other components are available	N/A

Subtotal for Source: _____ (Total / No. of non-N/A answers)

Calculation

Subtotal	Score

Access to Editing Tools	
Level of Expertise	
Meaningfully Editable	
Source Files Available	
Average	

Appendix B

ALMS Rubric (Inter-Rater Reliability)

Access to Editing Tools

    See Appendix C

Level of Expertise

	1	**2.5**	**4**	
What degree of text editing expertise is required?	Extensive text editing expertise required *(Example: Setting up a multi-column layout; extensive use of graphics and text. Example: newsletter layout)*	Some text editing expertise required *(footnotes, headers, tables; auto-generated fields such as Table of Contents)*	Minimal text editing expertise required (bolding, headings; double-spacing)	N/A
What degree of image editing expertise is required?	Extensive image editing expertise required (masking, clone stamp)	Some image editing expertise required (Example: multiple layers, filters [sepia, blur, etc])	Minimal image editing expertise required (cropping, brightness/contrast, red-eye reduction)	N/A
What degree of audio editing expertise is required?	Extensive audio editing expertise required *(mixing two or more audio sources; extensive sampling or altering of samples. Example:*	Some audio editing expertise required (audio filters, transition. Example: fade in/fade out for introductions or closing of a podcast)	Minimal audio editing expertise required (removing extraneous sound or silence and saving in the intended format for distribution)	N/A

	speeding up a clip and timing it)			
What degree of video editing expertise is required?	Extensive video editing expertise required *(mixing two or more video sources; extensive sampling or altering of samples. Example: Special effects such as Adobe After Effects)*	Some video editing expertise required *(Example: Introductory or closing transitions, filters. Example: fade in/fade out for introductions or closing of a video segment)*	Minimal video editing expertise required *(Example: removing extraneous portions of video and saving in the intended format for distribution)*	N/A

Meaningfully Editable

	1	2	3	4	
What portion of the text in the OER is editable? *(Example: Text stored independently, not embedded in an image)*	No text can be edited	Some text can be edited	A significant amount of text can be edited	All of the text can be edited	N/A
What portion of the images in the OER is editable? *(Example: Images stored independently, not embedded in a PDF)*	No images can be edited	Some images can be edited	A significant amount of the images can be edited	All of the images can be edited	N/A
What portion of the audio in the OER is editable? *(Example: Audio*	No audio can be edited	Some audio can be edited	A significant amount of the audio	All of the audio can be edited	N/A

stored independently, not embedded in a video)			can be edited	
What portion of the video in the OER is editable? *(Example: Video stored independently, not embedded in a Flash object)*	No video can be edited	Some video can be edited	A significant amount of the video can be edited	All of the video can be edited	N/A

Source Files

	1	**2.5**	**4**	
What portion of the source files for the text in the OER is available? *(Example: a Word document available for a PDF)*	Text source files are not available	Some text source files are available	All text source files are available	N/A
What portion of the source files for the images in the OER is available? *(Example: a Photoshop file with all layers available)*	Image source files are not available	Some image source files are available	All image source files are available	N/A
What portion of the source files for the audio in the OER is available? *(Example: raw, unedited audio files available)*	Audio source files are not available	Some audio source files are available	All audio source files are available	N/A
What portion of the source files for the video in the OER is available?	Video source files are not available	Some video source files are available	All video source files are available	N/A

(Example: files with timelines and edits from video software such as Final Cut Pro)			

Appendix C

Access Table (Final version)

(Scores are averages across Windows, Mac OS X and Linux operating systems)

Image formats

	JPG	PNG	GIF	TIFF	PSD	SVG
Is an appropriate software application pre-installed with the operating system?	4	4	4	4	1	1
Is an appropriate software application downloadable from the Internet?	4	4	4	4	3	4
How much does the application cost?	4	4	4	4	1	4
How large is the program to be downloaded?	3.67	3.67	3.67	3.67	1.5	3
Does the program depend on other software or programmin	3.67	3.67	3.67	3.67	3	4

g libraries?						
Is appropriate functionality available from a web-based service?	4	4	4	4	N/A	4
How much does the service cost?	4	4	4	4	N/A	4
Average	3.42	3.42	3.42	3.42	1.9	3.43

Video formats

	WMV	**AVI**	**MPG**	**SWF**	**FLV**	**MP4**	**MOV**
Is an appropriate software application pre-installed with the operating system?	2	2	1	1	1	2	2
Is an appropriate software application downloadable from the Internet?	3.33	3.33	3.33	2.67	3.33	3	3
How much does the application cost?	3	2.33	2.33	1	2	3	3
How large is the program to be downloaded?	2	1.67	1.67	1	1.33	2	2

Does the program depend on other software or programming libraries?	3	3	3	4	3	3	3
Is appropriate functionality available from a web-based service?	1	1	1	1	1	1	1
How much does the service cost?	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Average	2.39	2.22	2.06	1.61	1.94	2.33	2.33

Audio formats

	MP3	WAV	OGG	AIFF	WMA	RM
Is an appropriate software application pre-installed with the operating system?	2	2	2	1	1	1
Is an appropriate software application downloadable from the Internet?	4	4	4	1.67	1.67	1.33
How much does the application cost?	4	4	4	1	1	2.5
How large is the program	4	4	4	1	1	1

to be downloaded?						
Does the program depend on other software or programming libraries?	2	2	2	4	4	1.5
Is appropriate functionality available from a web-based service?	4	3	3	3.33	1	1
How much does the service cost?	4	4	4	4	N/A	N/A
Average	3.00	2.88	2.88	2.17	1.44	1.28

Text formats

	DOC	TXT	RTF	PDF (text)	PDF (img)	DOCX
Is an appropriate software application pre-installed with the operating system?	4	4	4	1	1	4
Is an appropriate software application downloadable from the Internet?	4	4	4	4	1	4
How much does the application cost?	4	4	4	1	1	4

How large is the program to be downloaded?	4	4	4	1	1	4
Does the program depend on other software or programming libraries?	4	4	4	4	1	4
Is appropriate functionality available from a web-based service?	4	4	3	1	1	4
How much does the service cost?	4	4	4	N/A	N/A	4
Are you able to cut-and-paste into a word processor or text editor?	4	4	4	4	1	4
Average	4.00	4.00	3.875	2.29	1.00	4.00