Peer Review Anew: Three Principles and a Case Study in Postpublication Quality Assurance

Users of online open access to educational materials can benefit from multiple reviews that identify the sources of the reviews as well as the intended uses and audiences for the materials.

By Christopher M. Kelty, C. Sidney Burrus, Fellow IEEE, and Richard G. Baraniuk, Fellow IEEE

ABSTRACT | Over the last 15 years, the Internet has enabled new modes of authorship, new forms of open licensing and distribution, and new forms of collaboration and peer production to flourish. But in turn, new anxieties have arisen, especially concerning quality assurance, peer review, reuse, and modification. New innovations are appearing in peer review, endorsement, the measurement of trust, and the understanding of reputation, but without any systematic analysis of the general principles of quality assurance and peer review in this new era. In this paper, we propose a general set of principles for understanding what peer review was in the past and how it should be applied today to different kinds of content and in new platforms for managing quality. The principles stress an analysis not only on the content in materials but also on their context of use. Our focus is on open educational resources, and we present a case study of the open education project Connexions’ lens system for quality assurance and review. However, the principles can be applied across multiple levels of knowledge production, including scholarship in engineering and science and reference materials in addition to educational publishing.

KEYWORDS | Connexions; lens; open access; open educational resources (OERs); peer review; publication; quality assurance

I. INTRODUCTION

The last 15 years have seen major shifts in the nature of knowledge production and circulation. New modes of authorship, new forms of open licensing and distribution, and new forms of collaboration and peer production have all flourished. New online education projects, scientific journals, and reference works have gained critical mass. But in turn, new anxieties have arisen, especially concerning quality assurance, peer review, reuse, and modification.

Twentieth-century peer review in engineering and science, as conducted by professional and learned societies along with commercial publishers, was designed to ensure that all published materials had met a certain standard of quality. It relies on voluntary labor, for the most part, and is by no means uniformly reliable as a measure of quality. It was nonetheless a significant improvement on the class- and status-based verification systems of earlier centuries.

In the twenty-first century, however, new challenges have presented themselves: the volume of materials demanding peer review today is enormous and growing daily, and the existing peer review system cannot scale to accommodate it. The results are long publication delays, difficulty in finding qualified reviewers, increased costs of publication, and the unavailability of content while it is under review.

Manuscript received xxxx; revised xxxx. This work was supported by the Hewlett Foundation, the National Science Foundation under Grant EEC-0538934, and William Marsh Rice University.

C. M. Kelty is with the Department of Anthropology, William Marsh Rice University, Houston, TX 77001 USA (e-mail: ckelty@rice.edu).

C. S. Burrus and R. G. Baraniuk are with the Department of Electrical and Computer Engineering, William Marsh Rice University, Houston, TX 77001 USA (e-mail: csb@rice.edu; richb@rice.edu).

Digital Object Identifier: 10.1109/JPROC.2008.xxxxxx
To make matters even more complex, there is no longer any capital barrier to publication: the Internet and new tools for publication have enabled not only academics but also anyone with access to the Internet to publish material. Authors and consumers of all kinds of materials have also begun to call for peer review, increasing the demand for the time and energy of relevant experts.

The availability of information resources such as open access scientific publications [1]–[5] and Wikipedia\(^1\) is of undeniable and extraordinary value, but the difficulty in evaluating the quality or correctness of this information has simply moved the bottleneck, not eliminated it, from publication to review and assessment. The problem is particularly challenging in the area of open educational resources (OERs), where open education projects make course and textbook materials freely available on the Internet. The scope of the OER world is so vast (basically all education from K–20 and beyond) and the number of potential contributors so large that the impossibility of traditional peer review is already evident. As an example, in the Connexions open access repository\(^2\) [6], content from a potentially endless number of disciplines can be updated by the original author or remixed in a new context by another continuously on a timescale of minutes or hours; such rapid change seems to put impossible demands on conventional peer review.

New innovations are appearing in peer review, endorsement, the measurement of trust, and the understanding of reputation, but without any systematic analysis of the general principles of quality assurance and peer review. What is needed is a general set of principles for understanding what peer review is and has been, and how it should be applied to different kinds of content and new platforms for managing quality. Our audience includes anyone involved in the design or improvement of such systems. We propose here a set of principles that can guide the design of systems for peer review in the age of open access, Internet publishing, and Web 2.0 [7]. We suggest a multidimensional set of criteria for postpublication evaluation of resources, rather than a single "yes-no" gate typical of the prepublication model. These principles stress an analysis not only on the content in materials but also on their context of use.

Our observations are based on both research into the changing dynamics of publication and collaboration [8] and our explorations and observations of postpublication peer review, in the context of our creation of "lenses" in the Connexions OER platform. While our principles are intended to make sense across all forms of knowledge production, our focus on educational resources has the advantage of both addressing a very pressing need and being a radical departure from the model of publication in engineering and science (including open access), since it enables and encourages the enthusiastic remixing and reuse of materials and ideas.

This paper is organized as follows. After a review of the OER movement in Section II, we discuss the challenges of peer review in the current milieu in Section III. We present and discuss the peer review principles in Section IV and apply them to the case of Connexions in Section V. We conclude with a discussion in Section VI.

II. THE RISE OF OPEN EDUCATIONAL RESOURCES

Educators across a wide spectrum of disciplines share a common set of values: that knowledge should be free and open to use and reuse; that collaboration across distances and across disciplines should be easier, not harder; that people should receive credit, accolades, and financial remuneration (if relevant) for contributing to education and research; and that concepts and ideas are linked in unusual and surprising ways and not just in the simple linear forms that textbooks and classroom lectures present [6]. Over the last decade, aided by technological advances, these values have crystallized into the growing and often grassroots-driven open education movement, which has the potential to fundamentally change the way authors, instructors, and students produce, share, and use educational materials worldwide (e.g., [9]–[11]; see [12] for an excellent overview).

Like the open access movement [1]–[5], "peer production" [13] (e.g., Wikipedia), and Web 2.0 [7], the open education movement is driven by the rise of the Internet and the transformed environment of knowledge production, not only in academia but also in corporate and popular domains. Inspired by developments in open source software such as the Linux operating system, the Apache Web server, and the Mozilla/Firefox Web browser [14], [15], the open education movement seeks to provide free access to high-quality educational materials with the legal right to reuse, modify, update, and redistribute those resources appropriate to local contexts. The key enablers are open licenses that make the materials legal to use and remix\(^3\) [16] and Web infrastructure that makes the materials globally available for virtually no cost.

Over the last few years, many educational institutions have implemented open education programs that make available repositories of teaching and learning materials. These can include text (course notes, curricula, and textbooks), images, audio, video, interactive simulations, problems and answers, and games [12]. The communication capabilities and connectivity of the Internet further enhance the value of these resources by allowing producers and users to collaborate, share materials with each another, and enhance their knowledge and understanding of the materials through these social interactions.

All open education programs are based on the principle of freely sharing learning resources. However,

---


\(^{3}\)http://www.creativecommons.org.
the structure—who produces the resources, what type of resources are shared, and how free and open they are—varies by program. Some programs, like MIT OpenCourseWare, are top-down-organized institutional repositories that provide open access to courses developed solely by that institution’s faculty and instructors. Others, like the Stanford Encyclopedia of Philosophy, provide open access to content contributed by faculty from many universities but restricted to a single discipline. Wolfram Mathworld, a highly accessed online mathematics resource, has been assembled and nurtured by single committed individual (E. Weisstein). Wikipedia and its offshoots Wikibooks, Wikieducator, and Wikiversity provide content developed by self-selected communities on a more-or-less anonymous basis. Finally, Connexions provides content developed and remixed by a global community of educators from all levels (K–12 to graduate school) and all manners of institutions.

Regardless of their structure, open education programs as a whole have the potential to change the traditional educational content creation and delivery models significantly in a number of ways. First, they accelerate the blurring of formal and informal learning processes by melding entertainment and collaboration (games, online chat, etc.) with learning. Anyone can go to one of these sites and learn about any subject that interests them at any time. With the knowledge and confidence acquired from the learning, they can then contribute to, recontextualize, or expand the base of existing knowledge.

Secondly, at a time when the effective use of knowledge is viewed as the key to economic success, some open education programs empower many groups that have been “shut out” of traditional publishing domains. These underserved groups include talented K–12 teachers and community college instructors, scientists and engineers who work in corporations, and the world majority who do not speak and write English.

Thirdly, open education programs can promote specific educational opportunities that are currently at the margin of emphases for mainstream educational institutions, such as lifelong and continuing education for individuals, the delivery of high-quality technical education (e.g., mathematics, engineering) in the languages of developing countries, on-the-job and refresher training for technology workers, and the collaborative creation of new intellectual content within a discipline or across disciplines.

### III. QUALITY ASSURANCE IN A DIGITAL WORLD

#### A. The Problem

While they represent tremendous opportunities, open education programs also face novel challenges and new anxieties. Perhaps the most obvious is the quality assurance of the open materials. How can materials produced in a grassroots fashion, by people with varying skill levels and degrees, for widely varied reasons, be adequately vetted for quality? The anxieties frequently aired about projects such as Wikipedia or other remakable and open-authorship projects suggest that they are threatened by the proliferation of massive amounts of low-quality dreck that will swamp the information environment and prove impossible to navigate.

Such an anxiety is based on an unexplored assumption: that the scholarly publishing infrastructure of the twentieth century produced high-quality material and that the system of peer review employed therein remains the best system for ensuring quality, if only we can find enough credentialed reviewers. This assumption may or may not be true—but it is clouded by a lack of understanding of how peer review used to work.

The twentieth-century publishing industry (including scholarly societies) integrated reviewing into the process of publishing. The reputation of a press, a journal, or a scholarly society became a proxy for the internal process of selecting reputable reviewers to review specific content for specific purposes. “Quality”—as a process, rather than an inherent feature of work—has therefore been hidden from view, concealed inside a well-developed publication infrastructure.

This hidden process was not necessarily a bad one; it was a solution to the problem of quality appropriate to the means of publication prior to the Internet, and one that took centuries to develop. Today, however, it is necessary to explore the assumptions that we hold about how review is related to quality in order to make visible how the process of credentialing, reviewing, and claiming authority can be made appropriate to the new means of publication that have emerged. Similarly, we should resist the temptation to throw the entire structure overboard in favor of exclusively automatic ranking systems; aggregated data about clicks, links, and page-views; or too much reliance on the “wisdom of crowds” in lieu of a careful rethinking of how peer review can be renewed and supplemented by such tools. We need to match novel modes of authorship, reuse, licensing, and distribution of materials with equally novel modes of reviewing, assessing, and sharing evaluations.

#### B. Three Pressures

There are three kinds of pressures that are most clearly facing the existing peer review system.

1. **Publication has changed from a hierarchical, capital-intensive, corporate process to cheap and easy distribution of new forms of digital objects.**
2. **The volume of material available has increased dramatically over the last decade.**
3. **There is an increasing demand for review of a greater scope of material, not only academic scholarship.**

---


These three issues have pushed the twentieth-century model of peer review—designed for a smaller scale, industrial-era commercial publication system in academia—to its breaking point.

1) New Forms of Digital Objects: Twenty-five years ago, print publication was the only way to get one's education and research results out to the globally distributed community of researchers, and it took limited but well-established forms: textbooks, monographs, journal articles, and reference materials. Today one can publish something with the mere push of a button—1-click publishing—in an increasing variety of formats ranging from articles to blog entries to video lectures to book-length manuscripts. Publication has been transformed from a focused, capital- and labor-intensive process to an individualized and notional one.

As individuals and institutions have experimented within this new publication landscape, two issues have become clear: a) simply placing something on the Internet is not the same as “publishing” it and b) the new objects we place on the Internet can be updated, transformed, and reused far more easily, and increasingly legally, than conventional published sources. Publishing—once an integrated activity—has begun to disintegrate into its component functions: acquisition, review, editorial analysis, design, copy-editing and typesetting, creation of printed texts, marketing, distribution, and so forth. Naturally, it has become more and more evident that what makes a work high quality involves more than simply making it available.

Making a digital object widely available, easy to edit, and easy to reuse legally (without explicitly asking permission) also means that it can be constantly changed. This is both a challenge and an opportunity: it renders the idea of “once and for all” review problematic but also enables objects to undergo novel forms of “postpublication” review and improvement. The debates around the quality of open access scholarly journals and the ArXiv preprint server have made this distinction clear: different kinds of value are attached to preprints and to peer-reviewed materials, even if they are equally openly available. The former is valuable primarily for communicating results and staking ownership over certain parts of a research field; the latter validates some results as more reliable and trustworthy than others. Increasingly, preprint experiments (like Nature Precedings) occupy a space something like the minor leagues in baseball—a place to make work available and hope that it gets noticed enough to be “published” in a more prestigious journal, even though its availability (or even content in many cases) will not have changed.

2) Volume: The exponentially growing amount of new scholarly research combined with the demand that it be reviewed by working researchers places ever increasing stress on the current peer review system. One response to the growing amount of research has been to expand the number of journals, including both commercial and open access journals. Every new journal, however, demands more uncompensated labor from academics, which in turn creates incentives for shallower forms of review. In addition, as the number of specialized journals increases, the prestige and legitimacy of the top journals also increases, and the competition for those slots becomes ever more fierce, and ever more valuable.

Traditional publishers reject much out of hand—the volume of potential candidates for publication has always been orders of magnitude larger than the amount the review and publication system could handle. With the advent of 1-click publishing, however, cart and horse have been reversed: we face now a situation of needing new ways to reject after the fact, or, to put it differently, new ways to make the high-quality material stand out above the rest, without simply placing yet more strain on the limited time of people who are deemed the most reliable judges of quality.

3) Scope: The growth in the volume of publication is accompanied by an expanding demand for review beyond the narrow domain of scholarly and scientific work. Educational materials, textbooks, reference materials, fiction, film, and music are all increasingly reviewed in some form prior to being officially published, and more often than not voluntarily reviewed by peers in the same field or domain. As the case of Wikipedia demonstrates, not all kinds of information are amenable to the same forms of peer review. The large scale of Wikipedia makes it difficult to handpick reviewers for a Wikipedia encyclopedia entry, but the scope of expertise necessary also outstrips the ability to identify and/or credential a set of appropriate reviewers. In addition, the constantly changing nature of Wikipedia entries renders a “once and for all review” untenable and less valuable. Wikipedia has created a new kind of reference material as a result; but it has not yet created a new review process appropriate to this new kind of knowledge production.

C. New Solutions?

The pressures created by new digital objects and the increasing volume and scope of the new publication landscape have not gone unrecognized, nor are they restricted to educational and scholarly content. Various technical innovations in dealing with review and quality management have emerged in the last ten years, especially in places where large-scale user-generated content has created a need for new modes, metrics, and markers of trust, moderation, endorsement, and aggregation of data. In the last five years, there has been tremendous enthusiasm for solutions based on large-scale user-generated data. Web 2.0 and the “wisdom of crowds” [7] are frequently lauded as one (if not the only) solution to the problem of quality. Google’s “page-rank” system is emblematic of these data- and statistics-driven approaches to assessing quality.

---

Web sites such as Slashdot,\textsuperscript{11} Del.icio.us,\textsuperscript{12} and Digg\textsuperscript{13} have been among the first to innovate with respect to the moderation, reputation, and approval systems that enable users to sort and filter content, view evaluations, and develop new markers of trust. Here the problem is not so much quality as it is the need to remove or filter out the massive amount of repetitive, low-quality material that comes with the increasing number of participants as well as the need to provide ways for users to control what they see according to more meaningful metrics.

Similarly, Amazon\textsuperscript{14} and eBay\textsuperscript{15} have been key innovators with respect to suggestions, reviews, and techniques for turning user-generated data into meaningful tools for differentiation and evaluation. Amazon’s review system commends itself as an example of how reviews can become their own kind of object, associated with individuals, and valuable in multiple ways (negative reviews often reveal things positive reviews do not). eBay’s seller-ratings system has evolved into a very powerful and economically significant aspect of the online market. Both systems are restricted to the respective Web sites, however, and the reviews remain the property of Amazon and eBay, not of the reviewers. For-profit Web 2.0 filters like Faculty of 1000\textsuperscript{16} and Squidoo\textsuperscript{17} have also begun experiments in leveraging widespread expertise of a similar sort.

All of these solutions are laudably democratic: they allow anyone to become a reviewer, and they do not distinguish between reviewers. Reviewers and sellers build up reputation for an online identity, and those reputations do not depend on offline credentials, education, or experience. Such solutions are all-or-nothing: either one joins the brave new world of Amazon reviewers and eBay sellers with an initial rating of zero or one remains outside the system altogether. But should we throw the baby (an existing and legitimate system of peer review) out with the bathwater (a pre-Internet publishing infrastructure)? Is there really an opposition between the populist approach of Web 2.0 and a supposedly elitist one in which old-school experts review and validate materials by hand?

In the sequel, we will explore some fallacies associated with current thinking about the nature of peer review and propose some principles that might be used to guide future innovation and design of systems away from such all-or-nothing approaches and towards the challenge of renewing peer review.

IV. NEW PRINCIPLES FOR REVIEW

The pressures described in the previous section have arisen in the wake of changing means of publication: the Internet, new tools for authorship, collaboration and distribution, new legal possibilities, and renewed ethical and political commitments to openness and free exchange. The increasing volume and scope of these new means of publication demand a new approach to peer review as well. This section offers a set of fallacies to avoid and principles to follow; they are offered in the spirit of guiding discussions about the forms peer review should take in the future and not as absolute rules. They are derived from our experiences in understanding the proliferation of open access resources, open source software, and open educational resources generally [8], [10], [20].

A key insight that governs all of these principles is that quality is not an intrinsic component of the content of a work but rather a feature of how that work is valuable to a specific community of users: its context of use. There is no “one size fits all” review system that will ensure quality across cutting-edge scientific research, cutting-edge criticism in the humanities, educational resources for high schools around the world, and reference materials like encyclopedias and almanacs.

Context of use can mean several things. It can indicate where a resource is being used by readers, such as in a classroom, in a laboratory or journal club, or as part of an encyclopedia. It can also describe the stage of an article from the perspective of authors, such as draft, revised version, or updated version. It can also refer, today especially, to contexts of reuse such as a translation, a “derivative work” for a different purpose, or a constantly updated resource like a Wikipedia article. The stress is on understanding the variety of contexts in which a resource exists and not only the end-user consumption of a resource. Depending on the context of use, some resources are good enough, while some are not; a review can make that difference explicit.

For instance, an academic article describing a result in materials chemistry is not inherently valuable; it must be recognized by a community of chemists who agree that it represents a novel result and an advancement of the field. At some stages of its life, such an article would benefit from close analysis by peer chemists, who can suggest changes for specific purposes; at other stages it might benefit from review by educators, translators, or scientists who find the result useful in another field. The same might be said of many educational and reference works, each of which has a particular life-cycle or production and use, that is, a variety of potential contexts in which it is valuable.

There is a tendency in information technology to treat all information resources alike because they can all be represented and distributed by computers, software, and networks. But, as Brown and Duguid make clear in The Social Life of Information [21], different kinds of information live different kinds of lives, and it is crucial that those differences be recognized in the design and implementation of information systems. Peer review should be capable of reflecting those differences and of making meaningful decisions about them.

distinctions amongst different types of resources and different contexts of use: in a classroom in the United States, in a journal article in Africa, or as part of an encyclopedia entry on the Internet.

Wikipedia is perhaps the most instructive example. Debates about Wikipedia often ignore its contexts of use. In terms of authorship, the rules for what makes an encyclopedia entry a good one are widely shared because of the historical ubiquity of the encyclopedia as a resource. Wikipedia’s success owes much to this wide recognition of what makes for a good encyclopedia entry [22]. What at first appears to be a free-for-all turns out to be governed implicitly by a widely shared context of use—a shared sense of what makes an encyclopedia entry valuable for readers. These implicit rules have evolved fairly rapidly into explicit rules governing the quality of the content (such as the “neutral point of view” and “no original research” rules) and into formal editorial hierarchies as well.

In terms of readers, the value of a Wikipedia article comes from being an up-to-date reference work, not necessarily an authoritative scholarly work or an effective educational resource. Specifying context(s) of use makes it possible to more precisely report on the quality of something. Only by ignoring the context of use is it possible to worry about the “quality” of Wikipedia in general.

The case of Amazon reviews also is instructive here: they are opinions of a finished work, not reviews of a work that might be changed. No Amazon reviewer’s comments are taken as advice to authors, whereas peer reviewers are sometimes expected to play this role. Here the key difference in context concerns the openness of a text. Authors of open educational resources that permit modification benefit more from reviews that propose changes or new directions; users of an open educational resource benefit more from reviews that stress its effectiveness in teaching; readers or students benefit from reviews that report on ease of use and value to a beginner. All of these reviews are valuable to different people. What would a system look like that can encompass and differentiate between all of them?

A. The Fallacy of Misplaced Novelty

How can one distinguish contexts of use from the perspective of an information resource or the system that manages it? The flattening effect of making all information resources available on the Internet can be confusing—it removes works from their contexts of use and creates a sense that all works are the same kind of stuff. It erases hard-won differences that are created by communities of users.

There is a false sense of novelty here: that since resources are available in a new medium, with new technology, the uses and values associated with them will also be new. This fallacy of misplaced novelty ignores the social life of information and the many contexts of use that have already developed around resources. In some cases, such a rejection can be liberating: Wikipedia appears to be a prime example. In others, however, it leads to the reinvention of the wheel. Instead of relying on the practices, expertise, tacit knowledge, and explicit skill of a community of practitioners of an art, it can lead to the need to recreate it, simply because the glare of a new technology blinds us to existing, valuable knowledge about a specific context of use.

What is more, the appearance of a new technology, such as a new form of archive or a new authoring system, does not cause scholars or writers to give up standards and practices that have been learned through a long process of pedagogy, training, and experience. The story of the QWERTY keyboard is often invoked in this respect as a parable of “lock-in” [23]: the particular layout of the QWERTY keyboard is fixed in place because of the investment of learning and skill that large numbers of people have made in using it; even though an arguably more efficient system, the Dvorak layout, exists. However, people do make massive changes: very few people are using typewriters today instead of a word processor on a computer—even though it still sports a QWERTY keyboard [24]. We cannot say that the former switch was too costly; only that the latter switch was more valuable to users.

Even if peer review as it has existed since the beginning of the twentieth century in academic circles is not the right tool for the job, whatever replaces it has to build on existing strengths. The principle of maximum bootstrapping says that designers of new systems should build on and adapt existing communities of expertise, existing norms for quality, and existing mechanisms of review: adapt existing practices, bootstrap them, rather than return to first principles; turn first to the various contexts of use of information to determine what counts as quality, for whom, and how it is assessed and displayed. Instead of building the rating system to end all rating systems, study how scholars, educators, or authors of reference works currently review and improve things and then use these findings as a guide to build new review systems.

As an example, consider the difference between scholarly articles and educational resources with respect to peer review. For most academics, peer review is a constant process. In the context of creation, review helps guide the writing of an article; responses can help anticipate critiques, locate blind spots, and, in the best cases, propose a better organization, a better experiment, or a better set of cases to look at. In the context of publication, review can help determine what counts as a novel result. Review helps condition scholarly works to be written and presented in particular styles appropriate to disciplines and journals. After publication, a review can attract or repel readers, or guide critics to a new approach.

To apply the same rules that work for academic publications to educational materials or reference materials would do authors and readers a major disservice. Novelty is not, by definition, the marker of a successful
reference work or educational resources. An educational resource written in the same fashion as a scientific article would most likely be a failure. And while a scholarly article can be used in a classroom and an educational resource can be used in a laboratory, different communities of people are qualified to make suggestions about these different contexts of use.

The principle of maximum bootstrapping of communities of review should guide how we design new peer review systems. It should be possible to tell, from a review, what its definition of quality is and for whom. Indeed, as the other principles will demonstrate, it is essential to be able to tell the difference, in order to allow users not only to find the right content but also to find the right reviews of content as well.

B. The Fallacy of Misplaced Finality

The fact that there are multiple contexts of use in which to assess the quality of a piece of scholarly material raises an interesting question, a paradox perhaps. As the number of reviews of a resource grows, and if those reviews are related each to a different context of use, then does their utility not diminish as contexts become more and more specific? Would we not prefer a more uniform metric that allows us to pick the one resource everyone agrees is best? Are we not faced with the new problem of trusting the reviews instead of the problem of trusting the resource? How do we judge and compare two resources if their reviews all address different, perhaps nonoverlapping, contexts of use?

The changing landscape of publication offers something that is both a challenge and a solution to this paradox: versions. Consider the GNU Linux operating system. Linux refers only to the kernel; there are dozens, if not hundreds, of distributions that pad the kernel with different tools for different environments, each with different trade names, corporations, and associated services. And yet all of them rely on the same core kernel source from which they build each version. All of these distributions differentiate themselves with respect to their intended context of use; some are good for high-end parallel clusters, others for microcode on an embedded processor. Deciding on a distribution often means deciding on a context first, and on quality second.

A similar case might be made for new forms of scholarly, reference, and educational materials; why not approach these resources as modifiable as well? The fallacy of misplaced finality results from thinking of resources as having a single, final form to which everyone must apply different criteria of quality, rather than as a resource that can be reused, reissued, or transformed. Uniquely identified versions of a resource can be differentiated more easily with respect to a context of use, not just a raw metric of quality.

In a conventional publishing setting, most published objects take a single final form—journal articles do not change; textbooks change only a little from edition to edition. But in a setting where it is costless to legally copy, modify, and create a new version of a resource for a new purpose, such a resource has no single final form but a number of distinct versions, each with a distinct identity (a unique object ID and location). Combined with the legal licenses often applied to new digital objects, it can also be legally very simple to create multiple objects that appear similar but actually have different legal status (e.g., for commercial versus noncommercial use).

All these versions can create an acute anxiety: what happens when versions change so rapidly or proliferate so quickly that there is no time to review everything? There may be no solution to this problem; it may be that the idea of every object and every idea’s receiving its due evaluation is a promise of the twentieth century that the twenty-first century cannot make good on. But it is also built on an assumption that there will eventually be one best way, one best result, one best teaching method, or one best encyclopedia entry. Such an assumption is not a safe one to make when finality is no longer a given, nor even necessarily a goal; the question of how peer review works and what it achieves is once again on the table.

The fact that resources have versions, for better or for worse, suggests that peer review and evaluation should mirror that fact: our second principle, the principle of objectified evaluations, suggests that reviews should be treated as their own kind of object, disassociated from a single resource, specifying context of use and potentially applicable to multiple versions or to only one version. One of the most valuable aspects of peer review comes at the stage of improving a resource, and for this, reviews must be specific, referring to a specific version, a specific community of readers, authors, or educators, and offering specific paths towards success. The only way to accomplish this in an age of constantly modified content is to start thinking about the meaning and impact of versions.

Evaluations are extremely valuable objects; they take work to construct and represent an investment of time that can be extremely useful in collectively and collaboratively vetting knowledge produced in society. As such, they should be treated as valuable objects—like high-quality metadata. They are neither part of the resource itself nor the private property of the reviewing agent. The impact of objectifying evaluations is that they might potentially be sorted according to context of use, evaluator, or evaluator’s institution, or aggregated in new ways, in a fashion similar to that pioneered in social bookmarking. If two reviewers review the same object, both reviews need not be relevant to a user of that object. It should be practically and technically possible to tell the difference between two reviews in order to make the reviews themselves into more computable objects.

Objectifying evaluations would also create a trace, or history, of the development of an object. In traditional publishing, peer review was hidden within the process of
publication precisely because it aimed at a final, unchanging product—it was not necessary for the reviews to be visible or to be debated after the fact. By contrast, seeing the evolution of a resource today—whether the Linux kernel or a scientific journal paper or a seventh-grade lesson plan—can provide a rich understanding of its development, why it takes the form it does, what has been tried and suggested, and what has been rejected or left unsaid.

What is more, making the review process visible also allows people to understand the necessity of a particular context of use. In the context of a scholarly journal, for instance, all of the reviewers are of putatively the same kind: peer scientists who are conducting similar kinds of research and are looking for results they can understand and use. The context of use for which they review is implicitly the lab or the theory, or the next article. The context is important to the quality: it needs to be novel, it needs to be replicable, and it needs to meet certain standards. All of these are aspects of scientific practice that are essential but informal, unstandardized, and resting largely in the domain of pedagogy and mentorship. Making the review process a more objectified and computable process can reveal how it occurs and how it is related to the life of an article or a lesson.

Objectifying evaluations does not mean quantifying them: they still need to be flexible, written, and specific to a resource. But making peer review more computable does produce a hybrid object: something between the raw data-driven power of influence of Google’s page-rank and handwritten notes in the margin of a text; something in between the informal phone conversation suggesting a new direction and the wisdom of crowds. One need not choose between these two directions but instead chart a course that draws on both, to create a new form of peer review in which evaluations are objectified, build on existing experience and legitimacy, and persist alongside resources themselves.\(^8\)

C. The Fallacy of Misplaced Focus

All of this reviewing of versions, however, seems to assume a nearly infinite pool of reviewers attuned to the specifics of different contexts of use. However, the quantity of available resources and the demand for uncompensated review may not allow peer review to scale. Depending on the size of the community of users, different dynamics of scale may occur. The familiar case of Wikipedia is actually an extreme example: the barrier to entry for writing a Wikipedia article is very low and the format is widely familiar, so the number of potential reviewers is extremely large. In a field like high-energy physics or bionanotechnology, the barrier to entry is high and so the pool of potential reviewers is small. Educational materials lie somewhere in between.

In traditional publication, it is the publishers, editors, scholarly societies, and journals that handpick reviewers. Some do a better job than others, but the process is concealed both behind the formal mechanism of blind and double-blind review and behind the organizational boundaries of the publisher. To insist that peer review works today is to insist that this process of handpicking appropriate reviewers is the right one. But is it?

One might think of this as the fallacy of misplaced focus: the idea that peer review requires the careful selection of specific reviewers to review specific content to achieve adequate quality assurance across the board. But the complexity of this scenario is unnecessary today. What publishers and scholarly societies controlled in an era before the Internet was a really excellent Rolodex (address book). More than any other entity in the publication landscape, journals, scholarly societies, and editors were essential nodal points in a network of related expertise. Journals and scholarly societies retain this expertise today, but they are no longer the only place to find it. There are many other people, entities, and networks of expertise who can be relied upon to select good reviewers.

If one follows the principle of maximum bootstrapping and the principle of objectified evaluation, they lead directly to the insight that it is possible to encourage everyone to review anything. Why restrict review to the small handful of reviewers carefully selected by existing publishers? The costs of scaling that method are too high. The principle of multiple magnifications suggests that reviewers should be able to self-select and that any entity interested in improving quality should be able to select and encourage reviews of specific kinds of content. Reviewers selected by existing organizations will automatically have more weight (the principle of maximum bootstrapping), but many other kinds of people can also offer relevant and useful reviews.

The strongest argument for allowing as many reviews and reviewers as possible and for standardizing some aspects of reviews as their own digital object is that the combination of reviews from different perspectives is far more powerful than a hand-selected expert review from a single source. The idea of “multiple magnifications” suggests that reviews can be combined, like lenses, to achieve different effects.

Such an approach does not require the existing peer review system to be altered—only that it be willing to become one kind of review among many, an economic proposition that most likely will not square with traditional publishers who seek exclusive control over material throughout its lifetime. However, in the case of open access journals, for instance, there is a tremendous opportunity in opening up peer review from a prepublication gate-keeping model to a more flexible pre- and postpublication system. A reviewer of this article offered an example: what if it were possible to combine the reviews by teachers of particular textbooks with reviews by students of particular textbooks to find the overlap between the two opinions?

\(^8\)Jensen provides an extensive list of the “new metrics of scholarly authority” that might be relevant in the future [25].
The principle of multiple magnifications implies several things: that more reviews are better, that more data about reviewers is better, and that reviewers can either be trusted in the traditional sense (handpicked by existing legitimate organizations) or develop a new kind of reputation, more in line with the kinds of reputations built up amongst Wikipedia authors, open source hackers, or Facebook “friends.” If the Internet has indeed changed the economics of publication and the definition of identity [26], [27] then there is no reason to suspect peer review will be immune.

The idea of a reviewer’s building up a novel reputation as a reviewer also implies that such peer review will no longer be blind. However, this need not imply that reviewers’ physical identities are known; only that multiple reviews are associated with the same pseudonym, which might allow a reader to “trust” that pseudonym more than an anonymous one—the corollary of pseudonymous review.19 Reviewers with no information and no other reviews thus appear similar to first-time sellers on e-Bay: devoid of trust. Lifelong reviewers with a long list of reviews in a particular field, by contrast, command attention. Reviewers who are “bootstrapped” from existing review organizations would stake their existing reputation on the review of new content, in new ways.

If evaluations are understood as unique objects that carry with them some information about the reviewer and the specified context of use, then it becomes possible to use multiple kinds of reviews at the same time to evaluate an object. Such reviews could combine automated data (in some ways, the least trustworthy data as compared with a thoughtful review by an expert in the field) with reviews by amateurs, students, practitioners, critics, experts, and institutions.

Combining a set of reviews of an object can create a particular view of that object that preserves a users’ concern with context of use and privileges their own evaluation of trustworthy agents. By allowing everyone to review anything, new digital objects can be associated with a variety of different evaluations, and new ones can be automatically generated, constantly or periodically solicited, and associated with particular versions. Quality shifts from being something a digital resource possesses intrinsically to something that is shared across a subset of reviewers and users with respect to an object.

By allowing everyone to review anything, it also becomes possible to begin to differentiate different levels of quality assurance: from the formal, standards-based quality schemes of scholarly societies and state and local governments, to the informal community-based intuitive and tacit understandings of quality, to the wisdom of crowds and Web 2.0-style metrics of quality and tools of evaluation to automated data. In all of these cases, more information about who is reviewing, at the behest of whom, and for what context of use will enhance the value of the subsequent evaluations.

The three principals are summarized in Table 1. Taken together, they are intended as a preliminary guide in the construction of systems of review and quality assurance for open access digital objects. The remainder of this paper presents a case in which they have been in play: the case of the “lens” architecture for the Connexions open education repository.

V. CASE STUDY: CONNEXIONS LENSES

The online open education repository Connexions has been under development at Rice University since 1999 [6], [9], [10]; all during this time, the question of how to enable authors and users to designate and easily find and access high-quality material has been a central concern. From the outset, the project has focused on ways for individual users to evaluate and rate educational materials as well as a means to direct new users to those materials that are deemed by others to be of high quality.

<table>
<thead>
<tr>
<th>Fallacy of Misplaced Novelty</th>
<th>Principle of Maximum Bootstrapping</th>
<th>Fallacy of Misplaced Finality</th>
<th>Principle of Objectified Evaluations</th>
<th>Fallacy of Misplaced Focus</th>
<th>Principle of Multiple Magnifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>The flattening effect of the Internet blinds us to existing contexts of use; the fallacy of assuming that new technologies automatically imply the need for new values.</td>
<td>Designers of new systems should build on and adapt existing communities of expertise, existing norms for quality, and existing mechanisms of review.</td>
<td>Assuming that resources have a single, final form to which everyone must apply different criteria of quality, rather than a resource that can be re-used, re-issued or transformed.</td>
<td>Treat reviews as their own kind of object, dis-associated from a single resource, specifying context of use, and potentially applicable to multiple versions.</td>
<td>Assuming that peer review requires the careful selection of specific reviewers to review specific content to achieve adequate quality assurance across the board.</td>
<td>More reviews are better, and more data about reviewers is better because multiple, combined views on an object are now possible. Corollary: review is not blind, but pseudonymous and persistent.</td>
</tr>
</tbody>
</table>

19For an interesting discussion of the impact of nonanonymous review in the British Medical Journal, see [28].
The structure of material in Connexions is unique: courses and textbooks are broken into small modules on the order of a chapter or section—just enough to present a single concept. By breaking up the material in this way, users gain flexibility in recombining existing materials into different collections. Modules and collections have distinct identities in the system and therefore can be evaluated separately or individually. A module should stand on its own (it should make sense to a reader who stumbles on it through a search engine), and collections can be as small as a short book or as large as a curriculum. As should be clear from the foregoing discussion, this change in technology has a real effect on what it means to evaluate and review material in Connexions. Traditional publishers, as well as some institution-based projects like MIT OpenCourseWare, generally stick with the age-old editorial review process, whereby material is vetted and reviewed before it is made publicly available.

Connexions recognized early on that a prepublication review process would not scale to the eventual large size and activity level of its repository, nor would it foster social networking or community [29]. So, rather than acting as a gatekeeper and making a single centralized accept/reject decision regarding each module or collection, Connexions admits all contributions and then opens up the editorial process to third-party reviewers and editorial bodies for postpublication review. Everything submitted to the repository and everything constructed out of it is unreviewed when it is published, which means that there are no a priori judgments of quality. One common reaction to this decision has been concern that the repository will be quickly filled with junk and that users will be unable to distinguish the good from the bad [30].

It is here that the principle of maximum bootstrapping has been applied. By following this principle, we assume that there are communities of users who are already able to distinguish good from bad content for a particular purpose. So for instance, there is a great deal of material on digital signal processing (DSP) in the Connexions repository, and much of it is deemed very high quality by the authors and their colleagues who make use of it. The community of working and teaching engineers, and the journals and scholarly societies they belong to, are the natural site for finding expertise and reviewers.

However, even if specialized users could recognize quality in the repository, they initially had no way of designating that quality on a module, and the Connexions repository admits all but clearly illegal contributions. Connexions’ solution to this problem was to develop a system called lenses that sort content according to quality assessments provided by third parties (see Fig. 1). Each lens has a different focus, and multiple lenses can be combined to change the focus, just as with optical lenses. A lens consists primarily of a designation of approval (with optional commentary) by some third party ranging from those whose identity is generally known, such as traditional editorial boards and professional societies to informal groups of colleagues, automated lenses based on popularity, the amount of (re)use, the number of incoming links, or other metrics.

The principle of objectified evaluation has driven the implementation; rather than assume that each module needs a single final evaluation determining its ultimate quality, a module can have any number of evaluations, which are implemented and stored separately from the module itself. Choosing a lens amounts to exhibiting trust in certain reviewers (or statistics) over others. So for instance, the IEEE Signal Processing Society is launching a reviewing and certification process for Connexions materials in the DSP area. The National Council of Professors of Educational Leadership (NCPEA) has launched a Connexions lens based on a peer review process involving both faculty from educational leadership programs and practicing principals and superintendents. Both of these societies are respected by their members and others, and their evaluations can carry a great deal of weight. By using the IEEE lens or NCPEA lens, it is possible to discard (i.e., not view) content that does not meet their standards.

At the other end of the spectrum, an individual teacher might review a number of different K–12 music modules, adding valuable commentary about what works and what...
VI. CONCLUSIONS

We have offered the principles of maximum bootstrapping, objectified evaluations, and multiple magnifications as guide stars to navigate toward as we develop new scalable systems for quality assessment and peer review in the Internet age. These new systems will likely take hold more readily in some arenas than others. Wikipedia has proved that community-based authoring, editing, and peer review of reference works like encyclopedias is possible on a large and global scale. The enormity of the content and context landscapes of open education makes traditional peer review infeasible and thus will (in our opinion) push users to systems more or less like the Connexions lenses. In engineering, science, and humanities publishing, however, the situation is less clear. While there is a clear need to develop new, more scalable peer review mechanisms, the somewhat conservative nature of the scholarly community could impede such experiments. A case in point is the journal Nature’s experiment with open peer review; the authors of only 5% of the papers that made it past the initial review agreed to make their papers available for open comment.  

Nevertheless, we are entering exciting times. The open education movement in particular has the potential to break the education world out of a once successful model (in which mass media determined the nature of our educational system: one textbook for everyone) to a situation of increasing richness and diversity, where resources, styles, and content can be tailored to more specific contexts. The ability to address quality issues is a key aspect of the movement. If we can maintain the openness of the growing number of open education programs and innovate on ways to guide the improvement of materials through cooperation and collaboration, then we might find radically new ways to educate ourselves and our children—new ways that match the complexity of the contemporary world and the many challenges it faces.

Acknowledgment

The authors offer thanks to D. L. Jones, V. Cevher, and D. B. Williams for many helpful comments on the manuscript while it was in preparation.


AUTHOR QUERY

No query.