Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Taking Learning Outcomes to the Gym: An Assignment-Based Approach to Developing and Assessing Learning Outcomes

Steve Joordens, Dwayne Paré
and Lisa-Marie Collimore
Advanced Learning Technologies Lab
University of Toronto Scarborough

Published by

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

Cite this publication in the following format:

Joordens, S., Paré, D., & Collimore, L-M. (2014). *Taking Learning Outcomes to the Gym: An Assignment-Based Approach to Developing and Assessing Learning Outcomes.* Toronto: Higher Education Quality Council of Ontario.

The opinions expressed in this research document are those of the authors and do not necessarily represent the views or official policies of the Higher Education Quality Council of Ontario or other agencies or organizations that may have provided support, financial or otherwise, for this project.
© Queens Printer for Ontario, 2014

Executive Summary

While there has been great interest and progress in terms of defining core learning outcomes related to the completion of various postsecondary programs, there has been far less progress in terms of elucidating powerful ways to assess these outcomes. Without clear assessment methods it is difficult to see how one could perform course or program redesign with these learning objectives in mind.

To date, attempts to address this "assessment of learning outcomes" gap have focused mostly on the use of qualitative tools that are given to students as they leave some institution or program, tools like the National Survey of Student Engagement (NSSE) questionnaire or the Collegiate Learning Assessment (CLA) index. But these tools rely on subjective report, are often difficult to administer repeatedly at scale, and typically only provide an "after education" snapshot of learning. This report highlights and validates a different assignment-based approach that has much greater potential to provide quantitative data with much higher resolution.

This assignment-based approach is illustrated via a muscle-building analogy to represent the preferred goal: the ability to simultaneously develop and assess skills on an assignment level. That is, when you pump as much weight as you can in the gym, you are simultaneously building the muscles needed to pump that weight and you are providing a clear measure of how strong those muscles currently are. Similarly, our most desired learning outcomes all have basic cognitive skills underlying them, skills that develop through repeated effective practice. If we could thus develop technologies that could simultaneously exercise and assess these skills quantitatively, then we could track their development on an assignment by assignment basis, providing a much more accurate index of the attainment of the learning outcomes associated with those skills.

The primary purpose of this report is to provide a concrete example of how an assignment-based approach can be instituted and to provide some initial data supporting the validity of exercising and assessing learning outcomes in this manner. We first describe briefly some of the core learning objectives virtually all educators see as critical, including critical thought, creative thought, self-reflective thought, and effective communication in both its receptive and expressive forms. We then draw on the example of a specific learning technology, peerScholar, to demonstrate that this assignment-based approach to assessing learning outcomes is indeed possible even within the current constraints of the higher education system. This potential is then demonstrated via two experiments, one that illustrates the development of self-reflective thought and another that illustrates the development of critical thought.

We conclude that the example we highlight with peerScholar could be extended to other learning technologies and, especially if learning technologies were built with the assessment of learning objectives in mind, we could ultimately create a toolbox that would allow us to simultaneously build and assess our progress as educators in an information-rich and powerful manner. To some extent, the entire value of the specification of learning objectives hinges on our ability to assess their development well, and it is our contention that educators should be looking towards assignment-based approaches to fill this need.

Table of Contents

List of Tables

List of Figures

Preamble

It happens to the best of us. We look at ourselves in the mirror and see a reflection we wish was better, stronger, more “in shape”. Feeling motivated, we plan how to get to where we want to be. The first step is defining what aspects of ourselves we wish to change. Do we just wish to lose fat? Are we also hoping to gain muscle mass? Is it time for plastic surgery? This planning stage is almost fun as we imagine the future us in somewhat abstract form and yet, in the back of our minds, we know the hard truth. The image of our future selves that we have devised will only come to be if we do the really hard work of figuring out how to get from here to there, and then implement those concrete steps in an effective manner.

In more recent years the public education system is also looking at itself in the mirror and similarly being unhappy with what it sees. Time has crept up in the form of larger classes and smaller budgets, and the resulting changes in how we educate often do not match up with our desired selves. So we begin by envisioning the selves we want to be, and these visions typically take the terms of relatively abstract learning outcomes: statements reflecting the knowledge and transferable skills we want our students to acquire under our tutelage. But again, while producing these images can be fun, to actually become those selves requires us to come up with a concrete plan of action, one that is possible, one that makes sense, one that will allow us to see progress. We need to take our lofty notions of learning outcomes to the gym.

If the kind reader will endure the analogy a little longer, physical transformation is achieved via specific exercises that are linked to the outcomes we desire, and one great thing about physical exercises is that they combine what an educator might call practice and assessment. For example, if I decide I want much larger biceps, then arm curls are a good place to start. If I regularly “practice” arm curls my biceps will indeed enlarge, and as they enlarge I will notice that I can curl heavier weight or do more curls at a given weight. That is, I am seeing the evidence of change: the “practice” is simultaneously providing an “assessment” of progress. The positive assessments motivate additional practice, producing a positive feedback loop whose inevitable end is… Arnold Schwarzenegger? Well, every analogy breaks down at some point, but if we exchange words like “biceps” and “pecs” for “critical thinking” and “verbal communication”, maybe that is our goal: to produce chiseled mind-builders.

The goal of this paper is to provide a relatively simple demonstration of how this same combination of practice and assessment can be accomplished within an education context in our quest to realize learning objectives. Specifically, in order to provide as concrete an example as possible, we will focus on peerScholar, an internet-based application that was designed to exercise some of the most desired learning outcomes (e.g., critical thought, creative thought, self-reflective thought, clear efficient communication, giving feedback, responding appropriately to feedback). It combines peer- and self-assessment within a formative learning process, and thus there is very good reason to believe that it gives students mental exercise with exactly the sorts of transferable skills in which we want them to be proficient upon graduation. But can it assess learning outcomes as it exercises them? That is, can peerScholar – and other tools like it – provide concrete assignments that both exercise and assess learning outcomes within a classroom context? And, in so doing, do these assignments provide an example of the next step educational institutions can take to begin realizing learning outcome goals in clear and powerful ways?

Learning Outcomes

Numerous educational stakeholders and organizations have put considerable effort into specifying the set of learning outcomes that reflect what skills our students should possess upon graduation.[1] More recently, the first author of this paper was a member of a HEQCO Learning Outcomes Tuning Committee comprised of university and college educators and spent well over a year coming up with a set of learning objectives meant to represent the skills and knowledge possessed by students graduating with an accreditation in the social sciences. While different groups ultimately arrive at different specific lists of learning outcomes, a few are consistently represented across virtually all lists. These include the following:

Content Knowledge

Clearly, whatever the subject matter of the course we are teaching, one of our primary goals as educators is to transfer the core aspects of that content – and hopefully some of the details as well – to our students. Content knowledge will be considered somewhat distinct from the rest of the members of this list because, unlike cognitive skills, knowledge can be gained from simple exposure. In addition, given that we assess knowledge regularly using conventional testing devices, the challenge we face as educators lies in the teaching and assessment of transferable cognitive skills.

Critical Thought

Perhaps the most important transferable skill we hope to impart to our students is the ability to think critically. Critical thought is multifaceted and includes the ability to analyze a piece to assess its quality, often by comparing it to other instances or by breaking it down according to some measure of quality (Foundation for Critical Thinking, 2013). The end goal is to assess quality, typically with an eye towards identifying aspects that are not as good, convincing or accurate as they could be. Thus, students with strong critical thinking skills are equipped to evaluate information in intelligent ways, knowing which aspects of it are convincing and which are better ignored.

Creative Thought

Where critical thinking ends, creative thought often takes over. Once something has been evaluated critically, one is often left with ideas of how it is suboptimal, which then prompts thoughts on how it could be improved. Thus creative thought involves the ability to see what something could be, and the costs and benefits of that re-imagined thing. Critical thought is associated with finding problems and creative thought with finding solutions. It is possible to be successful in life as only a critical thinker (e.g., movie critics), but those who can posit new and powerful solutions for current problems are those who change the world for the better, and we certainly hope our students will be world-changers.

[1] See, for example, the International Society for Technology in Education (https://www.iste.org/), or Partnership for 21st Century Skills (http://www.p21.org/).

Self-Reflective Thought

This is also sometimes called metacognition and it refers to the ability to analyze oneself critically; to have a firm idea of what one does and does not know, what one can and cannot do, and generally an ability to see both one's strengths and, more importantly, one's weaknesses. It is virtually impossible for any of us to improve ourselves without first having a good sense of what our strengths and weaknesses are. To some extent, self-reflective thought is tantamount to thinking critically about oneself and about the products of one's own work.

Communication Skills

At some level it does not matter how thoughtful someone is if they cannot communicate their thoughts well or listen and understand the communication of others. This can actually be seen as a set of sub-skills including effective written communication, effective verbal communication and the ability to understand both written and verbal communication. It can also be considered in more specific terms, such as the ability both to provide clear and useful feedback to others and the ability to evaluate and react appropriately to feedback on one's own work. But in the more general sense, it refers to the ability to transmit or receive thoughts and ideas in efficient and effective ways.

Collaborative Skills

Humans are highly social beings and no matter how thoughtful and how well one is able to communicate, it remains very difficult to accomplish anything of value alone. Thus it is also critical that our students learn to work within social groups, working collaboratively in ways that lead to positive outcomes for all.

Note that we have highlighted six commonly emphasized learning outcomes. This list is not meant to be exhaustive. For example, we would also like our students to be comfortable working with and interpreting numerical data, including those presented in tables and figures. We hope that our students graduate with a clearer social awareness, a strong sense of responsibility and better work habits than they had when they entered our institutions. However, the list above highlights what many seem to view as the core learning objectives and the ones that can be most challenging to teach, which is why we emphasize them here.

It is also very important to note that while the "content" learning objective is specific to the topic of a given course, the rest are general cognitive skills. This distinction is important to how we teach for the following reason. Human memory learns information differently than it does skills (Tulving, 1985). Information can be learned via simple exposure, although the more deeply one thinks about it the better the learning generally is (Craik & Lockhart, 1972). Skills, however, are only learned via repeated effective practice (Milner, 1962). This distinction is conveyed well with the following example. You can attend a two-hour session and learn a lot about, say, karate. But to be able to "do" karate – that is, to acquire the necessary skill base – takes regular repetition of the basic processes used to move the muscles in the desired manner.

Cognitive skills are no different. One cannot be told how to think critically or how to write well. Students must practice the skill repeatedly, preferably within a context that provides structure and guidance. This is exactly why these skills are so hard to teach within the current context of large class sizes and small budgets.

Traditional modes of instruction, ones that require the instructor to be involved in every step of learning, are expensive and can be logistically complex and, as a result, it is often this kind of instruction that is lost to other approaches that allow for easier training and assessment (i.e., the learning of content via lecture and multiple-choice testing).

There is a silver lining of sorts if we can find an effective way to give students the necessary practice with these cognitive skills. When skills are practiced enough, they become almost automatic and extremely robust (Samuels & Flor, 1997). Jimi Hendrix did not have to think about playing beautiful music; it just happened. Bruce Lee did not have to think about his kung fu; every reaction was reflexive and, once learned, not lost. Thus if we could give our students significant practice with the cognitive skills reflected by the previously listed learning outcomes, these skills could also become almost natural for them and would stick with them well after their time with us has ended.

But of course that leaves open two questions. How exactly do we teach the cognitive skills underlying these outcomes in the current economic climate? And can we assess the learning outcomes along the way in a manner that allows us to track our progress? Again, what we would ultimately like is some sort of "cognitive workout" that could simultaneously exercise and assess the development of the learning outcomes we have highlighted as critical.

peerScholar

With the basic context now presented, the goal of the remainder of this report is to focus on one specific technology and show how it can, indeed, deliver on the dual goals of helping our students achieve the learning outcomes we desire while also allowing us to assess their progress along the way. We chose peerScholar as the learning technology for two reasons. On the practical level, it is a technology developed by our Advanced Learning Technologies lab and, as such, we have the ability and the understanding of the code to mine the data in ways that might not be possible with someone else's technology. On the theoretical level, we simply know of no other technology that helps develop the breadth of learning objectives with the depth of experience provided by peerScholar. In fact, it was developed for exactly that purpose (Joordens, Desa & Paré, 2009).

The remainder of this section will explain the phases of a typical peerScholar assignment, highlighting the learning outcomes that each step supports. Generally speaking, peerScholar assignments require students to complete three sequential phases, each of which asks them to employ varying – though sometimes overlapping – cognitive skills. First, students create a composition in line with the provided instructions, then they assess a subset of the compositions submitted by their peers, then they see their peer feedback and they revise their original submission based on their analysis of the feedback peers gave to their work. Along the way students can also be instructed to rate the quality of their own work by giving it a mark at whatever stages as the instructor desires.

The Create Phase

In the first step of a peerScholar assignment, students are asked to submit some sort of digital composition, following the instructions provided by the course instructor. The learning outcomes supported in this phase are largely tied to the assignment the instructor defines. Students can be asked to think critically, creatively or reflectively, and they can express their composition in writing, images or movies (e.g., verbal presentations). Ultimately, whatever students compose will be graded, so the Create Phase provides the educator with a blank slate to exercise and assess any learning objective on which they wish to focus. In addition, the time students spend exercising the learning outcome in question is essentially multiplied in the phases that follow.

The Assess Phase

In Robert Persig's 1974 book *Zen and the Art of Motorcycle Maintenance*, the author describes an assignment he used in a class on rhetoric. Students wrote arguments with no names on them, then Persig shuffled them, taped them to the wall, and then asked the students to read each and decide which was best and why. Persig believed that the ability to sense quality was primitive and therefore that students would be able to tell which argument was better and then would undergo a personal and palpable learning experience as they tried to express why one argument seemed better than another.[2]

The underlying engine of peerScholar is very similar to the peer-learning process described by Persig. After submitting their composition, students log back into the system and are presented with a specified subset (say, six) of randomly selected and anonymously presented compositions submitted by their peers. Students are asked first to read each while mentally deciding which is of the highest quality. They then translate that vague feeling of quality into a score on some scale to make it a step more concrete. Finally, they are asked to make two comments related to each composition. First, they are asked to highlight something they really liked about the work. Second, and most critically, they are also asked to provide a constructive comment via which they clearly articulate the one thing the author could change that they think would result in the largest improvement in quality. When providing this feedback, students are encouraged to think first about the various ways in which the piece could be improved but then focus on the single most important change, and they are clearly instructed not just to highlight some problem but to also provide clear direction about how it could be addressed effectively.

In terms of exercising learning objectives, the process of assessing peer work clearly implicates critical thought and self-reflection. Students are analyzing, contrasting and evaluating, ultimately for the purpose of specifying a level of quality relative to the other instances to which they have access. These are all facets of the critical thinking concept. In addition, we are exposing students to the compositions submitted by their peers, and it is inevitable that they compare their own work to that of their peers, gaining a clear and palpable sense of where their work compares and how it can be improved. In fact, the exercising of self-reflection can be enhanced further by asking students to explicitly assess their own work as well, a possibility that can be included in any and all phases.

[2] The power of peer assessment to enhance learning fits well with the views of Vygotsky (1978) and has been supported empirically and theoretically via many research studies (e.g., King, 2002; Venables & Summit, 2003).

The need to translate their analyses into feedback further exercises students' communication and collaboration skills, while also requiring some level of creative thought. The communication aspect is relatively obvious, but it is the flavor of that communication, especially with respect to the constructive comment, that is important. Students are essentially being asked to help their classmates improve, which is clearly a form of collaboration. Moreover, the stipulation that students must also tell their peers how to improve is critical because this is where creative thought comes into play. Detecting a problem requires critical thought, solving it requires creative thought, expressing that solution requires communication, and doing it all in a truly helpful way embodies collaboration.

Before leaving the Assess Phase, it is important to return to a point raised earlier. All of the processes described above are occurring in the context of some specified assignment that itself could be configured to exercise any other learning objective. For example, students could be asked to write about some aspect of social responsibility. Now all of their critical thinking, creative thinking, self-reflective thinking, communication and collaboration is focused on social responsibility. Thus any initial self-directed thoughts about social responsibility are being multiplied by the need to think about, assess and comment on the thoughts of classmates.

With respect to logistics and resources, note that all of the intense learning highlighted above occurs in the absence of faculty involvement. Students are learning by assessing and commenting on the work of their peers and the system handles all the logistics. This fact is also true of the subsequent reflect/revise phase.

The Reflect/Revise Phase

In the first step of the Reflect/Revise Phase, students see the scores and feedback that their peers provided to their work. Again, this typically reflects six scores and six sets of positive and constructive comments as described in the previous section. Students are instructed in advance that no matter what they do in life, peers will comment on their behaviour. Sometimes the comments are helpful, sometimes they are not, and it is critical that students learn to consider and judge the feedback they receive. In line with this, students are asked to explicitly assess the usefulness of each set of comments by categorizing them as "not useful", "useful" or "very useful".

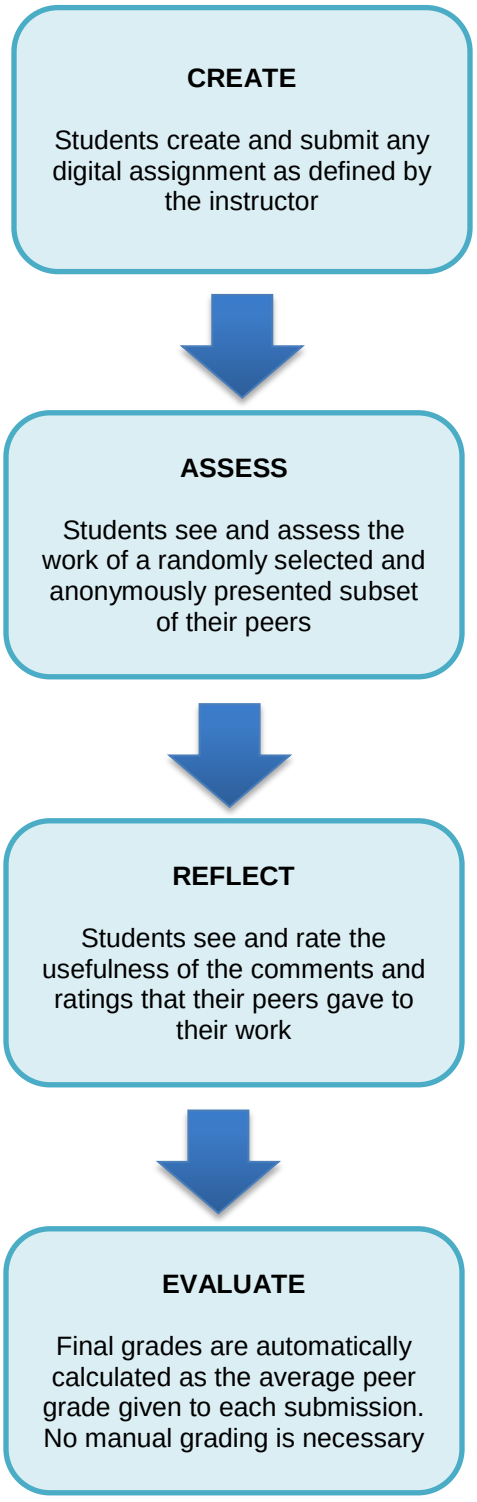
Before going any further, let us consider this process in terms of the stated learning outcomes. Students are assessing (i.e., thinking critically about) written comments (i.e., receptive communication) targeted at their work (i.e., self-reflection), work which hopefully was itself related to some desired learning outcome. Thus many of the learning outcomes exercised in the Assess Phase are being exercised again, though in the more personal context of one's own work. Repeated practice with any skill across various contexts enhances the generalization of skill learning.

A peerScholar assignment can and often does end at this point, and previous research from our lab (Paré & Joordens, 2008) and others (Cho, Schunn & Wilson, 2006) has shown that as long as at least five or six peer grades go into a final mark, that mark is as reliable as marks provided by a teaching assistant. Thus it is possible to exercise all of the learning objectives highlighted to this point without requiring any resources for

grading, which really is impressive indeed.3 This “partial peerScholar process” and its mapping to learning outcomes is illustrated in Figure 1.

[3] Despite the demonstrated validity of using an average peer grade as a final grade, and despite the rich learning that this process supports, in some institutions with unionized teaching assistants (e.g., the University of Toronto) having anyone other than teaching assistants providing grades has been argued successfully to violate labour law, even when no work hours are reduced. In these cases, the full peerScholar process must be used, which does require some grading resources – but no more than grading a typical essay, and the result is a much deeper learning experience.

Figure 1: Stages of a Partial peerScholar Assignment Mapping onto Common Core Learning Outcomes



Assignments can be created that focus on any specific learning outcome (e.g., students could be asked to be creative in some manner). Whatever outcome is the focus, thoughts related to that outcome will be amplified in every stage that follows (e.g., as they assess the creativity of their peers).

Providing a rating requires students to analyze, compare, evaluate and categorize, all aspects of critical thought.

Giving constructive comments to peers exercises creativity (how could it be better) and expressive communication.

To assess the usefulness of the comments directed at their own work students must exercise receptive communication, critical thought and self-reflective thought.

However, when resources are available for a final assessment by an expert, then it is possible to go a step further in a manner that provides even more exercise of critical learning objectives while also allowing more of them to be assessed quantitatively in the manner we will describe subsequently. Specifically, when the "full peerScholar process" is employed (Figure 2), students are required to submit a revised version of their work that is informed by the comments provided by their peers. The critical difference occurs in the third step of the depicted flowchart. After reading and assessing the comments, students are asked to do two things. First, they revise their work as they see fit, trying their best to improve it based on the feedback they received. Second, they write a short "reflection piece" in which they justify why they did and did not change their work in light of the specific comments given to them. Students are told that it is perfectly reasonable not to change their work in light of some comment as long as they can justify why they made that choice. In many ways then, this step, and the peerScholar process in general, embodies the sort of reflective practice championed by Shon and Argyris (1974, 1978) by asking students to put their learning into action in an iterative "learn then use" manner. Sometimes this approach is termed "formative assessment", and once again there is a strong research base supporting its power to enhance learning (e.g., Sadler, 1989).

These revision and reflection components put an emphasis on self-reflection and receptive communication. It is one thing to rate the usefulness of suggestions directed at oneself; it is something else to demonstrate the ability to sift through those comments and use them to evolve an idea or exposition in a way that improves its delivery. Thus, when the full process is employed, it ends by returning the student's focus to the improvement of their own work on the basis of their analysis of peer feedback.

Quantifying Learning Outcomes

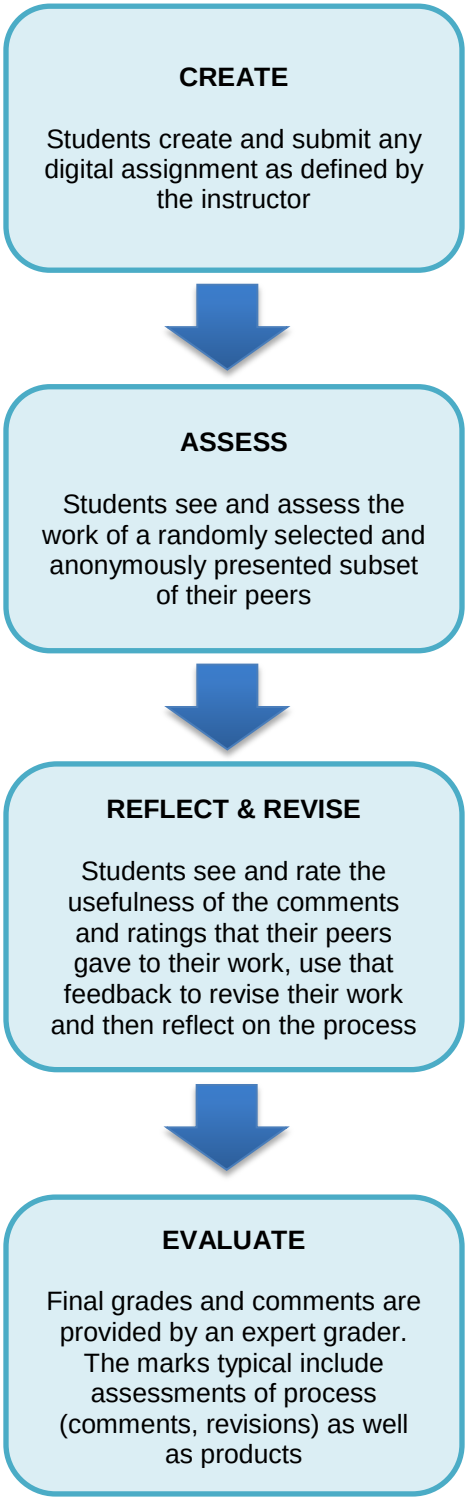
Returning to the initial analogy, our ultimate goal in this paper is to use peerScholar as an example of how the cognitive skills associated with critical learning outcomes can both be exercised and assessed simultaneously, just as doing arm curls both exercises and assesses the strength of one's biceps. The previous section focused on the exercise part of this formula and has hopefully provided a clear and compelling description of how an assignment employed within peerScholar exercises virtually all of the critical learning outcomes in a deep and interconnected way. That is, various components do not just exercise one or another outcome but often ask students to exercise various outcomes simultaneously and in combination, almost like doing arm curls while jogging! But exercise is one thing, assessment is another, and we now turn to the issue of how one can assess learning outcomes quantitatively within peerScholar.

As a first step, we must distinguish between outcomes that are measured automatically within the system and those that are quantified by a subject-matter expert (i.e., TA or instructor). The former will be possible to measure regardless of whether or not the full peerScholar process is employed, but the latter are only available using the full process. We will begin with the measures provided by a subject-matter expert and then move to the automated possibilities.

When the full peerScholar process is used, the final step of the process involves some subject-matter expert "grading" the assignment, and these grades are one way to obtain quantitative measures related to core learning objectives. Critically, a rich grading interface is available that supports the grading of both the products of the exercise and the underlying process. Typically, four components are graded: (1) the quality of

the initial draft composition, (2) the quality of the final composition, (3) the usefulness of the comments given to one's peers, and (4) the appropriateness of the revision.

Figure 2: Stages of a Full peerScholar Assignment Mapping onto Common Core Learning Outcomes



Assignments can be created that focus on any specific learning outcome (e.g., students could be asked to be creative in some manner). Whatever outcome is the focus, thoughts related to that outcome will be amplified in every stage that follows (e.g., as they assess the creativity of their peers).

Providing a rating requires students to analyze, compare, evaluate and categorize, all aspects of critical thought.

Giving constructive comments to peers exercises creativity (how could it be better) and expressive communication.

To assess the usefulness of the comments directed at their own work students must exercise receptive communication, critical thought and self-reflective thought.

To effectively use that feedback to revise they must exercise creativity, critical thought, reflective thought and expressive communication.

Recall our earlier statement that if an instructor wishes an assignment to target a specific learning outcome – social responsibility, for example – then they could ask the students to compose work that focused on that objective. The first two grades provided by the subject-matter expert then essentially provide a measure of the student's level – and short-term development – in terms of that learning outcome, at least as expressed through the medium used for composition (i.e., writing, art, video, etc.). The other two "grades" similarly reflect a student's ability to think critically, then creatively, then communicate their ideas for improvement to their peers (i.e., quality of comments to peers), and their ability to comprehend the communication of their peers, think critically and self-reflectively about it, then revise their work well (i.e., appropriateness of the revision). All of these relationships between component grades and core learning objectives are summarized in the top portion of Table 1.

Table 1. Quantitative Measures Derived from peerScholar and their Relation to Core Learning Objectives

Measure	***Related Learning Objectives***
Full peerScholar Process	
Quality of Draft Composition	Communication (in whatever form requested), plus any other learning outcome targeted by the assignment
Quality of Final Composition	Communication (in whatever form requested), plus any other learning outcome targeted by the assignment
Quality of Comments to Peers	Critical thought (identify main problem), creative thought (identify how best to improve it), communication (efficient and helpful)
Appropriateness of Revision	Receptive communication (understanding peer comments), self-reflection and critical thought (judging the usefulness of comments), communication (adjusting composition accordingly, and providing justifications in reflection piece)
Partial peerScholar Process	
Quality of Comments to Peers (average Usefulness Rating)	Critical thought (identify main problem), creative thought (identify how best to improve it), communication (efficient and helpful)
Quality of Composition (Average PeerMark)	Communication (in whatever form requested), plus any other learning outcome targeted by the assignment
Always Available	
Student-Peer Agreement (MAD Scores)	Critical thought
Self-Assessment Error (Self Grade vs. Actual Grade)	Reflective thought

However, it is also possible to obtain quantitative measures related to core learning objectives even when a subject-matter expert is not providing grades; that is, when the process is terminated after the rating of feedback, without students performing a revision or reflection, with their grade reflecting the average peer rating of their work. Once again, that average peer grade can be seen as reflecting a combination of a student's ability to communicate and their ability to demonstrate competency with respect to whatever learning objective was the focus of their composition. In the partial peerScholar process, this is the sole measure of this combined ability, but when the full peerScholar process is used it can also help inform the subject-matter specialists' opinion of the quality of the draft, in combination with their own reading and evaluation of said draft. So the full process provides what is likely a richer measure, but otherwise it targets the same basic outcomes.

Similarly, as highlighted in the middle portion of Table 1, the partial peerScholar process can also provide a somewhat less rich, yet still valid, measure of the quality of the comments a student gave to their peers. Recall that every comment a student gives ultimately receives a "usefulness" rating from their peers, and those ratings can be converted into numbers and averaged to provide a quantitative measure of a student's abilities to think critically to identify the main problem, think creatively about how the peer could best correct that problem, then communicate their thought in a clear, efficient and useful manner. Again, in the full peerScholar process, a subject-matter expert could use this measure in combination with their own subjective impressions of comments, but ultimately they would be trying to assess the same cognitive skills.

Thus both methods of implementing peerScholar can provide quantitative measures related to core learning objectives and they can do so as they exercise those same objectives. As described, these are composite measures in the sense that they do not purely measure single learning objectives but rather provide measures that reflect different combinations of cognitive skills. To some extent this is perhaps inevitable. Real-world tasks typically require one to combine skills in a dynamic and synergistic way, so given that these measures reflect task performance, it is not surprising that they reflect some combination. With that said, it would be fantastic if purer measures could be obtained.

In fact, there are two additional measures that can be generated automatically by peerScholar that do provide purer measures of critical and self-reflective thought, and they are obtainable regardless of how peerScholar is implemented. Of course, it is naive to argue that any single measure can reflect concepts as multifaceted as critical or reflective thought, so the more accurate statement is that these additional measures represent some aspect that is related to these outcomes. Caveat aside, at the very least they show the promise of using assignments that both exercise learning outcomes while providing an assessment clearly related to them. These measures are highlighted in the bottom portion of Table 1.

First, let us consider critical thought. Different people use the term a little differently in the literature, but most would agree that part of the concept at least refers to the ability to analyze, compare and evaluate with the goal to ultimately ascertain the quality of some argument. Of course, this is exactly what we ask our students to do in the Assess Phase of a peerScholar assignment, with the grade they assign reflecting their ultimate opinion of the quality of the given piece. It turns out that by making certain reasonable assumptions it is possible to quantify the accuracy of the grades a student provides and, in so doing, get a sense of their ability to think critically and ascertain the quality of the work they were assessing.

The measure of critical thought is computed as follows. A given student typically assesses six peers, each of whom is assessed by five other students. If we assume that the grade provided by averaging the five other student grades provides a good measure of the actual quality of the piece (Paré & Joordens, 2008), then if we compute the difference between the grade provided by a given student and the average of the other students who marked that peer, take the absolute difference, and sum these absolute differences and divide by the number of differences, we will arrive at what is termed a mean absolute deviation score. This formula is presented explicitly below:

$$\text{Critical thinking} = \frac{\sum_1^p |s - \bar{X}_g|}{p} \qquad [1]$$

Where p = number of peer assignments being graded
s = the grade the student in question gives to a specific assignment
$\bar{X}_g$ = the average of the other student grades for that assignment

Note that this measure is essentially an error score, in the sense that it gets larger when a given student is providing ratings that differ from the average peer grades. Thus a student with strong critical thinking skills should provide grades that are very close to the average peer grade, producing small average differences overall. If one preferred a critical thinking score for which higher levels suggested stronger critical thinking, then one could take the inverse of this quantity instead.

The computation of self-reflective thought would instead be driven primarily by the self-assessment scores that students can be asked to provide at any, or all, steps of the peerScholar process. Once again, if we assume that the average peer grade given to a student's work reflects the actual quality of that work (Paré & Joordens, 2008), then one can simply calculate the difference between the self-given grade and the peer average, with smaller differences reflecting better self-reflection skills.

Once again, this metric is expressed as an error, with larger scores meaning that students have a worse sense of the quality of their own work. Note that because one can ask students to self-assess in the Create, Asses and Revise/Reflect Phases, one can actually track changes in self-reflection abilities within an assignment.

Section Summary

The goal of this section was to describe explicitly how a tool designed to exercise learning outcomes could also provide assessments of those learning outcomes. Not all of the assessments are "pure", so we are not yet at the point of getting a breakdown of each particular learning outcome, but it is unclear that this is even a reasonable goal given that the cognitive skills underlying these outcomes often work together in some problem context. It is certainly clear, however, that something akin to the gym metaphor is possible: one can both exercise and assess at the same time, and one can do it at the assignment level.

Show Me the Data

Existence proof aside, to make this case as convincing as possible it would be really nice to know that some of the measures described above actually change in desirable ways. As students repeatedly assess their own work, does their measure of self-reflection improve? As students get practice in critical thought, does their critical thinking index get better? These questions were directly addressed across the following two experiments.

Experiment 1: Self-Reflection

As described, students doing a peerScholar assignment can be asked to provide a quantitative assessment of the quality of their own work, and they can be asked to do so at the end of any or all of the three phases. These self-assessments can be compared to a "real final grade", which might reflect either the average peer grade given to their work or a grade provided by an expert, depending on whether or not experts provide the final grades. The central question of this experiment was whether the accuracy of these self-assessments increases – or more specifically whether the difference between self-assessments and real final grade decreases – as a function of exposure to and assessment of peer work.

Participants

The participants were the 1,292 students within the Fall 2010 (A01) offering of Introduction to Psychology at the University of Toronto Scarborough who consented to allow their assignment data to be used for research.

Stimulus and Materials

The specific assignment was one in which the students were asked to form an argument either against or in favour of the current practice of performing certain research protocols on animals that we consider unethical when performed on humans. Thus, students were creating, then evaluating, and then revising short (approximately one- to two-page) argument pieces. All materials were presented within peerScholar as described previously in this report.

Procedure

All students were required to perform a peerScholar assignment, but the specifics of the assignment differed across two groups. At the time of this study, the partial peerScholar process was used for both groups (i.e., students were not asked to submit a revision) and the only difference was in terms of how often students were asked to self-assess. Specifically, half of the students (i.e., 646) were instructed to self-assess both after submitting their initial draft in the Create Phase and after assessing their peers' work in the Assess Phase; the remaining participants only self-assessed in the Assess Phase. This second group is not relevant to the current report but was included as part of a larger research project. Given that the current report is interested primarily in the change in self-assessment accuracy as a function of peer-assessment, we will focus only on the first group.

Students in this "double assessment" group first composed their draft, and after submitting it were simply asked to give themselves a mark out of 10 that reflected their assessment of the quality of their submission. Then, in the Assess Phase, they first evaluated, assessed and commented on the work of six of their peers. After completing that peer assessment they were again asked to assess their own work, prompted as follows: "You may recall the grade you gave to your work in the previous phase, but please try to think about your assignment with a fresh mind. Given all you have learned while grading your peers, look at your draft assignment again and let us know what grade you think it deserves."

Results and Discussion

The question of interest is whether the error in their ability to predict their own grade was reduced as a result of their exposure to and analysis of their peers' work. That is, by comparing their estimates to the average peer grade assigned to their work (see Paré & Joordens, 2008) and computing a difference score that essentially reflects their error in self-knowledge, we can assess whether this error declines once students see and assess the work of their peers.

It is interesting to note that students consistently overestimated their grades both before and after peer assessment. Specifically, the students' mean estimate after the Create Phase was 8.14 and after the Assess Phase was 7.74, both of which are higher than their actual average mark of 6.64. The mean associated with each phase was statistically significant compared to the average peer grade using a student's t-test, which was appropriate given that the scores underlying each mean were normally distributed and both were found to be reliably greater than the average peer grade, $ts(644) > 15.23$, $ps < .0001$. This is simply further evidence of a well-known social psychology phenomenon termed the "self-serving bias": seeing our performance as stronger than it actually was helps to keep our self-esteem reasonably high (Miller & Ross, 1975).

More relevant to the current report, though, the self-assessments before versus after peer-assessment (i.e., 1.5 vs. 1.1 respectively) were also compared to one another and the difference was also statistically reliable, $t(644) > 7.80$, $p < .0001$. This finding suggests that a single experience of assessing six peers was sufficient to improve these students' ability to assess the quality of their own work. That is, their ability to self-reflect was improved.

Experiment 2: Critical Thought

The concept of critical thought is multifaceted, and at some level it borders on egregious to argue that a student's ability to think critically can be captured by any relatively simple quantitative measure. And yet when one considers the sort of cognitive processes that an evaluator must draw on in order to assess the quality of a piece of work, especially in the context of other peer pieces, the terms one arrives at are those typically associated with the concept of critical thought. That is, they are being asked to evaluate, contrast, analyze, categorize, etc. Thus, although it may be a bold statement, we feel that the accuracy with which a student can grade a range of different works does provide a pretty good, if somewhat simple, measure of their critical thinking abilities. If this is the case, and if peerScholar really does help students develop their critical thinking abilities, then we would expect that the accuracy of their grading would also improve with experience. That is the question this experiment was designed to test.

Participants

The participants were the 1,279 students within the Fall 2011 (A01) and Winter 2012 (A02) offerings of Introduction to Psychology at the University of Toronto Scarborough who consented to allow their assignment data to be used for research. Of these, 737 performed peerScholar assignments at Times 1, 2 & 3 (the Repeated Practice Group), while 542 did their first and only peerScholar assignment at Time 3 (the Control Group). Specifically, students can take the first two parts of Introduction to Psychology separately and in any order. For those who first took A01 and then A02, Time 3 represented their third peerScholar assignment, but for those who took A02 without taking A01, Time 3 represented their first peerScholar assignment.

Stimulus and Materials

The specific assignment changed across the different levels of Time but, in each instance, students were asked to create a short, efficient argument either for or against some issue or practice. The assignments were created to be similar in terms of difficulty. All materials were presented within peerScholar as described previously in this report, and in each case students assessed, and were assessed by, six of their peers.

Procedure

At the time of this study, students in the first part of our Introductory Psychology course did two peerScholar assignments and those in the second part of the course did one. The process was the partial peerScholar process described earlier, and based on the assessments students provided we were able to compute the mean deviation scores related to critical thinking as outline in Equation [1]. Thus, the Repeated Practice Group provided three such scores, one associated with each of the levels of the Time variable. In contrast, the Control Group provided just one score associated with Time 3.

Results and Discussion

The goal of this experiment was to assess the impact of peerScholar on these critical thinking scores with two questions in mind. First, do these scores drop with repeated practice, as the gym metaphor would predict? This question can be examined by simply examining the scores for the Repeated Practice Group across the three Time levels. Second, if we do see a drop, can we be relatively certain that the drop is due to the practice and not simply due to being in a university environment for a term? A comparison of the Repeated Practice Group to the Control Group will provide a relatively clean answer to that question, given that both sets of students have been in university for a term, but only the Repeated Practice Group has previous practice with peerScholar.

With respect to the first question, a one-way analysis of variance (ANOVA) was conducted using the change in mean absolute deviation (MAD) scores (i.e., those computed using Equation [1]) as the dependent variable and the assignment time for the peer assessment group (Time 1, Time 2, Time 3) as the independent variable. A significant difference was found between the various assignment times, $F(2, 1472) = 24.7$, $p < 0.001$, $\eta 2 = 0.03$, see Table 1 for means for each factor. Given the significant finding, and based on our hypothesis, we further examined the data and found a significant linear trend, $F(2, 1472) = 24.7$, $p < 0.001$, $\eta 2 = 0.06$.

Table 2. Mean Absolute Deviation Scores Related to Critical Thought (lower values suggest closer convergence with final grade)

	Time 1	Time 2	Time 3
Repeated Practice	1.08	0.98	0.95
Control			1.04

These findings clearly indicate that participation in a peer assessment assignment enhanced students' ability to discriminate based on quality. Not only did the first exposure to the assessment phase of the assignment lead to enhanced quality-based discrimination skills, but so did each subsequent exposure. This finding supports our hypothesis that repeated exposure to peer assessment assignments would further develop students' ability to recognize quality and thereby enhance their critical thinking skills.

Next, we compared the MAD scores of the control group's first assignment (Time 3) to those of the peer assessment group's first assignment (Time 1) and the peer assessment group's third assignment (Time 3). At Time 3 for the control group vs. Time 1 for the peer assessment group, there was no significant difference found between MAD scores, $t(1277) = -1.876$, $p = 0.061$, n.s., whereas at Time 3 for the control group vs. Time 3 for the peer assessment group, there was a significant difference, $t(1277) = 3.573$, $p < 0.001$.

Thus, as was the case with the measure of self-reflection, the measure of critical thought that one can obtain from peerScholar changes as one would hope if the gym metaphor were correct. That is, repeated practice with the assessment of peers' work allows students to become better at judging the quality of that work, and to the extent that the judgments of quality that arise from comparison, evaluation, analysis and categorization reflect critical thought, the suggestion is that critical thought – or at least an index related to it – can both be exercised and assessed during the course of a single assignment.

Overall Summary

It is relatively easy for people to identify and describe desirable learning outcomes, but it is much harder for them to describe explicitly how they can be exercised and assessed. Some have suggested using subjective aggregate data that attempt to capture a student's experience on some wide time scale, as might be provided by surveys of student experience. We argue that it would be far more desirable both to exercise and assess learning outcomes on an assignment level, and from there extrapolate to any higher level one might wish.

The primary purpose of this report was to demonstrate that this goal is attainable. Focusing on one specific learning technology, peerScholar, we show how quantitative measures related to learning outcomes can be computed on an assignment level. We also describe some recent findings from our lab showing that these measures behave as one would hope in terms of showing improvements resulting from cognitive practice. Taken together, these provide what we see as very strong "existence proof" that assignment-based exercise and assessment is possible.

We should highlight that we are not arguing that everyone should therefore use peerScholar, whether or not we believe this to be true. We focused on it because we know it best as a product of our lab. Similar sorts of analyses could and should be done for other assignment contexts, which should result in a broad spectrum of tools one could use both to exercise and assess a range of learning objectives in different ways. Optimally, demonstrations of mastery with respect to specific learning objectives would be defined and tracked explicitly, perhaps through devices like Mozilla Badges or digital portfolios and, ultimately, these demonstrations of mastery might become more relevant than traditional grades. But to reach those kinds of goals we first need to link assignments to learning outcomes, and hopefully this report has shown convincingly that this is entirely possible. It is time to take that next step, a step into the gym of the mind.

References

Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and Reliability of Scaffolded Peer Assessment of Writing From Instructor and Student Perspectives. *Journal of Educational Psychology*, *98*, 891-901.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Foundation for Critical Thinking. (2013). Retrieved from http://www.criticalthinking.org/pages/critical-thinking-where-to-begin/796

Joordens, S., Desa, S., & Paré, D. E. (2009). The pedagogical anatomy of peer assessment: Dissecting a peerScholar assignment. *The Journal on Systemics, Cybernetics and Informatics, 5*(7), 11-15.

King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice*, *41*, 33-39.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, *82*(2), 213-225.

Milner, B. (1962). In P. Passouant (ed.), *Physiologie de l'hippocampe* (pp. 257-272). Paris: Centre national de la recherche scientifique.

Paré, D. E., & Joordens, S. (2008). Peering into large lectures: Examining peer and expert mark agreement using peerScholar, an online peer-assessment tool. *Journal of Computer Assisted Learning, 24*(6), 526-540.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119-144.

Samuels, S. J., & Flor, R. F. (1997). The Importance of Automaticity for Developing Expertise in Reading. *Reading & Writing Quarterly*, *13*(2), 107-121.

Schon, D. A., & Argylis, C. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco, CA: Jossey-Bass.

Schon, D. A., & Argylis, C. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison-Wesley.

Tulving, E. (1985). How many different memory systems are there? *American Psychologist, 40*, 385-398.

Venables, A., & Summit, R. (2003). Enhancing scientific essay writing using peer assessment. *Innovations in Education and Teaching International*, *40*, 281-290.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Higher Education Quality Council of Ontario

An agency of the Government of Ontario