



## AUTHORS

Rana Khan, Ph.D.  
University of Maryland,  
University College

Datta Kaur Khalsa, Ph.D.  
University of Maryland,  
University College

Kathryn Klose, Ph.D.  
University of Maryland,  
University College

Yan Zhang Cooksey, Ph.D.  
University of Maryland,  
University College

## CORRESPONDENCE

*Email*  
rana.khan@umuc.edu

## Abstract

Since 2001, the University of Maryland University College (UMUC) Graduate School has been conducting outcomes assessment of student learning. The current 3-3-3 Model of assessment has been used at the program and school levels providing results that assist refinement of programs and courses. Though effective, this model employs multiple rubrics to assess a wide variety of assignments and is complex to administer. This paper discusses a new outcomes assessment model called C2, currently being piloted in UMUC's Graduate School. The model employs a single common activity (CoA) to be used by all Graduate School programs. It is designed to assess four of the five student learning expectations (SLEs) using one combined rubric (ComR). The assessment activity, scored by trained raters, displays pilot results supporting inter-rater agreement. Pilot implementation of the C2 model has advanced its reliability and its potential to streamline current assessment processes in the Graduate School.

# Assessing Graduate Student Learning in Four Competencies: Use of a Common Assignment and a Combined Rubric

**U**niversity of Maryland University College (UMUC) has been involved in institutional assessment of student learning in both its undergraduate and graduate schools since 2001. According to Palomba and Banta (1999), assessment is “the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development” (p. 4). UMUC’s institutional assessment plan, consistent with Walvoord’s (2004) recommendations, aligns with its mission, core values, and strategic plans. The plan also provides an overarching conceptual framework that defines student learning outcomes, provides a roadmap for assessing student learning, and ensures the use of findings for the improvement of UMUC programs. In the Graduate School, the current model of assessment is based on a framework introduced in 2010. This framework measures five student learning expectations (SLEs) and consists of *three* rounds of assessment at *three* stages carried out over a *three* year period each spring semester and has been named the 3-3-3 Model. Though the current process is effective in systematically collecting data across the Graduate School, it is a complex process to administer. This paper describes two phases of a pilot study, the intent of which was twofold: (a) to simplify the current Graduate School assessment process and (b) to examine and refine a new model that employs a recently developed assessment instrument. This article contributes to educational literature that focuses on graduate school assessment methods and will assist assessment practitioners by sharing the authors’ experiences with piloting a new

assessment model. Details and results of the pilot study, including information on the current model, design of the new assessment model, online rater training, and interpretation of the pilot results follow.

### Graduate School Assessment Process-Current Assessment Model

In line with university priorities and strategies, UMUC's Graduate School has established a commitment to systematic assessment and the use of assessment results to improve student learning. The Graduate School views assessment as an ongoing and collaborative process driven by continuous reflection and improvement as described by Maki (2004). The current 3-3-3 assessment model employed by the Graduate School obtains evidence of student learning by assessing five student learning expectations (SLEs; Appendix A). The five SLEs include Communication (COMM), Critical Thinking (THIN), Information Literacy (INFO), Technology Fluency (TECH), and Content Knowledge (KNOW) and are expected of all UMUC graduate students.

The 3-3-3 model consists of three rounds of assessment carried out over a three-year period each spring semester, with each round assessing all five SLEs (See Figure 1). This model takes a "snapshot" of student learning at three points in a program lifecycle. Assessments are run within the first 9 credits, between 10 and 18 credits and at 19-36 credits, marking beginning, intermediate and advanced levels of study.

For each round, program directors, who manage courses in the Graduate School, identify the courses/sections that will conduct assessment activities. Within each course/section, class activities are chosen that will allow students to demonstrate their abilities in specific SLEs.

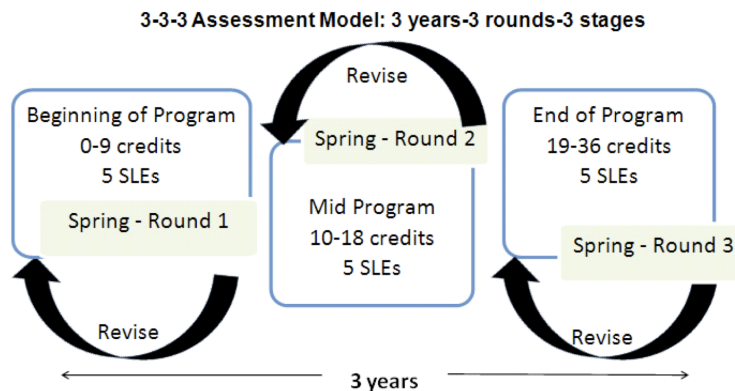


Figure 1. UMUC's 3-3-3 assessment model.

There are a variety of tools that may be used for assessing student learning, including standardized tests, interviews, surveys, external examiners, oral exams, rubrics, and e-portfolios (Prus & Johnson, 1994). UMUC's Graduate School has chosen to use rubrics to assess student learning for each SLE for reasons aligned with the thinking of Petkov and Petkova (2006), who cite ease of implementation, low costs, student familiarity, and applicability to a variety of performance criteria. Rubrics can also be used in both formative and summative evaluation. For use with its current 3-3-3 model, the Graduate School designed a set of analytic rubrics where rubric criteria align with each of the school's five SLEs. Each rubric describes student performance over four progressively increasing levels of attainment (unsatisfactory, marginal, competent & exemplary).

Consistent with the design recommendations offered by Moskal (2000) and Nitko (2001), each Graduate School rubric contains criteria that serve to identify and describe the separate dimensions of performance that constitute a specific SLE. Instructors are required to score each rubric criterion and sum the scores. For example, the Graduate School

**The Graduate School views assessment as an ongoing and collaborative process driven by continuous reflection and improvement.**

**While the current 3-3-3 model has served the Graduate School well and proven reliable in delivering useful data for our goals, it has limitations and challenges.**

has identified the criteria of Conceptualization, Analysis, Synthesis, Conclusions and Implications as dimensions of the Critical Thinking SLE. When assessing assignments associated with Critical Thinking, faculty assign a score to each criterion, which is then summed up. By assigning a score to each criterion, faculty and course/program administrators receive multidimensional information on student performance. In addition to providing insights on specific levels of student learning, the inherent design of analytic rubrics employed in the 3-3-3 model provides students with specific feedback via the criteria definitions. The feedback enables students to focus on areas where they need improvement. The analytic rubric lends itself to formative use of rubric information, as opposed to the more summative approach inherent in holistic rubrics (Mertler, 2001). In this way, UMUC faculty and administrators use the results derived from the rubric scores to inform improvements to their courses and programs. In line with the iterative approach to rubric design described by Wiggins (1998), the Graduate School has over the past three rounds of assessment refined its rubrics based on assessment findings and user feedback. An example of a rubric currently employed in the 3-3-3 model is contained in Appendix B.

**The primary resource needed for the development of the C2 model was time. The collaborative process took over a year from the time the idea was first proposed by the researchers to the Graduate School Assessment Committee to the time the pilot was conducted in Spring 2012.**

When Graduate School faculty carry out assessment activities in their classes, they are responsible not only for assigning a class grade to select assessment assignments, but must also score the assignments using the appropriate Graduate School rubrics. The faculty must record the students' rubric scores for each specific SLE criteria on a summary sheet and submit the sheet to the Graduate School. Faculty and administrators are later provided with a summary of the assessment findings and asked to develop action plans to address the most significant areas of weakness in their programs. This completes the assessment cycle by looping actionable improvements into the course/program.

An example of this loop-back into courses and programs is the implementation of an Accounting and Finance Research Module designed by UMUC's Library Services. Round 1 assessment findings indicated that, related to the SLE of Information Literacy, students in Accounting and Finance scored low on the criterion of Identification and were not able to competently differentiate between scholarly and trade journals when conducting research. Upon analyzing the findings, the program director asked UMUC Library Services to develop a resource exclusively for helping students understand how to evaluate the quality of publications used in their research. Subsequent findings in Rounds 2 and 3 showed improvement in the criteria of Identification among Accounting and Finance students.

The Graduate School completed its first 3-year assessment cycle under the 3-3-3 model in Spring 2012. While the current 3-3-3 model has served the Graduate School well and proven reliable in delivering useful data for our goals, it has limitations and challenges that include:

- extra grading workload for faculty who teach courses identified for assessment,
- no training or norming for faculty on rubric use,
- disparities in the types of assignments used for assessment across the Graduate School,
- misalignments between the assignments and rubrics, and
- inconsistencies in grading practices among faculty.

As described by Buzzetto-More and Alade (2006), the reflection that occurs in relation to the assessment cycle often stimulates discussion and suggestions for improvements, and plans for implementing change. With the completion of the cycle came the opportunity to review the current model, which led to the design of the C2 model and current pilot study discussed in this paper.

### **Graduate School Assessment Process-Proposed Assessment Model**

The limitations and challenges of the 3-3-3 model are not unusual in nature and relate to those described by those writing in the area of outcomes based assessment such as Banta (2002), Bresciani (2011), and Maki (2010). These challenges relate to understanding

the goals of assessment and having the resources and time necessary to carry out assessment activities. To address some of the aforementioned challenges, the authors proposed a new model called C2 to simplify the current annual process.

### Development of Common Activity (CoA)

In the C2 model, a single common activity (CoA) is used by all UMUC's Graduate School programs to assess four SLEs (COMM, THIN, INFO, and TECH). The CoA requires that students respond to a question in a short essay format to demonstrate their levels of performance in the four learning areas. Collaboratively developed with representatives of all the Graduate School departments, the question relates to commonly addressed program themes (i.e., technology globalization and leadership) and does not require prior knowledge of the topic. The CoA instructions present the essay question, clearly describe for students the steps for completing the task, and explain how the submission will be evaluated. Of note, the SLE, KNOW, was excluded from the model design. While it is a learning outcome expected of all students in the Graduate School, it is viewed as very program/discipline-specific and therefore, more appropriately assessed by other means, which may include standardized exams or special projects.

### Design of Combined Rubric (ComR)

A new rubric (ComR) was designed for use in the C2 model by initially combining relevant and established criteria from the current rubrics used in the 3-3-3 model, excluding those related to knowledge (KNOW). The researchers remained committed to the use of analytic rubrics in the C2 model for the benefits cited previously, including their ability to present a continuum of performance levels, provide qualitative information on observed student performance, and the potential for tracking student progress (Simon & Forgette-Giroux, 2001). The ComR rubric removed overlaps between the four existing rubrics. The steps in the design of the ComR involved:

- Consolidation of individual rubrics from four SLEs (COMM, THIN, TECH, INFO) into a single rubric (ComR) with fourteen criteria
- Review and revision based on feedback from the Graduate School Assessment Committee
- Use of ComR in Phase I to test content validity and alignment between ComR and the CoA
- Review and revision based on feedback from raters in Phase I to further consolidate ComR into nine criteria
- Application of the refined ComR in Phase II

The ComR rubric employed in Phase I is presented in Appendix C and Appendix D shows the refined ComR rubric used in Phase II.

The C2 model was designed to provide the means to evaluate multiple SLEs simultaneously and to score the common activity (CoA) by trained raters. Table 1 contrasts the new C2 model with the current 3-3-3 model.

### Allocation of Resources

The primary resource needed for the development of the C2 model was time. The collaborative process took over a year from the time the idea was first proposed by the researchers to the Graduate School Assessment Committee to the time the pilot was conducted in Spring 2012. Fortunately, all members of the committee were in agreement that the existing 3-3-3 assessment model needed to be simplified and improved, therefore it did not take much convincing for them to agree to participate in the pilot. The most time expended was in the development of the common activity (CoA) and the combined rubric (ComR). The essay question for the CoA was developed over a period of several months until a consensus was reached across the Graduate School. The ComR was created through

**Moskal and Leydens (2000) suggest that discussing differences in raters' scores helps improve reliability, as does making performance criteria more precise, though narrow criteria definitions may preclude applicability to other activities. Bresciani, Zelna and Anderson (2004) contend that norming ensures that raters understand the rubric in a similar manner, which promotes consistency in scoring, and thereby enhances reliability.**

Table 1

*Comparison of the Current 3-3-3 Model and the Combined Activity/Rubric (C2) Model*

<b>Current 3-3-3 Model</b>	<b>Combined Activity/Rubric (C2) Model</b>
Multiple Rubrics: one for each of 4 SLEs	Single rubric for all 4 SLEs
Multiple assignments across graduate school	Single assignment across graduate school
One to multiple courses/4 SLEs	Single course/4 SLEs
Multiple raters for the same assignment/course	Same raters/assignment/course
Untrained raters	Trained raters

an iterative process, which included sharing each draft edition and making adjustments until the committee was in agreement. Additional resources included a stipend paid to the seven hired raters trained for grading. The funds for the stipends were provided from a federal grant. These stipends resulted in a total cost of \$7,000.

### Implementation of C2 Model

The pilot study was conducted sequentially through two phases: Phase I and II. In Phase I, the ComR was used in three graduate programs to determine its reliability for grading the CoA. The three Masters' programs that were part of Phase I included Biotechnology, Master of Arts in Teaching, and Master of Education in Instructional Technology. The three programs were selected based on the interest and willingness of the degree program directors to participate in the pilot and their ability to easily incorporate the pilot activity into their courses. The CoA was explained in the syllabi of the courses selected for the pilot study and was scheduled to be completed during the first quarter of the semester.

### Raters' Training and Norming

**The C2 model appears to have simplified the assessment process. The new C2 assessment model implemented a common activity (CoA) and used a combined rubric (ComR) for the outcomes assessment process.**

Adding trained raters to the C2 model was done for the purposes of simplifying faculty workloads and improving scoring consistency. Program directors were asked to suggest faculty who could act as raters for the pilot papers. The faculty raters needed to fit the following guidelines: they were *not* teaching any of the pilot courses in Spring 2012, had experience teaching and grading in the participating programs, and therefore could easily become 'raters' for the pilot study. All seven recommended faculty members were contacted and 100% agreed to participate in the study. Contracts for the faculty raters were discussed, signed and processed with an agreed-upon timeline for training, scoring procedures and follow-up.

A total of 91 students completed the activity. The papers were collected, redacted of any identifiable information, and assigned a code number prior to being distributed to the raters. Raters were given a set of anchor papers, selected from the submissions, which provided the raters with samples of varying levels of student performance (Tierney & Simon, 2004). To strengthen reliability and yield a consistency in grading with the rubric, raters were required to participate in norming sessions (Trochim & Donnelly, 2006) prior to and after the scoring of the anchor papers. Since raters were geographically dispersed, the norming sessions were conducted online, both asynchronously and synchronously, to allow for flexibility and scalability. All raters actively engaged in the training and norming sessions, which provided them with the opportunity to practice scoring anchor papers and discuss in detail the interpretation and application of the combined rubric for grading. Moskal and Leydens (2000) suggest that discussing differences in raters' scores helps improve reliability, as does making performance criteria more precise, though narrow criteria definitions may preclude applicability to other activities. Bresciani, Zelna and Anderson (2004) contend that norming ensures that raters understand the rubric in a similar manner, which promotes consistency in scoring, and thereby enhances reliability.

Papers were assigned to raters in a discipline-specific manner in Phase I such that the raters from the Education department received and scored papers from students

in Education, while raters from Biotechnology graded papers from the Biotechnology program course.

## Inter-rater Reliability

In this study, each paper was randomly assigned to two independent raters and graded by them using the same scoring rubric. This process is called coding because the raters are creating the data when they assign scores (ratings) to each student paper. Stemler (2004) states that in any situation that involves judges (raters), the degree of inter-rater reliability is worthwhile to investigate, as the value of inter-rater reliability has significant implication for the validity of the subsequent study results. There are numerous statistical methods for computing a measurement estimate of inter-rater reliability (e.g., simple percent-agreement, Cohen's Kappa, generalizability theory, Pearson  $r$ , etc.) and each of them has advantages and disadvantages (Stemler, 2004). For example, Pearson  $r$  can be a useful estimator of inter-rater reliability only when one has meaningful pairings between two and only two raters (linear relationship between the two set of ratings). Cohen's Kappa is commonly used for calculating inter-rater reliability for qualitative (categorical) data (i.e., gender, age, education level, etc.). Its greatest advantage is taking into account chance agreement between two or more raters. However, Kappa assumes that all raters have similar training and experience. When raters have dissimilar training and experience, the Kappa statistic is likely to be underestimated (Crewson, 2005).

Intraclass Correlation Coefficients (ICC) were used in this study for the estimation of inter-rater reliability. An ICC is a measure of the proportion of a variance that is explained by the objects (i.e., students) of measurement (i.e., raters' ratings). ICC has advantages over bivariate correlation statistics, such as Pearson  $r$ , as it accounts for variability between multiple raters and among the multiple dimensions of the rubric. Reliability assessed by ICC is a scaled agreement index under ANOVA assumptions. As discussed in the works of McGraw and Wong (1996) and Shrout and Fleiss (1979), to select an appropriate form of the ICC, one has to make several decisions related to (a) which ANOVA model should be used to analyze the data (one-way or a two-way); (b) whether differences in raters' mean ratings relevant to the reliability of interest (ICC for consistency vs. absolute agreement) and (c) whether the unit of analysis is a mean of several ratings or single rating (ICC for average vs. single measurements).

In this study, each student paper was rated by a randomly selected group of two raters from a larger pool. In other words, the same two raters did not grade all the papers. No effort was made to disentangle the effects of the rater and student paper, but only the objects (students) were treated as a random factor. Therefore, a one-way random effects ANOVA model was used to calculate the ICC (measures of absolute agreement were selected, as consistency measures were not defined in this model). The "average measures" ICC was provided in the results, which indicates the inter-rater reliability when taking the mean of all ratings from multiple raters and multiple dimensions of the rubric. The ICC will approach 1.0 if there is less variance within item ratings. According to Nunnally (1978), an ICC of 0.7 is generally considered an acceptable level for the type of study employed in this pilot.

## Multiphase Approach

The researchers anticipated that the development of the C2 model would be a process of continuous improvement. For this reason, Phase II was performed and lessons learned from Phase I were applied that included further refining the ComR based on feedback provided by the raters and modifying the pilot process. Refining the rubric involved eliminating what the raters determined were redundant or overlapping criteria and clarifying criteria descriptions. In terms of modifying the pilot process, the same set of papers and raters from Phase I were used in Phase II, but the raters were given different subsets of papers and the papers were *not* assigned in a discipline-specific manner. This modification was made to allow us to gain insight into how well raters would handle rating papers from different disciplines, which is an ultimate goal in the Graduate School's full implementation.

**The pilot norming results emphasized the importance of providing a range of anchor papers that represented different levels of student performance in order to determine and discuss baseline scoring.**

## Results

In both Phases I and II, each paper was rated by two raters and the ICC was computed. Table 2 displays a value of 0.44 in ICC from the Phase I data, which means that approximately 44% of the time two independent raters assessed an item and then scored it with the same value. The ICC is lower than the generally acceptable level of 0.70. In an attempt to increase the relative low reliability (0.44) generated in Phase I, the authors refined and consolidated the ComR to remove redundancy, and thereby reduced the number of criteria from fourteen to nine. The authors carefully selected a different set of anchor papers than those used in Phase I that clearly represented different levels of student performance. In addition, in Phase II, a third rater was used for papers when the scores between two raters had discrepancies greater than 1 point in at least 3 criteria. These extreme scores were discarded before calculating the Phase II ICC.

Table 2

<i>Average Measures of ICC – Phase I &amp; II</i>		
	Intraclass Correlation Coefficients	
	Phase I	Phase II
Average Measures	0.44	0.75

By implementing the refinements and consolidations to the rubric and common activity, Phase II ICC provided a value of 0.75, meaning approximately 75% of the time two independent raters assessed an item and then gave it the same score (Table 2). Since the ICC for Phase II reached the generally acceptable level (0.70) of agreement among these raters, it provided confidence in the reliability of the C2 model.

## Discussion

As mentioned earlier, the present 3-3-3 Graduate School assessment model has some limitations. One of those is the increased faculty workload of grading a wide variety of assignments that are used for assessment across the Graduate School programs. With the 3-3-3 model, there can also be grading inconsistency and weak alignment between the assignment and the rubrics.

The C2 model appears to have simplified the assessment process. The new C2 assessment model implemented a common activity (CoA) and used a combined rubric (ComR) for the outcomes assessment process. It also addressed the concerns with the current 3-3-3 model in that it:

- shifted the faculty grading workload to external, trained raters,
- incorporated training and norming sessions to improve rubric consistency and use,
- eliminated assignment disparities by employing one common activity across the Graduate School, and
- provided tighter alignment between the assignment and rubric.

Rezaee and Kermani (2011) write that “raters’ inconsistencies in scoring can be attributed to different factors among which are raters’ gender, age, teaching experience, scoring experience, first language and scoring environment” (p. 109). Furthermore, Bresciani et al. (2004) report that low reliability among raters may be influenced by the (a) objectivity of the task or scoring, (b) complexity of the task, (c) group homogeneity of the raters, (d) work pace of the raters, and (e) number of assignments scored. A lower agreement among raters may result from various reasons such as ambiguity of the rubric criteria and activity instructions, misunderstanding of rubric criteria, preconceived notions held by raters, and using a small pool of raters. In Phase I the ICC of .44 was lower than the generally acceptable .70 level, indicating the potential presence of such issues for the participating raters. In Phase II, the authors addressed some of these issues in an attempt to improve the inter-rater reliability, the results of which, was an improved ICC of .75.

**Instructors often feel a pressure to align assessment scores with assignment grades, whereas raters can focus solely on the criteria under assessment. External raters may also possess more knowledge and understanding of the specific criteria under assessment.**

Although the effect of norming on inter-rater reliability may be disputed, the researchers recognized the importance of the norming process for refining the rubric and the activity. The pilot norming results emphasized the importance of providing a range of anchor papers that represented different levels of student performance in order to determine and discuss baseline scoring. Rater feedback during the norming process also informed further rubric consolidation. The iterative process of refining the CoA and ComR worked toward ensuring that the criteria for each SLE were discrete, not dependent on each other and directly assessable. As a result, the original combined rubric (ComR) with fourteen criteria was consolidated further in Phase II to nine criteria, again simplifying the use of the rubric and potentially contributing to better application and agreement among raters.

In addition, there appears from the pilot to be benefits in using external raters to score assessment activities as opposed to the teaching faculty. Instructors often feel a pressure to align assessment scores with assignment grades, whereas raters can focus solely on the criteria under assessment. External raters may also possess more knowledge and understanding of the specific criteria under assessment. In addition, providing a potential point of comparison between rater and teacher evaluations may serve in evaluating assessment findings.

### Limitations of this Study

Even though the main goals of this pilot study were met and simplification of the current Graduate School assessment process seems promising, there are limitations to this study and future research is needed to address them.

The use of a single assignment and rubric to evaluate multiple competencies may be construed as a limitation. As Maki (2004) points out, “Relying on one method to assess the learning described in outcome statements restricts interpretations of student achievement within the universe of that method” (p.156); using multiple measures to assess different learning outcomes, on the other hand, has its advantages. However, others have explored the possibility of combining various rubrics to evaluate multiple learning outcomes based on a single student assignment (Stanny & Duer, 2012). In addition, just as the trained raters provided feedback for the rubric in Phase I of this pilot study, the researchers expect to continue to receive feedback for further refinements in future phases of our studies.

Another limitation may result from the design of the study. In this pilot study no two raters graded all the same papers. This was intentional as eventually a pool of raters will be expected to grade all the papers that come out of the Graduate School. Having the same two or more raters grade all the papers will not be practical for implementation purposes. Consequently, one-way (or one-factor) random effect ANOVA model using objects (students) as the only effect was used to calculate ICCs. This approach limited the ability to evaluate the rater effect as a variable because specific raters and the interactions of raters with students were not disentangled. Intra-rater reliability, a measure of the rater’s self-consistency, also could not be defined in this study.

### Conclusions and Further Studies

This study describes the implementation of a unique assessment model, C2. Our findings indicate that this model may have a higher rate of reliability than the Graduate School’s current 3-3-3 model. Using the C2 model, several core learning competencies may be assessed simultaneously through a common assignment, a combined rubric, and trained raters across different graduate programs. This model is an attempt to improve the comparability of the data across programs, increase clarity of the process, decrease faculty workload, and therefore greatly simplify the outcomes assessment process. To evaluate both object (student) and rater effects, either the two-way random or mixed effects model, in which each student paper is rated by the same group of raters, may be used in future studies.

In order to further improve on the reliability of scores from the common activity and the combined grading rubric, Phase III of the C2 model will be applied to several

**This model is an attempt to improve the comparability of the data across programs, increase clarity of the process, decrease faculty workload, and therefore greatly simplify the outcomes assessment process.**



programs across the Graduate School in Fall 2012 in preparation for a potential graduate school-wide implementation. Post graduate school-wide implementation, the authors will focus on methods to establish the validity of the C2 model.

## References

- Banta, T. W. (2002). *Building a scholarship of assessment*. San Francisco, CA: Jossey-Bass.
- Bresciani, M.J. (2011). Identifying barriers in implementing outcomes-based assessments program review: A grounded theory analysis. *Research and Practice in Assessment*, 6, 5-16. Retrieved from <http://www.rpajournal.com/archive/>
- Bresciani, M.J., Zelna, C.L., & Anderson, J.A. (2004). *Assessing student learning and development: A handbook for practitioners*. Washington, DC: National Association of Student Personnel Administrators (NASPA).
- Buzzetto-More, N.A., & Alade, A. J. (2006). Best practices in e-assessment. *Journal of Information Technology Education*, 5, 251-269. Retrieved from <http://informingcience.org>
- Crewson, P. (2005). Fundamentals of clinical research for radiologists: Reader agreement studies. *American Journal of Roentgenology*, 184(5), 1391-1397. Retrieved from <http://www.ajronline.org>
- Maki, P. L. (2004). *Assessing for learning: Building a sustainable commitment across the institution*. Sterling, VA: Stylus.
- Maki, P. L. (2010). *Coming to terms with student outcomes assessment: Faculty and administrators' journeys to integrating assessment in their work and institutional culture*. Sterling, VA: Stylus.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46. Retrieved from <http://psycnet.apa.org>
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25). Retrieved from <http://pareonline.net/>
- Moskal, B. M. (2000). Scoring rubrics: What, when and how? *Practical Assessment Research & Evaluation*, 7(3). Retrieved from <http://pareonline.net/>
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubrics development: Validity and reliability. *Practical Assessment, Research, and Evaluation*, 7(10). Retrieved from <http://pareonline.net/>
- Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Merrill.
- Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implemented, and improving assessment in higher education*. San Francisco, CA: Jossey-Bass.
- Petkov, D., & Petkova, O. (2006). Development of scoring rubrics for IS projects as an assessment tool. *Issues in Informing Science and Information Technology Education*, 3, 499-510. Retrieved from <http://proceedings.informingscience.org>

- Prus, J., & Johnson, R. (1994). A critical review of student assessment options. *New Directions for Community Colleges*, 88, 69-83.
- Rezaee, A. A., & Kermani, E. (2011). Essay rater's personality types and rater reliability. *International Journal of Language Studies*, 5(4), 109-122. Retrieved from [www.ijls.net](http://www.ijls.net)
- Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 2, 420-428. Retrieved from <http://psycenet.apa.org>
- Simon, M., & Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research & Evaluation*, 7(18). Retrieved from <http://pareonline.net/>
- Stanny, C. J., & Duer, J. D. (2012). Authentic assessment in psychology: Using rubrics and embedded assessments to improve student learning. In D. S. Dunn, S. C. Baker, C. M. Mehrotra, R. E. Landrum, & M. A. McCarthy (Eds.), *Assessing teaching and learning in psychology* (pp. 19-34). Belmont, CA: Wadsworth, Cengage.
- Stemler, S.E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating. *Practical Assessment, Research & Evaluation*, 9(4). Retrieved from <http://pareonline.net/>
- Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research and Evaluation*, 9(2). Retrieved from <http://pareonline.net/>
- Trochim, M.K., & Donnelly, J. (2006). *Research methods knowledge base*. New York, NY: Cornell University.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.
- Walvoord, B. (2004). *Assessment clear and simple: A practical guide for institutions, departments, and general education*. San Francisco, CA: Jossey-Bass.


## Appendix A

## UMUC Graduate School Student Learning Expectations (SLEs)

STUDENT LEARNING EXPECTATIONS (SLEs)	
Written Communication (COMM)	Produce writing that meets expectations for format, organization, content, purpose, and audience.
Information Literacy (INFO)	Demonstrate the ability to use libraries and other information resources to effectively locate, select, and evaluate needed information.
Critical Thinking (THIN)	Demonstrate the use of analytical skills and reflective processing of information.
Technology Fluency (TECH)	Demonstrate an understanding of information technology broad enough to apply technology productively to academic studies, work, and everyday life.
Content/Discipline-Specific Knowledge (KNOW)	Demonstrate knowledge and competencies specific to program or major area of study.

## Appendix B

## Rubric for 3-3-3 Assessment Model

 University of Maryland University College Graduate School Writing Rubric for Outcomes Assessment Spring 2012					
CRITERIA	EXEMPLARY 4	COMPETENT 3	MARGINAL 2	UNSATISFACTORY 0-1	SCORE
<b>Context/Purpose</b> Considers the audience, purpose, and the circumstances surrounding the writing assignment(s).	Shows <b>superior understanding</b> of context, audience, and purpose that is <b>extremely appropriate</b> for the assignment(s).	Shows <b>good understanding</b> of context, audience, and purpose that is <b>mostly appropriate</b> for the assignment(s).	Shows <b>fair understanding</b> of context, audience, and purpose that is <b>somewhat appropriate</b> for the assignment(s).	Shows <b>insufficient or poor understanding</b> of context, audience, or purpose of the assignment(s).	—
<b>Content/Ideas/Support</b> Articulates and supports a main idea(s) that is consistent with context and purpose.	<b>Highly original</b> main idea(s) is <b>clearly articulated</b> and <b>strongly supported</b> by <b>predominantly current and relevant</b> evidence that may be researched based. Main idea(s) is <b>exceedingly consistent</b> with context and purpose.	<b>Mostly original</b> main idea(s) is <b>generally well articulated</b> and <b>sufficiently supported</b> by <b>mainly current and relevant</b> evidence that may be researched based. Main idea(s) is <b>generally consistent</b> with context and purpose.	Main idea(s) is <b>vague</b> , and/or <b>inadequately supported</b> , and/or <b>inconsistent</b> with context and purpose.	Main idea(s) is <b>hardly or not evident</b> and/or <b>lacks support</b> and/or <b>scarcely relates</b> to context and purpose.	
<b>Organization</b> Uses logical sequencing including introduction, transitions between paragraphs, and summary/ conclusion to develop main idea(s) and content.	Uses <b>highly logical sequencing</b> including introduction, transitions between paragraphs, and summary/ conclusion to <b>fully develop</b> main idea(s) and content.	Uses <b>mostly logical sequencing</b> including introduction, transitions between paragraphs, and summary/ conclusion to <b>generally develop</b> main idea(s) and content.	Uses <b>partially logical sequencing</b> . <b>Makes inadequate use</b> of introduction, and/or transitions between paragraphs, and/or summary/ conclusion. Main idea(s) and content are <b>incompletely developed</b> .	Uses <b>little or no logical sequencing</b> . Lacks introduction, and/or transitions between paragraphs and/or summary/ conclusion. Main idea(s) and content <b>remain undeveloped</b> .	
<b>Sources</b> Incorporates use of and identifies sources and/or research, according to APA and/or instructor guidelines.	Demonstrates <b>superior judgment</b> in selection, incorporation, and identification of <b>entirely appropriate quality and quantity</b> of sources and/or research that <b>fully meet or exceed</b> established guidelines.	Demonstrates <b>good judgment</b> in selection, incorporation, and identification of <b>mainly appropriate quality and quantity</b> of sources and/or research that <b>mostly meet or exceed</b> established guidelines.	Demonstrates <b>limited judgment</b> in selection and/or incorporation and/or identification of sources and/or research. Quality and/or quantity, and/or appropriateness <b>partially meet</b> established guidelines.	Demonstrates <b>little or no judgment</b> in selection and/or incorporation and/or identification of sources and/or research. Quality and/or quantity and/or appropriateness <b>do not meet</b> established guidelines.	
<b>Word Usage/ Grammar/Spelling/ Punctuation</b> Uses wording, grammar, spelling and punctuation accurately and correctly.	Demonstrates <b>virtually error-free</b> grammar, spelling and punctuation.	Demonstrates <b>very few errors</b> in grammar, spelling and punctuation.	Demonstrates <b>numerous errors</b> in grammar, spelling and punctuation.	Demonstrates <b>unacceptable amount and/or type of errors</b> in grammar, spelling and punctuation.	


## Appendix C

### ComR Rubric for Phase I

 University of Maryland University College Graduate School of Management and Technology COMBINED Rubric for Outcomes Assessment for Spring 2012					
CRITERIA	EXEMPLARY 3.1-4.0	COMPETENT 2.1-3.0	MARGINAL 1.1-2.0	UNSATISFACTORY 0-1.0	Score
<b>Conceptualization</b> Identifies and describes nature of idea(s) or issue(s) in relation to research and assignment context.	Shows a superior ability to identify and describe basic and complex issues with exceptional depth and clarity within context for full understanding.	Shows a good ability to identify and describe basic and complex issues with sufficient depth and clarity within context. Omissions do not seriously impede understanding.	Shows fair ability to identify and describe basic and complex issues within context with some depth and clarity. Ambiguities and omissions impede understanding.	Shows insufficient or no ability to identify basic and complex issues. Lack of clarity or depth impedes understanding.	0.00
<b>Analysis</b> Considers pros/cons; compares/contrasts in logical examination of issue(s) and source data.	Analyzes information in a highly organized and logical manner. Is exceptionally consistent and accurate in identifying embedded hypotheses, biases, causalities, and conclusions.	Analyzes information in a mostly organized and logical manner. Is generally consistent and accurate in identifying embedded hypotheses, biases, causalities, and conclusions.	Analyzes information in a somewhat organized and logical manner. Is slightly inconsistent and/or inaccurate in identifying embedded hypotheses, biases, causalities, and conclusions.	Analyzes information in a disorganized and illogical manner. Is inconsistent and/or inaccurate in identifying embedded hypotheses, biases, causalities, and conclusions.	0.00
<b>Synthesis</b> Integrates key concepts from research and analyses in coherent manner to form a cohesive response.	Consistently incorporates analyses with other information to connect key concepts in a highly coherent way. Provides strong base for further application and perspective.	Usually incorporates analyses with other information to connect key concepts in a mostly coherent way. Provides adequate base for further application and perspective.	Occasionally incorporates analyses with other information to connect key concepts in a partially coherent way. Provides minimal base for further application and perspective.	Rarely or never incorporates analyses with other information to connect key concepts. Work is incoherent. Provides no base for further application and perspective.	0.00
<b>Conclusion</b> Integrates analysis and synthesis to formulate a new perspective or position that is appropriate to the conceptualization of the question or assignment.	Integrates prior criteria in a highly effective manner demonstrating an original, well-reasoned, and justifiable perspective(s).	Integrates prior criteria in a mostly effective manner demonstrating a generally original, well-reasoned, and justifiable perspective(s).	Integrates prior criteria in a partially effective manner demonstrating weakness in originality, reasoning, and justifiable perspective(s).	Integrates prior criteria in an ineffective manner. Lacks an original, well-reasoned, or justifiable perspective(s).	0.00
<b>Implications</b> Based upon the positions, perspectives or conclusions, determines practices or processes and/or the need for further study.	Suggests highly appropriate considerations or actions for practice, policy and future research.	Suggests mostly appropriate considerations or actions for practice, policy and future research.	Suggests somewhat appropriate considerations or actions for practice, policy and future research.	Suggests inappropriate or fails to make considerations or actions for practice, policy and future research.	0.00
<b>Evaluation</b> Identifies appropriate resources by critically assessing reputation and quality of information.	Thoroughly analyzes information sources for currency, relevance, accuracy, authority and objectivity.	Sufficiently analyzes information sources for currency, relevance, accuracy, authority and objectivity.	Partially analyzes information sources for currency, relevance, accuracy, authority and objectivity.	Insufficiently analyzes information sources for currency, relevance, accuracy, authority and objectivity.	0.00
<b>Incorporation</b> Use information to accomplish specific purpose.	Expertly synthesizes and presents information to fully achieve a specific purpose with clarity and depth.	Sufficiently synthesizes and presents information to fully achieve a specific purpose with some clarity and depth.	Partially synthesizes and presents information with little clarity or depth.	Inadequately synthesizes and presents information with little or no clarity or depth.	0.00
<b>Ethical Use</b> Understands and complies with institutional policies related to access and use of information, demonstrating academic integrity.	Fully demonstrates understanding of ethical and legal guidelines for published, confidential and proprietary information.	Mostly demonstrates understanding of ethical and legal guidelines for published, confidential and proprietary information.	Partially demonstrates understanding of ethical and legal guidelines for published, confidential and proprietary information.	Fails to demonstrate understanding of ethical and legal guidelines for published, confidential and proprietary information.	0.00
<b>Context/Purpose</b> Considers the audience and purpose of the assignment.	Shows superior understanding of context, audience, and purpose that is extremely appropriate for the assignment(s).	Shows good understanding of context, audience, and purpose that is mostly appropriate for the assignment(s).	Shows fair understanding of context, audience, and purpose that is somewhat appropriate for the assignment(s).	Shows insufficient or poor understanding of context, audience, or purpose of the assignment(s).	0.00
<b>Content/Ideas/Support</b> Articulates and supports a main idea(s) that is consistent with context and purpose.	Highly original main idea(s) is clearly articulated and strongly supported by predominantly current and relevant evidence that may be researched based. Main idea(s) is exceedingly consistent with context and purpose.	Mostly original main idea(s) is generally well articulated and sufficiently supported by mainly current and relevant evidence that may be researched based. Main idea(s) is generally consistent with context and purpose.	Main idea(s) is vague, and/or inadequately supported, and/or inconsistent with context and purpose.	Main idea(s) is hardly or not evident and/or lacks support and/or scarcely relates to context and purpose.	0.00
<b>Organization</b> Uses logical sequencing as required of the assignment to develop main ideas and content.	Uses highly logical sequencing including introduction, transitions between paragraphs, and summary/ conclusion to fully develop main idea(s) and content.	Uses mostly logical sequencing including introduction, transitions between paragraphs, and summary/ conclusion to generally develop main idea(s) and content.	Uses partially logical sequencing. Makes inadequate use of introduction, and/or transitions between paragraphs, and/or summary/ conclusion. Main idea(s) and content are incompletely developed.	Uses little or no logical sequencing. Lacks introduction, and/or transitions between paragraphs and/or summary/ conclusion. Main idea(s) and content remain undeveloped.	0.00
<b>Grammar/Spelling/Punctuation</b> Uses wording, grammar, spelling and punctuation accurately and correctly.	Demonstrates virtually error-free grammar, spelling and punctuation.	Demonstrates very few errors in grammar, spelling and punctuation.	Demonstrates numerous errors in grammar, spelling and punctuation.	Demonstrates unacceptable amount and/or type of errors in grammar, spelling and punctuation.	0.00
<b>Technology Management</b> Creates accurate electronic document with appropriate layout, formatting, and accuracy.	Shows exceptional skills in creating accurate electronic documents appropriate for the assignment or context.	Shows good skills in creating accurate electronic documents appropriate for the assignment or context.	Shows fair skills in creating accurate electronic documents appropriate for the assignment or context.	Shows minimal or no skills in creating accurate electronic documents appropriate for the assignment or context.	0.00
<b>Information Retrieval</b> Utilizes technology to research, evaluate, inform, and communicate information retrieved from appropriate resources.	Uses technology extremely effectively to research, evaluate, inform, and communicate information from very appropriate resources.	Uses technology very effectively to research, evaluate, inform, and communicate information from mostly appropriate resources.	Uses technology somewhat effectively to research, evaluate, inform, and communicate information from appropriate resources.	Uses technology ineffectively or not at all to research, evaluate, inform, and communicate information from often inappropriate resources.	0.00

## Appendix D

## ComR Rubric Phase II

 <b>University of Maryland University College</b> COMBINED Rubric for Outcomes Assessment for Spring 2012, The Graduate School					
CRITERIA	EXEMPLARY 3.1-4.0	COMPETENT 2.1-3.0	MARGINAL 1.1-2.0	UNSATISFACTORY 0-1.0	Score
<b>Conceptualization/Content/Ideas</b> Identifies and articulates the main idea(s) or issue(s) in a way that is appropriate for the audience, research, context, and purpose of the assignment.	Identifies and articulates the main ideas/issues as appropriate with <b>exceptional</b> depth and clarity for full understanding with no ambiguities.	Identifies and articulates the main ideas/issues as appropriate with <b>sufficient</b> depth and clarity. Ambiguities and omissions do not seriously impede understanding.	Identifies and articulates the main ideas/issues within context with <b>some</b> depth and clarity. Ambiguities and omissions impede understanding.	<b>Insufficiently</b> identifies and articulates the main ideas/issues. Lack of clarity or depth impedes understanding.	0.00
<b>Analysis/Evaluation</b> Determines essential components and characteristics of the idea(s) or issue(s) while considering connections and significance.	Examines information in a <b>highly</b> logical and accurate manner and <b>extensively</b> exposes relationships, causalities, and importance of the ideas/issues.	Examines information in a <b>mostly</b> logical and accurate manner and <b>sufficiently</b> exposes relationships, causalities, and importance of the ideas/issues.	Examines information in a <b>somewhat</b> logical and accurate manner and <b>insufficiently</b> exposes relationships, causalities, and importance of the ideas/issues.	Examines information in an <b>illogical and inaccurate</b> manner and <b>fails</b> to expose relationships, causalities, and importance of the ideas/issues.	0.00
<b>Synthesis /Support</b> Integrates key concepts from research and analyses in a coherent manner to form a cohesive response.	<b>Consistently</b> incorporates analyses with other information/research to connect key concepts in a <b>highly</b> coherent way.	<b>Usually</b> incorporates analyses with other information/research to connect key concepts in a <b>mostly</b> coherent way.	<b>Occasionally</b> incorporates analyses with other information/research to connect key concepts in a <b>partially</b> coherent way.	<b>Rarely or never</b> incorporates analyses with other information/research to connect key concepts. Work is <b>incoherent</b> .	0.00
<b>Conclusion/Implications</b> Formulates a new perspective or position based upon consequences for practice, policy and/or the need for future study.	Forms a conclusion in a <b>highly</b> effective manner demonstrating an original, well-reasoned, and justifiable perspective(s) that <b>extensively</b> considers potential implications.	Forms a conclusion in a <b>mostly</b> effective manner demonstrating a generally original, well-reasoned, and justifiable perspective(s) that <b>sufficiently</b> considers potential implications.	Forms a conclusion in a <b>partially</b> effective manner demonstrating weakness in originality, reasoning, and justifiable perspective(s) that <b>insufficiently</b> considers potential implications.	Forms a conclusion in an <b>ineffective</b> manner. Lacks an original, well-reasoned, or justifiable perspective(s) with <b>no</b> consideration of potential implications.	0.00
<b>Selection/Retrieval</b> Chooses appropriate resources identified through online searches and critically assesses the quality of the information according to the criteria in the assignment.	Displays <b>thorough</b> evidence that information sources were chosen and assessed according to assignment expectations.	Displays <b>mostly complete</b> evidence that information sources were chosen and assessed according to assignment expectations.	Displays <b>incomplete</b> evidence that information sources were chosen and assessed according to assignment expectations.	Displays <b>very little</b> or <b>no</b> evidence that information sources were chosen and assessed according to assignment expectations.	0.00
<b>Organization</b> Uses logical sequencing as required of the assignment to develop the main ideas and content.	Uses <b>highly</b> logical sequencing including introduction, transitions between paragraphs, and summary/conclusion to <b>fully</b> develop the main idea(s) and content.	Uses <b>mostly</b> logical sequencing including introduction, transitions between paragraphs, and summary/conclusion to <b>generally</b> develop the main idea(s) and content.	Uses <b>partially</b> logical sequencing. Makes <b>inadequate</b> use of introduction, and/or transitions between paragraphs, and/or summary/conclusion. Main idea(s) and content are <b>incompletely</b> developed.	Uses <b>little</b> or <b>no</b> logical sequencing. Lacks introduction, and/or transitions between paragraphs and/or summary/ conclusion. Main idea(s) and content remain <b>undeveloped</b> .	0.00
<b>Writing Mechanics</b> Uses wording, grammar, spelling and punctuation accurately and correctly.	Contains <b>virtually no</b> errors in grammar, spelling and punctuation; any errors in writing mechanics and word usage do not interfere with reading or message.	Demonstrates <b>some</b> errors in grammar, spelling, punctuation and/or word usage that <b>somewhat</b> interfere with reading or message.	Demonstrates <b>numerous</b> errors in grammar, spelling, punctuation and/or word usage. These errors distract from the reading and weaken the message.	Demonstrates <b>excessive</b> errors in grammar, spelling, punctuation and word usage. These errors display an inability to communicate the message.	0.00
<b>APA Compliance</b> Follows APA style that includes headings, citations and a reference page.	Employs <b>very accurate</b> APA style.	Employs <b>mostly accurate</b> APA style.	Employs <b>mostly inaccurate</b> APA style.	Employs <b>little</b> or <b>no</b> APA style.	0.00
<b>Technology Application</b> Creates accurate electronic document according to specifications of the assignment.	Creates an electronic document that <b>complies</b> with <b>all</b> of the assignment specifications.	Creates an electronic document that <b>mostly</b> complies with the assignment specifications.	Creates an electronic document that <b>partially</b> complies with the assignment specifications.	Creates an electronic document that <b>minimally</b> complies or shows <b>no</b> evidence of compliance with the assignment specifications.	0.00