

Analysis and validation of a rubric to assess oral presentation skills in university contexts

Rafael García-Ros¹

¹ Department of Developmental & Educational Psychology,
University of Valencia

Spain

Correspondence: Rafael García-Ros. Departamento Psicología Evolutiva y de la Educación. Avda. Blasco Ibáñez, 21. 46010-Valencia. Spain. E-mail: Rafael.Garcia@uv.es

© Education & Psychology I+D+i and Editorial EOS (Spain)

Abstract

Introduction. The main objective of this study was to analyze users' perceptions and convergent validity of peer- and teacher summative assessment using a rubric for students' oral presentation skills in a university context.

Method. Peer- and teacher-assessment convergence was analyzed from an analytical and holistic perspective. Students' perceptions of validity and usefulness were determined from a questionnaire developed ad-hoc for this study.

Results. Students perceive that the rubric is useful for explaining and clarifying assessment criteria, planning the development of projects, and evaluating their results. They also highlight its validity, integrating the key criteria to consider in the development and presentation of their projects. The measures of agreement between peer- and teacher assessment using the rubric were significant, from both an analytical and a holistic perspective -correlation of .89-.

Conclusion. Results show the perceived usefulness and validity of the rubric to promote and support high-level cognitive processes in the development of projects. The rubric is a valid tool for the assessment and rating of student presentations by peers. These findings are discussed in terms of their instructional implications.

Keywords: Rubrics, higher education, reliability and validity, user perceptions, peer assessment, convergent validity, self-regulated learning.

Received: 10/10/11

Initial acceptance: 10/10/11

Final acceptance: 04/11/11

Análisis y validación de una rúbrica para evaluar habilidades de presentación oral en contextos universitarios

Resumen

Introducción. Este trabajo analiza la validez convergente entre la evaluación sumativa de las habilidades de comunicación oral con apoyo visual de estudiantes universitarios efectuada por el profesorado y por los pares a través de la aplicación de una rúbrica.

Método. Se analiza el nivel de convergencia entre las valoraciones efectuadas por el profesorado y por los pares desde una perspectiva analítica –criterio a criterio- y holística –puntuación global-. La percepción de validez y utilidad se determina a partir de las valoraciones efectuadas por los estudiantes sobre ambos aspectos en relación a la rúbrica desarrollada.

Resultados. Los estudiantes señalan que la rúbrica resulta útil para explicitar y clarificar los criterios de valoración, planificar el desarrollo de su trabajo y evaluar los productos resultantes. También destacan su validez para evaluar este tipo de proyectos, integrando los criterios clave a considerar en su desarrollo. El grado de acuerdo en su aplicación entre profesores y pares es significativo, tanto en desde una perspectiva analítica como holística, especialmente entre las valoraciones globales de los proyectos de trabajo –correlación de .89-.

Conclusión. Se destaca la percepción de utilidad y validez de la rúbrica para promover y dar soporte a procesos cognitivos de alto nivel en el desarrollo de este tipo de proyectos de trabajo, así como para su evaluación y para el desarrollo de competencias importantes en el ejercicio profesional. La rúbrica constituye una herramienta válida para valorar y calificar las presentaciones de los estudiantes a través de los pares. Estas conclusiones se discuten en términos de sus implicaciones y repercusiones instruccionales.

Palabras Clave: Rúbricas, enseñanza universitaria, fiabilidad y validez, percepción de los usuarios, evaluación a través de los pares, validez convergente, aprendizaje autorregulado.

Recibido: 10/10/11

Aceptación inicial: 10/10/11

Aceptación final: 04/11/11

Introduction

In the past few years, there has been a considerable increase in interest in learning-centered teaching models in our university context. To a large extent, they have been promoted by the European Higher Education Area, which emphasizes the importance of integrating the teaching, learning and assessment processes in carrying out a great variety of open, authentic and realistic tasks –e.g., project work, doing research, case analysis, etc. These types of tasks facilitate the acquisition of the general and specific competences considered in the different degree programs.

From this perspective, it is necessary to consider alternative, unconventional or innovative assessment techniques which make it possible to (a) use assessment as one more instructional resource to promote active learning, the activation of high-level cognitive processes and the acquisition of competences, (b) provide students with performance criteria to be achieved when performing the tasks, favoring the development of skills of realistic self-regulation and self-assessment of their work, (c) receive specific feedback about how to improve their performance levels, and (d) provide instructors with information about the learning results reached by their students. Traditional assessment strategies are clearly insufficient for achieving these aims, given that (a) the results of traditional exams can help students to monitor their learning to a certain extent, but they do little to promote it, and (b) proposing authentic and realistic tasks involves considering a wide range of possible satisfactory responses and activities that would be difficult to reflect in traditional exams (Huba & Freed, 2000).

Rubrics

Rubrics are non-conventional assessment tools that can be defined as guides for evaluating the quality of work and performance level achieved by students on a wide variety of complex tasks, specifying the criteria to be considered and levels of quality for each of them (from insufficient to excellent) (Andrade & Du, 2005). Currently, rubrics are a widespread assessment method in compulsory education, especially in the Anglo-Saxon setting – see the large number of web resources available in Dornish and Sabatini (2006)-, although their use is still quite limited in our context.

Perhaps one of the issues that best explains university teachers' resistance to introducing this type of assessment tool is the controversy over its reliability and validity for judging students' work –according to some authors, along with scant teacher training or training in assessment techniques–, a question that should continue to be the focus of research (Malini & Andrade, 2010). From this perspective, it is not surprising that numerous previous studies have focused on (a) analyzing the intra-rater and inter-rater reliability in its application to assessing students' work (e.g., Thaler, Kazemi & Huscher, 2009), (b) the consistency between the ratings made by the students when assessing their classmates' or their own work (e.g., Roblyer & Wiencke, 2003; Sadler & Good, 2006), and (c) the consistency between the criteria used by the students and teachers in applying rubrics (e.g., Hafner & Hafner, 2003; Kocaküla, 2010; López-Pastor, Fernández-Balboa, Santos & Fraile, 2011; Stellmack, Konheim-Kalkstein, Manor, Massey & Schmitz, 2009). Their basic conclusions are that rubrics make it possible to increase the consistency between inter-rater and intra-rater valuations, helping teachers to make valid judgments about students' performance level when assessing complex skills (Jonsson & Svingby, 2007).

The interest in the study of rubrics has also focused on analyzing whether their use promotes learning or improves instruction (e.g., Andrade & Du, 2005; Andrade, Du & Micek, 2010; Kocaküla, 2010; Popham, 1997), and whether they facilitate self- and peer-assessment of students' work (e.g., Andrade, 2001; Baron & Keller, 2003; Cho, Schunn & Wilson, 2006; Magin & Helmore, 2001; Schafer, Swanson, Bené & Newberry, 2001). The basic conclusions of these studies are that rubrics can promote a higher quality of learning by focusing attention on the assessment criteria in elaborating the projects, and that they help students to perform a much more precise self-assessment of their own work. They also allow teachers to make more valid judgments about the level of acquisition of complex skills, and provide more specific feedback about students' performance (Jonsson & Svingby, 2007). However, various studies have also shown that many students may not consider all of the criteria included (e.g., Andrade & Du, 2005), that they may be reluctant to use rubrics to evaluate their classmates' work (Norcini, 2003), or that it is important to involve the students in their development to facilitate their comprehension and application (Huba & Freed, 2000; Stix, 1997; Taggart, Phifer, Nixon & Wood, 2001).

Along the same lines, the research has also focused on analyzing students' perceptions (and those of the teachers) of the instructional benefits and utility of rubrics – characteristics, appropriateness, validity, usefulness...-, given their determinant role in the way learning and study are approached (e.g., García-Ros & Pérez-González, 2011; Sander, 2005; Struyven, Dochy & Janssens, 2005). The conclusions highlight that: (a) both students and teachers indicate their usefulness for explaining and clarifying the evaluation criteria – *transparence*-; (b) the teachers indicate that rubrics promote the development of reflexive practices and make it possible to obtain more information about their effectiveness, while helping to provide higher quality feedback and serving as support for students' self-assessment of their work; and (c) rubrics cause students to be more involved in task development (Jonsson & Svingby, 2007; Schamberr & Mahoney, 2006). However, different studies also show that students can perceive rubrics more as a tool to satisfy teachers' demands than as a representation of the criteria and standards of quality to be considered in their work (Andrade & Du, 2005), or that they can manifest doubts about their usefulness for self-assessing their work and better interpreting the feedback received (Baron & Keller, 2003). Thus, the instructional efficacy of rubrics can be seriously affected if, for example, students think that they do not contemplate the key criteria for carrying out the task, or that they are not useful for improving the result of their work, or that they do not allow an adequate grading of the quality of their work.

Description of the rubric developed

Following the principles highlighted in the previous research on how to increase the usefulness, appropriateness and instructional efficacy of rubrics (e.g., Andrade, 2001, 2005; Jonassen, Peck & Wilson, 1999; Tierney & Marielle, 2004), the following elements are specified in the rubric developed in this study, which was designed to evaluate oral presentation skills with visual aids (see Appendix 1):

- *Evaluation criteria.* The rubric incorporates fourteen different criteria for evaluating students' work (e.g., “Visual aids: relevance and appropriateness of the images, graphs and outlines used”).
- *Performance levels.* The rubric considers four performance levels graduated according to their suitability for each criterion (unsuitable, basic/to be improved, sufficient and quite appropriate).

- *Description of performance levels.* The performance levels are clearly described and reflect the differences between them.
- *Grading strategy.* The rubric incorporates a quantitative rating for each criterion and performance level, making it possible to perform an analytic (criterion by criterion, assigning to each a score between 0 and 3) and holistic (sum of all the scores obtained on all the criteria, with a range between 0 and 42) rating of the presentations made by the students.

Objectives

Based on the preceding proposals, the main objective of this study is to analyze and validate a rubric designed to assess skills of oral presentation with visual aids in university contexts. In this way, an analysis is carried out of the consistency between the teachers' and students' ratings of their classmates' work, as well as the relationship between the grades derived from them. In addition, an analysis is also made of the students' perceptions of the utility and validity of the rubric as a support for developing their oral presentations with visual aids.

Method

Participants

The participants in the study were 64 students in the Educational Psychology degree program in the 2008-2009 academic year. Of them, 83% are females and 17% males, with an age range from 22 to 49 years (mean of 29.5 years and s.d. of 5.2). Only 17% are full-time students, while 83% are also involved in some type of professional activity. The distribution by sex, age and time dedication is characteristic of this type of university studies at the University of Valencia.

Procedure

After describing to the students the project to be developed in groups (elaboration and oral presentation with visual aids of the principles and application in the classroom of an instructional model), the criteria described in the rubric were presented and discussed. In order to familiarize the students with the rubric's application, facilitate its comprehension and elim-

inate any reluctance, it was first used and discussed in assessing two previous presentations made by the teacher. Next, fourteen work groups developed and performed their presentations, which were assessed by means of the rubric by the teacher and classmates. Finally, the questionnaires on the use and perception of validity of the rubric were filled out individually and anonymously after all the presentations had been made.

Measures

A questionnaire was used that had been designed to evaluate the perception of usefulness and validity of the rubric developed in this study. Along with its application in assessing the students' presentations, the rubric served to obtain the subjects' scores on the different variables considered in the study.

Perception of usefulness and validity. Following the principles and orientations of prior research (e.g., Moskal, 2003; Moskal & Leydens, 2000), a questionnaire was developed that was designed to evaluate the perception of the rubric's usefulness in developing the work project (10 items) and the perception of its validity (11 items). A five-point Likert-type response scale was used, from "completely inappropriate" (1) to "completely appropriate" (5)-.

Rating the presentations. Professor and students evaluated the quality of the presentations using the rubric, both from an analytical and a holistic point of view. The analytical rating makes it possible to determine the level of agreement, criterion by criterion, between teacher and peers. The holistic score makes it possible to determine the existence of a significant relationship between the global ratings made by both.

Data analysis

The perception of the usefulness and validity of the rubric is determined by the basic descriptors of the students' responses to the different items on the questionnaire designed for this purpose.

The degree of agreement in the application of the rubric between professor and students is calculated with two different estimation procedures, one conservative and the other lax, both accepted and commonly used in this area (Tinsley & Weiss, 2000). Using the conservative procedure, there is said to be agreement between professor and students when they give exactly the same rating (level) on the criteria on the rubric. The lax procedure considers

that there is agreement between professor and students when they give ratings on the criteria that differ from each other by one level at the most (adjacent agreement). In both cases, the degree of agreement or convergence is determined by the *kappa* statistic. The level of association between the global ratings made by the professor and by the peers (sum of the scores on the different criteria, with a maximum of 42 points) is also estimated using the *Pearson correlation coefficient*.

Results

Perception of usefulness and validity of the rubric

Table 1 shows the basic descriptors of the responses to the questionnaire about the perception of the usefulness of the rubric. The ratings are higher than the theoretical mean of the response scale (3.0) on all items, with the exception of “reduce my anxiety in doing the work”. The students especially highlight its usefulness for “having greater knowledge about the grading criteria” (mean of 4.4), “evaluating the appropriateness of the work presented” (mean of 4.1), “developing appropriate expectations about what is required” (mean of 4.0) and “planning the elaboration of the project and the presentation” (mean of 4.0).

Table 1. Basic descriptors about perception of usefulness of the rubric

<i>It was useful to me for ...</i>	Mean	s.d.
1.- Developing appropriate expectations about what is required of me	4.0	(0.80)
2.- Planning the elaboration of the work and its presentation	4.0	(0.83)
3.- Revising what I was doing to adjust it to the established criteria	3.7	(0.74)
4.- Evaluating the suitability of the work presented	4.1	(0.75)
5.- Guiding the development of the work	3.8	(0.79)
6.- Making decisions within the group about how to carry out the project	3.8	(0.74)

<i>I think it has allowed me to ...</i>	Mean	s.d.
1.- Improve the “product” or final result of my work	3.9	(0.89)
2.- Facilitate the development of the project	3.5	(0.79)
3.- Reduce my “anxiety” about doing the work	2.7	(0.98)
4.- Know the evaluation criteria better	4.4	(0.61)

As table 2 shows, the students rate the validity of the rubric as adequate, especially with regard to “integrating the key elements in the development of the task” (mean of 3.9), “developing criteria that I will apply in future situations” (mean of 3.8) and “makes it possible to evaluate important competences for the educational psychologist” (mean of 3.8).

Table 2. Basic descriptors about the perception of the rubric’s validity

<i>I think the rubric...</i>	Mean	s.d.
1.- Integrates the key elements in the development of the task performed	3.9	(0.81)
2.- Makes it possible to evaluate important competences for the educational psychologist	3.8	(0.72)
3.- Allowed me to develop criteria that I will apply in future situations	3.8	(0.67)
4.- Is a reliable tool (adequately measures the quality of the work)	3.6	(0.83)
5.- Clearly highlights the levels considered in each criterion	3.6	(0.82)
6.- Incorporates very specific criteria that are not very useful in other situations	2.6	(0.92)
7.- Facilitates a fair comparison of the groups’ projects	3.5	(0.96)
8.- Presents biases or stereotypes (gender, age, personal situation,...)	2.3	(0.70)
9.- Has helped me to acquire the criteria for a correct performance	3.6	(0.78)
10.- Incorporates too many criteria	3.2	(0.84)
11.- Integrates criteria that are too generic	2.6	(0.93)

Consistency between the teacher's and students' ratings

The consensus estimates, the percentage of exact agreement between teacher and peers, is 66% ($kappa = .36, p < .001$), while the level of adjacent agreement is 98% ($kappa = .80, p < .001$). The analysis of the differences in the global rating awarded to the presentations by the teacher and peers (sum of the scores on the different criteria on a scale from 0-42) indicates that 70% of them differ by four points or less, and that the greatest discrepancy is 7 points.

After comparing the teacher and peer ratings on each of the blocks of criteria and transforming the results into a scale from 0 to 10, a significant relationship was observed ($r = .89, p < .001$) between the scores derived from the teacher's ratings (range 5.6-9.2, mean of 7.3 and standard deviation of 1.1) and those of the classmates (range 6.2-9.3, mean of 7.8, standard deviation of 1.0). The scores associated with the peers' ratings were generally higher (average of 0.5 points) than those of the teacher, with a difference of more than one point in four groups, and the maximum difference being 1.4 points.

Discussion and conclusions

The first objective of the study was to analyze the perception of utility and validity of the rubric developed in this study. The students highlight that it was of use to them for planning, developing and evaluating their own projects, especially to clarify and have greater knowledge about the criteria for grading their work, develop appropriate expectations about the level of performance to achieve, plan the development of their presentations, and assess the appropriateness of the final products. In sum, the students highlight the usefulness of the rubric for promoting and providing support for high-level cognitive processes, favoring the academic self-regulation processes in the development of their work projects (Zimmerman & Schunk, 2001). Furthermore, the students perceive the rubric as a valid and reliable tool for assessing this type of projects, pointing out that it integrates criteria and key elements to consider in their elaboration, making it possible to assess important competences for professional practice and allowing them to develop skills that they will use in future situations.

These conclusions basically coincide with those from previous studies focused on other competences and university degrees (e.g., Cothran, 2003; Hafner & Hafner, 2003; Rohrbach, 2008; Struyven, Dochy & Janssen, 2008). However, based on the ratings obtained, future studies could aim to (a) reduce the number of criteria incorporated, coinciding with the principle that having fewer criteria is more practical for developing the learning activities (e.g., Thaler et al., 2009), and (b) analyze the moderate perception of the rubric's usefulness in reducing the anxiety produced by elaborating the work projects – perhaps provoked by the fact that oral presentations continue to be one of the tasks on which university students show higher levels of anxiety (García-Ros & Pérez-González, 2011b)-, which is highlighted in previous studies (e.g., Andrade & Du, 2005) .

The second objective was directed toward analyzing the convergent validity between the application of the rubric by the teacher and the peers in rating the presentations of their classmates, as well as the grade obtained based on these ratings. This question is especially relevant given that it implies considering the rubric's suitability for supporting peer-assessment, saving time for teachers in correcting the projects, and having an indicator of the degree to which the students have interiorized its criteria. On the other hand, skills for judging peers' work critically and objectively are basic competences in any professional field (Magin & Helmore, 2001) that should be developed in university degree programs.

From an analytical perspective (analysis criterion by criterion), the degree of agreement between teacher and peers is significant, showing a degree of convergence even higher than in the few previous studies designed to determine the usefulness of rubrics as an aid in peer evaluation of students' work (e.g., Sadler & Good, 2006; Stellmack et al., 2009). More specifically, Jonsson and Svingby (2007) point out that the values of the exact level of agreement between raters fall in the range of 55-75% -in this study, 66%-, that the level of agreement between adjacent levels is 90% -in this study, 98%- and that the studies that use the kappa index present values of .45-.75 –this study obtains significant values of .36 for the exact criterion and .80 for adjacent levels-.

From a holistic perspective, the relationship between the teacher's and peers' global ratings of the presentations also reached a high-magnitude relationship. More specifically, the

majority of the studies that analyze this question obtain values between .55 and .75 -in their meta-analytic study, Falchicov and Goldfinch (2000) show an average value of .69-, while in the present study it reached a value of .89. However, the score derived from the peer ratings was slightly higher than that of the teacher (average difference of 0.5 points).

These results support the idea that the students apply the rubric in a similar way to that of the teacher, and that it can be a valid instrument for peer rating and scoring of student presentations, also facilitating the comprehension of the quality criteria involved in the work projects (Lu & Law, 2011). However, the results also highlight the students' tendency to overrate their classmates' work compared to the teacher's ratings, an issue also found in one previous study related to oral presentation skills (Magin & Helmore, 2001), but not in others (e.g., Freeman, 1995; Hugues & Large, 1993). Involving the students in the development of the rubrics, and providing additional information and feedback about their application to the students who show the greatest divergence with the teacher's application, can facilitate a better comprehension of the criteria considered and improve the exactitude of their ratings (Cho et al., 2006; Kocaküla, 2010; Lindblom-Yläne, Pihlajamäki & Kotkas, 2006). All of these aspects highlight the importance of continuing to focus the research on identifying measures that can increase the validity and reliability of rubrics for teacher-, self- and peer-assessment purposes (Malini & Andrade, 2010).

References

- Andrade, H. (2001). Using rubrics to promote thinking and learning. *Educational Leadership*, 57 (5), 13-18.
- Andrade, H. (2005). Teaching with rubrics: The Good, the Bad, and the Ugly. *College teaching*, 53 (1), 27-31.
- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10 (3). Retrieved May 2009. Available from:
<http://pareonline.net/getvn.asp?v=10&n=3>.
- Andrade, H.; Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice*, 17 (2), 199-214.
- Baron, J., & Keller, M. (2003). *Use of rubrics in online assessment*. Evaluation and Assessment Conference, 23-25 November 2003.
- Cho, K.; Schunn, Ch. D., & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891-901.
- Cothran, D. J. (2003). *Students's use of and perspective on rubrics*. Educational Research, Risks and Dilemmas. NZARE/AARE Conference. Auckland, NZ. Retrieved October 2009. Available from: <http://www.aare.edu.au/03pap/cot03119.pdf>.
- Dornisch, M., & Sabatini, A. (2006). Limitations of web-based rubric resources: Addressing the challenges. *Practical Assessment Research & Evaluation*, 11(3). Retrieved May 2010. Available from: <http://pareonline.net/getvn.asp?v=11&n=3>.
- Falchicov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70 (3), 287-322.
- Freeman, M. (1995). Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education*, 20 (3), 289-299.
- García-Ros, R., & Pérez-González (2011a). Assessment preferences of preservice teachers: analysis according to academic level and relationship with learning styles and motivational orientation. *Teaching in Higher Education*, 16 (6), 719-731.

- García-Ros, R., & Pérez-González, F. (2011b). Predictive and incremental validity or self-regulation skills on academic success in the university. *Journal of Psychodidactics*, 16 (2), 231-250.
- Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509-1528.
- Huba, M.E., & Freed, J.E. (2000). *Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning*. Needham Heights, Allyn & Bacon.
- Hugues, I., & Large, B. (1993). Staff and peer-group assessment of oral communication skills. *Higher Education*, 18 (3), 379-385.
- Jonassen, D., Peck, K., & Wilson, B. (1999). *Learning with technology*. Upper Saddle River, NJ: Merrill/Prentice Hall.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2 (2), 130-144.
- Kocaküla, M. S. (2010). Development and application of a Rubric for Evaluating Students' Performance on Newton's Laws of Motion. *Journal of Science Education Technology*, 19 (2), 146-164.
- Lindblom-Yläne, S.; Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer and teacher-assessment of student essays. *Active Learning in Higher Education*, 7 (1), 51-62.
- López-Pastor, V. M.; Fernández-Balboa, J. M.; Santos, M. L., & Fraile, A. (2011). Students' self-grading, professor's grading and negotiated final grading at three university programmes: analysis of reliability and grade difference ranges and tendencies. *Assessment & Evaluation in Higher Education*, published online ahead of print publication, doi: 10.1080/02602938.2010.545868.
- Lu, J., & Law, N. (2011). Online peer assessment: effects of cognitive and affective feedback. *Instructional Science*, published online ahead of print publication, doi: 10.1007/s11251-011-9177-2.
- Magin, D., & Helmore, Ph. (2001). Peer and teacher assessment of oral presentation skills: How reliable are they? *Studies in Higher Education*, 26 (3), 287-298.
- Malini, Y., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35 (4), 435-448.

- Moskal, B. (2000). Scoring rubrics: what, when and how? *Practical Assessment, Research & Evaluation*, 7(3). Retrieved May 2009. Available from: <http://pareonline.net/getvn.asp?v=7&n=3>.
- Moskal, B. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research & Evaluation*, 8 (14). Retrieved May 2009. Available from: <http://pareonline.net/getvn.asp?v=8&n=14>.
- Moskal, B., & Leydens, J. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10). Retrieved May 2009. Available from: <http://pareonline.net/getvn.asp?v=7&n=10>.
- Norcini, J. J. (2003). Peer assessment of competence. *Medical Education*, 37(6), 539-543.
- Popham, J. W. (1997). What's wrong-and what's right-with rubrics. *Educational Leadership*, 55(2), 72-75.
- Roblyer, M. D., & Wiencke, W.R. (2003). Design and use of a rubric to assess and encourage interactive qualities in distance courses. *The American journal of distance education*, 17 (2), 77-98.
- Rohrbach, S. (2008). *Educational assessment in emerging areas of design*. Design Research Society Biennial Conference 2008, Sheffield, UK.
- Sadler, P. M., & Good, E. (2006). The Impact of Self- and Peer-Grading on Student Learning. *Educational Assessment*, 11(1), 1-31.
- Sander, P. (2005). Researching our students for more effective university teaching. *Electronic Journal of Research in Educational Psychology*, 3 (1), 113-130.
- Schafer, W. D.; Swanson, G.; Bené, N., & Newberry, G. (2001). Effects of teacher knowledge of rubrics on student achievement in four content areas. *Applied Measurement in Education*, 14 (2), 151-170.
- Schamberr, J. F., & Mahoney, S. L. (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects collaborative learning groups. *Journal of General Education*, 55 (2), 103-137.
- Stellmack, M. A.; Konheim-Kalskstein, Y.; Manor, J.; Massey, A. R., & Schmitz, J. A. (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. *Teaching of Psychology*, 36 (2), 102-107.
- Stix, A. (1997). *Creating rubrics through negotiable contracting and assessment*. US Department of Education. Retrieved May 2009. Available from: <http://www.interactiveclassroom.com/006%20-%20Creating%20Rubrics.pdf>.

- Struyven, K.; Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, 30 (4), 325-341.
- Struyven, K.; Dochy, F., & Janssens, S. (2008). The effects of hands-on experience on students' preferences for assessment methods. *Journal of Teacher Education*, 59 (1), 69-88.
- Taggart, G.; Phifer, S.; Nixon, J., & Wood, L. (2001). *Rubrics: A handbook for construction and use*. Lanham, MD: Scarecrow Press.
- Thaler, N.; Kazemi, E., & Huscher, C. (2009). Developing a rubric to assess student learning outcomes using a class assignment. *Teaching of Psychology*, 36 (2), 113-116.
- Tierney, R., & Marielle S. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, 9 (2). Retrieved May 2009. Available from: <http://pareonline.net/getvn.asp?v=9&n=2>
- Tinsley, H.E., & Weiss, D. J. (2000). Interrater reliability and agreement. In H.E. Tinsley and S.D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling*, 95-124. San Diego: Academic Press.
- Zimmerman, B. J., & Schunk, D. H. (Eds.) (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd Ed.) Mahwah, NJ: Lawrence Erlbaum.

Appendix 1. Rubric for evaluating oral presentations with visual aids

1.- Basic principles of the instructional model (30%)

<i>CRITERION</i>	<i>Quite suitable (3)</i>	<i>Sufficient (2)</i>	<i>Basic/To be improved (1)</i>	<i>Insufficient (0)</i>
<i>Content of model: Proficiency and comprehension</i>	Shows a thorough and profound understanding of the topic	Shows a good understanding of the topic	Shows a good understanding of some parts of the topic, but not others	Does not master or understand the topic
<i>Content of model: Planning and organization of the content</i>	A well-planned talk, links the topics in a logical and coherent way	Follows a generally adequate outline, although sometimes gets lost or needs to pay more attention to details	Difficult talk to follow with incoherent connections, which indicates poor planning and lack of effort	Did not plan at all
<i>Content of model: Integration of important elements</i>	Effectively incorporates the important elements of the topic in the entire presentation	Effectively incorporates the important elements of the topic during most of the presentation	Very little integration of the basic elements of the topic	Insufficient integration

2.- Application and exemplification of the instructional model (30%)

<i>CRITERION</i>	<i>Quite suitable (3)</i>	<i>Sufficient (2)</i>	<i>Basic/To be improved (1)</i>	<i>Insufficient (0)</i>
<i>Practical application: Coherence with theoretical model</i>	The example used is coherent with the theoretical model	The example used is coherent with the theoretical model, although it would need to be adjusted in some specific aspects	The example developed is coherent with the model, although various aspects are not sufficiently included	The example used is not coherent with the model
<i>Practical application: Model clarification and audience motivation</i>	The example is clarifying and motivating for the audience	The example does not sufficiently clarify or does not motivate the audience	The example does not sufficiently clarify and does not motivate the audience	The example selected is confusing and does not fit the model

3.- Visual Aids (20%)

<i>CRITERION</i>	<i>Quite suitable (3)</i>	<i>Sufficient (2)</i>	<i>Basic/To be improved (1)</i>	<i>Insufficient (0)</i>
<i>Visual aids: Number and explanation of slides</i>	Sufficient number of slides and a relevant explanation of them	Sufficient number of slides, although the explanation of some of them is not precise	Uses an insufficient number of slides, but tries to explain them	Does not meet the requirements for number of slides and correctness of their explanation.
<i>Visual aids: Appropriateness of slides</i>	All of them are related to the topic and connected to the talk, aiding the comprehension of the material and being interesting/catching one's attention	All of them are related to the topic and aid the comprehension of the material, although some are not connected to the talk or are boring	Although the majority of them are related to the topic, they do not aid comprehension or motivation of the audience	Incorporates slides that are irrelevant to the topic, do not facilitate understanding and do not motivate the audience.
<i>Visual aid: Legibility</i>	No more than 5 words per line or 5 lines per slide, letters that are legible to the whole	Words legible to the entire audience, more than 5 lines per slide. Correct color con-	Words legible to approximately 85% of the audience, too much text on many	Illegible, over-use of color, abuse of text and paragraphs

	audience, correct color contrast	trast.	slides. Incorrect color contrast.	
<i>Visual aids: Relevance and appropriateness of images/graphs/outlines</i>	Relevant, agree with contents, and have correct resolution	Relevant, agree with contents most of the time, but low resolution	Agree with content, although not relevant and with low resolution	Does not use them or they are distracting
<i>Visual aids: Sequencing, animation/slide transitions</i>	Important animation points, help to focus attention on what is important and avoid distraction	Logical order, animation points not important, or transitions that interfere with attention	Logical order, although lacks animation and only applies transitions on some slides, or does both incorrectly	No order, excessive or no use of animation, and transition that produces distraction

4.- Communication skills in the presentation (20%)

<i>CRITERION</i>	<i>Quite suitable (3)</i>	<i>Sufficient (2)</i>	<i>Basic/To be improved (1)</i>	<i>Insufficient (0)</i>
<i>Verbal communication skills: Volume and tone of voice</i>	Loud enough and with a tone of voice that can be heard by the entire audience	Loud enough to be heard by the entire audience and adequate tone of voice about 90% of the time	Loud enough to be heard about 80 % of the time by everyone and/or inappropriate tone (monotone, boring)	Too low to be heard by everyone and excessively monotonous
<i>Verbal communication skills: Speaking clearly</i>	Speaks clearly almost all of the time.	Speaks clearly 90-95% of the time, although makes mistakes on the pronunciation of some words	Speaks clearly 85-90% of the time and pronounce many words incorrectly	Not very clear. Often mumbles and pronounces words incorrectly.
<i>Verbal communication skills: Pauses and use of "fillers"</i>	Uses pauses correctly and at the end of sentences. Uses complete sentences more than 95% of the time	Uses pauses correctly, although introducing some unexpectedly/incorrectly, uses fillers on various occasions (um..., uh..., well...etc). Uses incomplete sentences more than 5% of the time	Inappropriate number of pauses- too many or too few-, introduces some in unexpected moments or uses fillers frequently (um..., uh..., well...etc). Uses many incomplete sentences	Does not uses pauses in a thoughtful way. Uses too many fillers. Many incomplete sentences.
<i>Non-verbal behavior: Posture and eye contact</i>	Back straight, relaxed and confident posture, establishes eye contact with everyone	Back straight and establishes eye contact with everyone, although there are some indicators of tension/lack of interest	Sometimes maintains a straight posture and maintains eye contact with the audience at least 90% of the time	Does not maintain eye contact and body posture is incorrect

[This page intentionally left blank]