

## Reliability and validity test of a Scoring Rubric for Information Literacy

Jos van Helvoort<sup>a</sup>, Saskia Brand-Gruwel<sup>b</sup>, Frank Huysmans<sup>c</sup>, Ellen Sjoer<sup>a</sup>

<sup>a</sup>The Hague University of Applied Sciences, Sustainable Talent Development Research Group

<sup>b</sup>Open University of the Netherlands

<sup>c</sup>University of Amsterdam

### Abstract

**Purpose:** The main purpose of the research was to measure reliability and validity of the Scoring Rubric for Information Literacy (Van Helvoort, 2010).

**Design/methodology/approach:** Percentages of agreement and Intraclass Correlation were used to describe interrater reliability. For the determination of construct validity, factor analysis and reliability analysis were used. Criterion validity was calculated with Pearson correlations.

**Findings:** In the described case, the Scoring Rubric for Information Literacy appears to be a reliable and valid instrument for the assessment of information literate performance.

**Originality/value:** Reliability and validity are prerequisites to recommend a rubric for application. The results confirm that this Scoring Rubric for Information Literacy can be used in courses in higher education, not only for assessment purposes but also to foster learning.

**Keywords** Information literacy, Student performance measurement, Scoring rubrics, Reliability, Validity

**Paper type** Research paper

### Introduction

A scoring rubric is a grading tool that is often used for the rating of authentic student work. Jonsson and Svingby define it as “criteria for rating important dimensions of performance, as well as standards of attainment for those criteria” (2007). Angell (2015) and Carbery and Leahy (2015) remark that rubrics are also popular in library and information science literature and that they are often mentioned in the context of assessing student assignments.

In the context of the measurement of information literacy skills, rubrics have the benefit of supporting the assessment of the student’s real performance in resolving information problem-solving tasks, while other popular evaluation methods like multiple choice tests are more appropriate for the measurement of knowledge and understanding (Cameron et al., 2007). Rubrics follow the trend in higher education towards authentic assessment, as Knight remarked, “a process that measures how students apply their knowledge to real-time tasks” (Knight, 2006). They are supposed to combat subjectivity and unfairness during the grading process (Bresciani et al., 2009). Other benefits of scoring rubrics are their appropriateness for the supply of detailed feedback, the possibility to inform students about the expectations of their instructors, and the usefulness of rubrics for peer- and self-assessment (Oakleaf, 2008; Oakleaf, 2009; Reddy and Andrade, 2010; Belanger et al., 2015).

Keeping in mind the importance of information problem solving in today’s higher education (Brand-Gruwel et al., 2005), Van Helvoort developed a Scoring Rubric for Information Literacy, an instrument for the grading of the processes of information problem solving by students in higher education (Van Helvoort, 2010). Part time adult students at the department of Information Studies at The Hague University of Applied Sciences reported on the appreciation of the rubric in 2012 (Van Helvoort, 2012). This happened again in 2013 on the way academic staff of this university used the rubric for student performance measurement and instruction (Van Helvoort, 2013). When used by two graders, the rubric seemed to be a reliable and valid grading instrument (Van Helvoort, 2010 and 2016) but

**This article is © Emerald Group Publishing and permission has been granted for this version to appear here ([www.hbo-kennisbank.nl](http://www.hbo-kennisbank.nl)). Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.'**

this conclusion was based on two studies with rather small numbers of student participants (N=27 and N=19). To confirm this claim for reliability and validity, and to find a more robust basis for the Scoring Rubric for Information Literacy, a third study with a larger scaled test was launched in the department of Media Studies at the University of Amsterdam. This research was part of the PhD research by Jos van Helvoort (Van Helvoort, 2016).

The properties of the scoring rubric that are being investigated are the traditional requirements for assessment instruments: interrater reliability, construct validity and criterion validity. Interrater reliability (IRR) is regarded as the “level of agreement between a particular set of judges on a particular instrument at a particular time” (Stemler, 2004). It is an important property of a ‘fair’ grading instrument in the opinion of both graders and students, particularly when the scores are used for pass/fail decisions.

In relation to validity, the following values are distinguished: construct validity, criterion validity and content validity. Content validity refers to the question of whether all the intended content is referred to in the scoring instrument (Moskal and Leydens, 2000). In the Scoring Rubric for Information Literacy, this is ensured during the development process of the rubric by a number of review sessions with fellow teachers from different faculties. Content validity is not investigated anymore in the current research where the goal was to test the properties of the existing scoring rubric.

Construct validity refers to the question of whether all of the criterions of the grading instrument are relevant for the construct of interest (Moskal and Leydens, 2000). It is determined by a factor analysis that measures whether different criterions refer to one or more dimensions. Together with factor analysis, a reliability analysis is often executed which measures the internal consistency of such a group of related criterions. A high degree of internal consistency for the different criterions is therefore an indication of the reliability of the total instrument (Pinto and Sales, 2015). Lastly, criterion validity refers to the question of whether the scores with an assessment instrument correlate with the scores of another instrument that is supposed to measure the same construct (Cronbach and Meehl, 1955).

In the present paper we answer four research questions. Question 1 refers to the interrater reliability of the rubric, questions 2 and 3 both refer to the construct validity, and question 4 refers to the criterion validity of the scoring rubric.

1. What is the interrater reliability of the Scoring Rubric for Information Literacy when 80 student papers are graded by two different graders?
2. What is the homogeneity of the gradings when using the scoring rubric?
3. What is the internal consistency of the criterions when the gradings are done using the scoring rubric?
4. What is the correlation between the gradings that are done using the scoring rubric and those using the alternative instrument that was formerly used by the department of Media Studies themselves?

### **Scoring Rubric for Information Literacy (Van Helvoort, 2010)**

The Scoring Rubric for Information Literacy (Appendix A) consists of seven criterions. The first five criterions refer to properties of the knowledge product that the students have created. Such products can for instance be a research paper, an advisory report or a poster presentation. Criterions 6 and 7 refer to parts of the research *process*, the search terms that were used (6) and the databases, search engines or other resources where the search was executed (7). To grade these last two

**This article is © Emerald Group Publishing and permission has been granted for this version to appear here ([www.hbo-kennisbank.nl](http://www.hbo-kennisbank.nl)). Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.'**

criteria, it is necessary that students are asked to deliver a 'search process report' or a description of their 'search strategy' (Van Helvoort and Joosten, forthcoming).

Figure 1 gives a snapshot of one of the rubric's criteria, in this case criterion 5 on the creation of new knowledge.

Criterion		Professional behaviour	Insufficient behaviour	Grade 1-20=			
5	Creation of new knowledge out of relevant information	<input type="checkbox"/> The student product makes clear that the student analyzed information from different resources and that – based on this analysis – he / she formulated new insights, hypotheses or applications. Scope note: practice shows that students succeed in analysing and comparing several information sources, but are not capable of synthesizing the retrieved data into a new insight, hypothesis or application. If so, this criterion should be graded as “sufficient” or “poor”.	In the student product the student <input type="checkbox"/> did not reproduce the content of the retrieved information correctly or clearly and / or <input type="checkbox"/> paid no attention whatsoever to the analysis of the information sources found and / or <input type="checkbox"/> used only one information source without discussing the relevance or the reliability of the content, although there is reason for doubt.				
Score:		0 very good	0 good	0 sufficient	0 poor	0 bad	0 very bad

**Figure 1**  
Criterion 5 of the Scoring Rubric for Information Literacy

Figure 1 shows the description of professional behaviour for each criterion in column 3 and of insufficient behaviour in column 4. Graders can use the check boxes and mark or circle text phrases to make it clear which description, in their opinion, is applicable to the student product or the search strategy. Those checks and marks can be regarded as the feedback which is provided to the students.

Each criterion table ends with a 6 point Likert scale to give a score. Those scores are formulated in words because these—together with the descriptions of the professional and insufficient behaviour—are more informative for students than the grades which have a certifying role. If a teacher wants to give a grade, this is possible in the last column. The ranges for the grades are 1-10 or 1-20 for each criterion. This depends on the weight which is given to a criterion. As one can see in appendix A, criteria 1, 3 and 5 are regarded as more important than the other ones.

The scores on the 6 point Likert scale can be—conforming to the Dutch grading system—translated to the following grades: Very good = 10/20; good = 8/16; sufficient = 6/12; poor = 5/10; bad = 3/6; very bad = 1/2.

## Methodology

### Participants

For budgetary and work load reasons the test was restricted to 80 student papers. Those 80 were randomly selected from a group of 119 available papers. All students of those 119 papers had given permission for their work to be used anonymously in the research.

### Assignment

The assignment that was used for the test at the department of Media studies at the University of Amsterdam is part of the undergraduate course of Media History. Each student has to individually conduct an information review on a historic media topic and to formulate a research question and theoretical framework for further research. Each of the reports would be approximately 2500 words excluding the reference list.

### Comparison of the two assessment instruments

The department of Media Studies has its own grading instrument for the assignment which, contrary to the scoring rubric, is more or less a simple checklist with only criteria. Table 1 maps the criteria of the department's assessment instrument with the criteria in the Scoring Rubric for Information Literacy. One can see that most of the criteria from the scoring rubric are used

**This article is © Emerald Group Publishing and permission has been granted for this version to appear here ([www.hbo-kennisbank.nl](http://www.hbo-kennisbank.nl)). Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.'**

somewhere in the department's grading instrument. The main exception is the use of search terms. In the assignment, students are not asked to report about the exact search terms. Some students did report this but not all of them, and therefore this criterion could not be used in the research. The contrary occurred with the criteria concerning evaluation and planning in the department's grading instrument. Those criteria are not used in the scoring rubric and thus could not be used in the comparison.

**Table 1**

*Mapping the criteria in the Media Studies grading instrument and those in the Scoring Rubric for Information Literacy*

Media Studies grading instrument	Points	Scoring Rubric for Information Literacy	Points
Problem area	10	Orientation	20
Problem accounting	5		
Theoretical framework	20	Creation of new knowledge out of relevant information	20
Research methods	30		
. Search strategy (where and how?)		Search terms / keywords	10*
. Reliability of resources		Quality of the primary sources (books, journal articles, websites etc.)	20
. Relevance of resources			
. Variety of search engines and databases		Use of secondary sources	10
Evaluation	10*		
Planning	5*		
Accuracy	20		
. Reference lists		Reference list	10
. Correctly citing and paraphrasing		In text-citations	10

\*These criteria could not be mapped and therefore are not used in the test

### *Procedure*

For the grading process itself, two assistants were hired who had recently graduated from the department of Media Studies. They had previously helped undergraduate students with these assignments but had never actually been engaged with the grading of the student papers.

To apply a counterbalance with the aim of avoiding the hindering effects of the sequence of grading, the papers were randomly selected and divided into three groups. Twenty papers were in group T that was used for the training of the two graders. Forty papers were assigned to group A, and 40 other papers to group B. Before the start of the actual grading processes there were training meetings for each assessment instrument. During the training sessions the recommendations as given by Holmes and Oakleaf (2013) were followed. At the end of the meetings the two graders reached a high level of agreement on the scores. For the scoring rubric they attained absolute agreement on all criteria but one, for which they reached adjacent agreement.

During the actual grading processes, the two graders worked on different groups of papers. The grading work was distributed according to the scheme in Table 2.

**Table 2***Time scheme for the grading of student papers*

<b>Grader 1</b>	<b>Grader 2</b>	<b>Period</b>
Group A Scoring Rubric	Group B Scoring Rubric	Week 1-3
Group B Media Studies Instrument	Group A Media Studies Instrument	Week 4-6
Group B Scoring Rubric	Group A Scoring Rubric	Week 7-9
Group A Media Studies Instrument	Group B Media Studies Instrument	Week 10-11

The graders noted their scores on paper grading forms and transferred them to Excel forms which were then sent to the researchers. The maximum for the total scores using the scoring rubric was 90 points because criterion 6 (search terms) was eliminated. The maximum for the total scores with the Media Studies instrument was 85 points. The total score for each paper was recalculated to a grade—conforming to the Dutch grading system—on a 10 points scale from 1 to 10. After each round, feedback and training sessions with the graders were organised to exchange experiences and to answer and discuss any questions that had arisen.

### *Data analysis*

Interrater reliability for the scores using the scoring rubric (research question 1) is presented by the calculation of the percentages of agreement and those of 'adjacent agreement' between the two raters. 'Adjacent agreement' means that the graders did not differ more than one point on the 6 point Likert scale or the 10 points scale for the final gradings. The use of adjacent agreement might give a too positive picture of results but it is in practice often used because exact agreement is hard to realize (Stemler, 2004). Ballator and others (1999) indicate that with a 6 points Likert scale 80% is an acceptable level of adjacent agreement.

The percentages of agreement have the advantage that they are easy to understand. The problem is however that they do not correct for agreements that would be expected by chance (Hallgren, 2012). When there is enough data available, for interval data such as on the scoring rubric, it is therefore recommended to use Intraclass Correlation (Hallgren, 2012). This can be calculated with SPSS, and because there were two raters who graded each the same set of papers, we chose the type two-way mixed (ICC(3)) (Landers, 2011). Both absolute agreement and consistency were calculated and presented. The norm that we used regards 0.60 or higher as an indication of 'good interrater reliability', and 0.75 as 'excellent' (Cicchetti, 1994).

Homogeneity (research question 2) is determined using factor analysis in SPSS (extraction method 'Principal axis factoring') and internal consistency (research question 3) using the procedure reliability analysis (Cronbach's Alpha and Item-Total Statistics). For the determination of the correlation between the scores using the two different instruments (research question 4), we employed Pearson Correlation ( $r$ ).

## **Results**

### *Interrater Reliability (research question 1)*

Table 3 gives the simple numbers and percentages for absolute and adjacent agreement. Adjacent agreement is for all criterions 80% or higher, except for the grades on the 10 points scale. That those scores are overall rated as 'good' is confirmed by the coefficients for ICC(3) in Table 4. Values for all criterions except the first one ('Orientation') are above 0.60. For the total scores it even exceeds 0.75. This excellent level is hardly missed for the final grading on the scale 1–10.

**Table 3**

*Absolute and adjacent agreement between grader 1 and grader 2 when they use the scoring rubric*

Criterion	N =	Absolute agreement	Adjacent agreement
1 Orientation (1-6)	80	30 (38%)	64 (80%)
2 Reference list (1-6)	80	31 (39%)	66 (83%)
3 Quality of primary sources (1-6)	80	34 (43%)	67 (84%)
4 In text citations (1-6)	80	23 (29%)	71 (89%)
5 Creation of new knowledge (1-6)	80	29 (36%)	67 (84%)
7 Use of secondary sources (1-6)	80	31 (39%)	65 (81%)
Final grading (1-10)	80	25 (31%)	61 (76%)

**Table 4**

*Intraclass Correlation Coefficient Two-Way Mixed (ICC(3)) for the scores using the scoring rubric*

Criterion	ICC(3) type A (absolute agreement)	ICC(3) type C (consistency)
1 Orientation (1-6)	.566	.578
2 Reference list (1-6)	.618	.701
3 Quality of primary sources (1-6)	.708	.706
4 In text-citations (1-6)	.749	.751
5 Creation of new knowledge (1-6)	.635	.632
7 Use of secondary sources (1-6)	.639	.659
Total score (9-90)	.763	.763
Final grading (1-10)	.736	.736

*Homogeneity (research question 2) and Internal consistency (research question 3)*

The homogeneity of the scoring rubric is determined using factor analysis in SPSS (extraction method Principal Axis Factoring). The analysis shows for both graders that the six criteria that were applied are together one homogenous factor. The results for grader 1 shows that only one of the six potential factors had an Eigenvalue higher than 1. The results for grader 2 shows a second factor with an Eigenvalue higher than 1 (1.005). However, when in this case two factors are extracted while the rotation method Oblimin with Kaiser Normalisation is applied, it occurs that criterion 'Orientation' as well as criterion 'Secondary sources' each load higher than 0.300 on both factors. Furthermore, there is obviously a break in the scree plot after factor number 1. These are all signs that it is recommended to keep only one factor.

Table 5 shows that in both graders' case, if all six criteria are loaded on that only one factor, each criterion has a value higher than 0.4. For grader 1 there is no value below 0.6. This all makes it plausible that all the six criteria together refer to only one underlying construct.

**Table 5***Factor matrixes for the scoring rubric when one factor is extracted*

Criterion	Grader 1	Grader 2
1 Orientation	.772	.817
2 Reference list	.717	.521
3 Quality of primary sources	.940	.881
4 In text-citations	.723	.452
5 Creation of new knowledge	.822	.765
7 Use of secondary sources	.658	.805

Extraction method Principal Axis Factoring

The suggestion that the six criteria refer to only one underlying construct is confirmed by the item-total statistics in Tables 6 and 7. Cronbach's Alpha for both graders is 'good' (>0.8; Gliem and Gliem, 2003) and that for grader 1 is almost 0.9. None of the item-total correlations come below 0.4 and only the table for grader 2 has two criteria that would improve Cronbach's Alpha lightly if they were deleted. These are the same criteria that in the factor analyses loaded less high on the one factor that was distinguished. However, the improvements in the last column of Table 7 are so small (0.005 for criterion 2 'Reference list' and 0.017 for criterion 4 'In text-citations') that they don't mean very much.

**Table 6***Item-total statistics for the scoring rubric for grader 1*

Criterion	Corrected item-total correlation	Cronbach's Alpha if item deleted <sup>a</sup>
1 Orientation	.707	.881
2 Reference list	.686	.883
3 Quality of primary sources	.869	.853
4 In text-citations	.689	.883
5 Creation of new knowledge	.766	.872
7 Use of secondary sources	.629	.895

<sup>a</sup> Cronbach's Alpha: 0.896**Table 7***Item-total statistics for the scoring rubric for grader 2*

Criterion	Corrected item-total correlation	Cronbach's alpha if item deleted <sup>a</sup>
1 Orientation	.744	.807
2 Reference list	.460	.857
3 Quality of primary sources	.782	.797
4 In text-citations	.422	.869
5 Creation of new knowledge	.723	.810
7 Use of secondary sources	.740	.812

<sup>a</sup> Cronbach's Alpha: 0.852

*Correlation between the gradings using the scoring rubric and those using the Media Studies grading instrument (research question 4)*



The correlations between the scores using the two assessment instruments cannot be discussed when we are not sure about the qualities of the second assessment instrument, the instrument of the department of Media Studies. Table 1 shows five criteria of the Media Studies grading instrument that could be used for this research. Tables 8 and 9 give the item-total statistics for both graders when they graded the 80 papers using the Media Studies instrument.

**Table 8**

*Item-total statistics for the Media Studies instrument for grader 1*

Criterion	Corrected item-total correlation	Cronbach's alpha if item deleted <sup>a</sup>
1 Problem area	.759	.837
2 Problem accounting	.655	.859
3 Theoretical framework	.809	.764
4 Research methods	.836	.772
5 Accuracy	.776	.774

<sup>a</sup> Cronbach's Alpha: 0.843

**Table 9**

*Item-total statistics for the Media Studies instrument for grader 2*

Criterion	Corrected item-total correlation	Cronbach's alpha if item deleted <sup>a</sup>
1 Problem area	.526	.767
2 Problem accounting	.513	.794
3 Theoretical framework	.817	.633
4 Research methods	.753	.711
5 Accuracy	.603	.721

<sup>a</sup> Cronbach's Alpha: 0.778

It appears that for both graders, Cronbach's Alpha is a bit lower than in the case of the scoring rubric, but the results in Tables 8 and 9 show that the graders were also very consistent in their grades when they used the Media Studies instrument. None of the criteria have a Corrected item-total correlation of under 0.500.

The Intraclass Correlation Coefficient (type Two-Way Mixed (ICC(3)) for the final gradings using the Media Studies instrument are even a bit higher than those with the scoring rubric: .787 for absolute agreement and .784 for consistency, while they are both .736 in the case of the scoring rubric. Lastly, the factor analysis shows that all five criteria in the Media Studies instrument refer to only one dimension. Those are all indications that the final gradings using the Media Studies instrument can be used to compare them to those from the scoring rubric.

The Pearson correlation matrix for the final gradings in Table 10 shows that the correlations were 0.93 for grader 1 ( $p < 0.01$ ) and 0.76 for grader 2 ( $p < 0.01$ ). When the total scores are compared, the correlations are even higher: 0.97 for grader 1 and 0.77 for grader 2 (Table 11). All these values are regarded as 'strong' and/or 'very strong' (Evans, 1996).



**Table 10**

*Pearson correlation matrix for the final gradings using the Scoring Rubric for Information Literacy and those using the Media Studies grading instrument for grader 1 and 2*

	ScR Grader 1	ScR Grader 2	MS Grader 1	MS Grader 2
ScR Grader 1	1	0.587	0.927	0.626
ScR Grader 2		1	0.626	0.758
MS Grader 1			1	0.657
MS Grader 2				1

**Table 11**

*Pearson correlation matrix for the total scores using the Scoring Rubric for Information Literacy and those using the Media Studies grading instrument for grader 1 and 2*

	ScR Grader 1	ScR Grader 2	MS Grader 1	MS Grader 2
ScR Grader 1	1	0.622	0.971	0.629
ScR Grader 2		1	0.660	0.768
MS Grader 1			1	0.665
MS Grader 2				1

It is also remarkable that the correlations that were not expected – for instance that between ScR Grader 1 and MS Grader 2 – appear to still be rather strong (more than 0.60). However, this correlation is indeed less strong than when the two scoring instruments are applied by the same grader.

### Conclusions and discussion

In this case of the department of Media Studies at the University of Amsterdam where the Scoring Rubric for Information Literacy was tested with two graders, the scoring rubric appears to be a reliable and a valid assessment instrument. Interrater Reliability (investigated with research question 1) is proven by ‘good’ scores for the Intraclass Correlation. Only the first criterion (Orientation) had a result below 0.60. An explanation for this is that the criterion Orientation makes it possible for graders to hold—to a certain extent—their own interpretation for the criterion. The only other criterion that provides graders some opportunity to use their own interpretation is criterion 5 (Creation of new knowledge).

Construct validity is proven by a set of homogenous criterions (research question 2) that are highly correlated (research question 3). The high internal consistency of the criterions is furthermore an indication that the whole rubric, in this situation, was a reliable assessment instrument. However, one of the restrictions is that there was one criterion that could not be researched in this case (search terms). Finally, criterion validity is made plausible because of the high correlations between the final gradings using the Scoring Rubric for Information Literacy and those using the department’s grading instrument.

In relation to validity, it should be remarked that the high correlation between the scores using the two instruments is not an irrefutable argument that those instruments indeed refer to the construct information literacy: they could both measure the same construct but differ from ‘information literacy’. However, the reference to the construct information literacy is very plausible because these correlations are underpinned by:

- the attention for the content validity during the construction process of the rubric; and
- the high internal consistency of both instruments that appeared in the Item Total Statistics (to be found in the results of research question 3).

**This article is © Emerald Group Publishing and permission has been granted for this version to appear here ([www.hbo-kennisbank.nl](http://www.hbo-kennisbank.nl)). Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.'**

After considering all of the results from this research and from former research (Van Helvoort, 2010 and 2016), it can be concluded that two graders can use the Scoring Rubric for Information Literacy as a reliable and valid instrument for the measurement of information literacy performance. This conclusion is based on research in different education situations.

An advantage of the scoring rubric is that it can be applied in multiple situations, though minor modifications will often be needed. Appliance in different courses makes it possible to report the results to a broader public and benchmark results from different institutes or departments.

The focus in this research was on the traditional requirements of assessment instruments. As mentioned in the introduction, rubrics also have the function to stimulate learning because of their possibilities to supply exhaustive feedback and to set clear goals for an assignment, and their appliance for peer- and self-assessment. In former research (Van Helvoort, 2012) it was confirmed that the Scoring Rubric for Information Literacy indeed functioned as such an assessment tool for the stimulation of learning.

## References

- Angell, K. (2015), "The application of reliability and validity measures to assess the effectiveness of an undergraduate citation rubric", *Behavioral & Social Sciences Librarian*, Vol. 34, No. 1, pp. 2-15.
- Ballator, N., Farnum, M. and Kaplan, B. (1999), *NAEP 1996 trends in writing: Fluency and writing conventions*, National Center for Education Statistics, Washington, DC.
- Belanger, J., Zou, N., Mills, J., Holmes, C. and Oakleaf, M. (2015), "Project RAILS: Lessons learned about rubric assessment of information literacy skills", *Portal: Libraries and the Academy*, Vol. 15, No. 4, pp. 623-644.
- Brand-Gruwel, S., Wopereis, I., and Vermetten, Y. (2005), "Information problem solving by experts and novices: Analysis of a complex cognitive skill", *Computers in Human Behavior*, Vol. 21, No. 3, pp. 487-508.
- Bresciani, M., Oakleaf, M., Kolkhorst, F., Nebeker, C., Barlow, J., Duncan, K. and Hickmott, J. (2009), "Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines", *Practical Assessment, Research & Evaluation*, Vol. 14, No. 12, pp. 1-7., available at <http://www.pareonline.net/getvn.asp?v=14&n=12> (accessed 10 September 2016).
- Cameron, L., Wise, S. and Lottridge, S. (2007), "The development and validation of the information literacy test", *College & Research Libraries*, Vol. 68, No. 3, pp. 229-236.
- Carbery, A. and Leahy, S. (2015), "Evidence-based instruction: Assessing student work using rubrics and citation analysis to inform instructional design", *Journal of Information Literacy*, Vol. 9, No. 1, pp. 74-90.
- Cicchetti, D. (1994), "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology", *Psychological Assessment*, Vol. 6, No. 4, p. 284.
- Cronbach, L. and Meehl, P. (1955), "Construct validity in psychological tests", *Psychological Bulletin*, Vol. 52, pp. 281-302.
- Evans, J. (1996), *Straightforward statistics for the behavioral sciences*, Brooks/Cole Publishing, Pacific Grove, CA.

**This article is © Emerald Group Publishing and permission has been granted for this version to appear here ([www.hbo-kennisbank.nl](http://www.hbo-kennisbank.nl)). Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.'**

- Gliem, R. and Gliem, J. (2003), "Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales", in *2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education*, pp. 82-88.
- Hallgren, K. (2012), "Computing inter-rater reliability for observational data: An overview and tutorial", *Tutorials in Quantitative Methods for Psychology*, Vol. 8, No. 1, pp. 23.
- Helvoort, J. van (2010), "A scoring rubric for performance assessment of information literacy in Dutch Higher Education", *Journal of Information Literacy*, Vol. 4, No. 1, pp. 22-39.
- Helvoort, J. van (2012), "How adult students in information studies use a scoring rubric for the development of their information literacy skills", *Journal of Academic Librarianship*, Vol. 38, No. 3, pp. 165-171.
- Helvoort, J. van (2013), "How faculty in The Hague University of Applied Sciences uses the scoring rubric for information literacy", *Communications in Computer and Information Science*, Vol. 397, pp. 436-442.
- Helvoort, J. van (2016), *Beoordelen van informatievaardigheden in het hoger onderwijs: Academisch proefschrift ter verkrijging van de graad van doctor aan de Universiteit van Amsterdam*, Den Haag, De Haagse Hogeschool.
- Helvoort, J. van and Joosten, H. (forthcoming), "The Scoring Rubric for Information Literacy as a tool for learning", in Sales, D. and Pinto, M. (Eds.), *Pathways into Information Literacy and Communities of Practice: Teaching Approaches and Case Studies*, Chandos, Oxford [?].
- Holmes, C. and Oakleaf, M. (2013), "The official (and unofficial) rules for norming rubrics successfully", *The Journal of Academic Librarianship*, Vol. 39, No. 6, pp. 599-602.
- Jonsson, A. and Svingby, G. (2007), "The use of scoring rubrics: Reliability, validity and educational consequences", *Educational Research Review*, Vol. 2, No. 2, pp. 130-144.
- Knight, L. A. (2006), "Using rubrics to assess information literacy", *Reference Services Review*, Vol. 34, No. 1, pp. 43-55.
- Landers, R. (2011), "Computing intraclass correlations (ICC ) as estimates of interrater reliability in SPSS", *NeoAcademic*, Vol. 2011, No. 16 november.
- Moskal, B. and Leydens, J. (2000), "Scoring rubric development: Validity and reliability", *Practical Assessment, Research & Evaluation*, Vol. 7, No. 10.
- Oakleaf, M. (2008), "Dangers and opportunities: A conceptual map of information literacy assessment approaches", *Portal: Libraries and the Academy*, Vol. 8, No. 3, pp. 233-253.
- Oakleaf, M. (2009), "Using rubrics to assess information literacy: An examination of methodology and interrater reliability", *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 5, pp. 969-983.
- Pinto, M. and Sales, D. (2015), "Uncovering information literacy's disciplinary differences through students' attitudes: An empirical study", *Journal of Librarianship and Information Science*, Vol. 47, No. 3, pp. 204-215.
- Reddy, Y., and Andrade, H. (2010), "A review of rubric use in higher education". *Assessment & Evaluation in Higher Education*, Vol. 35, No. 4, pp. 435-448.
- Stemler, S. (2004), "A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability", *Practical Assessment, Research & Evaluation*, Vol. 9, No. 4.

## Appendix A

### Scoring rubric for Information Literacy student product

Name teacher / grader:

Name/ID-No. student:

Criterion		Professional behaviour		Insufficient behaviour			
1	Orientation	<input type="checkbox"/> The student product makes clear that the student did a good orientation on the topic and that he/she formulated his/her own focus on the topic or research question. This is also expressed by the fact that the student formulated one or more good research questions.		<input type="checkbox"/> The student product makes clear that the student used the question as it was originally formulated in the assignment or student task. The student him/herself did not further explore the question as such. An example of this behaviour is that the student did not define the core key terms and that these terms are supposed to be clear while they are at least multi interpretable.		Grade 1-20=	
Score:		0 very good	0 good	0 sufficient	0 poor	0 bad	0 very bad

Criterion		Professional behaviour		Insufficient behaviour			
2	Reference list	<input type="checkbox"/> The student product has a reference list that is complete and the citation style is used correctly. With the reference list it is easy to identify the documents that the student used.  Remark: the last point is more important than a correct bibliographic description in accordance with a standard citation style. However, for the score 'very good' the citation style must also be used correctly.		<input type="checkbox"/> There is no reference list in the student product and / or <input type="checkbox"/> The reference list is not complete (documents that are cited in the text are not listed in the reference list) or <input type="checkbox"/> Important bibliographic data (title, author, year of publication) are missing.  An example that often recurs in educational practice: for internet resources only the URL is mentioned.		Grade 1-10=	
Score:		0 very good	0 good	0 sufficient	0 poor	0 bad	0 very bad

Criterion		Professional behaviour		Insufficient behaviour			
3	Quality of the primary sources (books, journal articles, websites etc.)	<input type="checkbox"/> The reference list of the student product makes clear that the student has used relevant, reliable (preferably authentic) and up-to- date information sources that discuss the topic or the question from different points of view.		<input type="checkbox"/> The information sources the student has used are insignificant, outdated or not relevant enough. An example of 'insignificance' is that the student only used Internet-sites as an information source. And / or ... <input type="checkbox"/> The information sources the student used are one-sided (too much from one point of view). The student has, for instance, only used government information(.gov-sites) or publications from one particular author.		Grade 1-20=	
Score:		0 very good	0 good	0 sufficient	0 poor	0 bad	0 very bad

## Scoring rubric for Information Literacy

Name teacher / grader:

Name/ID-No. student:

Criterion		Professional behaviour		Insufficient behaviour		Grade 1-10=	
4	In text-citations	<input type="checkbox"/> In the text of the product it is made clear what information sources the student has used. In the case of a digital student product this is also true for images and audiovisual information.		<input type="checkbox"/> The student has used someone else's work (text fragments, images, audiovisuals) in his / her own product without reference to the original source. Even if this was done unintentionally, strictly speaking this is plagiarism.			
Score:		0 very good	0 good	0 sufficient	0 poor	0 bad	0 very bad

Criterion		Professional behaviour		Insufficient behaviour		Grade 1-20=	
5	Creation of new knowledge out of relevant information	<input type="checkbox"/> The student product makes clear that the student analyzed information from different resources and that – based on this analysis – he / she formulated new insights, hypotheses or applications. Scope note: practice shows that students succeed in analysing and comparing several information sources, but are not capable of synthesizing the retrieved data into a new insight, hypothesis or application. If so, this criterion should be graded as “sufficient” or “poor”.		In the student product the student <input type="checkbox"/> did not reproduce the content of the retrieved information correctly or clearly and / or <input type="checkbox"/> paid no attention whatsoever to the analysis of the information sources found and / or <input type="checkbox"/> used only one information source without discussing the relevance or the reliability of the content, although there is reason for doubt..			
Score:		0 very good	0 good	0 sufficient	0 poor	0 bad	0 very bad

## Scoring rubric for Information Literacy

### Search Strategy

Search Strategy

Criterion		Professional behaviour			Insufficient behaviour			Grade 1-10=
6	Search terms / keywords	<input type="checkbox"/> The student used search terms that are relevant for the topic or the research question. He / she used relevant synonyms, search terms in English and from the professional jargon.			<input type="checkbox"/> The student used search terms that are too general (non-professional) and / or <input type="checkbox"/> the student did not use relevant synonyms, associated terms or search terms in English.			
Score:		0 very good	0 good	0 sufficient	0 poor	0 bad	0 very bad	
Criterion		Professional behaviour			Insufficient behaviour			Grade 1-10=
7	Use of secondary sources	<input type="checkbox"/> The student used a variety of secondary sources (search engines, books for tracking citations, scholarly journals, databases, social networks). If necessary he / she used an interlibrary loan to obtain the materials needed.			<input type="checkbox"/> The student only used information sources that are easily accessible. For instance: he / she only used <ul style="list-style-type: none"><li>• The “quick search”-box of a general search engine and / or</li><li>• Materials provided by his / her professor.</li></ul>			
Score:		0 very good	0 good	0 sufficient	0 poor	0 bad	0 very bad	

Total score (maximum 100) =

Final grading (1-10) =

This article is © Emerald Group Publishing and permission has been granted for this version to appear here ([www.hbo-kennisbank.nl](http://www.hbo-kennisbank.nl)). Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.'

