

Assessors' approaches to portfolio assessment in Assessment of Prior Learning procedures

Desiree Joosten-ten Brinke^{a*}, Dominique M.A. Sluijsmans^a and Wim M.G. Jochems^{a,b}

^a*Educational Technology Expertise Centre, Open University of the Netherlands, Heerlen, The Netherlands;* ^b*Eindhoven School of Education, Eindhoven University of Technology, Eindhoven, The Netherlands*

In an effort to gain better understanding of the assessment of prior informal and non-formal learning, this article explores assessors' approaches to portfolio assessment. Through this portfolio assessment, candidates had requested exemptions from specific courses within an educational programme or admission to the programme based on their prior learning. The assessors judged the portfolios according to set rating criteria, and subsequently discussed their approaches. Their decision-making processes, perception of portfolio use in the Assessment of Prior Learning (APL), deciding factors in portfolio assessment and use of the rating criteria were key elements in this discussion. The results show that they do use the rating criteria as an indicator in decision-making, but have mixed perceptions regarding the fairness of APL portfolio assessment. They perceive the portfolio evidence in combination with sound argumentation as the deciding elements in portfolio assessment.

Keywords: prior learning assessment; portfolio assessment; assessment criteria

Portfolio assessment is a complex task (Baume and Yorke 2002; Driessen et al. 2006; Tigelaar et al. 2005; Van der Schaaf and Stokking 2008) given that the content is often descriptive, context-bound, personal and requires much interpretation (Delandshere and Petrosky 1998; Moss 1994). In procedures in which predominantly prior informal and non-formal learning is assessed (henceforth indicated as Assessment of Prior Learning and abbreviated as APL), the portfolio is the most common method for presenting the evidence of prior learning (Bjørnavold 2001; Clarke and Warr 1997).

In this article, we investigate assessors' approaches to portfolio assessment in the context of APL. Assessment of formal learning, established in certificates and diplomas, is beyond the scope of this article. Approach is defined as the steps taken in the decision-making process, the perception of portfolio use in APL in terms of fairness, usability and relevance, the deciding factors in portfolio assessment, and the use of rating criteria. We first elaborate on the background of and quality criteria for portfolio assessment. Then we describe portfolio design in the context of our research on APL at the Open University of the Netherlands and present a study in which assessors' approaches to portfolio assessment are investigated. Finally, we use the findings of this study as a starting point for recommendations about portfolio assessment in APL.

*Corresponding author. Email: desiree.joosten-tenbrinke@ou.nl

Quality of portfolio assessment

It is argued that in measuring learning of knowledge, skills and attitudes, it is often better to use a combination of assessment methods (e.g. Baartman et al. 2006; Dierick and Dochy 2001; Sluijsmans, Straetmans, and Van Merriënboer 2008). Evans (1993) states that academic staff is responsible for selecting the most appropriate assessment methods for judging the submitted evidence of prior learning. In the context of APL, a variety of assessment methods can be used, such as interviews, demonstrations and simulations (Michelson and Mandell 2004). However, portfolio assessment is the most common method. A portfolio is a compilation of work by a learner to demonstrate acquired knowledge, skills and competences, and includes the learner's reflections on this work (Barrett 2003). It is seen as one of the best instruments for visualising and evaluating competences acquired in informal or non-formal contexts (Bjørnavold 2001). Portfolios actively engage learners in understanding the relationship between the culture of academic knowledge in higher education and that required in the workplace (Michelson and Mandell 2004). At the same time, they are well received by tutors and assessors as a structured approach to the presentation of evidence (Clarke and Warr 1997), and allow full appreciation of the holistic nature of competences and the personal character of individuals' work over long periods of time.

Composing a portfolio requires that learners assess their own prior learning, and present it by way of evidence and argument. Several types of evidence are appropriate; Barrett (2003) distinguishes between artefacts, reproductions, attestations (references) and productions (documents specially prepared for the portfolio). This evidence should meet quality criteria such as educational relevance, transferability, validity, authenticity, currency and sufficiency (see for a detailed description of these criteria Joosten-ten Brinke et al. 2008).

Research on portfolio use in APL mainly has focused on the difficulties of APL candidates in gathering and presenting evidence of prior learning (e.g. Colley, Hodgkinson, and Malcolm 2002; Cretchley and Castle 2001; Shapiro 2003). In most studies, the perspective of the learner is the main focus of study. However, the assessors' role in assessing this evidence of prior learning is underexposed. And, for example, it is unclear how assessors can gauge the level of candidates' prior learning (Trowler 1996). Osman (2006) argues that in general, assessors feel competent in the assessment of formal learning, which is directly linked to the academic educational programme, but in contrast, they are unfamiliar with the assessment of prior informal and non-formal learning. What strategy do they use in assessing this prior informal and non-formal learning?

Research in the context of teacher education has shown that objective and unambiguous portfolio assessment is difficult (Tigelaar et al. 2005; Van der Schaaf, Stokking, and Verloop 2005), and how assessors can gauge the level of candidates' prior learning is unclear (Trowler 1996). Although portfolio assessment should meet quality requirements such as reliability and validity, in practice it is often difficult to sufficiently address these criteria. Portfolio reliability (i.e. the extent to which its assessment remains consistent over repeated measurements under identical conditions) is a complex issue given its interpretative, context-bound and personal character (Delandshere and Petrosky 1998; Moss 1994). An essential step in portfolio assessment is the need to improve agreement among assessors as to the rating criteria and the use of rating forms (Tigelaar et al. 2005). However, this agreement is not sufficient

for establishing assessment's validity (Linn 1994; Messick 1995). Johnston (2004) suggests using reliability and validity within an interpretative approach, which allows discussion of values and standards, and bridging between the local context and the curriculum. Negotiating with co-assessors over the outcomes will improve the reliability and validity. This interpretative approach is based on the idea that there is not one absolute, objective answer in assessing portfolios. Still, portfolio assessment particularly affects consequential validity; interpretation of assessment scores may not damage candidates' future learning paths. Score validity in portfolio assessment is threatened when assessors omit major steps in the decision process, evaluate portfolios without consideration of important criteria or apply extraneous assessment criteria (Heller, Sheingold, and Myford 1998). Driessen et al. (2006) found in their research on validity of portfolio assessment in the context of medical education that assessors are able to use only relevant criteria in their judgements and to neglect irrelevant extraneous criteria.

It remains unclear as to whether the assessors use these rating criteria and forms appropriately, even if they are trained in their use (Baume, Yorke, and Coffey 2004; Tigelaar et al. 2005; Van der Schaaf and Stokking 2008). Assessors' actual decision-making processes, defined as the processes to seek an appropriate, but not necessarily optimal, solution to a problem (Simon 1957), in APL are largely undocumented. According to Heller, Sheingold, and Myford (1998), the process of portfolio assessment consists of three essential components: evaluating individual texts in the portfolio one at a time, evaluating across texts in the portfolio and articulating a rating, weighing relative amounts or importance of evidence as needed. Van der Schaaf, Stokking, and Verloop (2005) showed, in the context of teacher education, that judgements were influenced by previous ratings and experiences, and that despite agreeing on ratings there remains a difference in portfolio interpretation. In addition, intuition, as a domain-specific competence to reach an appropriate decision, is mentioned as a crucial component of a decision-making process by professionals (Harteis and Gruber 2008). Intuition makes use of knowledge resources secured through individuals' professional experiences. Baume, Yorke, and Coffey (2004) analysed assessors' rationale behind their portfolio judgements in a course completion setting in which portfolio assessment was the only assessment method. They also concluded that assessors interpret the same rating criteria in different ways. Some are rather stringent, stating that evidence must be clearly and accurately labelled and in the proper place, while others are prepared to dig about for evidence within the portfolio. However, studies by Van der Schaaf and Stokking (2008) and Baume, Yorke, and Coffey (2004) show that assessors are capable of articulating the reasons behind their judgements, which helps us gain more insight into decision-making in portfolio assessment. Johnston (2004) mentions several aspects influencing the decision-making process. First, the way a final grade is calculated. Are there assessment scales and how is the agreement between raters? Second, the nature of assessor training. Assessors should have a marking rubric, discuss their grading standards before marking and have a consistent view over what class of degree the marks represent, especially in areas where it is hard to get agreement. Third, issues related to local and external assessors. Evidence about the use of local graders is mixed. Familiarity with the learning and assessment context appears to be significant in assisting 'reliable' grading. Fourth, holistic versus individual element scoring and finally, the way assessors actually use assessment criteria. These five aspects will be subject of this study to the assessors' approach to portfolio assessment in APL.

APL portfolio assessment at the Open University of the Netherlands

The presented study was conducted in the context of the Open University of the Netherlands (OUNL). The OUNL caters to lifelong learners of 18 years and over with no admission requirements. Learners at the OUNL are adults who have many learning experiences obtained from informal and non-formal learning. In a previous study (Joosten-ten Brinke, Sluijsmans, and Jochems, 2009), one-third of the population of learners indicated that they thought that their prior learning was at the required academic level and exemptions were appropriate. Initially, the OUNL developed an APL procedure primarily based on the credit exchange model (Butterworth 1992; Trowler 1996) in which learners received credit points if competences acquired either informally or non-formally matched the learning outcomes of an accredited educational programme. Since 2006, though, the procedure has shifted to a more developmental approach (Butterworth 1992). The credit exchange approach implies that proven competences are exchanged for course credits by allowing exemption from part of the educational programme. A developmental approach, however, focuses more on reflection of the achieved competences in relation to future learning (Butterworth 1992).

Based on this second approach, a general portfolio template was developed containing the following components (Bloor and Butterworth 1990): (1) summary of the APL application; (2) overview of the competences; (3) reflective writing on how the candidate's experience produced learning which meets the programme criteria; and (4) evidence to support the application. If content experts deemed it necessary, the basic elements of the template were then refined for each educational programme according to its specific domain and standards. Likewise, the rating forms and criteria were first designed for all educational programmes and subsequently refined to meet the domain-specific criteria of each educational programme. The portfolios were to demonstrate candidates' prior learning with respect to relevant knowledge, skills or competences of the educational programme. No particular kinds of entries or standardised pieces of work were required for the portfolios. Rather, candidates were to select those pieces of work that, taken together, would provide strong evidence of their achievement with respect to the educational programme. Such a portfolio is named by Heller, Sheingold, and Myford (1998) as a standard-based, non-prescriptive assessment portfolio design. Table 1 shows the rating form used by the assessors. The criteria enclosed in this form are derived from Joosten-ten Brinke et al. (2008). Each numbered criterion is a verbal description and involves sub-criteria. The assessors have to judge whether the verbal descriptions correspond with the described learning experiences in the portfolio by giving an insufficient or sufficient decision. Sadler (2005) refers to this method of grading as the 'qualitative assessor judgement'. Other common grading policies (e.g. scores on different assessment are added together and then projected to a scale, or checklists for achievement of course objectives) were less appropriate because of the personal contexts of the APL candidates. Their portfolios are never perfectly characterised by the assessment criteria, so the assessor makes a judgement whether it is sufficient or not.

During the APL procedure, candidates fill out the portfolio template for their intended educational programme and submit it, in duplicate, for assessment. The assessors, these are at least two academic employees of the educational programme, fill in the rating form and then decide whether it contains sufficient information for a final decision on the provision of exemptions. If not (i.e. as one of the criteria of the

Table 1. The rating form for portfolio assessment in APL.

| Portfolio assessment form for [name candidate] | |
|---|-------------------------|
| 1. The candidate describes his/her prior learning experiences (describing situation, tasks, activities and results). Candidate reflects on these experiences and provides information on the value of this experience for the educational programme. | Insufficient/sufficient |
| <i>Assessor comments</i> | |
| 2. The candidate's competences, knowledge or skills are on the level of higher vocational education. ^a | Insufficient/sufficient |
| <i>Assessor comments</i> | |
| 3. The evidence should be educationally relevant, transferable from the experiential to academic environment, valid (i.e. focused on the appropriate competences, knowledge and skills specified by the educational programme), authentic (i.e. the candidate must have undertaken what is claimed), recent and sufficient. | Insufficient/sufficient |
| <i>Assessor comments</i> | |
| 4. Domain-specific criteria | Insufficient/sufficient |
| <i>Assessor comments</i> | |
| If failing one of the criteria, additional information is necessary. This additional information will be gained by (tick where appropriate). | |
| Give candidate an additional assignment: (Fill in type of assignment and passing requirements) _____ | |
| <i>Assessment conversation</i> | |
| Fill in subjects (questions) and appropriate answers for passing _____ | |

The assessors final decision is:

^aA description of 'higher vocational education' is derived from the European Qualifications Framework (European Union 2008, 16): 'Advanced knowledge of a field of work or study, involving a critical understanding of theories and principles; advanced skills, demonstrating mastery and innovation, required to solve complex and unpredictable problems in a specialised field of work or study; [ability to] manage complex technical or professional activities or projects, taking responsibility for decision-making in unpredictable work or study contexts; and [ability to] take responsibility for managing professional development of individuals and groups'.

rating form is negatively judged), the assessors can request (1) an assessment conversation (i.e. a criterion-oriented interview), (2) an additional assignment, or (3) an assessment conversation and an additional assignment. Before the additional assignment and the assessment conversation, the assessors have to communicate how these additional assessments support their decision-making. Based on the portfolio and any additional assessments, a decision is then made as to exemptions. In summary, the APL assessment process involves decisions both on the portfolio's quality in relation to the rating criteria and on the need for additional assessments.

In conclusion, portfolio assessment is a difficult task for assessors; assessors' approaches differ and research on assessors' approaches in the specific APL context lacks. Therefore, the present study focuses on the question of 'What are assessors' approaches to portfolio assessment in APL?' Our approach involves investigation of

four aspects: (1) the decision-making process; (2) perception of portfolio use in terms of fairness, usability and relevance; (3) the deciding factors in assessment; and (4) the use of rating criteria.

Method

Participants

Ten assessors affiliated with six educational domains – educational science ($n = 2$), cultural science ($n = 2$), computer science ($n = 2$), Dutch law ($n = 1$), management science ($n = 2$), and psychology ($n = 1$) – were involved in the portfolio assessment. Each assessor had judged three to six portfolios before the interview. For each candidate at least two assessors judged the portfolio. The assessors are experienced and qualified teachers (more than 10 years) with content knowledge, skills in constructing tests, knowledge of criteria and standards in their domain, judgemental skills and feedback skills.

Materials

To investigate assessors' approaches, a retrospective, open-ended interview and questionnaire was administered.

Retrospective interview

This interview consisted of open-ended questions about the following topics: the deciding factors in evaluating a portfolio, the fairness of portfolio assessment alone, the extent to which the portfolios meet the rating criteria and the whole decision-making process (based on the aspects of Johnston [2004]: procedure to reach the final decision, level of judgement, assessment method, cooperation and agreement between assessors, assessor types, use of scales [holistic or elementary] and criteria, moment for final decision-making). To illustrate their responses with examples from the portfolios, each assessor had two assessed portfolios at hand; one with additional assessments and one without.

Questionnaire

The questionnaire consisted of 42 items divided into three parts on a five-point Likert scale ('strongly disagree' to 'strongly agree'). The items were derived from previous research; the first part (10 items; see Table 2), for example, is derived from Baume, Yorke, and Coffey (2004), who investigated assessors' rationale in the decision-making process. For example, the statement 'It doesn't matter too much where it is or how it's labelled as long as it's there' (which served in Baume, Yorke, and Coffey as a rationale for a judgement) on our questionnaire became 'In assessing the portfolio, it doesn't matter too much where the evidence is or how the evidence is labelled as long as it's there'.

The second part of the questionnaire (19 items; see Table 3) is based on Van der Schaaf's (2005) questionnaire in which portfolio use (relevance, usability and fairness) is investigated. The questions were adapted to the context of our study – for example, the question 'The portfolio elements are relevant for teaching students research skills',

Table 2. Means and standard deviations on the items derived from Baume, Yorke, and Coffey (2004).

| Item | Item | <i>N</i> | <i>M</i> | <i>SD</i> |
|------|---|----------|----------|-----------|
| B1 | It does not matter much how it is labelled as long as it is there (recoded). | 8 | 3.13 | 1.25 |
| B2 | In assessing the portfolio, it doesn't matter too much where the evidence is or how the evidence is labelled as long as it's there (recoded). | 10 | 4.10 | .88 |
| B4 | A course requirement must be addressed explicitly to achieve a positive judgement. | 10 | 4.50 | .71 |
| B5 | A judgement can be lifted to reflect other good material in the portfolio. | 10 | 3.60 | .84 |
| B6 | The assessor should be prepared to use their common sense and judgement. | 9 | 4.11 | 1.36 |
| B7 | A claim (APL application) must be made in addition to the provision of evidence. | 10 | 4.50 | .71 |
| B8 | Assertion without evidence is just acceptable. | 10 | 2.60 | 1.08 |
| B9 | An assertion must be clear. I must not read into what is given. | 10 | 4.40 | 1.08 |
| B10 | An overall judgement is more important than a slavish adherence to the rules. | 9 | 3.67 | .71 |

became 'The portfolio elements are relevant to the competences of the educational programme'. Van der Schaaf (2005) distinguished reliable Likert scales for the relevance of the portfolio elements (Cronbach's $\alpha = .85$; 8 items, $M = 4.20$; $SD = .49$; $N = 8$), usability (Cronbach's $\alpha = .89$; 7 items, $M = 4.13$; $SD = .59$; $N = 8$) and fairness (Cronbach's $\alpha = .91$; 4 items, $M = 4.03$; $SD = .75$; $N = 8$).

The third part of the questionnaire (13 items; see Table 4) is derived from research by Joosten-ten Brinke, Sluijsmans, and Jochems (2009) which measures learners' perceptions of portfolio use. Examples of the items on portfolio assessment to be answered on the five-point Likert scale are 'The structure of the portfolio was clear', and 'The candidates were capable to deliver evidence at the required level'. After removing four items from the analyses (one item based on Baume, Yorke, and Coffey, one on Van der Schaaf and two on Joosten-ten Brinke, Sluijsmans, and Jochems, 2009), the reliability of the whole questionnaire was .76 (38 items; $N = 10$).

Procedure

The portfolio assessments took place in summer 2007. After completing the assessments, the assessors were interviewed by the researcher in a pre-structured and audio-taped session. The interviews averaged 33.5 minutes ($SD = 6.38$ minutes). During the interviews, the assessors had two of their last-assessed portfolios at hand for reference. Directly after the interview they filled in the questionnaire.

Data analysis

The audio-taped interviews were transcribed, and two researchers (i.e. the first two authors) coded the transcriptions into categories relating to information about the assessors in general and the decision-making process in specific. The first column of

Table 3. Means and standard deviations on the items derived from Van der Schaaf (2005).

| Item | Item | <i>N</i> | <i>M</i> | <i>SD</i> |
|------|---|----------|----------|-----------|
| V1 | The portfolio elements are relevant for the assessments of the competences in the educational programme. | 10 | 4.80 | .42 |
| V2 | The portfolio elements are in congruence with the content standards. | 10 | 4.30 | .95 |
| V3 | The portfolio elements reflect the activities during informal and formal learning related to the educational programme. | 10 | 4.60 | .52 |
| V4 | The portfolio elements fit the content standards. | 10 | 3.90 | .57 |
| V5 | The portfolio reflects the salient tasks of the learners' work. | 10 | 3.50 | .71 |
| V6 | Less competent learners will score lower on the portfolio assessment than learners who are more competent in the domain of study. | 10 | 4.30 | .48 |
| V7 | The portfolio shows the learner's competences. | 10 | 4.00 | .47 |
| V8 | The portfolio shows the work conducted by the learner in relevance with the educational programme. | 10 | 4.10 | .32 |
| V9 | The information given by a portfolio is sufficient for raters to judge the competences. | 10 | 3.60 | .70 |
| V10 | The information given by a portfolio is sufficient for raters to give candidates feedback on their strong and weak points. | 9 | 3.67 | .50 |
| V11 | The portfolio elements suit the learners' everyday practice. | 10 | 4.60 | .52 |
| V12 | It is clear to learners how to develop a portfolio. | 9 | 3.56 | .73 |
| V13 | It is clear to learners what the content of the portfolio is. | 10 | 3.70 | .48 |
| V14 | It is clear to learners what the assessment standards are. | 10 | 3.70 | .68 |
| V15 | It is clear to learners how the assessment results are used. | 10 | 3.80 | 1.03 |
| V16 | The portfolio causes learners to be judged incorrectly (recoded). | 7 | 4.00 | 1.00 |
| V17 | It is fair to use the portfolio model for prior learning assessment. | 10 | 4.10 | .57 |
| V18 | It is fair to give learners feedback on their prior learning based on the portfolio elements. | 10 | 4.30 | .68 |

Table 5 shows these categories. This coding was undertaken according to nominal scales, and agreement between the two researchers analysed using Cramer's V. No significant difference was found between them; therefore, the coding of just one researcher was used for further analysis. The coded data were analysed by calculating frequencies, while the transcriptions, which served as explanations and examples, were analysed qualitatively. The questionnaires were analysed by mean scores and standard deviations. Mean scores of three or lower represent non-supportive opinions; scores higher than three represent supportive opinions. Scores between two and four with a standard deviation higher than 1.0 remain inconclusive. These items are therefore qualitatively analysed on their meaning.

Results

Table 5 shows the interview results concerning information about the assessors and the decision-making process. Three different types of assessors could be distinguished: course examiners, APL assessors and assessment committees.

Table 4. Means and standard deviations on the items derived from Joosten-ten Brinke, Sluijsmans, and Jochems (2009).

| Item | Item | <i>N</i> | <i>M</i> | <i>SD</i> |
|------|--|----------|----------|-----------|
| J2 | The structure of the portfolio is clear. | 10 | 4.40 | .52 |
| J3 | Portfolio assessment is suitable for measuring candidates' prior learning. | 10 | 4.10 | .57 |
| J4 | The rating criteria are clear. | 10 | 3.90 | .57 |
| J5 | All APL candidates have the right to an assessment conversation in addition to portfolio assessment. | 10 | 2.80 | 1.75 |
| J6 | APL candidates did not pay enough attention to portfolio composition. | 9 | 3.78 | 1.09 |
| J7 | Some portfolios provide sufficient information for assessors to make a final decision. | 10 | 4.10 | .99 |
| J8 | Some portfolios provide insufficient information for assessors to make a final decision. | 10 | 4.00 | .94 |
| J10 | The prior learning level could be judged through portfolio assessment. | 10 | 3.70 | .68 |
| J11 | Interviews not necessarily based on the outcome of portfolio assessment were held. | 5 | 2.00 | 1.73 |
| J12 | An assessment conversation always yields additional information for judgement. | 8 | 2.75 | 1.04 |
| J13 | The assessment conversation led to adjustments to the final decision. | 4 | 3.75 | 1.26 |

Two assessors assessed at individual course level (psychology and Dutch law) and eight assessed course clusters. They established that the candidates usually submitted clear applications, indicating that they were clear as to their own objectives for taking part in APL.

Assessors' decision-making processes

Eight assessors (excluding assessors B and H) indicated that they first individually assess the portfolios, then discuss this assessment with a co-assessor and finally collaboratively with the two assessors or in a team decide whether additional assessments are necessary. After the portfolio assessment, four (D, E, F and J) decided in favour of an assessment conversation, and four (B, F, I and J) in favour of an additional assignment. Whether the candidate was to receive exemptions was decided at different moments. Six assessors (A, B, C, G, H and I) reached the final decision in collaboration with co-assessors in a general session. For two of these six (B and H), the decision was made based on the assessor's initial judgement. The other four assessors (D, E, F and J) reached the final decision directly after the assessment conversation, based on discussion with the second assessor. Half of the assessors amended their initial decision on the basis of additional assessments or the second opinion of a colleague (I and J).

Part 1 of the questionnaire involved items on the argumentation used in the decision-making process. Table 2 gives an overview of the mean scores and standard deviations of the items derived from Baume, Yorke, and Coffey (2004). The answers to the Baume, Yorke, and Coffey items on judgement rationale are similar to those on

Table 5. Interview results: general assessor information and the decision-making process.

| Interview question | Assessor* | | | | | | | | | |
|--|-----------|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J |
| <i>Level of judgement</i> | | | | | | | | | | |
| 1 = Judgement at course level | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 2 |
| 2 = Judgement of course clusters | | | | | | | | | | |
| <i>Assessor types</i> | | | | | | | | | | |
| 1 = Course examiners | 3 | 1 | 3 | 2 | 2 | 2 | 3 | 1 | 3 | 2 |
| 2 = APL assessors | | 3 | | 3 | 3 | 3 | | 3 | | 3 |
| 3 = Assessment committee | | | | | | | | | | |
| <i>Procedure</i> | | | | | | | | | | |
| 1 = Individually assess portfolio then make group decision | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |
| 2 = The sum of the results of each course examiner | | | | | | | | | | |
| <i>Assessment methods</i> | | | | | | | | | | |
| 1 = Portfolio assessment | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 = Assessment conversation | | 3 | | 2 | 2 | 2 | | | 3 | 2 |
| 3 = Assignment | | | | | | 3 | | | | 3 |
| <i>Moment for final decision-making</i> | | | | | | | | | | |
| 1 = In general session with fellow assessors | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 |
| 2 = Directly after assessment conversation | | | | | | | | | | |

*The assessors have been labelled A–J. References to these letters are given in the text.

the following items: that requirements be addressed explicitly to achieve a positive judgement (B4); applications must be made in addition to evidence provision (B7); assertions must be clear; assessors must not read into what is given (B9); and assessors should be prepared to use common sense and judgement (B6). To the assessors, it does matter how evidence is labelled (B1) and where it is placed in the portfolio (B2). Assertion without evidence (B8) is viewed in a negative light.

Assessors' perceptions of portfolio use in APL

The interviews show that three assessors (C, D and E) believe it is unfair to make decisions based solely on the portfolio. The other assessors take the opposite view. Assessor I said 'That is just the point. You are not influenced by other factors, such as knowing the person. There is a certain distance'. Although the majority see this as fair, they also provided the following reasons for requesting additional assessments: candidates failed to relate their experience to the learning objectives (assessor B); candidates lacked academic writing style (G and J); and candidates failed to clarify ambiguity (D, E, H and I). Assessor I did not see any benefits in organising an assessment conversation given its lack of objectivity 'even when there are multiple conversations with the same candidate'. The questionnaire results also show the difference between the assessors' opinion on the role of the assessment conversation. Items J11, J12 and J13, concerning the assessment conversation, show high standard deviations ('Interviews not necessarily based on the outcome of portfolio assessment were held', $SD = 1.73$; 'An assessment conversation always yields additional information for judgement', $SD = 1.04$; 'The assessment conversation led to adjustments to the final decision', $SD = 1.26$).

Part 2 of the questionnaire included items on portfolio relevance, usability and fairness. Table 3 gives an overview of the mean scores and standard deviations of the items derived from Van der Schaaf (2005), who distinguished three scales. The scale for relevance of the portfolio elements (items V1–V8) has a mean score of 4.23 and a standard deviation of .30; that for usability (V9–V15) has a mean of 3.66 and a standard deviation of .32; and that for fairness (V16–V18) has a mean of 3.92 and a standard deviation of .57. The mean scores are comparable with those of Van der Schaaf. Item V15 (It is clear to learners how the assessment results are used) shows an inconclusive result. A mean score of 3.80 indicates that most assessors find it clear to learners how the assessment results are used. A standard deviation of 1.03 however indicates that there are assessors who disagree with this perception.

Table 4 gives an overview of the mean scores and standard deviations on the items derived from Joosten-ten Brinke, Sluijsmans, and Jochems (2009). The assessors found the portfolio structure and the rating criteria clear, and portfolio assessment suitable for measuring candidates' prior learning. However, they indicated that the portfolio does not always provide sufficient information, and there are mixed perceptions with respect to the assessment conversation as a candidate's right, indicated by a high standard deviation ($SD = 1.75$).

The deciding factors in portfolio assessment

Assessor F found the arguments to be the most important part of the portfolio; three assessors (A, B and C) rated evidence as most important; the other six saw the combination of both argument and evidence as most important. The assessors judge

the portfolios as convincing in the event of overlap between job and curriculum content (A, C and I), the descriptions of experience being rendered in terms of the learning objectives (B and G), theoretical foundations to the argumentation (B), in terms of the duration, complexity and level of experience (C). Candidates having already reached the level of higher vocational education was found to be a supporting element, but not decisive. Moreover, the evidence (artefacts, reproductions, attestations and productions) in itself could be convincing, particularly scientific articles. Other characteristics mentioned included portfolio style, accuracy and structure.

Use of the rating criteria

The same general rating criteria were used for all educational domains. Table 6 presents the assessors' perceptions of the extent to which these criteria were met.

The assessors experienced few problems in using the criteria, although some (transferability, recent and higher vocational education) were difficult to interpret. The former is often rated as 'not transferable' or 'difficult to say'. Assessor F reported, 'That is a difficult criterion; candidates like to study programmes that will help them acquire theory related to their practice. So at a stretch there is some transfer, but it remains difficult to judge'. Remarkably, some of the assessors asked, 'What is recent?' This often depends on the domain. Assessor A said broadly, 'It was not made before World War II, so I regard it as recent'. The word is usually interpreted as 'learned in the past and still in use'.

The definition of 'at least at vocational higher education level' is also variable, although the assessors attest to being aware of its written definition in the rating criteria. Three assessors (D, I and J) explain that this is more intuitive than rational. In an attempt to describe vocational education level, assessors A, B, C, G and I mention different candidates' characteristics, such as holding a position normally only reached through having completed education at that level or higher; already having a certificate at that level; being an independent problem solver; working with theoretical models; or using theoretical portfolio argumentations. In addition, the nature of the evidence, its similarities with the curriculum, the CV, the writing style and the linguistic usage all help to decide whether vocational higher education level has been achieved. Two assessors (B and H) mentioned additional criteria: academic writing style and theoretical foundation.

Table 6. General rating criteria and assessors' perception of the extent to which these are met in candidates' portfolios.

| The evidence is ... | Most of the time | Sometimes | No | I don't know |
|-------------------------------------|------------------|-----------|----|--------------|
| Relevant | 8 | 2 | | |
| Valid | 6 | 3 | | 1 |
| Authentic | 8 | 1 | | 1 |
| Recent | 4 | | | 3 |
| Transferable | 1 | 1 | | 3 |
| Sufficient | 7 | | | 1 |
| At least at higher vocational level | 6 | | | 3 |

Discussion and conclusion

The purpose of this study was to gain insight into assessors' approaches to APL portfolio assessment to provide guidelines for its appropriate use. Although assessors' perceptions in a previous study were found to be positive (Joosten-ten Brinke, Sluijsmans, and Jochems, 2009), how they actually deal with portfolio assessment remains an issue of interest. We therefore conducted interviews with 10 assessors and administered a questionnaire to investigate how they reach their final decisions, how they perceive the portfolio use in terms of fairness, usability and relevance, what the deciding factors are in portfolio assessment and how they use the rating criteria.

First, it appears that the decision-making process is identical for assessors in the same domain, but differs for those in different domains. Assessors in all domains can opt to request additional assessments; in this study, the assessors using the same arguments in their decision-making processes as the assessors in Baume, Yorke, and Coffey's study, such as 'candidates should address criteria explicitly to achieve a positive judgement', 'applications must be made in addition to evidence provision', 'assertions must be clear'. Further, the decision-making is influenced positively if evidence is labelled in a proper way and the portfolio is logical and conveniently arranged.

The positive mean scores in the second part of the questionnaire show that the assessors found portfolio assessment relevant, fair and useful, while the third part also shows that assessors tend towards a supportive perception of the assessment. The low scores on three statements ('APL candidates have the right to an assessment conversation', 'Interviews not necessarily based on the outcome of portfolio assessment were held' and 'An assessment conversation always yields additional information for judgement') may indicate that the assessment conversations were not superfluous, but rather provided additional information. An assessment conversation, however, is not a candidate's right; it simply provides another avenue for assessors in gathering information.

The questionnaire showed that assessors perceive the possibility to request additional assessments not only as fair but also helpful, especially when the portfolio is ambiguous or shows a lack of academic writing skills. However, this last argument may be negated given that academic writing skills and style as well as appropriate linguistic usage and theoretical argumentation is a necessary characteristic of a vocational higher education level. The assessors deemed portfolio assessment suitable for judging the level of prior learning, but considered the criterion of a minimum higher vocational education level difficult. Thus, in future practice it might be useful to translate this criterion into the characteristics mentioned earlier, such as holding a position normally only reached through having completed education at that level, or having already worked with theoretical models. The observed need for assessors to independently interpret such criteria will then conceivably reduce, and transparency for the candidates will increase. The finding that differing interpretations exist in APL portfolio assessment is in line with the work of Van der Schaaf and Stokking (2008) and supports the interpretative approach of Johnston (2004). A debate between co-assessors about their interpretations is necessary.

Several other criteria also require elucidation. Based on differing definitions of the term 'recent', we concluded that this criterion is not clear. Surprisingly, however, the questionnaire results do not reflect this: the item 'the rating criteria are clear' showed general assent. This may indicate that assessors find these criteria

transparent, but that they interpret them differently. These and the interview results confirm conclusions from Baume, Yorke, and Coffey (2004), Tigelaar et al. (2005), and Van der Schaaf and Stokking (2008) that although assessors use rating forms and criteria, they may not, in fact, use them appropriately. As Baume, Yorke, and Coffey reported, the criteria are often interpreted differently. It is important to communicate the interpretation of the criteria with the co-assessors and candidates. The interviews, however, give some insight into the assessors' understanding and use of the criteria, while in turn rendering them more transparent for the assessors and thus also the candidates. A disciplinary difference between the understandings of the criteria is obvious.

Although the criteria are largely derived from literature, refining specific elements according to domain appears to have been useful. The definition of recent evidence could in Dutch law, for example, specify that 'the evidence should meet present regulations'. For cultural science, it may be that 'the evidence is still related to present work activities'. In both cases, the evidence may be older than 10 years but still in use. In line with the research of Sadler (2005) and Johnston (2004), we conclude that in any event, it should be clear how the criteria are to be used in relation to the final decision. Moreover, the interviews suggest that holistic portfolio judgement overruled any analytical judgement regarding separate criteria. Johnston (2004) also mentions the issue of holistic versus elementary judging as an important issue. The convincing portfolio characteristics may play a role in this, though they directly relate to the general criteria. For example, the assessors found portfolios convincing when the candidate's job overlapped the curriculum, and when the duration, complexity and level of experience were deemed adequate. This seems to support the idea that the criteria should be as specific, not as general, as possible.

An important issue in APL portfolio assessment is that assessors may request further assessment in cases where they judge this to be necessary. This brings about more efficient assessment than situations in which the final decision is taken based solely on a portfolio. Still, our interview results show that 70% of the assessors deem making decisions on a portfolio basis alone as fair, with some arguing that assessment conversations decrease objectivity. This contrasts with the quality framework for competency assessment programmes put forth by Baartman et al. (2006) which includes quality criteria such as reliability and validity, but takes also alleges that a professional judgement is more important than objective, standardised measures. One criterion in their framework – reproducibility of decisions – means that assessment quality will increase if different perspectives are used to reach a final decision. In the present study, this view is not supported by assessor I.

With regard to the set up of this study, its main shortcoming was the small number of interviewed assessors. Therefore, we only could report descriptive statistics. This restricted size, however, enabled us to collect more in-depth data on their approaches to APL portfolio assessment than would otherwise have been feasible, and therefore could be used to improve new methods of portfolio assessment. Another issue concerns the methodology of this study. We chose to start with the interviews and to give the questionnaire after the interview, because interviews provide the opportunity to gain more qualitative information about the assessors' decision-making process. The small number of assessors does not deliver reliable and informative quantitative information. Further, presenting the questionnaire before the interview could have biased the interviews results. In retrospect, however, presenting the questionnaires before the interviews could have led to more in-depth interviews.

This study has shown that although the assessors' approaches to portfolio assessment differed and the need to interpret criteria was observed, the majority of assessors found the process fair. Moreover, this fairness will continue to increase in line with growing criteria transparency. Additional research on assessors' perceptions and approaches is still needed. However, if we accept the existence of interpretation in portfolio assessment and trust in assessors' competences, portfolio assessment in APL will enjoy a positive future.

Acknowledgements

We like to thank the anonymous reviewers for their useful comments.

Notes on contributors

Desiree Joosten-ten Brinke studied applied educational science at the University of Twente, Enschede, the Netherlands. Since 1990, she has worked as an assistant professor at the Educational Technology Expertise Center of the Open University of the Netherlands. Her research interests encompass assessment of prior learning and computer-based assessment.

Dominique M.A. Sluijsmans is an assistant professor at the Educational Technology Expertise Centre of the Open University of the Netherlands and a lector (associate professor) at the Professional University of Arnhem-Nijmegen, The Netherlands. Her research interests encompass sustainable assessment and lifelong learning, with a focus on the design of sustainable assessment and the involvement of learners through self- and peer-assessment.

Wim M.G. Jochems is the dean of the Eindhoven School of Education and a full professor in educational innovation at Eindhoven University of Technology, Eindhoven, the Netherlands. His research interests focus on teacher training and professional learning, especially in the domain of science and technology education.

References

- Baartman, L.K.J., T.J. Bastiaens, P.A. Kirschner, and C.P.M. Van der Vleuten. 2006. The wheel of competency assessment: Presenting quality criteria for Competency Assessment Programmes. *Studies in Educational Evaluation* 32: 153–77.
- Barrett, H.C. 2003. Electronic portfolio. In *Educational technology: An encyclopedia*, ed. A. Kovalchick and K. Dawson, 271–6. Santa Barbara, CA: ABC-CLIO.
- Baume, D., and M. Yorke. 2002. The reliability of assessment by portfolio on a course to develop and accredit teachers in higher education. *Studies in Higher Education* 27: 7–25.
- Baume, D., M. Yorke, and M. Coffey. 2004. What is happening when we assess, and how can we use our understanding of this to improve assessment? *Assessment & Evaluation in Higher Education* 29: 451–77.
- Bjørnavold, J. 2001. Making learning visible: Identification, assessment and recognition of non-formal learning. *European Journal* 22: 24–32.
- Bloor, M., and C. Butterworth. 1990. The accreditation of prior learning on in-service education courses for teachers. *Aspects of Educational Technology* 22: 77–82.
- Butterworth, C. 1992. More than a bite at the APEL. *Journal of Further and Higher Education* 16: 39–51.
- Clarke, J.B., and J. Warr. 1997. Academic validation of prior and experiential learning: Evaluation of the process. *Journal of Advanced Nursing* 26: 1235–42.
- Colley, H., P. Hodgkinson, and J. Malcolm. 2002. *Non-formal learning: Mapping the conceptual terrain*. A consultation report. Leeds: University of Leeds Lifelong Learning Institute.
- Cretchley, G., and J. Castle. 2001. OBE, RPL and adult education: Good bedfellows in higher education in South Africa? *International Journal of Lifelong Education* 20: 487–501.

- Delandshere, G., and A.R. Petrosky. 1998. Assessment of complex performances: Limitations of key measurement assumptions. *Educational Researcher* 27: 14–24.
- Dierick, S., and F.J.R.C. Dochy. 2001. New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation* 27: 307–29.
- Driessen, E.W., K. Overeem, J. Van Tartwijk, C.P.M. Van der Vleuten, and A.M.M. Muijtjens. 2006. Validity of portfolio assessment: Which qualities determine ratings? *Medical Education* 40: 862–6.
- European Union. 2008. *Recommendation of the European parliament and of the council on the establishment of the European Qualifications Framework for lifelong learning*. Brussels: European Union.
- Evans, N. 1993. *Making sense of lifelong learning: Respecting the needs of all*. London: RoutledgeFalmer.
- Harteis, C., and H. Gruber. 2008. Intuition and professional competence: Intuitive versus rational forecasting of the stock market. *Vocations and Learning* 1: 71–85.
- Heller, J.I., K. Sheingold, and C.M. Myford. 1998. Reasoning about evidence in portfolios: Cognitive foundations for valid and reliable assessment. *Educational Assessment* 5: 5–40.
- Johnston, B. 2004. Summative assessment of portfolios: An examination of different approaches to agreement over outcomes. *Studies in Higher Education* 29: 395–412.
- Joosten-ten Brinke, D., D.M.A. Sluijsmans, S. Brand-Gruwel, and W.M.G. Jochems. 2008. The quality of procedures to assess and credit prior learning: Implications for design. *Educational Research Review* 3: 51–65.
- Joosten-ten Brinke, D., D.M.A. Sluijsmans, and W.M.G. Jochems. 2009. The quality of Assessment of Prior Learning (APL) in university programmes: Perceptions of candidates, tutors and assessors. *Studies in Continuing Education* 31, no. 1: 61–76.
- Linn, R. 1994. Performance assessment: Policy promises and technical measurement issues. *Educational Researcher* 23: 4–14.
- Messick, S. 1995. Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice* 14: 5–8.
- Michelson, E., and A. Mandell, eds. 2004. *Portfolio development and the assessment of prior learning: Perspectives, models, and practices*. Sterling, VA: Stylus.
- Moss, P.M. 1994. Can there be validity without reliability? *Educational Research* 23: 5–12.
- Osman, R. 2006. RPL: An emerging and contested practice in South Africa. In *Re-theorising the recognition of prior learning*, ed. P. Andersson and J. Harris, 51–76. Leicester: National Institute of Adult Continuing Education.
- Sadler, D.R. 2005. Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education* 30: 175–94.
- Shapiro, J. 2003. Exploring teachers' informal learning for policy on professional development. Unpublished PhD diss., RAND. <http://www.rand.org/publications/RGSD/RGSD174/> (accessed April 19, 2004).
- Simon, H.A. 1957. *Models of man*. New York: Wiley.
- Sluijsmans, D., G. Straetmans, and J.J.G. Van Merriënboer. 2008. Integrating authentic assessment with competence-based learning: The protocol portfolio scoring method. *Journal of Vocational Education and Training* 60: 157–72.
- Tigelaar, D.E.H., D.H.J.M. Dolmans, I.H.A.P. Wolfhagen, and C.P.M. Van der Vleuten. 2005. Quality issues in judging portfolios: Implications for organizing teaching portfolio assessment procedures. *Studies in Higher Education* 30: 595–610.
- Trowler, P. 1996. Angels in marble? Accrediting prior experiential learning in higher education. *Studies in Higher Education* 21: 17–30.
- Van der Schaaf, M.F. 2005. Construct validation of teacher portfolio assessment. PhD diss., University of Utrecht, Utrecht, the Netherlands.
- Van der Schaaf, M.F., and K.M. Stokking. 2008. Developing and validating a design for teacher portfolio assessment. *Assessment & Evaluation in Higher Education* 33: 245–62.
- Van der Schaaf, M.F., K.M. Stokking, and N. Verloop. 2005. Cognitive representations in raters' assessment of teacher portfolios. *Studies in Educational Evaluation* 31: 27–55.

Copyright of Assessment & Evaluation in Higher Education is the property of Routledge and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.