



**REVIEW**

**August 2003**

---

**What is known about  
successful models of  
formative assessment for  
trainee teachers during  
school experiences and  
what constitutes effective  
practice?**

**This review is supported by the Teacher Training  
Agency (TTA) to promote the use of research  
and evidence to improve teaching and learning**

*Review conducted by the Assessment of ITT Students Review Group*

# **AUTHORS**

This review was undertaken by the Centre for Research into Education and Teaching (CREATe) within the School of Education at Anglia Polytechnic University (APU). It was conducted following the procedures for systematic reviews developed by the EPPI-Centre and in collaboration with Dr Nicholas Houghton and Professor Diana Elbourne, and with help from other members of the EPPI-Centre education team.

## **Principal authors**

Professor Janet Moyles, School of Education, APU  
Richard Yates, School of Education, APU

# **REVIEW TEAM MEMBERSHIP**

Alan Bradwell, School of Education, APU  
Tim Cooper, School of Education, APU  
Alison Feist, Anglo-European School, Ingatestone, Essex  
Ann Lahiff, School of Education, APU  
Jenny Lansdell, School of Education, APU  
Jill Martin, Anglo-European School, Ingatestone, Essex  
Professor Janet Moyles, School of Education, APU  
Andy Scott-Evans, St Thomas of Canterbury, Essex  
Alison Shilela, School of Education, APU  
Douglas Stuart, School of Education, APU  
Richard Yates, School of Education, APU

# **ADVISORY GROUP**

Professor Diana Elbourne, EPPI-Centre, Institute of Education  
Dr Nicholas Houghton, EPPI-Centre, Institute of Education  
Paul Moses, Teacher Training Agency (TTA)

# **ACKNOWLEDGEMENTS**

The authors would like to thank the above and members of the EPPI-Centre team for their guidance and support during the conduct of the review. They would also like to thank Maxine Levy for her administrative support.

## Funding

This review was supported by the Teacher Training Agency (TTA). The TTA were represented as an affiliated member of the review team and could be subject to evaluative comments in relation to evidence related to policies it has and is promoting.

## LIST OF ABBREVIATIONS

APU	Anglia Polytechnic University
B. Ed	Bachelor of Education
BEI	British Education Index
BIDS	Bath Information and Data Services
CEP	Career Entry Profile
CERUK	Current Educational Research in the United Kingdom
DfES	Department for Education and Skills
EPPI-Centre	Evidence for Policy and Practice Information and Co-ordinating Centre
ERIC	Educational Resources Information Centre
ESRC	Economic and Social Research Council
GTP	Graduate Teacher Programme
HE	Higher Education
HEI	Higher Education Institution
ICT	Information and Communication Technology
IT	Information Technology
ITT	Initial Teacher Training
NQT	Newly Qualified Teacher
NUT	National Union of Teachers
OfSTED	Office for Standards in Education
PDP	Professional Development Profile
QCA	Qualifications and Curriculum Authority
QTS	Qualified Teacher Status
RTP	Registered Teacher Programme
SCITT	School Centred Initial Teacher Training
SCRE	Scottish Centre for Research in Education
SOSIG	Social Science Information Gateway
TTA	Teacher Training Agency
UK	United Kingdom
URT	User Review Team
US	United States
USA	United States of America

This report should be cited as: Moyles J, Yates R (2003) What is known about successful models of formative assessment for trainee teachers during school experiences and what constitutes effective practice? In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

© Copyright

Authors of the systematic reviews on the EPPI-Centre Website (<http://eppi.ioe.ac.uk/>) hold the copyright for the text of their reviews. The EPPI-Centre owns the copyright for all material on the Website it has developed, including the contents of the databases, manuals, and keywording and data-extraction systems. The Centre and authors give permission for users of the site to display and print the contents of the site for their own non-commercial use, providing that the materials are not modified, copyright and other proprietary notices contained in the materials are retained, and the source of the material is cited clearly following the citation details provided. Otherwise users are not permitted to duplicate, reproduce, re-publish, distribute, or store material from this Website without express written permission.

# TABLE OF CONTENTS

SUMMARY .....	1
Background .....	1
Aims and review question .....	1
Results .....	1
1. BACKGROUND .....	3
1.1 Aims and rationale for current review .....	5
1.2 Definitional and conceptual issues .....	6
1.3 Policy and practice background .....	7
1.4 Research background .....	8
1.5 Authors, funders and other users of the review .....	8
1.6 Review questions .....	9
2. METHODS USED IN THE REVIEW .....	11
2.1 User involvement .....	11
2.2 Identifying and describing studies .....	11
2.3 In-depth review .....	16
3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS .....	18
3.1 Studies included from searching and screening .....	18
3.2 Characteristics of the included studies .....	21
3.3 Identifying and describing studies: quality assurance results .....	24
4. IN-DEPTH REVIEW: RESULTS .....	25
4.1 Selecting studies for in-depth review .....	25
4.2 Further details of studies included in the in-depth review .....	27
4.3 Synthesis of evidence .....	27
4.4 In-depth review: quality assurance results .....	33
4.5 Nature of actual involvement of users in the review and its impact .....	33
5. FINDINGS AND IMPLICATIONS .....	35
5.1 Summary of principal findings .....	35
5.2 Strengths and limitations of this systematic review .....	37
5.3 Implications .....	38
6. REFERENCES .....	41
6.1 Studies included in the descriptive map .....	41
6.2 Full bibliography after first screening stage .....	44
6.3 Other references used in the text of the report .....	60
APPENDIX 2.1: Search strategy for electronic databases .....	62
APPENDIX 2.2: APU Assessment review-specific keywording sheet .....	63
APPENDIX 3.1: Keywords of studies included in the in-depth review .....	64
APPENDIX 4.1: Studies excluded from data-extraction .....	88
APPENDIX 4.2: Details of studies included in the in-depth review .....	91

# SUMMARY

## Background

The focus underpinning this systematic review is the assessment of student teachers on Initial Teacher Training (ITT) courses. The background to the review is complex. Assessment has been the subject of much scholarly debate in recent years, the emphasis being on the processes, aims and outcomes of assessment. Comparative studies have been undertaken and have offered up alternative models for assessment, creating a new level of awareness and self-consciousness in the United Kingdom (UK). A number of issues dominate the subject and are central to this review. They include gender (Moyle and Cavendish, 2002), the emphasis on fair testing (Martin, 1997), the new idea of multiple assessments (Willis and Davies, 2002), the tension between standardisation and individualisation (Reyes and Isele, 1990), the demands for replicability (Hartsough 1998) and for non-discriminatory practice (Chambers and Roper, 2000), and so forth. A division, often awkward to uphold, between formative assessment and summative assessment has been drawn, the former being championed as a fairer, more personalised form of assessment than the latter (Adams, 1995).

## Aims and review question

The aim of this review is to explore and examine models for formatively assessing student teachers within the context of school-based experience, focusing on specific locations within the English-speaking international community and isolating studies published between 1987 and 2002. One of the aims of the review is to identify components of effective practice as well as effective models for formatively assessing student teachers. 'Effective' is defined in the review as pertaining to 'validity' (assessing the right criteria) and 'reliability' (assessing in a transparent, consistent and replicable way). A number of outcomes were expected: a systematic review and synthesis of existing research in the topic area; a database of evidence extracted from existing research; a descriptive map offering an orientation on the topic area; a small body of trustworthy and relevant studies; and, an indication of the gaps in research in the topic area.

## Review question

- What is known about successful models of formative assessment for trainee teachers during school experience and what constitutes effective practice?

## Results

The portfolio is identified as the most successful and effective formative assessment tool currently available that has been analysed and evaluated. Having followed the stages of the EPPI-Centre process, two studies, both focusing on portfolios, were

data-extracted and synthesised (Berg and Curry, 1997; Willis and Davies, 2002). Although the content and trustworthiness of the two studies was generally low, and indicative of the general lack of good research in this area of assessment, there were some useful findings. Portfolios were found to increase personal and professional growth and development, allow teachers to express themselves creatively, provide an unprecedented insight into the mind of the student teacher, create a strong bond between the assessed and the assessor, as well as to increase the confidence, reflective capacity and self-awareness of the trainee.

Whilst the review confirms the need for further investigation of, and experimentation with, portfolios, a few problem areas were identified. Firstly, excessive use of the portfolio may cause the education community to lose sight of pedagogical ability and focus attention overly on cognitive ability and clarity, as well as reflective capacity. Secondly, there is a tension between the time the portfolio takes to undertake properly and its ultimate worth (Reis and Villaume, 2002). This time-versus-worth dynamic may have been best explored using a controlled trial. Neither study, however, uses one. In conclusion, the review emphasises the general lack of quality research in the field and highlights the need for more focused research to be undertaken.

# 1. BACKGROUND

This chapter outlines the theoretical, policy, practice, and research background to the systematic review, including details of the authors and other users of the review, and the development and aims of the review question.

The focus underpinning this review is the assessment of initial teacher trainees on ITT courses. A definition of academic assessment that both reflects the understanding of the User Review Team (URT) and has been a useful working definition is 'the process of identifying collective aims for cumulative learning in terms of knowledge, skills, abilities, values, and/or attitudes and determining whether those aims have been met' (San Francisco State University, retrieved on 13 February 2003). Specifically, the review concentrates on how forms of assessment are devised and applied to trainee teachers<sup>1</sup> in the school-based element of ITT courses.

As a concept, assessment in education has been problematised in the UK. The excess of United States (US) literature on assessment suggests that this is also the case across the Atlantic (see section 3.2). Discourse about assessment, its processes, aims, and outcomes, has become increasingly prevalent (Fairbrother, 2000). Comparative studies of educational models have heightened awareness in the UK, offered up alternatives, and led to the UK system being brought into question, why it exists as it does and why it has its own particular characteristics (Bitner and Kratzner, 1995). A main intention of the review has, therefore, been to examine different models of assessment as applied to trainee teachers.

Some structuralist theories have also affected the way that assessment in general is perceived, offering, for example, the understanding of assessment where the female typically favours the *process* of instruction whilst the male favours the actual *product* of the final assessment (Stobart and Gipps, 1997). Theories such as this have contributed to a climate in which the fairness and effectiveness of different forms of assessment have been questioned and fed a growing belief that assessment should be more personalised, negotiable and adapted to the needs of individual students. With this process-versus-product/female-versus-male structuralist model in mind, the URT became interested in assessment models that emphasised both the *process* and the *product* of learning and evaluating. This was one of many interests that led the URT to its eventual focus on formative assessment.

Not all the theories and beliefs of recent years reflect the reality of assessment within education. If, as some structuralist theories propose, males do in general favour summative to formative assessment, one would expect attainment levels to be higher amongst boys than they are in light of the summative nature of most national examinations. As it stands, Higher Education Institutions (HEIs) are having problems with male recruitment and evidence suggests that this is due to under-achievement and the failure of males to strive (Chambers and Roper, 2000; Moyles and Cavendish, 2002). This may not necessarily be the result of changes in assessment

---

<sup>1</sup> 'Trainee teachers' is the term used by the TTA. However, in many studies, the term 'student teacher' is used. In this review, the two terms are used synonymously except where specific to individual studies.

models. However, the increase in emphasis placed on coursework and the high standard of performance of females in recent years may go some way to proving the validity of the process-versus-product/female-versus-male structuralist model.

Summative assessment has been criticised for its lack of flexibility, its inability to cover the whole course of instruction and for its standardised nature, potentially favouring one particular type of trainee teacher (Moyle and Cavendish, 2002).<sup>2</sup> Minority pressure, gender equity and political correctness have combined, resulting in an emphasis on fair testing. Fair testing has tended to be interpreted as opposing discrimination; of being non-discriminatory. Assessment is, however, by nature a system devised for the purpose of discriminating between people. Moreover, if assessment is to treat all candidates fairly and position all candidates at an equal advantage, it cannot treat everyone as though they were equal, if equal is used synonymously with 'the same' (i.e. from the same cultural, religious and social background, of one gender and of one age). Different people have different expectations of assessment.

The National Foundation for Educational Research (NFER) website, for example, features a definition of 'fair testing' that is flawed. The NFER concede that assessment is designed for the purpose of discriminating. However, the claim is made that 'the test should not discriminate between sub-groups on any basis unrelated to the purpose of the test' (retrieved on 16 January 2003). This is the contradiction. Teachers, like all professionals, are not generic. Trainee teachers from different cultures, religions and backgrounds, of different ages and of different sexes, are not all the same. A single fixed test cannot fairly assess such a varied body of participants. The NFER assert that 'there should be no unfair advantage to any sub-group based on attributes such as gender or ethnic group,' but an unfair advantage is extended to a particular type of candidate because of the failure of most assessment to distinguish between ages, genders and ethnicities.

Formative assessment promises a more personalised, tailored and negotiated form of assessment (Adams, 1995). This continual, fashioning form of assessment during ITT courses, which notes the ability of the candidate to absorb instruction and to measure vocational and professional suitability, ranks as an area worthy of inquiry. Summative assessment perceives the trainee teacher as a generic type; formative assessment, if applied effectively, favours the trainee teacher as an individual, assessing potential by way of individual merit, albeit currently against set criteria (TTA Standards for ITT: TTA, 2002). The interest of the URT lay essentially in formative assessment insofar as this could be isolated from other forms. This interest, in turn, meant that the focus of the review centred on the school experience of trainee teachers, for it is during their practice placements that trainee teachers mainly experience the continual, formative processes of assessment and evaluation, as opposed to those towards the end of an ITT course.

---

<sup>2</sup> For a systematic review on summative rather than formative assessment, see Harlen W and Crick RD (2002) A systematic review of the impact of summative assessment and test on students' motivation for learning. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

The main interests of the URT at the start of the review could be qualified as:

- formative assessment as a way of emphasising the *process* as well as the *product* of learning, teaching and evaluating
- the notion of ‘multiple assessments’ and the reality of achieving it
- the notion of ‘fair testing’

## 1.1 Aims and rationale for current review

### 1.1.1 Aims

The aim of this review is to explore and examine models for formative assessment of trainee teachers within the context of school-based experience. It focuses on specific locations within the English-speaking international community and studies published between 1987 and 2002. The period of 15 years, whilst not looking back so far that the studies describe an outdated system, offered a manageable timeframe. Though it is not a specific aim of the review to recommend models of assessment, we attempted to identify components and patterns of effective practice. In preferring evaluative studies to descriptive ones, certain models of assessment were deemed more ‘effective’ than others. ‘Effective’, in the context of assessment models, can be understood as double-pronged and defined as (i) the ‘validity’ of that which is being assessed and (ii) the ‘reliability’ of the assessment model in measuring the student teacher (Martin, 1997).

The review question was influenced by the concept of reflection (Schön 1983) and developed in the wake of the Career Entry Profile (CEP), a portfolio-style document that monitors professional development, identifies competencies, and supports Newly Qualified Teachers (NQTs) towards Statutory Induction. Unlike the UK, most of the US education systems depend heavily upon portfolios. The UK Government has now committed itself to their use, but research is still needed into whether they have been operationalised effectively.

We expected the following outcomes from our literature review:

- a systematic review of existing research on formative assessment of trainees in ITT
- a database of evidence extracted from existing studies reporting empirical research
- a descriptive map offering an orientation on assessment in ITT with specific reference to trainees’ school-based experiences and formative models
- a range of studies targeted at different audiences, such as practitioners and policy-makers

- An indication of gaps in the literature and research evidence in the areas of formative assessment of ITT trainees which require further examination.

At the end of the two-and-a-half month, truncated EPPI-Centre review process, the URT aimed to have an informed and informative literature review, bringing tried and tested models of assessment to the fore and, in the process, noting areas of inquiry that have so far eluded research. The studies the URT sought tended, therefore, towards the evaluative rather than the descriptive. Hence the URT's interest in effective practice, how it is defined and what its components are.

### **1.1.2 Rationale**

Before undertaking any new research, policy or practice, it is advisable first to be informed about the research, policies and practices that already exist in relation to the topic. Such information may be known by experts in the field or may have already been summarised within a literature review, but these have not traditionally been based upon systematic methods to ensure substantial and unbiased searching and processing of potentially relevant studies. The URT, therefore, was committed to using a systematic approach. For this reason, this review was allied to the EPPI-Centre review procedures. Systematic reviews are also in alignment with the general move in the UK towards evidence-informed policy and practice. In systematising the review, the URT synthesised the results of primary research and used explicit and transparent methods. The review presents research that is accountable, replicable, and updateable, and involves users.

## **1.2 Definitional and conceptual issues**

There are theoretical issues about assessment that informed the conceptual framework of the in-depth review. A number of studies expounded theories and ideas that were characteristic of a general move towards 'authentic assessment' (Willis and Davies, 2002; p 18). Criteria for assessing trainee teachers are often expressed as standards or competencies. However, checklists that define good teaching in behaviourist terms have come under continual criticism (Martin, 1997). Checklists invariably include high-inference as well as low-inference criteria. High-inference criteria are often subject to the 'halo effect', in which assessors mark according to personal likes (Reyes and Isele, 1990). To avoid this, competencies need to be expressed in behaviourist terms. Developing reliable and valid low-inference criteria for assessing teaching has therefore become something of a priority, as has the need to discover an assessment procedure that assesses a 'replicable finding' (Hartsough *et al.*, 1998).

Implicit within this area of assessment, therefore, is the friction between (i) the development of low-inference criteria that express competence in behaviourist terms because they allow an objective and replicable assessment, and (ii) the development of high-inference items that are liable to make the assessment more subjective and untraceable, but at the same time do not consign the art of teaching to a few simple behaviours.

Two models of assessment come to the fore: the scientific measurement model and the judgmental model (Martin, 1997). Martin raises a question that the User Review Team approach implicitly in the review question: should assessment influence as well as reflect instruction and learning? Formative assessment individualises the assessment programme to the extent that it offers the opportunity for the different degrees of help that are necessary in raising all trainee teachers to the same level of competence to be realised. The intention of the URT was to identify any 'best' models for achieving this, whilst at the same time identifying a means of assessing trainees that is itself more negotiable and personalised.

The ultimate reason for all assessment theorising and development is to 'minimise the number of wrong classifications' (Martin, 1997; p 339). Another issue is again outlined by Martin: the problem of 'MacNamara's Fallacy'. This refers to the common shortcoming of many assessment models, namely their failure to make the important *measurable* rather than the measurable *important*.

A number of different types of assessment exist. Ipsative assessment is a type of assessment to some degree borne of the current climate. 'Ipsative' refers to a way of assessing in which achievement is measured according to past reference and performance. It is cumulative and developmental. It is the academic equivalent of the 'personal best' of an athlete. Ipsative assessment, in the early stages of the review at least, was an additional interest of the URT.

Assessment has many different functions: selection and grading, diagnosis and remediation, motivation, and recording and reporting (Mahoney and Knox, 2000). In this in-depth review, the URT aimed to identify a formative assessment model or instrument that made all of these functions concurrently possible.

### 1.3 Policy and practice background

The current backdrop for ITT assessment comprises a number of agents. There are the Skills Tests that must be passed in order for the DfES to grant Qualified Teacher Status (QTS) to the trainee. There is the Career Entry Profile (CEP: TTA, 2002), which has been developed mainly for the purpose of integration and continuous professional development during Statutory Induction of trainee teachers (arrangements for which were introduced in 1999 and are now subject to Into Induction 2002: TTA, 2002). There is Qualifying to Teach (TTA, 2002), a document that contains standards that are essential to the assessment of trainees. There are also the recommended, though not compulsory, Professional Development Profiles (PDPs), which are part of a general move towards portfolio-based assessment.

At APU, for example, the PDP is used 'as a reflective journal in which [trainees] can record professional progress and set targets for future development' (APU, 2003 Undergraduate Prospectus; p.94). This profile contributes to the CEP and, to an extent, utilises a formative model of assessment. The URT sought to establish whether other such profiles or portfolios have proven successful, and, if so, identify successful formats and benefits. The URT also wanted evidence as to whether a hybrid model that comprised national standards and a portfolio component had been tried and tested in other countries. In recent years, QTS has been awarded for the

ability to meet a set of competencies or standards (for example, the DfES Circulars 9/92, 14/93, and 4/98). The URT was curious as to whether there existed within current research a way that could perhaps, through portfolios, communicate more creatively the abilities that these checklists assess.

## 1.4 Research background

The research background to formative assessment is particularly limited in relation to the review question, hence the poor yield of studies in terms of number and quality. Chapter 3 describes the field of assessment in ITT.

## 1.5 Authors, funders and other users of the review

The URT comprised and represented a cross-section of the community involved in ITT, covering almost all the potential user groups. Some potential users, such as students, NQTs, Local Education Authority (LEA) representatives, school governors, and parents, were not included due to the limited time available to assemble the URT. Ironically, students were unavailable at the time the review began owing to the pressures of their work placements. Those involved in the review and the nature of their involvement in ITT are listed below. With this constitution, the ultimate intention was to create an interface useful to all possible user groups. None was under-represented in the construction of the question and everyone participated in the discussions to formulate the review question.

The review was funded by the Teacher Training Agency (TTA).

The User Review Team (URT) comprised the following members:

- Alan Bradwell, Education Librarian and academic liaison between ITT students and the library at Anglia Polytechnic University (APU)
- Tim Cooper, a Research Administrator at APU
- Alison Feist, a supervising teacher for ITT students at a local maintained secondary school
- Ann Lahiff, the learning and teaching advisor in the School of Education at APU, concerned with the professional development of teachers and trainers across the entire age range, from pre-school and early years to post-compulsory education and training
- Jenny Lansdell, Deputy Dean of the School of Education at APU, responsible for both primary and secondary ITT and the assessment of it
- Jill Martin, Deputy Head of a local maintained secondary school, in charge of ITT partnerships with higher education institutions (HEIs)

- Professor Janet Moyles, Director of Research in the School of Education at APU
- Andy Scott-Evans, Deputy Head of a local Church of England (Voluntary Aided) junior school, in charge of ITT partnerships with HEIs
- Alison Shilela, Associate Dean of the School of Education at APU
- Douglas Stuart, a Researcher at APU
- Richard Yates, a Researcher at APU

Advisory members to the URT were as follows:

- Professor Diana Elbourne from the EPPI-Centre
- Dr Nicholas Houghton from the EPPI-Centre
- Paul Moses from the TTA

## 1.6 Review questions

The review question agreed by the URT is:

- What is known about successful models of formative assessment for trainee teachers during school experiences and what constitutes effective practice?

In effect, the question is two-tiered. The intention was first to identify all those studies that focused on models of formative assessment and secondly, during the in-depth review, to determine which of them was replicable, valid and reliable, and indicated effective practices. Objectives that are implicit in the question are explored in section 4.3.

### 1.6.1 Definitions

The definitions of the four key components of the question are outlined below.

#### **Models**

By this, we refer to the form taken by the type of assessment related in the reviewed study. 'Models' in this context has *no* connection with any form of representation (e.g. 'modelling' or 'model-making'). In many ways, it covers everything about how the assessment is *carried out*, including its variables, its physical manifestations (e.g. portfolios) and particularly its characteristics. A model is a *system* or a *process*.

#### **Formative assessment**

By this, we refer to the diagnostic use of assessment, the continual process of application and alteration, of feedback between student teachers, teachers and

tutors over the course of instruction as opposed to after the period of instruction, requiring no overall or retrospective judgement. It refers to assessment that both *forms* and *informs* the progress of the student.

### ***Trainee teachers***

This term is intended to be a synonym for *student teachers*; its usage is mostly for the sake of parity and alignment with the TTA and the term used in some ITT courses. It is also favoured because it implies the onsite *training* of the student teacher rather than their HEI-based *education*. Whether these two terms can be comfortably separated is a topic for extended discussion beyond the scope of this review.

### ***School-based experience***

By this, we refer to the practice of school placement experience, field experience (US), or practicum (Australia), that is, the sustained period in which the trainee attends a school to perform actual or simulated teaching.

## 2. METHODS USED IN THE REVIEW

This chapter details each stage of the review to ensure that it is accountable, replicable and updateable.

### 2.1 User involvement

#### 2.1.1 Approach and rationale

The User Review Team was involved in the process for a number of reasons. Firstly, it was crucial to the success of the review that it raised questions that users wanted answering rather than questions that may have had more limited practical application or relevance. Secondly, in undertaking a process in which the review question is (at least in the early stages of research) continually changing and is informed and refined by the research itself, it was important to have valid interest groups contributing to the character of the question. User involvement, therefore, was one form of quality assurance.

#### 2.1.2 Methods used

User involvement was operationalised in the following ways. First, the screening stage when the inclusion and exclusion criteria were applied to full texts took place in URT face-to-face meetings. These meetings were arranged so that as many members of the URT as possible were present. The meetings were conducted weekly, for the sole purpose of screening texts. Studies that were not screened during these meetings were sent by post to paired reviewers. Second, users were also involved in the keywording and data-extraction stages of the review. During keywording, much of the data, whilst they originated from and were generated by members of the URT, had to be input into the EPPI-Centre website by the Administrator because of lack of time for URT members to be familiarised with accessing the website. That said, all but two members of the URT received training in inputting keywording data and data-extraction using the EPPI-Centre website facilities directly and would have been able to input data had more time been available. Prior to submission to the Teacher Training Agency (TTA), the draft review was distributed to all members of the URT for preliminary quality assurance.

### 2.2 Identifying and describing studies

#### 2.2.1 Defining relevant studies: inclusion and exclusion criteria

To ensure that only papers focusing on the review question were selected for mapping, an explicit list of inclusion and exclusion criteria were developed to exclude inappropriate papers.

- *Included (1)*: Papers that were published after 1987. *Rationale*: The period of 15 years, whilst not looking back so far that the papers described an outdated system, promised a manageable timeframe within the short period available for this review.
- *Included (2)*: Papers that were conducted within Europe, the US, Australia and English-speaking Canada. *Rationale*: Achieving consistency in ideology, culture and practice (i.e. comparing like with like) was one of the foremost concerns of the URT. The URT determined that those from other English-speaking areas of the world, e.g. Hong Kong, would be too culturally and educationally different to be either informative or comparative.
- *Included (3)*: Papers written in English. *Rationale*: The timescale was limited and the first (and main) language within the URT was English.
- *Excluded (4)*: Papers that focused on the CEP. *Rationale*: The URT was encouraged by the TTA to exclude these.
- *Excluded (5)*: Papers that focused on the assessment of students or teachers but not student teachers. *Rationale*: The population focus of the review was student teachers or 'trainee teachers' (the TTA's term for student teachers).
- *Excluded (6)*: Papers that focused on an 'evaluation' or 'assessment' of the ITT system rather than the forms of assessment used by ITT courses to assess trainees. *Rationale*: The search terms ASSESS\* and EVALUAT\* are neither synonymous nor consistent, especially in American usage, therefore papers that did not deal with the assessment process were excluded.
- *Excluded (7)*: Papers that focused on the school experience of the student teachers but not the assessment of it. *Rationale*: There was a genus of papers in which the school experience was the educational setting, but this did not presuppose any focus on assessment. These papers were excluded.
- *Excluded (8)*: Papers that focused more on mentoring and supervising than assessment. *Rationale*: There was a whole corpus of papers that took these two areas as their main focus. This did not presuppose any focus on assessment. The TTA had also encouraged the URT not to concentrate on mentoring. These papers were excluded.
- *Excluded (9)*: Papers that focused on pre-student rather than pre-service teachers. *Rationale*: The population of the review was student teachers or 'trainees'.
- *Excluded (10)*: Papers that focused on portfolios, but not as assessment tools. *Rationale*: Although portfolios, coming under the rubric of self-assessment, are a tool for formative assessment, some portfolio-focused papers did not explore them as assessment instruments. These papers were excluded.
- *Excluded (11)*: Papers that dealt with reflective practice but not the assessment of it. *Rationale*: There was a whole corpus of papers that took reflection as their

main focus, but this did not presuppose any focus on the assessment of reflection. These papers were excluded.

- *Excluded (12):* Papers that were digests from the ERIC database. *Rationale:* Firstly, these texts were generally reviews and were therefore secondary research. Secondly, regarding the length of the review process and the financial constraints, these texts were deemed unobtainable. These papers were excluded.
- *Excluded (13):* Papers that were duplicates. *Rationale:* Allowance had to be made for human error during the searching, screening and data-entry stages of the review process. Identifying duplicates early on was one way in which the budget could be more effectively utilised. These papers were excluded.

### 2.2.2 Identification of potential studies: search strategy

Initially relevant sources were searched using a controlled vocabulary which placed our searches firmly within the topic area of assessment. Due to the time constraints, electronic databases were used for initial searches. Handsearching was not undertaken. The databases, Ask ERIC, ERIC and BEI via BIDS Education, and Education Line, were searched. Government organisations and agencies, the TTA, QCA and DfES, supplemented these sites, but from them no extra relevant papers were uncovered on assessment in ITT.

A controlled vocabulary was employed to attempt an initial search that was as far-reaching and comprehensive as possible, using variations on different search terms (which were often truncated) such as STUDENT TEACH\*, PRESERVICE TEACH\*, BEGINNING TEACH\*, and TESTING, EVALUATION and ASSESSMENT. The more complex and numbered the combination of search terms, the smaller the yield of the search. The two largest searches, which between them covered all of the studies yielded from the earlier, smaller searches, became the basis for the literature review.

The two large searches were as follows: (1) An Ask ERIC search using the search terms STUDENT TEACHER EVALUATION and specifying post-1987 publications. This search deliberately used an Americanism because ERIC is an American database; the term EVALUATION was therefore preferential to ASSESSMENT. (2) The second large search was through BEI using the same search term STUDENT TEACHER EVALUATION, because the result of supplanting the search term EVALUATION with ASSESSMENT was a very low yield. These yields suggest that the bulk of papers on assessment are American rather than British. The searches were then extended to allow for the possibility that vital papers had been overlooked. This was achieved through smaller additional searches. These searches included two through Ask ERIC, the first comprising the search terms PORTFOLIO, ASSESSMENT and STUDENT TEACHER, the second comprising TEST, STUDENT TEACHER, TRAINING and ASSESSMENT. One more search was assimilated into the total of searches, this time using the Education Line database, the search term being ASSESSMENT.

Following initial searches, the URT determined that it would be a worthwhile measure to take the precaution of formulating a comprehensive search strategy that

would highlight any missing key papers. This was completed by breaking the question into four sections. Using ERIC via BIDS, all the conceptual synonyms that are used to describe each particular field or 'family' within UK, US, Australian and Canadian education were determined. These were entered and combined using BIDS within their families (using the Boolean operator OR) and then combined together (using the Boolean operator AND), so that each paper, were it to show up in this particular yield, would feature a conceptual synonym from each of the four families that the question is composed of. Appendix 2.1 shows this search strategy.

Other internet sites were searched with similar intentions in mind. These sites comprised CERUK, Regard, SOSIG, SCRE, ESRC, Child Data, SARA, Zetoc, OPAC and COPAC.

### **2.2.3 Screening studies: applying inclusion and exclusion criteria**

Following searching citations were screened on the basis of abstracts and titles and included or excluded according to inclusion and exclusion criteria 1 to 11. These papers were then put onto a database (DB1). A second stage of screening was then carried out on all papers in database 1. This process involved scanning papers again and including or excluding them according to inclusion and exclusion criteria 1 to 13. The final two criteria were devised at the stage in the process when full texts were to be acquired.

### **2.2.4 Characterising included studies**

For articles meeting the inclusion and exclusion criteria the full text articles were obtained and keyworded using the *EPPI-Centre Core Keywording Strategy* (version 0.9.6: EPPI-Centre, 2003). The EPPI-centre Core Keywording strategy included the following categories:

- the origin of the report
- the publication status of the report
- other linked reports
- the language of the report
- the country of the report
- the main topic focus of the study
- the programme name related to the study
- the population focus of the study
- the educational setting of the study
- the type of study the report describes

Additional keywords were developed to create an APU Assessment review-specific keywording sheet. On this sheet, it was necessary to gauge how much the study concentrated on a number of aspects:

- the school experience
- assessment rather than the assessors and assessed

- the degree to which it focused on formative rather than summative assessment

Keyword categories included the following:

- the type of assessment on which the study focuses
- the main aspect within assessment on which the study focuses
- the type of learner on which the study focuses
- the type of teaching staff on which the study focuses
- the phase of ITT on which the study focuses
- the type of research described in the report

Further inclusion criteria 14 to 18 were developed before the process of keywording and applied during keywording. These also informed the APU Assessment review-specific keywording sheet. A Criterion 19, being the cut-off point time limit for obtaining full copies of papers for keywording, was applied after criteria 14 to 18.

Inclusion and exclusion criteria 14 to 18 are as follows:

- Criterion 14: *Included*: Only papers that reported an empirical study (see section 3.2). *Rationale*: On account of the nature of the review question and the evidence-based approach of the EPPI-Centre, only empirical studies were judged to be suitable in terms of trustworthiness and in answering the review question (see APU Assessment review-specific keywording sheet, section 10c).
- Criterion 15: *Included*: Only studies that focus on the assessment model. *Rationale*: This was applied because the URT agreed that assessment in its broadest sense could be divided into three aspects, comprising the actual assessment itself and the model used (*the process*), the assessors (*the subject*) and the *assessed* (*the object of the assessment conducted by the assessors*). The URT intended to look at formative assessment models. The other two aspects were therefore rejected (see APU Assessment review-specific keywording sheet, section 6c).
- Criterion 16: *Included*: Only studies that took place at the post-developmental stage of assessment. *Rationale*: The review question demanded that an assessment instrument, if it were to be examined through data-extraction, must be tested, not at the developmental/experimental stage of its life (see *EPPI-Centre Core Keywording Strategy* (version 0.9.6: 2003), section 10a. Unwanted studies were excluded via the 'Development of Methodology' keyword).
- Criterion 17: *Included*: Only studies that focus on the assessment of the school-based experience. *Rationale*: The review question focuses on only this phase of ITT because this was where issues about 'authentic assessment' (Willis and Davies, 2002) and teaching rather than academic ability come into sharpest focus (see APU Assessment review-specific keywording sheet, section 9b).
- Criterion 18: *Included*: Only studies that focused on formative not summative assessment. *Rationale*: The review question demanded that only formative,

fashioning forms of assessment were of interest (see APU Assessment review-specific keywording Sheet, section 6b).

## 2.2.5 Identifying and describing studies: quality assurance process

### ***Application of inclusion and exclusion criteria***

First, during the application of the inclusion and exclusion criteria, the first ten abstracts were screened by the URT to ensure parity of approach. Second, 20 abstracts then had the inclusion and exclusion criteria applied by independent members of the URT and were then compared with other URT members' results. Finally, 20 abstracts then had the inclusion and exclusion criteria applied by EPPI-Centre personnel, who then compared results with the URT.

### ***Keywording***

First, during keywording the first two papers were worked on with the URT to ensure parity of approach. Second, the rest of the papers were keyworded by independent members of the URT, and then compared with other URT members' results. Finally, ten papers were keyworded by EPPI-Centre personnel who then compared results with the URT.

## 2.3 In-depth review

### 2.3.1 Moving from broad characterisation (mapping) to in-depth review

Following the keywording process and the development of the systematic map, the URT could have:

- gone straight to the 'in-depth review'
- undertaken a two-stage process in which narrower inclusion criteria were first applied, followed then by an 'in-depth review'

The second choice was chosen because the studies included in the descriptive map were not all deemed of high relevance, and so their inclusion in the in-depth review would not have been worthwhile.

Moving from the broad characterisation of the field to the in-depth review involved devising a series of searches using keywords that would identify relevant studies for data-extraction and inclusion in the in-depth review. Having combined these searches and therefore having identified the studies most likely to answer the review question, a further screening stage was included for the purpose of assessing the relevance of each of these studies. Having determined this, further inclusion and exclusion criteria were applied. Studies that met all of the criteria and were deemed of high relevance, though not necessarily of high trustworthiness, were identified and data-extracted.

### **2.3.2 Detailed description of studies in the in-depth review**

The studies identified for possible inclusion in the in-depth review were data-extracted using the *EPPI-Centre Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research* (version 0.9.5: EPPI-Centre, 2002).

### **2.3.3 Assessing quality of studies and weight of evidence for the review question**

In light of the data-extraction outcomes, the quality and weight of evidence for the studies were assessed. To do this, they were rated according to three dimensions: (A) soundness of the study (i.e. internal coherence and implementation of design), (B) relevance of research design and analysis employed to the review question, and (C) external relevance. On the basis of how a study was rated on each of these three, an overall weighting (D) was given (see section 4.3).

### **2.3.4 Synthesis of evidence**

The relevant information for this section was identified during data-extraction. In section 4.3, the studies included in the in-depth review are assessed and synthesised with reference to the weight of evidence they were awarded during their data-extractions.

### **2.3.5 In-depth review: quality assurance process**

Of the data-extraction, 100 percent was carried out by the URT and two reports were data-extracted by an EPPI-Centre representative who then compared results with the URT.

### 3. IDENTIFYING AND DESCRIBING STUDIES: RESULTS

This chapter outlines the search strategy employed to identify studies for the systematic review and describes the nature and extent of research within the field.

#### 3.1 Studies included from searching and screening

The number of studies identified by the search process and included in different stages of the process of the review are shown in Figure 3.1. Table 3.1 illustrates the process of identifying, obtaining and describing papers for the review. Incorporating all of the searches, the total number of papers prior to the next stage of screening (which involved scanning the abstracts and titles and excluding those outside of the geographic area and not in English) was 668.

**Table 3.1:** Flow of papers from searching to inclusion in map

Number of prospective studies in the field	?
Number of 'hits' using a controlled vocabulary	668
Reports that met criteria 1 to 11 on basis of abstract	233
Reports excluded because they were duplicates	9
Reports not received or unobtainable	21
Reports excluded because they were ERIC digests	109
Reports excluded through reapplication of criteria 1 to 11	12
Total full reports excluded	151
Total full reports keyworded	82

**Note:** The figure for the first row is not known as it is the sum of every paper on the ERIC, BEI, and Education Line databases.

The total number of papers included after the application of inclusion and exclusion criteria 1 to 11 was 233. Of these, 109 papers were then excluded because they were digests, seven because the abstract or full citation had not indicated that they were outside of the timeframe, one because it was not actually about trainee teachers, and nine because they were duplicates. The papers from Education Line (N=4) were not deemed relevant for further inclusion once they had been exposed to criteria 1 to 11. Unfortunately, 21 papers were either not received by the deadline (10 March 2003) or were unobtainable.

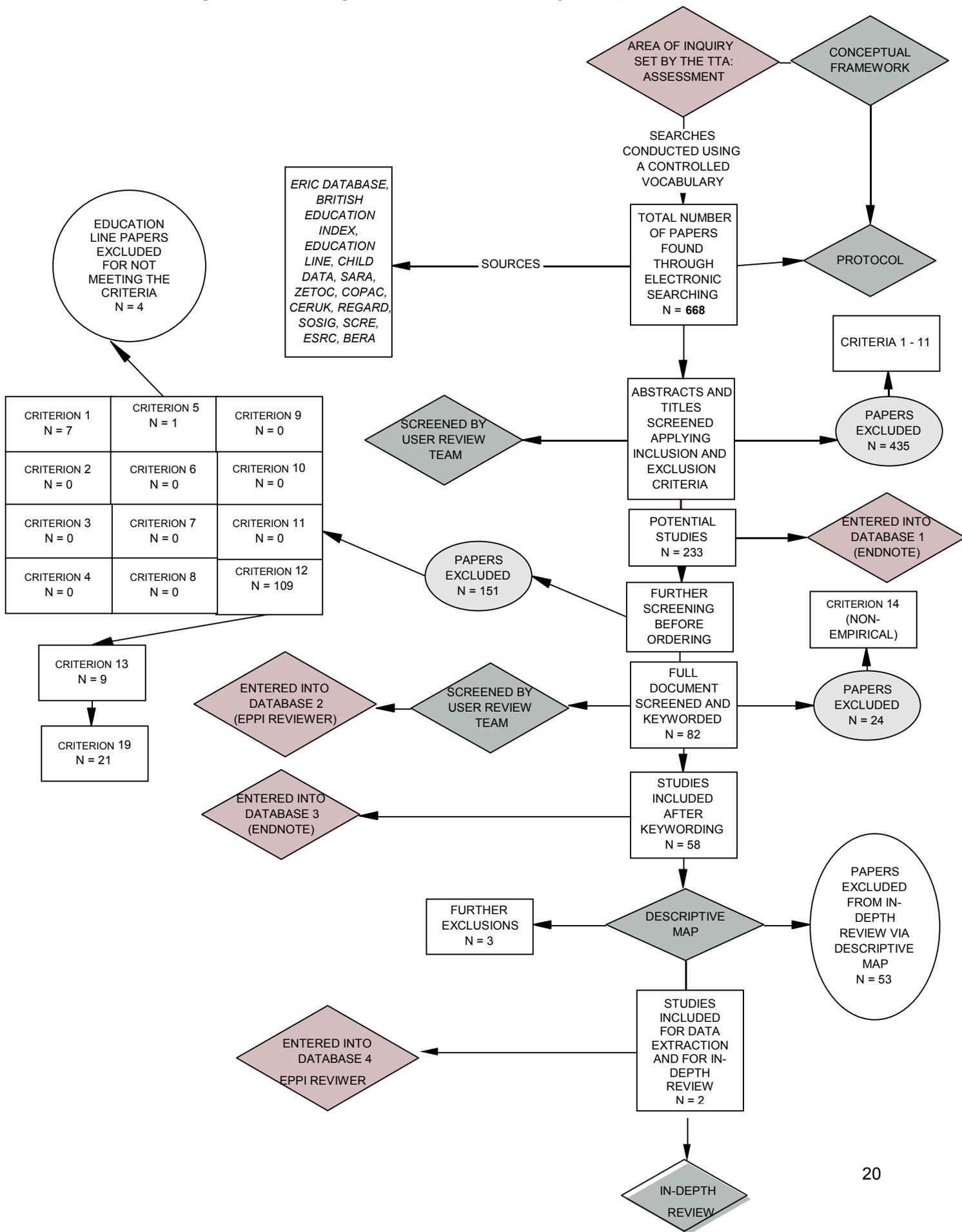
Table 3.2 presents the origin, by database, of all the papers.

**Table 3.2:** Origin of papers by electronic database

<b>Database</b>	<b>Found</b>	<b>Checked</b>	<b>Keyworded</b>	<b>Empirical</b>
ERIC	573	204	60	42
British Education Index	66	25	22	16
Education Line	29	4	0	0
<b>Total (mutually exclusive)</b>	<b>668</b>	<b>233</b>	<b>82</b>	<b>58</b>

No additional studies were found through the searches of CERUK, Regard, SOSIG, SCRE, ESRC, Child Data, SARA, Zetoc, OPAC and COPAC.

**Figure 3.1: Filtering of studies from searching to map**



## 3.2 Characteristics of the included studies

Eighty-two included papers were keyworded using the *EPPI-Centre Core Keywording Strategy* (version 0.9.6: EPPI-Centre, 2003). Whilst the *EPPI-Centre Core Keywording Strategy* (version 0.9.6: EPPI-Centre, 2003) was designed for characterising only papers based on empirical research (i.e. *reports* of actual *studies*), the URT uncovered a number of papers that had no empirical foundation, and were instead conjectural, philosophical or positional. This may have been symptomatic of the subject area, which often appears to feature large numbers of proposals and suggestions for models for assessment purposes without an empirical base. Of the 82 keyworded papers, 58 (71%) were empirical and 24 (29%) non-empirical. Only 58 papers, therefore, qualified as ‘reports’ according to the working definition used by EPPI-Centre (see *EPPI-Centre Core Keywording Strategy*, version 0.9.6, 2003; section 7). Although all 82 items were left on EPPI-Reviewer, only 58 could reasonably be keyworded under section 10 of the EPPI-Centre Educational Keywording Sheet; these feature in the descriptive map. All studies included in the map were written in English and of published status.

*For full bibliographic details and keywords for the 58 included studies, see Appendix 3.1.*

Table 3.3 shows the number and proportion of studies according to the country in which they were conducted. Most studies were conducted in the US (35: 60.3%). The second largest exponent of studies was the UK (13: 23.4%). Three studies fell outside of the geographic area (two from China, one from Nigeria). This was because the country of origin was not indicated on either the abstract or full citation, and so the full text was ordered and read. Keywording was completed on all full texts that were obtained within the timeframe; this included the three studies that would normally have been excluded by Criterion 2. These figures may reflect bias within the bibliographic database sources searched towards reports published within the US and the UK.

**Table 3.3:** Frequency report: In which country/countries was the study carried out? (Coding was mutually exclusive.)

Country	Number	Percentage (%)
United States	35	60
United Kingdom	13	23
Australia	2	3
Canada	2	3
China	2	3
US, Luxembourg, Germany and Austria	1	2
United Kingdom and Germany	1	2
Ireland	1	2
Nigeria	1	2
<b>Total (number of studies = 58)</b>	<b>58</b>	<b>100</b>

Table 3.4 describes the educational settings of the 58 studies. A study could be conducted in more than one setting, and in most studies that focused on the school

experience of trainees both sectors were featured. This is reflected by 22 studies focusing on both the primary and secondary school settings.

**Table 3.4:** Frequency report: What is/are the educational setting(s) of the study? (Coding was not mutually exclusive.)

Attribute	Number	Percentage of studies (%)	Percentage of all coding (%)
Higher education institution	39	67	43
Primary school	22	38	24
Secondary school	22	38	24
Special needs school	3	5	3
Home	2	5	3
Nursery school	2	5	3
<b>Total (number of studies = 58)</b>	<b>90</b>	<b>158 (100/58)</b>	<b>100 (100/90)</b>

Table 3.5 shows the population focus of the studies, whilst Table 3.6 shows the sex of the learners, being as they were the overwhelming population focus of the 58 studies (55: 74.3%). A number of these studies also focused on teaching staff (14: 18.9%). The majority of the studies focusing on learners reported mixed sex participants (50: 90.9%). A small number of studies were undertaken using only female trainees (5: 9.0%), whilst none was undertaken involving only males.

**Table 3.5:** Frequency report: What is/are the population focus/foci of the study? (Coding was not mutually exclusive.)

Attribute	Number	Percentage of studies (%)	Percentage of all coding (%)
Learners	55	95	74
Teaching staff	14	24	19
Other population focus	4	7	5
Senior management	1	2	1
<b>Total (number of studies = 58)</b>	<b>74</b>	<b>127 (100/58)</b>	<b>99 (100/74)</b>

**Table 3.6:** Frequency report: Sex of learners (Coding was mutually exclusive and depended upon the population focus of the study.)

Attribute	Number	Percentage of studies (%)	Percentage of all coding (%)
Female only	5	9	9
Male only	0	0	0
Females and males	50	86	91
<b>Total (number of studies = 58)</b>	<b>55</b>	<b>95 (100/58)</b>	<b>100 (100/55)</b>

The loading of gender in these studies may be an indication of dated research in the field, for the absence of studies using exclusively male participants is perhaps surprising in light of the attention males have received in recent years regarding assessment and attainment (Stobart and Gipps, 1997). On the other hand, with only

14 percent of primary schoolteachers reported to be male, and only a few more males reported to be secondary teachers, it is perhaps *not* surprising that five studies deal exclusively with females and none with males. The gender loading may reflect the average loading of an ITT course, or may simply have been accidental, i.e. no males happened to be on the ITT courses selected for investigation.

Table 3.7 shows how the studies were weighted when assessment was conceptualised as a system involving three aspects. In this system, the assessment was the *process*, the assessors were the *subject* and the assessed were the *object*. The majority of the studies (32: 51.6%) focused on the assessment process and would therefore be included within Criterion 15.

**Table 3.7:** Frequency report: Main agent within assessment on which the study focuses (Coding was not mutually exclusive.)

Attribute	Number	Percentage of studies (%)	Percentage of all coding (%)
Assessors	5	9	8
Assessment	32	55	52
Assessed	25	43	40
<b>Total (number of studies = 58)</b>	<b>62</b>	<b>107 (100/58)</b>	<b>100 (100/62)</b>

Table 3.8 shows the overall distribution of reports according to study type. Many of the studies were evaluative and therefore more suited to answering the review question than the other types of study. Development of methodology (6:10.3%) described studies in which the principal focus was on the development of the assessment instrument, drawing on empirical research to construct a model of assessment. These six studies did not meet Inclusion Criterion 16.

**Table 3.8:** Frequency report: Which type(s) of study does this report describe? (Coding was not mutually exclusive.)

Attribute	Number	Percentage of studies (%)	Percentage of all coding (%)
Evaluation: Researcher-manipulated	17	29	29
Evaluation: Naturally occurring	16	28	28
Description	13	22	22
Development of methodology	6	10	10
Exploration of relationships	5	9	9
Review: Other review	1	2	2
<b>Total (number of studies = 58)</b>	<b>58</b>	<b>100 (100/58)</b>	<b>100 (100/58)</b>

Table 3.9 cross-tabulates types of assessment with the phase of the ITT course. These phases were not intended to be comprehensive; their selection was deliberate. 'Initial pre-student recruitment' was included because a genus of studies focusing on predicting trainee teacher success was identified during previous screening stages. 'School-based experience' was a key component within the review question; in order to be included beyond the descriptive mapping stage the study would have to be focused on this and take a formative model as its particular

interest. The phrases 'Exit practices', 'Career Entry Profile' and 'Statutory Induction' all related to the request of the TTA to avoid studies that focused on the CEP. Thirty-five studies of the 58 would therefore be included under Inclusion Criterion 17.

**Table 3.9:** Cross-tabulation (x-axis: phase of ITT that is the focus of the study; y-axis: type(s) of assessment on which the study focuses)  
(Total number of studies = 58. Coding was not mutually exclusive.)

Type of assessment / phases of ITT	Initial pre-student recruitment	School-based experience	Exit practices	Career entry profile	Statutory induction
Formative assessment	0	35	2	0	3
Summative assessment	0	12	3	0	2
Ipsative assessment	0	4	1	0	0
Self-assessment	0	13	1	0	0
Portfolio assessment	0	8	2	0	0

### 3.3 Identifying and describing studies: quality assurance results

#### 3.3.1 Application of inclusion and exclusion criteria

The first ten titles and/or abstracts were completed with the URT to ensure parity of approach. In addition, 20 titles and/or abstracts then had the inclusion and exclusion criteria applied by independent members of the URT and were then compared with other URT members' results. Finally, 20 titles and/or abstracts, having been screened by members of the URT, then had the inclusion and exclusion criteria applied by EPPI-Centre personnel, who compared results with the URT. The quality assurance result was that the EPPI-Centre representative was more exclusive. Of the 20 abstracts screened by the EPPI-Centre, ten of which the URT had included and ten excluded, 14 were excluded. The ten URT exclusions were included in this figure. Whilst the URT had an exclusion rate of 50 percent, the EPPI-Centre applied a 70 percent exclusion rate. This satisfied the URT's aim to be over-inclusive.

#### 3.3.2 Keywording

The first two papers were keyworded with the URT to ensure parity of approach. The rest of the papers were then keyworded by independent members of the URT and compared with another URT member's result. Finally, ten papers were keyworded by EPPI-Centre personnel who then compared results with the URT. The quality assurance result was that section 10a of the EPPI-Centre Educational Keywording Sheet (version 0.9.6) was the only area where EPPI-Centre and the URT differed. As a result, section 10a on all papers was re-keyworded by the research members of the URT in close contact with EPPI-Centre personnel to ensure agreement.

## 4. IN-DEPTH REVIEW: RESULTS

This chapter details the studies included in the systematic review, drawing on the data-extracted using *EPPI-Centre Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research* (version 0.9.5: EPPI-Centre, 2002).

### 4.1 Selecting studies for in-depth review

Table 4.1 is a representation of the combined search strategy used by the User Review Team to exclude studies. This strategy was devised combining keywords from both the EPPI-Centre Educational Keywording Sheet (version 0.9.6) and the APU Assessment review-specific keywording sheet. The third column refers to the sections on these two sheets, which must be viewed together. The review-specific keywords were devised specifically for the purpose of identifying reports that would answer the review question. The strategy was a means by which 58 reports were reduced to a number manageable in a systematic review lasting less than three months.

**Table 4.1:** Search strategy

Search	Keyword	Section	Number of hits
1	Assessment	6	50
2	Formative assessment	6b	43
3	Assessment	6c	32
4	Learners	8	53
5	Mixed sex	8b	50
6	Undergraduate trainees OR Post-graduate trainees	8c	48
7	Primary school OR Secondary school	9	34
8	School-based experience	9b	47
9	Evaluation OR Evaluation: naturally occurring OR Evaluation: researcher-manipulated	10a	33
10	Empirical	10c	58
11	1 AND 2 AND 3 AND 4 AND 5 AND 6 AND 7 AND 8 AND 9 AND 10		5

Search 1 identified all the studies in which the main topic was assessment. Search 2, using the review-specific keywords, determined those studies that focused on formative assessment more than any other from, such as summative or ipsative. Search 3 pooled all the studies that concentrated on assessment models rather than the assessors and/or the assessed. Search 4 ensured that the main population focus of the studies included was student teachers; Search 5 that the sample used was of mixed gender; and Search 6 that the trainee teachers were undergraduates

or post-graduates. Searches 7 and 8 both ensured that the studies retrieved focused on the school-based experience and therefore concentrated on performance-based assessment studies with a formative slant. Search 9 ensured that only evaluative studies were included. Search 10 was to double-check that only empirical studies had been included.

The *five* studies remaining after the descriptive map are as follows:

- Berg M and J Curry (1997) Portfolios: what can they tell us about student teacher performance? *Social Studies Review* **36**: 78-84.
- Brucklacher B (1998) Cooperating teachers' evaluations of student teachers: all 'A's'? *Journal of Instructional Psychology* **25**: 67-72.
- Fishman AR and EJ Raver (1989) 'Maybe I'm just NOT teacher material': dialogue journals in the student teaching experience. *English Education* **21**: 92-102.
- Fitzgibbon A (1994) Self-evaluative exercises in Initial Teacher Education. *Irish Educational Studies* **13**: 145-164.
- Willis EM and MA Davies (2002) Promise and practice of professional portfolios. *Action in Teacher Education* **23**: 18-27.

The five studies that seemed to meet the inclusion criteria for the in-depth review according to the keyword codings were then re-examined in detail. Three of the studies were then judged not to meet the second set of inclusion criteria (criteria 14 to 18). It was necessary to reconsider the studies that were included using a keyword search because some of the review-specific keywords required a high level of inference. Three of the studies could have been differently and perhaps more suitably keyworded.

The keywords that affected the keyword search in the way described above were all review-specific. This is not therefore an admission of inaccurate keywording of keywords from the *EPPI-Centre Core Keywording Strategy*, version 0.9.6 (2003).

The *two* included studies were as follows:

- Berg M and J Curry (1997) Portfolios: what can they tell us about student teacher performance? *Social Studies Review* **36**: 78-84.
- Willis EM and MA Davies (2002) Promise and practice of professional portfolios. *Action in Teacher Education* **23**: 18-27.

The *three* excluded studies were as follows:

- Brucklacher B (1998) Cooperating teachers' evaluations of student teachers: all 'A's'? *Journal of Instructional Psychology* **25**: 67-72.

- Fishman AR and EJ Raver (1989) 'Maybe I'm just NOT teacher material': dialogue journals in the student teaching experience. *English Education* **21**: 92-102.
- Fitzgibbon A (1994) Self-evaluative exercises in Initial Teacher Education. *Irish Educational Studies* **13**: 145-164.

Details of these three excluded studies are given in Appendix 4.1.

To summarise, having developed a map that describes the field of assessment of trainees during their initial teacher training (ITT), a combined search was undertaken using EPPI-Reviewer. The result of this combination search was five reports. On further inspection, three of the number were judged to be unfit for data-extraction. Two reports remained. Prior to data-extraction, therefore, the paucity of studies that took as their main focus formative assessment, trainee teachers and school-based experiences, was palpable. Only two studies were identified that were closely relevant. Assessment in ITT, therefore, falls into an obvious gap in assessment research.

## 4.2 Further details of studies included in the in-depth review

For detailed descriptions and analysis of the studies, see Appendix 4.1.

The Berg and Curry (1997) study investigated portfolios as tools for increasing professional and personal development whilst at the same time providing a means for assessment of trainee teachers. It focused on 30 trainees, monitoring and assessing their portfolios at three times during one year. The portfolio structure was devised by a team of professors, supervisors and cooperating teachers. Four headings were identified, within which questions were asked of the portfolio contents and answers assessed. Ratings were given according to three categories.

The Willis and Davies (2002) study similarly investigated portfolios as tools for increasing professional development, whilst at the same time providing a means for trainee teacher assessment. It focused on 93 trainees and, instead of determining the effectiveness of the portfolio through interventions, the study drew upon data from a questionnaire requiring trainee teachers to reflect on the portfolio component of their ITT courses in retrospect.

## 4.3 Synthesis of evidence

To determine how the evidence in the two studies relates to the question and the aims of the review, it is necessary to examine the five objectives that are implicit in the review question:

1. To DESCRIBE the best studies that focus on formative models for assessing trainees

2. To evaluate only studies that make POSITIVE claims about the models they examine
3. To evaluate a number of DIFFERENT models to compare and contrast
4. To determine what the different models actually ASSESS
5. To evaluate how good teaching is CONCEPTUALISED

Objectives 1 and 2 were achieved. Objectives 3, 4 and 5 proved to be more problematic.

Using EPPI-Centre questions for review-specific weight of evidence, a number of questions were asked of the data-extracted studies. Table 4.2 summarises the outcomes of these questions, following which is a more extensive explanation of these questions.

**Table 4.2:** Weight of evidence

Study	A	B	C	D
Berg and Curry (1997)	Low	Medium	Low	Low
Willis and Davies (2002)	Medium	Medium	High	Medium

### 4.3.1 Weight of evidence A

*Weight of evidence A: The trustworthiness of the findings of the studies in answering the study questions, taking account of all quality assessment issues*

The Berg and Curry (1997) study would have been difficult to undertake differently. If one wishes to research the use of portfolios in trainee teacher development and growth, then getting trainees to keep a portfolio and assessing it is the obvious choice of method. No information, however, is given about why the particular contents of the portfolios were appropriate nor why they were used. Without significantly more information about the subjective nature of the analysis, bias is inherent in this study. Furthermore, the results are not generalisable without much more contextual information. In US and San Diego terms, one feels certain that there is the basis of a good study. However, the evidence provided is scant both conceptually and empirically, making the findings of the study of low trustworthiness. Results, discussion, conclusions and implications all flow into one another and have to be interpreted.

There is an element within this study that the 'desired' outcomes, which are only implicitly stated, were achieved. There is, however, no argument presented as to whether or why some students were less effective in their portfolios, for example. It is hard to agree or disagree with the findings and conclusions without further information or evidence.

In the Willis and Davies (2002) study there is a loose connection between the climate described in the introduction and the actual study, but no sound justification,

reasoning or apologetic is discernible. Decisions are rarely made explicit, nor are they given a rationale. The questionnaire seems to have been the best, perhaps only way, to have drawn data retrospectively on the portfolio component of the ITT course. Interviews would have been an alternative, for better or worse.

There is no deliberation on the possible shortcomings and flaws of the questionnaire and the collection of it in the Willis and Davies (2002) study. One major shortcoming may be that the Likert scale is four-pronged, instead of the more common five-pronged scale. One school of thought advocates that there is no room for indifference on a four-pronged scale, as everything is either agreeable or disagreeable, to different degrees, and there is no middle ground. Another school of thought, however, advocates that the four-pronged Likert scale is better than the five-pronged one, because it forces participants off the fence. With middle ground, as well, there is the potential for drawing conclusions either way, and seeing them as positive or negative. Use of the four-pronged scale is not given a rationale.

The data-extractor (the questionnaire) used in the Willis and Davies study is based mainly on a set of former survey questions. To some degree, it builds on a tried and tested base. This could be a minor validity assurance measure if the reasoning behind this appropriation was explained.

Although the findings of the study are for the most part reported in the section 'Survey Results', there are obvious omissions. Table 1, for example, does not present all the questions and therefore omits a lot of the data. Because the entire process is not adequately described, the research method and design is not easy to follow; therefore the chance that bias and error have distorted the findings is quite high.

Regarding the generalisability of the findings of the Willis and Davies study, there is nothing to suggest that the sample would seriously differ from any other taken from an average ITT course. The findings, therefore, could safely be applied to other courses, especially within the US.

Although the research design and method is not made explicit, and this naturally raises questions about the reliability and validity of the findings, the data and the assertions in the Willis and Davies study made from the data were fairly trustworthy, insofar as the study is a simple one, without a huge margin of error.

Regarding the data, therefore, the Willis and Davies study is fairly sound. Quantitative data without complex statistical analysis applied cannot really be interpreted in too many ways, so this is fairly trustworthy. The qualitative data is the untrustworthy component of the report. As we cannot see the original questionnaire in its entirety, we cannot fairly assess the validity of the questions, but the report does draw a clear distinction between the findings and the conclusions drawn from them. Formulating a conclusion independently from the conclusion in the report is therefore possible.

Since the study is not completely traceable, high trustworthiness for the findings of the study cannot be awarded, but because it presents most of the data before it interprets it and draws conclusions based on common sense, it can be awarded

medium trustworthiness. The Willis and Davies study, therefore, scores a higher level of trustworthiness regarding findings than the Berg and Curry (1997) study.

The concluding section of the Willis and Davies (2002) study, more than anything, synthesises the findings rather than interprets them. The section on page 25, which offers possible implications and recommendations, is the most interpretative part. No real justification is offered, however, for the concluding section, but the conclusions are, at the same time, not implausible. The conclusions drawn from the findings are the same as those the reviewer would make.

### **4.3.2 Weight of evidence B**

*Weight of evidence B: The appropriateness of the research design and the analysis of the studies for addressing the review question*

The Berg and Curry (1997) study addresses one particular model of formative assessment, that of portfolios. The development and use of these portfolios is described within the study in some detail. How the research was conducted is also outlined with reasonable clarity. The research design, however, is vague in terms of the sample and the context of the research, as well as in terms of how the data were analysed. The findings and conclusions of the study concentrate mainly on what the trainees said in their portfolios, using their statements and comments as indicative of professional and personal development. A better means of data-collection for the purpose of this review would have been interviews after completion of the portfolios to determine how the trainees evaluated the use of portfolios as a means of formative assessment of their skills and understandings. Owing to the limited appropriateness of the research design and analysis for addressing the question for this systematic review, the study was of limited trustworthiness. However, the study does describe a formative model of assessment (Objective 1) and does give it a positive evaluation (Objective 2). The Berg and Curry (1997) study, therefore, scores a medium level of trustworthiness.

Similarly, the Willis and Davies (2002) study offers the portfolio as a model for assessing trainees formatively during their school experience and the rest of the course. The study focuses on the presentation or 'exhibition' of portfolios, although in this context portfolios could be seen as summative, because the presentation of them takes place at the end of the course. That they are deemed formative by the authors is discernible from the fact that key concepts and actions are incorporated within the formative sphere. Examples of these are reflection, self-evaluation, professional development, self-confidence and communication skills. This particular study also describes an assessment model (Objective 1) and makes a positive evaluation of the portfolio, so it does describe a 'successful model' of formative assessment (Objective 2).

The research design and analysis of the Willis and Davies (2002) study are suited to the review question because they combine qualitative and quantitative data, and they also establish from the trainees directly their views on portfolios as a means of formative assessment. Transcribed interviews might have been a better option, if the data were reported satisfactorily, but a questionnaire is a suitable method, and from what is deducible from the findings, the questions asked were ones that the User

Review Team (URT) thought worth answering. The Berg and Curry (1997) study is perhaps less informative because of its failure to collect data relating to portfolios *per se*, although neither of the studies draws conclusions that are completely free from bias.

The URT wanted to determine what is known about formative models, and in the Willis and Davies (2002) study, possible benefits and shortcomings are suggested. The second part of the review question, 'what constitutes effective practice...' is not answerable using this study, however, because it does not focus on an exact model for assessing but explores an assessment tool. The Berg and Curry (1997) study does use a more exact model, but fails to report in sufficient detail. The criteria used for assessing portfolios in the Willis and Davies study are not presented in sufficient detail for the URT to determine how that particular college personnel conceptualise good teaching or to answer the question 'What constitutes good teaching?' What the college personnel in the Berg and Curry (1997) conceptualise as good teaching is easier to determine (p 80). For example, regarding portfolios, a good teacher should have the...

- ability to reflect clearly on own growth and change
- ability to discuss strengths and weaknesses of performance
- ability to address cross-cultural and language development infusion.

Both studies employ appropriate research designs to a degree, and both draw conclusions that make positive evaluations of portfolios as a formative assessment model. Both studies, therefore, satisfy the first two objectives of the review question. It is, however, the quality of the reporting of the research designs and analyses that are problematic. If everything that had been learned by the researchers had been reported, then it would be possible to answer more comprehensively 'what is known...' and 'what constitutes effective practice', but unfortunately what is reported is limited.

In the Willis and Davies (2002) study, the questionnaire used to collect the data is not represented in its entirety. In the Berg and Curry (1997) study, the methods of data analysis used are not stated and the 'reflective questions' that the assessors asked of the portfolios are only represented by four broad 'areas' (p 80). General conclusions can be drawn about portfolios from both studies, but no exact models for either (i) the portfolios, (ii) the assessment of them, or (iii) acquiring feedback about them are offered that are trustworthy, with adequate validity and reliability measures taken. Both studies score a medium level of trustworthiness in this area.

### 4.3.3 Weight of evidence C

*Weight of evidence C: The relevance of the particular focuses of the studies, including their conceptual focus, context, sample and measures for addressing the review question*

In the Berg and Curry (1997) study, the portfolio is relevant to the research question, but the conceptual basis of portfolios is only minimally revealed and the context is not described or explained. The authors do not give information on the sample although we know there were 30 students involved. The stage in their training is

unknown, as is the duration or level of their course. We do know that the portfolios were kept for one academic year and that they were monitored through three stages (initial, medial and towards the end of the year) but we are not told much more. It is frustrating that a study that potentially has direct relevance to the research question is so poorly reported and fails to collect feedback from the trainees.

The relevance of the Willis and Davies (2002) study to the review question is high because it focuses on undergraduate trainee teachers on a course that balances theory and practice (p 19) whilst seeking to investigate an instrument that may eventually lead to a more 'authentic assessment' (p 18). The context, sample and conceptual focus, therefore, are perfectly suited to the review question. The problem is that the means by which the teacher educators and peers evaluated the portfolios is not revealed, which may have helped the URT more readily to conceptualise what the college personnel perceived good teaching to be. The report reveals that the portfolios and the presentation of them were examined 'on the rubric of professionalism, organisation of presentation, delivery, and responses to questions', but no more detail is given.

#### 4.3.4 Weight of evidence D

*Weight of evidence D: Taking into account the quality of the execution, the appropriateness of the design and the relevance of the focuses of the studies in answering the review question...*

Overall, both studies would have been highly relevant had the reporting of research design and methodology been better. The Berg and Curry (1997) study, in US and San Diego terms, is probably a good study because interested parties there would understand the various aspects of portfolio use in that context. However, the evidence provided is scant in both conceptual and empirical terms, meaning that it can have only low trustworthiness for the purposes of answering the review question, satisfying only two of the five objectives. The findings need considerable interpretation in order to answer the review question, as there is little depth to exploring how the outcomes were reached and how the conclusions drawn relate to summative assessment specifically. Many of the findings have to be interpreted and conclusions drawn by the reader.

The Willis and Davies (2002) study is relevant and the research design and methodology is generally appropriate. However, the validity of the conclusions is questionable as the analysis is sometimes vague and there are gaps in the reporting.

It is arguable whether either study is in any way replicable. It is also regrettable that neither study used controlled trials to determine whether the portfolios either enhanced professional development or merely recorded natural development over the course of the training.

With regard to the background to this review (Chapter 1), both studies offer models of assessment that emphasise the equal weighting of process and product in learning, teaching and evaluation (Stobart and Gipps, 1997). However, neither study offers a standardised form nor individual items for the portfolio that are subject to validity and reliability measures. Neither study engages in the current debate about

fair testing nor the notion of multiple assessments, although both imply that the need for authentic assessment motivated the studies. The Willis and Davies (2002) study refers directly to 'authentic assessment' (p 18), whilst the Berg and Curry (1997) study states that the portfolio would 'also embody an attitude that assessment is dynamic and that the best representations of student teacher performance are based on multiple sources of assessment collected over time in authentic settings' (p 78).

In regard to the aims of this review (section 1.1.1), neither the validity nor the reliability of the assessment models are adequately explored (Martin, 1997). Owing to the lack of reliability measures, a way of making an assessment that is a 'replicable finding' is not forthcoming (Hartsough, 1998). As a result, 'minimising the number of wrong classifications' is not possible from the results of the studies either (Martin, 1997). Both studies do, however, support the belief that portfolios can be used for the purpose of selection and grading, diagnosis and remediation, motivation, and recording and reporting (Mahoney and Knox, 2000).

#### 4.4 In-depth review: quality assurance results

Each study was data-extracted using the *EPPI-Centre Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research* (version 0.9.5: EPPI-Centre, 2002) by two separate reviewers in isolation from one another. The extractions were then compared, discrepancies were discussed and changes were negotiated. A third version of the data-extraction was then amended to reflect the agreement of both reviewers. Both studies registered an excellent inter-rater reliability score. The Berg and Curry (1997) study was 0.85 whilst the Willis and Davies (2002) study was 0.91. The double-reviewed data-extractions were then quality assured by an independent member of the User Review Team. The agreed ('triple-reviewed') data-extractions were then quality assessed by an EPPI-Centre team representative. There were again few discrepancies. The Willis and Davies (2002) study had an inter-rater reliability score of 0.92, as did the Berg and Curry (1997) study. To work out the degree of correlation between the coding of the two raters, we used Cohen's Kappa. Table 4.3 represents these scores.

**Table 4.3:** Inter-rater reliability

<b>Study</b>	<b>Internal kappa score (between URT personnel)</b>	<b>External kappa score (between APU and EPPI- Centre personnel)</b>
Berg and Curry 1997	0.85	0.91
Willis and Davies 2002	0.92	0.92

#### 4.5 Nature of actual involvement of users in the review and its impact

The User Review Team was involved in the review at all stages: the screening of abstracts, the screening of full texts, keywording, and data-extraction. They brought

to the review their expertise and ensured that the questions asked and the consequent findings were of use. Some challenges were experienced by the teachers and tutors within the team because of the speed with which all the processes had to occur in the review period (less than three months). For this reason, the majority of the inputting of the studies into EPPI-Reviewer were undertaken by the URT Administrator to save time and inconvenience for the teacher and tutor members of the team. All team members, however, checked their data inputs for accuracy.

## 5. FINDINGS AND IMPLICATIONS

This chapter details the findings and implications of the in-depth review and the descriptive map.

### 5.1 Summary of principal findings

In this section, findings are drawn from the three studies that were excluded prior to data-extraction, as well as the two that were included.

On account of the limited trustworthiness of all studies, mainly due to missing details in the report and failure to detail the methods for data gathering and analysis (see section 4.2), the benefits and problems below must be treated with caution.

The two data-extracted studies focused on portfolios.

#### 5.1.1 Benefits of the excluded studies

Portfolios are distinct from *profiles* in that they represent a collection of work and reflections based on a loose, interpretative structure rather than checklists of responses to competencies and statements related to teaching. They offer a more constructivist approach to assessment.

It is also argued by the authors of the two studies that the portfolios are a successful means of assessing teacher trainees during school experience insofar as they:

- are generally perceived to be 'worthwhile' by trainee teachers (Willis and Davies, 2002, p 20; cf. Reis and Villaume, 2002, in which the 'time versus worth' problem is explored in relation to portfolio assessment)
- perpetuate professional growth and development (Berg and Curry, 1997, p 82)
- allow trainee teachers to express themselves creatively as teachers (Willis and Davies, 2002, p 20). The majority of trainee teachers felt that their portfolio 'uniquely represented themselves' (Willis and Davies, 2002, p 22).
- provide insight into teaching and the impact it has on pupil learning (Berg and Curry, 1997, p 82; Willis and Davies, 2002, p 20)
- help trainee teachers to work towards making meaning comprehensible for all pupils (Berg and Curry, 1997, p 82)
- allow trainee teachers to reflect on what is 'fun' and 'motivating' to pupils (Berg and Curry, 1997, p 82)
- allow trainee teachers to reflect on their practice *per se* (Willis and Davies, 2002, p 20)
- create a 'strong bond' between the assessors and the assessed (Berg and Curry, 1997, p 84)
- encourage all involved in the assessment process to 'join as collaborators in learning' (Berg and Curry, 1997, p 84)

- are more popular with trainee teachers when they have ‘ownership and decision making about the portfolio categories’ (Willis and Davies, 2002, p 20)
- create awareness of personal and professional growth (Willis and Davies, 2002, p 20)
- increase self-confidence in terms of presentation (Willis and Davies, 2002, p 22) and improve trainee teachers’ speaking and interviewing skills (Willis and Davies, 2002, p 24)
- prepare trainee teachers for their ‘job search’ (Willis and Davies, 2002, p 22)
- help trainee teachers, when used in relation to National Standards, to reflect on those Standards and how they could be incorporated and represented in the portfolio (Willis and Davies, 2002, p 22)
- encourage self-perception as lifelong learners (Willis and Davies, 2002, p 24)

### **5.1.2 Problems with the excluded studies**

According to the authors of the studies, portfolios are problematic insofar as they:

- create problems for trainee teachers regarding format, selection and design (Willis and Davies, 2002, p 24)

### **5.1.3 Findings drawn from the excluded studies**

- Dialogue journals – in which both the assessor and the assessed keep journals – enrich the experience of both the trainee teacher and the cooperating teacher (Fishman and Raver, 1989).
- As stated by Willis and Davies (2002), the majority of trainee teachers find portfolios useful and claim their use increases reflection and self-awareness (Fitzgibbon, 1994).

### **5.1.4 Gaps in research**

During the course of the review, and especially after producing the descriptive map, what was most revealing was what the systematic review did not find, rather than what it did. As can be seen in section 3.2, the bulk of the studies were of US origin, with only 13 loosely relevant studies undertaken in the UK. Whilst the majority of the studies did focus loosely on formative assessment (40 in total), only ten of them involved portfolios. Moreover, whilst seven studies in the UK featured formative assessment to some degree, no UK studies focused on portfolio assessment. A likely explanation for this is the recent introduction of the CEP (arrangements for which were made only in 1999).

The User Review Team stands by the statement made in their original bid (and reiterated in section 1.1.1) therefore, that whilst the UK system has only recently started to explore the possibilities of portfolios, profiles, reflective journals and dialogic journals for purposes of assessment and professional development, the US systems have been dependent upon them for some time. That the two data-extracted studies were both conducted in the US suggests a real lack of good UK studies.

An unfortunate absence is that of the controlled trial. Whilst only 17 studies involved a controlled trial, neither of the data-extracted studies featured this type of research. It would have been useful to have had a study that started off with two groups together, with one exposed to a portfolio component or similar, and the other not.

### **5.1.5 Findings drawn from the map**

Most studies of at least loose relevance to the review question were conducted in the US (35: 60.3%). The second largest exponent of studies was the UK (13: 23.4%). These figures may, however, reflect bias within the bibliographic database sources searched towards reports published within both of these countries. The exact physical settings of these studies were evenly conducted at primary and secondary schools (22: 24.4% for each).

Learners of various types were the overwhelming population focus of the 58 studies. With inclusion Criterion 5 specifying trainee teachers, it comes as no surprise that a number of these studies also focused on teaching staff. The majority of the studies reported on mixed sex participants (50: 90.9%) but a small number were undertaken using only female trainees (5: 9%), whilst none were undertaken involving only males.

The loading of gender in these studies may be an indication of dated research in the field, for the absence of studies using exclusively male participants is perhaps surprising in light of the attention males have received in recent years regarding assessment and attainment (Stobart and Gipps, 1997). On the other hand, with only 14% of primary school teachers reported to be male, and only a slightly larger percentage of males reported to be secondary teachers, it is perhaps not surprising that five studies deal exclusively with females and none with males. The gender loading may reflect the average loading of an ITT course, or may simply have been accidental (i.e. no males happened to be on the ITT courses selected for investigation).

The majority of studies were evaluative (33: 56.8%) and therefore more suited than other types to answering the review question. Development of methodology studies, which might also have been useful, were relatively few (6: 10.3%). These studies focused principally on the development of the assessment instrument, drawing on empirical research to construct a model of assessment.

## **5.2 Strengths and limitations of this systematic review**

### **5.2.1 Strengths of this systematic review**

The main strength of this review is that it identifies a number of factors affecting the use of portfolios as a formative assessment tool. However, it also reveals the need for significant further research into an area that has already received much recent attention, namely formative assessment. The general conclusions it draws from the data-extracted studies are informative in relation to portfolios, if not of high

trustworthiness. More reliable and carefully conceived studies need to be conducted, however, to inform the UK education system better about the use of portfolios and other means of formative assessment of trainee teachers.

## 5.2.2 Limitations of this review

The main limitation of this review is that it cannot confidently recommend, without further research being undertaken, any successful models of formative assessment, nor can it offer any further insight into the debate of what constitutes effective formative assessment practices. Problems may stem from the review question itself, as well as from the research field. A systematic EPPI-Centre education review typically takes twelve months to complete; this particular review was undertaken in less than three months with a relatively small URT. Time to explore and change the question under review was, therefore, equally limited. Had more time been spent on exploring the various studies on formative assessment initially uncovered, it might have been possible to generate a different question with the potential to offer a richer outcome.

No research was uncovered that engaged in the debate about 'gendered' assessments, nor were any studies unearthed that contribute to the debate about structuralist models of learning (see Chapter 1).

A further limitation was that the map of research reported in Chapter 3 included three studies beyond the geographical inclusion criteria.

## 5.3 Implications

### 5.3.1 Policy

Evidence suggests that what appears to be a general move in the UK towards the incorporation of US-style reflective journals within ITT courses is potentially an advantageous move. They appear to increase professional and personal growth and allow teacher educators further insight into the depth of their trainee teachers' pedagogical knowledge. The decision of many higher education institutions (HEIs), such as Anglia Polytechnic University, to introduce a portfolio component to their ITT course, is supported by encouraging (but contestable) evidence that they may provide a suitable way forward. In the current push for 'authentic assessment', portfolios may offer 'the promise of identifying both a broader and more in-depth picture of an emerging teacher's thinking and behaviour than other, more traditional forms of assessment' (Berg and Curry, 1997, p 84).

### 5.3.2 Practice

A corollary of the increased use of portfolios appears to be heightened confidence, metacognition and reflection, as well as professional and personal development in trainees. The introduction of portfolios may have positive implications with regard to the transition from good studentship to good practice in the classroom. 'In today's rapidly expanding field of hires, an assessment model that can shape professional

growth is very useful in creating the seamless fabric between preservice training and practice into the classroom' (Berg and Curry 1997:84).

### 5.3.3 Research

The absence of trustworthy research in all areas of ITT assessment, especially formative and portfolio assessment, is palpable. If portfolios are to become increasingly widespread for assessment purposes, their validity and reliability, and their ability to satisfy a more 'authentic assessment', must first be proven. There is also a noticeable absence of research that engages either the debate about 'gendered' assessments or the debate about structuralist models of learning.

## 5.4 Additional points

An EPPI-Centre systematic review would normally be scheduled as a twelve-month project. Because there are no dispensable stages in the process, this particular review had to condense twelve months work into three months. No stage could be bypassed. Putting this into operation was difficult and generated a large amount of work. One of the options available to the User Review Team in order to make the review more manageable was disengagement from the EPPI-Centre process.

Indeed, following the EPPI-Centre process resulted in a relatively small yield of usable studies. One of the available options was to take a step back and depart from the process at the descriptive mapping stage prior to data-extraction. This would have meant the presentation of more findings and a larger review, drawing on more studies. In order to address this issue and reach agreement, the URT scheduled a meeting at which they agreed to continue in alignment with the EPPI-Centre. Some of the reasons for this included the following:

- *The impossibility of making the review exhaustive:* Because of the limited timeframe, performing an exhaustive review was not possible. The need for a systematic review, therefore, was great. If the review was neither exhaustive nor fully systematic, it could not be informative. If the review was not exhaustive but was replicable and reliable, at least a large surface area of the field of research would be covered. Furthermore, what had and had not been synthesised would be explicit.
- *The need for evidence-informed policy and practice:* Policy and practice in education should ideally be informed by research that is not just empirical but also rigorous and trustworthy. By disengaging from the EPPI-Centre process, the URT would have perpetuated the view that was emerging of weak research in the field of assessment.
- *The need for transparency:* The team resolved that there would be no justification for making only part of the process transparent, and leaving the latter, perhaps most important stages of the review process, obscure or potentially subjective.

- *The need for standardised quality assessment:* To follow the EPPI-Centre process to the descriptive mapping stage and then depart from it would have denied the URT the opportunity to weigh the evidence in the studies in a standardised way. The weight of evidence stage of the EPPI-Centre process allows transparent and replicable judgements to be made about the quality of the studies, the appropriateness of their research design, and the relevance of their focus. An overall weight is then awarded. It is on this basis that recommendations are made.
- *Not wishing to distort the field and undermine keywording:* The URT were confident that the retrieval system they had created with keywords could accurately describe the field and a search using it would include only relevant studies. Ignoring the keywording could have distorted the nature and extent of the field and resulted in weak research being perpetuated.
- *The need for synthesising primary research:* At the end of the EPPI-Centre review process, the evidence is synthesised. If the data had not been extracted in a systematic way, ensuring consistent extraction across the board, the synthesis of evidence might have been less reliable.
- *The need for detailed characterisation of studies:* The difference between keywording and data-extraction is that the former employs limited coding, describing the characteristics of the studies. Data-extraction, however, describes in more depth, assessing findings and methodological quality.
- *The need for consensus judgements:* In continuing the systematic process throughout, the URT made use of the utility of 'double reviewing' at the data-extraction stage. Moreover, aligning with the EPPI-Centre during the latter stages of the review opened the extraction of data up to another stage of peer reference for the purposes of quality assurance.

In short, had the User Review Team departed from the EPPI-Centre process, the review may indeed have synthesised more studies, but it might also have been misleading and would certainly not have had the same level of accountability and replicability.

## 6. REFERENCES

### 6.1 Studies included in the descriptive map

*\*Indicates studies included in the in-depth review.*

Anderson RS, DeMeulle L (1998) Portfolio use in twenty-four teacher education programs. *Teacher Education Quarterly* **25**: 23-31.

Arroya AA, Sugawara AI (1993) A scale of student teaching concerns (SSTC). *Early Child Development and Care* **93**: 41-56.

Ashcroft K, Tann S (1988) Beyond building better checklists: staff development in a school experience program. *Assessment and Evaluation in Higher Education* **13**: 61-72.

Baillie LE (1994) Paradigms lost: the role of reason in reflection. *Journal of Educational Thought* **28**: 245-259.

Basom M (1994) Pre-service identification of talented teachers through non-traditional measures: a study of the role of affective variables as predictors of success in student teaching. *Teacher Education Quarterly* **21**: 47-57.

\*Berg M, Curry J (1997) Portfolios: what can they tell us about student teacher performance? *Social Studies Review* **36**: 78-84.

Brooker R (1998) Improving the assessment of practice teaching: a criteria and standards framework. *Assessment and Evaluation in Higher Education* **23**: 5-24.

Brucklacher B (1998) Cooperating teachers' evaluations of student teachers: all 'A's'? *Journal of Instructional Psychology* **25**: 67-72.

Buchanan D, Jackson S (1998) Supporting self-evaluation in Initial Teacher Education. *Scottish Educational Review* **30**: 110-124.

Calderhead J, James C (1992) Recording student teachers' learning experiences. *Journal of Further and Higher Education* **16**: 3-12.

Cassidy JW (1993) A comparison between students' self-observation and instructor observation of teacher intensity behaviours. *Bulletin of the Council for Research in Music Education* **115**: 15-29.

Costigan AT III (2000) Teaching the culture of high stakes testing: listening to new teachers. *Action in Teacher Education* **23**: 35-42.

Di X, Lee SJ (2000) The impact of an alternative evaluation for group work in teacher education on students' professional development. *Action in Teacher Education* **22**: 43-58.

- Dutt K (1997) Assessing student teachers: the promise of development portfolios. *Teacher Educator* **32**: 201-215.
- Dutt-Donter K, Gillman DA (1998) Students react to portfolio assessment. *Contemporary Education* **69**: 159-165.
- Edwards A (1997) Guests bearing gifts: the position of student teachers in primary school classrooms. *British Educational Research Journal* **23**: 27-37.
- Everton TC, Hopper BA, Thwaites EA (1999) Student teachers in primary schools: the views of pupils. *Teacher Development* **3**: 65-78.
- Fahey PA, Fingon JC (1997) Assessing oral presentations of student teacher showcase portfolios. *Educational Forum* **61**: 354-359.
- Fishman AR, Raver EJ (1989) 'Maybe I'm just NOT teacher material': dialogue journals in the student teaching experience. *English Education* **21**: 92-102.
- Fitzgibbon A (1994) Self-evaluative exercises in initial teacher education. *Irish Educational Studies* **13**: 145-164.
- Fueyo V, Brown TW, Hansen JH (1998) Pre-professional accomplished practice indicators: a metric for learning to teach. *Action in Teacher Education* **19**: 131-148.
- Gerlach GJ, Milward RE (1989) A new perspective for strengthening teaching skills: pre-teacher assessment. *Foreign Language Annals* **22**: 145-148.
- Goos M, Moni K (2001) Modelling professional practice: a collaborative approach to developing criteria and standards-based assessment in pre-service teacher education courses. *Assessment and Evaluation in Higher Education* **26**: 73-88.
- Hartsough CS, Perez KD, Swain CL (1998) Development and scaling of a preservice teacher rating instrument. *Journal of Teacher Education* **49**: 132-139.
- Hayes D (1999) A matter of being willing? Mentors' expectations of student primary teachers. *Mentoring and Tutoring* **7**: 67-79.
- Ingersoll GM, Kinman D (2002) Development of a teacher candidate performance self-assessment instrument. *Action in Teacher Education* **23**: 4-9.
- Isele F (1992) The role of research in evaluating the effectiveness of an elementary teacher's performance: a national study of evaluation criteria. *Contemporary Education* **63**: 130-136.
- Johnson CJ, Shewan CM (1988) A new perspective in evaluating clinical effectiveness: the UWO clinical grading system. *Journal of Speech and Hearing Disorders* **53**: 328-340.
- Johnson J (1999) Professional teaching portfolio: a catalyst for rethinking teacher education. *Action in Teacher Education* **21**: 37-49.

- Jones M (2000) Becoming a secondary teacher in Germany: a trainee perspective on recent developments in Initial Teacher Training in Germany. *European Journal of Teacher Education* **23**: 65-76.
- Kain DL (1999) On exhibit: assessing future teachers' preparedness. *Action in Teacher Education* **21**: 10-23.
- Klenowski V (2000) Portfolios: promoting teaching. *Assessment in Education: Principles, Policy and Practice* **7**: 215-236.
- Lyle S (1996) The education of reflective teachers? A view of a teacher educator. *Journal of Teacher Development* **5**: 4-11.
- McLaughlin HJ (1991) The reflection on the blackboard: student teacher self-evaluation. *International Journal of Qualitative Studies in Education* **4**: 141-159.
- Martin S (1997) Two models of educational assessment: a response from Initial Teacher Education: if the cap fits... *Assessment and Evaluation in Higher Education* **22**: 337-343.
- Naizer GL (1997) Validity and reliability issues of performance portfolio assessment. *Action in Teacher Education* **18**: 1-9.
- Preece PFW (1993) The assessment of teaching practice performance. *Research in Education* **49**: 23-27.
- Rakow SJ (1999) Involving classroom teachers in the assessment of preservice intern portfolios. *Action in Teacher Education* **21**: 108-115.
- Ramanathan H, Wilkins-Canter E (2000) Preparation of cooperating teachers as evaluators in early field experiences. *Action in Teacher Education* **22**: 101-111.
- Reis NK, Villaume SK (2002) The benefits, tensions and visions of portfolios as a wide-scale assessment for teacher education. *Action in Teacher Education* **23**: 10-17.
- Reyes DJ, Isele Jr F (1990) What do we expect from elementary student teachers? A national analysis of rating forms. *Action in Teacher Education* **12**: 8-13.
- Riner PS (1993) The reality of failure: two case studies in student teaching. *Teacher Education and Practice* **9**: 39-48.
- Ryan LB, Krajewski JJ (2002) The journey toward becoming a standards driven and performance based teacher preparation program: one college's story. *Action in Teacher Education* **23**: 59-68.
- Scruggs TE, Mastropieri MA (1993) The effects of prior field experience on student teacher competence. *Teacher Education and Special Education* **16**: 303-308.
- Shannon DM, Boll M (1998) State-mandated assessment of preservice teachers in Alabama. *Professional Educator* **21**: 31-44.

Sharp S (1997) A factorial study of student performance in initial education. *Studies in Higher Education* **22**: 363-378.

Simbo FK (1989) The effects of micro teaching on student teachers' performance in the actual teaching practice classroom. *Educational Research* **31**: 195-200.

Simmons C, Wild P (1992) New forms of student teacher learning. *Educational Review* **44**: 31-40.

Smith PL (2001) Using multimedia portfolios to assess preservice teacher and P-12 student learning. *Action in Teacher Education* **22**: 28-39.

Stroh L (1991) High school student evaluation of student teachers: how do they compare with professionals? *Illinois School Research and Development* **27**: 81-92.

Sumison J, Fleet A (1996) Reflection: can we assess it? Should we assess it? *Assessment and Evaluation in Higher Education* **21**: 121-130.

Tse K-S, Chung C-M (1995) The relationship between personality-environment congruency and teaching performance in student teachers. *Chinese University Education Journal* **23**: 69-81.

Unrau NJ, McCallum RD (1996) Evaluating with KARE: the assessment of student teacher performance. *Teacher Education Quarterly* **23**: 53-76.

Vaughan G (1992) Profiling: a mechanism for professional development of students? *Cambridge Journal of Education* **22**: 163-175.

Wardlow G (1987) The teaching performance of graduates of teacher education programs in vocational and technical education. *Journal of Vocational Education Research* **12**: 57-68.

Wepner SB (1997) 'You can never run out of stamps': electronic communication in field experiences. *Journal of Educational Computing Research* **16**: 251-268.

Wile JM (1999) Professional portfolios: the 'talk' of the student teaching experience. *Teacher Educator* **34**: 215-231.

\*Willis EM, Davies MA (2002) Promise and practice of professional portfolios. *Action in Teacher Education* **23**: 18-27.

## 6.2 Full bibliography after first screening stage

Adams A (1995) A portfolio assessment procedure for preservice teachers for admission to the professional year: the project so far. Paper presented at the Annual Meeting of the Mid-South Educational Research Association. Nashville, TN, USA: November.

Alban T, Proffitt TD, SySantos C (1998) Defining performance based assessment within a community of learners: the challenge and the promise. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. New Orleans, LA, USA: February 25-28.

Anderson JO, Bacher D, Baer M (2001) Using portfolio assessment to study classroom assessment practice. Paper presented at the Annual Meeting of the American Educational Research Association. Seattle, WA, USA: April 10-14.

Anderson RS, DeMeulle L (1998) Portfolio use in twenty-four teacher education programs. *Teacher Education Quarterly* **25**: 23-31.

Arroya AA, Sugawara AI (1993) A scale of student teaching concerns (SSTC). *Early Child Development and Care* **93**: 41-56.

Artzt AF, Armour-Thomas E (2002) *Becoming a Reflective Mathematics Teacher: A guide for observations and self-assessment. Studies in Mathematical Thinking and Learning Series*. New Jersey, USA: Lawrence Erlbaum Associates, Inc.

Ashcroft K, Tann S (1988) Beyond building better checklists: staff development in a school experience program. *Assessment and Evaluation in Higher Education* **13**: 61-72.

Ashelman P (1996) Portfolio assessment: an early childhood and family studies department model. In: Hujala E (Ed) *Childhood Education: International Perspectives*. Finland: Association for Childhood Education International, Oulu University.

Backlund P, Black D (1988) Team evaluation of student teacher communication competencies through videotape review. Paper presented at the Annual Meeting of the Association of Teacher Educators. San Diego, CA, USA: February 13-17.

Baillie LE (1994) Paradigms lost: the role of reason in reflection. *Journal of Educational Thought* **28**: 245-259.

Baker BR (1996) The role of the professional development school to prepare teachers of young children. A description of two models: the Hillcrest Professional Development School and the La Vega Primary School. Texas, USA: Baylor University.

Baker M, Hedges LE (1991) The identification of components underlying the summative evaluation of student teachers and component influence upon students' final grade. A summary of research. Ohio, USA: Ohio University.

Bare C, Hill S (1992) Perceptions of Secondary Student Teaching Experiences. Monograph Number 13. Canada: University of Saskatchewan.

Bartell CA, Candace MJA (1998) Portfolio conversation: a mentored journey. *Teacher Education Quarterly* **25**: 129-139.

- Bartell CA, Kaye C (1998) Teaching portfolios and teacher education. *Teacher Education Quarterly* **25**: 5-8.
- Basom M (1994) Pre-service identification of talented teachers through non-traditional measures: a study of the role of affective variables as predictors of success in student teaching. *Teacher Education Quarterly* **21**: 47-57.
- Baturo A, Nason R (1996) Student teachers' subject matter knowledge within the domain of area measurement. *Educational Studies in Mathematics* **31**: 235-268.
- Beardon T, Reiss M (1990) Profiling in Initial Teacher Training: building the bridge. *Links* **16**: 16-19.
- Belcher C, Garten TR (1998) Student teaching assessment redesign. Paper presented at the Annual Conference of the Association of Teacher Educators. Dallas, TX, USA: February 13-17.
- Berg M, Curry J (1997) Portfolios: what can they tell us about student teacher performance? *Social Studies Review* **36**: 78-84.
- Biddle J (1992) Portfolio development in teacher education and educational leadership. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. San Antonio, TX, USA: February 25-28.
- Bitner T, Kratzner R (1995) A primer on building teacher evaluation instruments. Paper presented at the Annual Meeting of the Midwest Educational Research Association. Chicago, IL, USA: October 11-14.
- Blome AC (1987) Collaborative approach: developing a teacher evaluation process. *NASSP Bulletin* **71**: 36-43.
- Briscoe C (1989) Assessment of performance of field experiences at Clemson University. South Carolina, USA: Clemson University.
- Brooker R (1998) Improving the assessment of practice teaching: a criteria and standards framework. *Assessment and Evaluation in Higher Education* **23**: 5-24.
- Brown S (2000) How student teachers think about their classroom teaching. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA, USA: April 24-28.
- Brucklacher B (1998) Cooperating teachers' evaluations of student teachers: all 'A's'? *Journal of Instructional Psychology* **25**: 67-72.
- Buchanan D, Jackson S (1998) Supporting self-evaluation in Initial Teacher Education. *Scottish Educational Review* **30**: 110-124.
- Byra M (1994) Supervisory conferences: promoting inquiry and reflection in preservice teachers. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA, USA: April 4-8).

- Calderhead J, James C (1992) Recording student teachers' learning experiences. *Journal of Further and Higher Education* **16**: 3-12.
- Capie W (1979) *Selected Investigations of the Reliability of the Teacher Performance Assessment Instruments (A Working Paper)*. Atlanta, GA, USA: Department of Education.
- Capie W (1980) Using science achievement to validate student teachers competencies. Paper presented at the Annual Meeting of the National Association for Research in Science Teaching. 53rd, Boston, MA, USA: April 11-13.
- Carney JM, Jay JK (2002) Translating theory into practice: the dilemmas of teacher portfolios. Paper presented at the Annual Meeting of the Southwest Educational Research Association. 25th, Austin, TX, USA: February 14-16.
- Cassidy JW (1993) A comparison between students' self-observation and instructor observation of teacher intensity behaviours. *Bulletin of the Council for Research in Music Education* **115**: 15-29.
- Cellucci PMC, Ragland GG (1988) Competency evaluations of prospective special education teachers. South Carolina, USA. ERIC document number ED302548.
- Christensen PS (1988) The nature of feedback student teachers receive in post-observation conferences with the university supervisor: a comparison with O'Neal's study of cooperating teacher feedback. *Teaching and Teacher Education* **4**: 275-286.
- Clay D, Cohen M (1998) Assessment in restructured teacher preparation. *Restructuring Texas Teacher Education Series* **4**. Texas: USA.
- Coll RK, Taylor N (2002) Assessment of work-based learning: some lessons from the teaching profession. *Asia-Pacific Journal of Co-operative Education* **3**: 5-12.
- Cooper CC (1988) Design for excellence in teacher education: a developmental model for student assessment, selection and retention. *Journal of Teacher Education* **33**: 15-20.
- Copenhaver R, Waggoner JE (1997) Promoting preservice teachers' professional growth through developmental portfolios. *Teacher Educator* **33**: 103-111.
- Costigan AT III (2000) Teaching the culture of high stakes testing: listening to new teachers. *Action in Teacher Education* **23**: 35-42.
- Crotty T, Allyn D (2001) Evaluating student reflections. Wisconsin: USA. ERIC document number ED459174.
- Darling-Hammond L, Snyder J (2000) Authentic assessment of teaching in context. *Teaching and Teacher Education* **16**: 523-545.
- Darling-Hammond L, Wise AE, Klein SP (1999) *A License to Teach: Raising standards for teaching*. San Francisco, CA, USA: Jossey-Bass, Inc.

Di X, Lee SL (2000) The impact of an alternative evaluation for group work in teacher education on students' professional development. *Action in Teacher Education* **22**: 43-58.

Diamantes T (1996) Using portfolios to assess student teacher progress. Kentucky: USA. ERIC document number ED405360.

Diez ME (1998) Changing the Practice of Teacher Education: Standards and assessments as a lever for change. Washington, DC, USA: AACTE (American Association of Colleges for Teacher Education) Publications.

Diez ME, Hass JM (1997) No more piecemeal reform: using performance-based approaches to rethink teacher education. *Action in Teacher Education* **19**: 17-26.

Drake P (2002) Testing teachers: perspectives on a pilot numeracy skills test in England. *Assessment in Education: Principles, Policy and Practice* **9**: 205-220.

Dutt-Donter K (1997) Assessing student teachers: the promise of development portfolios. *Teacher Educator* **32**: 201-215.

Dutt-Donter K, Gillman DA (1998) Students react to portfolio assessment. *Contemporary Education* **69**: 159-165.

Dybdahl CS, Shaw DG (1997) Teacher testing: reason or rhetoric? *Journal of Research and Development in Education* **30**: 248-254.

Ediger M (1987) Evaluating Student Teacher Performance. Kirksville, MO, USA: Truman State University.

Edwards A (1997) Guests bearing gifts: the position of student teachers in primary school classrooms. *British Educational Research Journal* **23**: 27-37.

Emme MJ (1996) Three dimensional assessment and the art of portfolio building. *Art Education* **49**: 66-72.

Everton TC, Hopper BA (1999) Student teachers in primary schools: the views of pupils. *Teacher Development* **3**: 65-78.

Fahey PA, Fingon JC (1997) Assessing oral presentations of student teacher showcase portfolios. *Educational Forum* **61**: 354-359.

Fairbrother RW (2000) What is happening in assessment? What should be happening? Paper presented at the Annual Conference of the British Educational Research Association. Cardiff University, Wales: September 7-9.

Fennell HA (1993) Student teachers' evaluation of their preparation for internship course: a case study. Ontario: Canada. ERIC document number ED361335.

Fishman AR, Raver EJ (1989) 'Maybe I'm just NOT teacher material': dialogue journals in the student teaching experience. *English Education* **21**: 92-102.

Fitzgibbon A (1994) Self-evaluative exercises in Initial Teacher Education. *Irish Educational Studies* **13**: 145-164.

Fleming LE (1999) Portfolios in teacher education. Paper presented at the Annual Meeting of the Mid-Western Educational Research Association. Chicago, IL, USA: October 13-16.

Foehr RP (1992) Validating secondary teachers: programs and research at Illinois State University. ERIC document number ED353598.

Freeland K, Willis M (1999) *Integrated College Methods Courses*. Morehead, KY, USA: Morehead State University.

Freeze CR (1988) Assessment of performance in teaching of student teachers. South Carolina: USA. ERIC document number ED289839.

Freidus H (1998) Narrative practices: portfolios in teacher education. Paper presented at the Annual Meeting of the American Educational Research Association. San Diego, CA, USA: April 13-17.

Fueyo V, Brown TW (1998) Pre-professional accomplished practice indicators: a metric for learning to teach. *Action in Teacher Education* **19**: 131-148.

Garrett JL, Dudt K (1998) Using video conferencing to supervise student teachers. In: *SITE 98: Society for Information Technology & Teacher Education International Conference (9th, Washington, DC, March 10-14, 1998) Proceedings*. Norfolk, VA, USA: Association for the Advancement of Computing in Education (AACE).

Gatlin L, Jacob S (2002) Standards-based digital portfolios: a component of authentic assessment for preservice teachers. *Action in Teacher Education* **23**: 28-34.

Georgi D, Crowe J (1998) Digital portfolios: a confluence of portfolio assessment and technology. *Teacher Education Quarterly* **25**: 73-84.

Gerlach GJ, Milward RE (1989) A new perspective for strengthening teaching skills: pre-teacher assessment. *Foreign Language Annals* **22**: 145-148.

Goos M, Moni K (2001) Modelling professional practice: a collaborative approach to developing criteria and standards-based assessment in pre-service teacher education courses. *Assessment and Evaluation in Higher Education* **26**: 73-88.

Grant GE, Huebner TA (1998) The portfolio question: a powerful synthesis of the personal and professional. *Teacher Education Quarterly* **25**: 33-43.

Green P (1991) No surprises: the guidelines approach to evaluation of pre-service teachers. ERIC document number ED336338.

Guffy T, Mann G (1999) Performance assessment team: a retention program revisited. Paper presented at the Annual Meeting of the Association of Teacher Educators. 79th, Chicago, IL, USA: February 13-17.

Guillaume AM, Yopp HK (1995) Professional portfolios for student teachers. *Teacher Education Quarterly* **22**: 93-101.

Haggard CS, Lang DC (1988) The impact of weighting student teacher evaluation variables. Paper presented at the Annual Meeting of the Association of Teacher Educators. San Diego, CA, USA: February 13-17.

Hambrick-Dixon PJ (1999) Meeting the Challenges of Urban School Reform: Assessment portfolios for teachers' professional development. Paper presented at the Annual Meeting of the American Educational Research Association. Montreal, QC, Canada: April 19-23.

Hanhan SF (1988) A qualitative and qualitatively different format for the evaluation of student teachers. *Action in Teacher Education* **10**: 51-55.

Harlen W, Crick RD (2002) A systematic review of the impact of summative assessment and test on students' motivation for learning. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Harris LB (1997) using performance-based assessment to improve teacher education. Paper presented at the Annual Meeting of the Association of Teacher Educators. 77th, Washington, DC, USA: February 15-17.

Hartsough CS, Perez KD (1998) Development and scaling of a preservice teacher rating instrument. *Journal of Teacher Education* **49**: 132-139.

Hauser J, Zimmerman G (1996) Self-enhancing seminars and experiential journals for cooperating teachers: a professional growth model. In: Baker B (ed) *Spawning New Ideas: A Cycle of Discovery. Conference Proceedings of the Annual International Conference of the Association for Experiential Education (24th, Spokane, WA, September 26-29, 1996)*. Boulder, CO, USA: Association for Experiential Education.

Hayes D (1999) A matter of being willing? Mentors' expectations of student primary teachers. *Mentoring and Tutoring* **7**: 67-79.

Hayes ER, Kretschmann KJ, Berry MA (1994) *Portfolio Assessment in Adult Basic Skills Education. Final Report*. Madison, WI, USA: Madison Area Technical College.

Heyman EL (2002) A comparison of the alternative and traditional licensure programs at Metropolitan State College of Denver. Denver, CO, USA.

Hill DM (2002) Electronic portfolios: teacher candidate development and assessment. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. 54th, New York, NY, USA: February 23-26.

Hill FH (1991) Assessing the relationship between reflective practice, content knowledge and teaching effectiveness of student teachers. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, IL, USA: April 3-7.

Hogge J, Murrell J (1991) Exploring values underlying the assessment of teaching competence: an application of social judgement theory. Paper presented at the Annual Meeting of the Mid-South Educational Research Association. Lexington, KY, USA: November 15.

Holt D (1997) Integrating preparation and practice through a technology-based approach to portfolios for professional development using CD-ROM technology. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. 49th, Phoenix, AZ, USA: February 26-March 1.

Hougham P (1992) Improving student teachers' strategies for asking a range of both high level and low level questions through video evaluations. Florida, USA: Nova University.

Hudson J, Latham AM (1996) Student assessment is quality assessment? *Mentoring and Tutoring* **4**: 47-52.

Hughes M (1993) Relationships among various demographic variables and measures of teacher competence: making sense of a variety of bi-variate correlations. Paper presented at the Annual Meeting of the Southwest Educational Research Association. Austin, TX, USA: January 28-30.

Hunt DE (1967) A situational testing approach to assessing teacher trainees. ERIC document number ED011259.

Hus LA, Bergeron B (1997) Portfolios and Program Assessment: Addressing the challenges of admission to a pre-service program. ERIC document number ED414257.

Ingersoll GM, Kinman D (2002) Development of a teacher candidate performance self-assessment instrument. *Action in Teacher Education* **23**: 4-9.

Isele F (1992) The role of research in evaluating the effectiveness of an elementary teacher's performance: a national study of evaluation criteria. *Contemporary Education* **63**: 130-136.

Johnson CJ, Shewan CM (1988) A new perspective in evaluating clinical effectiveness: the UWO clinical grading system. *Journal of Speech and Hearing Disorders* **53**: 328-340.

Johnson J (1999) Professional teaching portfolio: a catalyst for rethinking teacher education. *Action in Teacher Education* **21**: 37-49.

Jonas JE (1994) The effect of experience upon teacher adaptability among preservice teachers. ERIC document number ED375132.

Jones M (2000) Becoming a secondary teacher in Germany: a trainee perspective on recent developments in Initial Teacher Training in Germany. *European Journal of Teacher Education* **23**: 65-76.

- Kain DL (1999) On exhibit: assessing future teachers' preparedness. *Action in Teacher Education* **21**: 10-23.
- Kemis MR (1990) Teaching potential and its relationship to multiple ratings of student teaching performance and other factors. Paper presented at the Annual Meeting of the Midwestern Educational Research Association. Chicago, IL, USA: October 17-20.
- Klenowski V (2000) Portfolios: promoting teaching. *Assessment in Education: Principles, Policy and Practice* **7**: 215-236.
- Laitsch D (1998) *State Evaluation of Teacher Preservice Programs in Texas and Florida. An Issue Paper*. Washington, DC, USA: AACTE Publications.
- Lamson S, Thomas KR, Aldrich J, King A (2001) Assessing pre-service candidates' web-based electronic portfolios. Missouri: USA. Paper presented at the Annual Meeting of the Association of Teacher Educators. Portland, OR, USA: August 5-8.
- Lester S (1998) Assessment versus self-directed learning: a way forward? DfEE Conference paper on standards and the assessment of capability. London: DfEE.
- Limback ER, Mansfield JB (2002) Student teacher assessment: a multi-dimensional process. *Delta Pi Epsilon Journal* **44**: 50-58.
- Lin Q (2002) Beyond standardisation: testing and assessment in standards-based reform. *Action in Teacher Education* **23**: 43-49.
- Lyle S (1996) The education of reflective teachers? A view of a teacher educator. *Journal of Teacher Development* **5**: 4-11.
- Lyons N (1998) Reflection in teaching: can it be developmental? A portfolio perspective. *Teacher Education Quarterly* **25**: 115-127.
- McConney A, Ayres RR (1998) Assessing student teachers' assessments. *Journal of Teacher Education* **49**: 140-150.
- McGrath I (1996) Participant action plans and the evaluation of teachers' courses. *Edinburgh Working Papers in Applied Linguistics* **7**: 85-99.
- McKay JW (1994) A report on the assessment procedures used in the evaluation of student teaching at the middle school level. Paper presented at the Academic Exchange Conference of the Dalian Institute of Education. 4th, Dalian, China: November 8-11.
- McKinney M (1998) Preservice teachers' electronic portfolios: integrating technology, self-assessment, and reflection. *Teacher Education Quarterly* **25**: 85-103.
- McLaughlin HJ (1991) The reflection on the blackboard: student teacher self-evaluation. *International Journal of Qualitative Studies in Education* **4**: 141-159.
- Mahoney B, Knox J (2000) Slaves to the rhythm: an approach to ITT beyond competence. *Journal of Design and Technology Education* **5**: 265-269.

Marso RN, Pigge FL (1991) The identification of academic, personal, and affective predictors of student teaching performance. Paper presented at the Annual Meeting of the Mid-western Educational Research Association. Chicago, IL, USA: October 16-19.

Martin DV (1987) Use of stimulated recall in videotape analysis of student teacher development. *Action in Teacher Education* **9**: 71-83.

Martin S (1997) Two models of educational assessment: a response from Initial Teacher Education: if the cap fits... *Assessment and Evaluation in Higher Education* **22**: 337-343.

Martin S, Cloke C (2000) Standards for the award of Qualified Teacher Status: reflections on assessment implications. *Assessment and Evaluation in Higher Education* **25**: 183-190.

Mathies B, Uphoff JK (1992) The use of portfolio development in graduate programs. A conference paper. ERIC document number ED343855.

Melnick SA (1989) Cooperating teachers: what do they see in the classroom? Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, USA: March 27-31.

Metzler MW, Tjeerdsma BL (2000) The physical education teacher assessment project. *Journal of Teaching in Physical Education* **19**: 395-556.

Mitchell J (1994) Negotiating the practice of teaching. Paper presented at the Annual Meeting of the Australian Teacher Education Association. 24th, Brisbane, QL, Australia: July 3-6.

Montgomery K (1997) Student teacher portfolios: a portrait of the beginning teacher. *Teacher Educator* **32**: 216-225.

Murray-Harvey R, Slee PT, Lawson MJ, Silins H, Banfield G, Russell A (1999) Under stress: the concerns and coping strategies of teacher education students. *European Journal of Teacher Education* **23**: 19-35.

Nagel NG (1993) Authentic learning - authentic assessment: let's begin with tomorrow's teachers. ERIC document number ED412184.

Naizer GL (1997) Validity and reliability issues of performance portfolio assessment. *Action in Teacher Education* **18**: 1-9.

Nweke WC (1991) What type of evidence is provided through the portfolio assessment method? Paper presented at the Annual Meeting of the Mid-South Educational Research Association. 20th, Lexington, KY, USA: November 12-15.

Orland-Barak L (2002) The impact of the assessment of practice teaching on beginning teaching: learning to ask different questions. *Teacher Education Quarterly* **29**: 99-122.

- Parkes KR (1989) The performance and subsequent employment status of PGCE students: academic and attitudinal predictors. *British Educational Research Journal* **15**: 231-248.
- Peters WH (2000) Through the looking-glass portfolio: the journey of preservice teachers in becoming reflective practitioners. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA, USA: April 24-28.
- Pickle JG, Perl MF (1991) Analysing content on student teacher evaluation forms: a format for collaborative discussion. *Critical Issues in Teacher Education* **1**: 23-31.
- Piper CH (2000) Electronic portfolios in teacher education reading methods courses. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA, USA: April 24-28.
- Pissalidis C, Walker T (1998) Observational methods for evaluating changes in student teaching as a result of a large scale teacher intervention program. Paper presented at the Annual Meeting of the American Educational Research Association. San Diego, CA, USA: April 13-17.
- Preece PFW (1993) The assessment of teaching practice performance. *Research in Education* **49**: 23-27.
- Quick BN, Dasovich JA (1994) The role of the supervisor: meeting the needs of early childhood preservice teachers. Paper presented at the Annual Meeting of the Mid-South Educational Research Association. Nashville, TN, USA: November 9-11.
- Rakow SJ (1999) Involving classroom teachers in the assessment of preservice intern portfolios. *Action in Teacher Education* **21**: 108-115.
- Ramanathan H, Benedetti T, Thorson A (1997) Cooperating teachers: to grade or not to grade. A one act play. ERIC document number ED417160.
- Ramanathan H, Wilkins-Canter E (1997) Training for cooperating teachers and university supervisors in their role as evaluators in early field experiences. Paper presented at the Annual Meeting of the Mid-western Educational Research Association. Chicago, IL, USA: October 15-18.
- Ramanathan H, Wilkins-Canter E (2000) Preparation of cooperating teachers as evaluators in early field experiences. *Action in Teacher Education* **22**: 101-111.
- Rasmussen J (1996) Generating more light than heat: a departmental approach to performance assessment. ERIC document number ED395542.
- Reed TM (1987) Computerised observation system (COS) for field experiences. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. Washington, DC, USA: February 13.

- Reis NK, Villaume SK (2002) The benefits, tensions and visions of portfolios as a wide-scale assessment for teacher education. *Action in Teacher Education* **23**: 10-17.
- Reyes DJ, Isele Jr F (1990) What do we expect from elementary student teachers? A national analysis of rating forms. *Action in Teacher Education* **12**: 8-13.
- Riggsby D (1995) Electronic portfolio: assessment, resume or marketing tool? ERIC document number ED387115.
- Riner PS (1993) Evaluating student teaching by multiple lines of evidence. *Teacher Education Quarterly* **20**: 29-47.
- Riner PS (1993) The reality of failure: two case studies in student teaching. *Teacher Education and Practice* **9**: 39-48.
- Rogers GL (1995) The student teaching portfolio. *Music Educators Journal* **82**: 29-31.
- Ross EW (1996) The role of portfolio evaluation in social studies teacher education. *Social Education* **60**: 162-166.
- Rubin RB, Feezel JD (1984) Elements of teacher communication competence: an examination of skills, knowledge and motivation to communicate. Paper presented at the Annual Meeting of the Speech Communication Association. 70th, Chicago, IL, USA: November 1-4.
- Runyan CK, Sparks R (2000) A journey of change: redefining and assessing a multifaceted teacher training program. ERIC document number ED446052.
- Ryan CW (1997) Teacher education field experiences: impact on self-esteem of professional year program interns via electronic portfolios. Paper presented at the Annual Meeting of the Eastern Educational Research Association. 20th, Hilton Head, SC, USA: February 20.
- Ryan JM, Kuhs TM (1993) Assessment of preservice teachers and the use of portfolios. *Theory into Practice* **32**: 75-81.
- Ryan LB, Krajewski JJ (2002) The journey toward becoming a standards driven and performance-based teacher preparation program: one college's story. *Action in Teacher Education* **23**: 59-68.
- Salzman JA (2000) Talking the same language: training mentors to use the Pathwise Performance Assessment (PAP) with induction-year colleagues. Paper presented at the Annual Meeting of the American Educational Research Association. New Orleans, LA, USA: April 24-28.
- Salzman SA (1991) Selecting the qualified: predictors of student teacher performance. Paper presented at the Annual Meeting of the Association of Teacher Educators. New Orleans, LA, USA: February 16-20.

Schallock HD, Myton DV (1988) A new paradigm for teacher licensure: Oregon's demand for evidence of success in fostering learning. *Journal of Teacher Education* **39**: 8-16.

Scriven M (1996) Assessment in teacher education: getting clear on the concept. *Teaching and Teacher Education* **12**: 443-450.

Scruggs TE, Mastropieri MA (1993) The effects of prior field experience on student teacher competence. *Teacher Education and Special Education* **16**: 303-308.

Shannon DM (1991) Teacher evaluation: a functional approach. A conference paper. Alabama: USA.

Shannon DM (1994) An evaluation approach for the development of preservice teachers. A conference paper. ERIC document number ED329582.

Shannon DM, Boll M (1998) State-mandated assessment of preservice teachers in Alabama. *Professional Educator* **21**: 31-44.

Sharp S (1997) A factorial study of student performance in initial education. *Studies in Higher Education* **22**: 363-378.

Sharpe DM (1969) Isolating relevant variables in student teaching assessment. Final report. ERIC document number ED028999.

Siders JA (1988) *Prototype for Automated Teacher Performance Assessment*. Hattiesburg, MS: University of Southern Mississippi.

Simbo FK (1989) The effects of micro teaching on student teachers' performance in the actual teaching practice classroom. *Educational Research* **31**: 195-200.

Simmons C, Wild P (1992) New forms of student teacher learning. *Educational Review* **44**: 31-40.

Slick SK (1997) Assessing versus assisting: the supervisor's roles in the complex dynamics of the student teaching triad. *Teaching and Teacher Education* **13**: 713-726.

Smith ER, Smith PK (1991) Assessment of student teachers: an alternate model. *Teacher Education Quarterly* **18**: 39-46.

Smith PL, Harris CM (2001) Using multimedia portfolios to assess preservice teacher and P-12 student learning. *Action in Teacher Education* **22**: 28-39.

Snyder J, Lippincott A (1998) The inherent tensions in the multiple uses of portfolios in teacher education. *Teacher Education Quarterly* **25**: 45-60.

Snyder JA (1993) Beyond assessment: university and school collaboration in portfolio review and the challenge to program improvement. *Action in Teacher Education* **15**: 55-60.

Stahlhut R (1991) Teacher education: student teacher views versus those of cooperating teachers. Paper presented at the Annual Meeting of the Association of Teacher Educators. 71st, New Orleans, LA, USA: February 16-20.

Stolworthy RL (1987) Assessing and interpreting student progress: evaluating the competency in the preparation of teachers. ERIC document number ED290797.

Stolworthy RL (1987) Managing classroom behavior effectively: measuring the competency in the preparation of teachers. ERIC document number ED282883.

Stolworthy RL (1988) Evaluating teacher competencies: an analysis of the variance existing in different assessments. ERIC document number ED294846.

Stolworthy RL (1988) Preservice teacher education: the application of Scheffe's LSD matrix of multiple comparisons to evaluation efforts. ERIC document number ED299250.

Stolworthy RL (1990) Preservice teacher evaluation. ERIC document number ED328530.

Stolworthy RL (1990) Teacher education: the application of Fisher's LSD matrix in the evaluation of preservice teachers. ERIC document number ED318771.

Stolworthy RL (1990) Teaching competencies in preservice teacher education: a study of the analysis of variance in the evaluative efforts. ERIC document number ED324283.

Stolworthy RL (1991) Methodological applications of theory to practice by preservice secondary school mathematics teachers. ERIC document number ED342721.

Stolworthy RL (1991) Preservice teacher evaluation: an analysis of the variance. ERIC document number ED338574.

Stolworthy RL (1992) The application of academic content to practice by preservice secondary school history and social studies teachers. ERIC document number ED350255.

Stone BA (1998) Problems, pitfalls and benefits of portfolios. *Teacher Education Quarterly* **25**: 105-114.

Stroh L (1991) High school student evaluation of student teachers: how do they compare with professionals? *Illinois School Research and Development* **27**: 81-92.

Struyk LR, McCoy HL (1993) Pre-service teachers' use of videotape for self-evaluation. *Clearing House* **67**: 31-34.

Sudzina MR, Knowles JG (1992) Personal characteristics and contextual conditions of student teachers who 'fail': setting a course for understanding failure in teacher education. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, CA, USA: April 20-24.

Sumison J, Fleet A (1996) Reflection: can we assess it? Should we assess it? *Assessment and Evaluation in Higher Education* **21**: 121-130.

Swetnam L (1997) Teacher candidate portfolios: a continuum of assessment for professional development and institutional accountability. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. 49th, Phoenix, AZ, USA: February 26-March 1.

Tam PTK (1992) Quality control mechanisms for appraising the teaching labour force. *Chinese University Education Journal* **20**: 25-35.

Tanner DE (1987) The theory and practice 'double bind' in student teaching. *Teacher Education and Practice* **4**: 45-49.

Taylor CS, Nolen SB (1995) A question of validity: a model for teacher training in assessment. Paper presented at the Annual Meeting of the National Council on Measurement in Education. San Francisco, CA, USA: April 19-21.

Trube MB, Madden PM (2001) National standards, state goals, and the university's vision to align to provide a framework for the pre-service teacher portfolio. ERIC document number ED454169.

Tse K-S, Chung C-M (1995) The relationship between personality-environment congruency and teaching performance in student teachers. *Chinese University Education Journal* **23**: 69-81.

Unrau NJ, McCallum RD (1996) Evaluating with KARE: the assessment of student teacher performance. *Teacher Education Quarterly* **23**: 53-76.

Vaughan G (1992) Profiling: a mechanism for professional development of students? *Cambridge Journal of Education* **22**: 163-175.

Vavrus M (1999) Preservice teacher performance assessment during student teaching: consequences and pitfalls. Paper presented at the National Symposium of the National Association of State Directors of Teacher Education and Certification. Bellevue, WA, USA: March 25-27.

Veenman S, Leenders Y, Meyer P, Sanders M (1992). Evaluation of a pre-service training programme based on direct instruction. Paper presented at the European Conference on Educational Research (ECER). University of Twente, Enschede, Netherlands: June 22-25.

Verkler KW (2000) Let's reflect: the professional portfolio as a prerequisite for internship. *Action in Teacher Education* **22**: 116-121.

Vezzuto LA (1992) Beyond minimum competence: the appraisal of teaching practice: a proposal submitted to the Teaching Practice Committee, University of Brunei. ERIC document number ED377173.

- Vollmer ML (1993) Teacher assessment: a continuing controversy. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. San Diego, CA, USA: February 24-27.
- Waimon MD (1976) The use of teaching performance tests to assess the ability of prospective teachers to relate lower-order learning in PBTE programs. Paper presented at the Annual Meeting of the American Educational Research Association. 60h, San Francisco, CA, USA: April 19-23.
- Walker L (1992) Perceptions of preservice teacher efficacy. Paper presented at the Annual Meeting of the Mid-south Educational Research Association. Knoxville, TN, USA: November 11-13.
- Wardlow G (1987) The teaching performance of graduates of teacher education programs in vocational and technical education. *Journal of Vocational Education Research* **12**: 57-68.
- Webre E (2000) Portfolio Use in the Reading Practicum. ERIC document number ED440957.
- Webster V, Byrne C (1994) Grading the practicum. Paper presented at the Annual Conference of the Australian Teacher Education Association. 24th, Brisbane, QL, Australia: July 3-6.
- Weiner L, Bloom A, Hall A, Juncker J, Obi R, Pagano A, Rand M (1999) Clarifying the multiple linkages between curriculum and instruction in programs of urban teacher preparation and state educational policy promoting school reform. ERIC document number ED431728.
- Wepner SB (1997) 'You can never run out of stamps': electronic communication in field experiences. *Journal of Educational Computing Research* **16**: 251-268.
- Wiersma W (1983) Assessment of teacher performance; constructs of teacher competencies based on factor analysis of observation data. ERIC document number ED230586.
- Wile JM (1999) Professional portfolios: the 'talk' of the student teaching experience. *Teacher Educator* **34**: 215-231.
- Wilkerson JR (2000) Standards-based accountability as a tool for making a difference in student learning: a state and an institutional perspective on standards-based accountability. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. 52nd, Chicago, IL, USA: February 26-29.
- Williams JL (1995) Differences between cooperating teachers and student teachers in their assessment of student teacher performance: potential threats to a successful relationship. Paper presented at the Annual Meeting of the Association of Teacher Educators. 75th, Detroit, MI, USA: February 18-22.

Willis EM, Davies MA (2002) Promise and practice of professional portfolios. *Action in Teacher Education* **23**: 18-27.

Wolf K, Dietz M (1998) Teaching portfolios: purposes and possibilities. *Teacher Education Quarterly* **25**: 9-22.

Zeichner K, Wray S (2001) The teaching portfolio in US teacher education programs: what we know and what we need to know. *Teaching and Teacher Education* **17**: 613-621.

Zheng B, Webb L (2000) A new model of student teacher supervision: perceptions of supervising teachers. Paper presented at the Annual Meeting of the Mid-south Educational Research Association. Bowling Green, KY, USA: November 15-17.

### **6.3 Other references used in the text of the report**

Adams A (1995) A Portfolio Assessment Procedure for Preservice Teachers for Admission to the Professional Year: The project so far. Paper presented at the Annual Meeting of the Mid-South Educational Research Association. Nashville, TN, USA: November.

Bitner T, Kratzner R (1995) A primer on building teacher evaluation instruments. Paper presented at the Annual Meeting of the Midwest Educational Research Association. Chicago, IL, USA: October 11-14.

Chambers G, Roper T (2000) Why students withdraw from Initial Teacher Training. *Journal of Education for Teaching* **26**: 25-44.

EPPI-Centre (2002) *Review Guidelines for Extracting Data and Quality Assessing Primary Studies in Educational Research* (version 0.9.5). London: EPPI-Centre, Social Science Research Unit.

EPPI-Centre (2003) *EPPI-Centre Core Keywording Strategy* (version 0.9.6). London: EPPI-Centre, Social Science Research Unit.

Fairbrother RW (2000) What is happening in assessment? What should be happening? Paper presented at British Educational Research Association Annual Conference. Cardiff, Wales: September 7-10.

Harlen W, Crick RD (2002) A systematic review of the impact of summative assessment and test on students' motivation for learning. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Mahoney B, Knox J (2000) Slaves to the rhythm: an approach to ITT beyond competence. *Journal of Design and Technology Education* **5**: 265-269.

Martin S (1997) Two models of educational assessment. *Assessment and Evaluation in Higher Education* **22**: 337-343.

Moyles J, Cavendish S (2002) Male students in primary ITT: a failure to strive, thrive or survive? Paper presented at British Educational Research Association Annual Conference. Exeter: September 11-14.

Reyes DJ, Isele F (1990) What do we expect from elementary student teachers? An analysis of ratings forms. *Action in Teacher Education* **12**: 8-13.

Schön D (1983) *The Reflective Practitioner*. New York, USA: Basic Books.

Stobart G, Gipps C (1997) *Assessment: a teacher's guide to the issues*. 3<sup>rd</sup> edition. London: Hodder and Stoughton.

Teacher Training Agency (2002) *Career Entry Profile for Newly Qualified Teachers*. London: TTA.

Thompson A, Gitlin A (1995) Creating spaces for reconstructing knowledge in feminist pedagogy. *Educational Theory* **45**: 125-150.

Willis EM, Davies MA (2002) Promise and Practice of Professional Portfolios. *Action in Teacher Education* **23**: 18-27.

### ***Electronic media***

National Foundation for Educational Research, entered 16 January 2003 from [www.195.194.2.100/aboutus/amd5.asp](http://www.195.194.2.100/aboutus/amd5.asp)

San Francisco State University website, entered on 13 February 2003 from [www.sfsu.edu/~acadplan/sld002.htm](http://www.sfsu.edu/~acadplan/sld002.htm)

## APPENDIX 2.1: Search strategy for electronic databases

MODELS		FORMATIVE ASSESSMENT		STUDENT TEACHERS		SCHOOL EXPERIENCE
OR Lesson Observation Criteria	AND	OR Student Teacher Evaluation	AND	OR Preservice Teachers	AND	OR Student Teaching
OR Measures (individuals)		OR Assessment		OR Student Teachers		OR Practice Teaching
OR Tests		NOT Summative Assessment		OR Beginning Teachers		OR Microteaching
OR Standards		OR Appraisal		OR Education Majors		OR Practicums
		OR Grading				OR Teaching Experience
		OR Formative Assessment				OR School Experience
		NOT Holistic Assessment				
		OR Informal Assessment				
		OR Peer Evaluation				
		OR Performance-Based Assessment				
		OR Measurement				
		OR Portfolio Assessment				
		OR Self-Evaluation (individuals)				
		NOT Student Evaluation				
		OR Vocational Evaluation				
		OR Testing				
		OR Evaluation				
		OR Teacher Competency Testing				

ERIC Yield = 27

ERIC Yield post 1987 = 21

## APPENDIX 2.2: APU Assessment review-specific keywording sheet

<p><b>6b. Type(s) of assessment on which the study focuses</b> (Apply only if you have circled 'assessment' in Section 6) <b>Circle more than one if necessary</b></p> <p>Formative assessment Summative assessment Ipsative assessment Self-assessment Portfolio assessment</p> <p><b>6c. Main agent within assessment on which the study focuses</b> (Apply only if you have circled 'assessment' in section 6) <b>Circle only one</b></p> <p>Assessors (subject of assessment) Assessment (process of assessment) Assessed (object of assessment)</p>	<p><b>8c. Type of learner</b> (Apply only if you have circled 'learners' in Section 8) <b>Circle more than one if necessary</b></p> <p>Undergraduate trainees Post-graduate trainees <b>SCITT trainees</b> GTP trainees RTP trainees</p> <p><b>8d. Type of teaching staff</b> (Apply only if you have circled 'teaching staff' in Section 8) <b>Circle more than one if necessary</b></p> <p>University tutor Supervising teacher</p>	<p><b>9b. Phase of ITT that is the focus of the study</b> <b>Circle only one</b></p> <p>Initial pre-student recruitment School-based experience Exit practices Career Entry Profile Statutory Induction</p> <p><b>10c. Type of research</b> <b>Circle only one</b></p> <p>Empirical Non-empirical</p> <p>Definition of empirical research: empirical – <i>adj.</i> 1 based or acting on observation or experiment, not on theory (<i>The Concise Oxford Dictionary</i>); i.e. being or including a report based on data gathered first-hand</p>
--	---	---

## APPENDIX 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
<i>ABEI 003, Klenowski V (2000) Portfolios: promoting teaching</i>	Details <i>Hong Kong</i>	Assessment	Learners	21 and over <i>assumed</i>	Mixed sex <i>assumed</i>	Primary school  Secondary school	Evaluation: Naturally occurring	Summative assessment	Assessment			School-based experience
<i>ABEI 004, Goos M, Moni K (2001) Modelling professional practice: a collaborative approach to developing criteria and standards-based assessment in pre-service education courses</i>	Details <i>Australia</i>	Assessment  Organisation and management  Teaching and learning	Learners	17-20 <i>deduced</i> 21 and over <i>deduced</i>	Mixed sex <i>deduced</i>	Higher education institution	Description	Formative assessment	Assessment	Under-graduate trainees		

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
<i>ABEI 006, Everton TC (1999) Student teachers in primary schools: the views of pupils</i>	Details UK: England	Assessment  Other topic focus <i>students assessed by the pupils they teach</i>	Learners	17-20 <i>deduced</i> 21 and over <i>deduced</i>	Mixed sex <i>deduced</i>	Higher education institution <i>Homerton Primary school</i>	Evaluation: Naturally occurring	Formative assessment  Self-assessment	Assessed	Under-graduate trainees  Post-graduate trainees		School-based experience
<i>ABEI 008, Hayes D (1999) A matter of being willing? Mentors' expectations of student primary teachers</i>	Details UK: England	Teacher careers	Learners	17-20 <i>assumed</i> 21 and over <i>assumed</i>	Mixed sex <i>assumed</i>	Primary school	Evaluation: Naturally occurring					School-based experience
<i>ABEI 009, Brooker R (1998) Improving the assessment of practice teaching: a criteria and standards framework</i>	Details Australia	Assessment	Learners	17-20 <i>assumed</i> 21 and over <i>assumed</i>	Mixed sex <i>assumed</i>	Secondary school	Evaluation: Researcher-manipulated	Formative assessment	Assessed	Under-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
<i>ABEI 011, Martin S (1997) Two models of educational assessment: a response from Initial Teacher Education: if the cap fits...</i>	Details UK: England	Assessment	Learners	21 and over	Mixed sex <i>Assumed</i>	Secondary school	Review: Other review	Formative assessment	Assessed	Post-graduate trainees		
<i>ABEI 012, Sharp S (1997) A factorial study of student performance in Initial Teacher Education</i>	Details UK: England	Assessment	Learners	17-20 <i>assumed</i> 21 and over <i>assumed</i>	Mixed sex <i>assumed</i>	Higher education institution	Evaluation: Naturally occurring	Formative assessment  Summative assessment	Assessment	Under-graduate trainees		
<i>ABEI 013, Sumison J, Fleet A (1996) Reflection: can we assess it? Should we assess it?</i>	Details Australia	Assessment	Learners	17-20 <i>assumed</i> 21 and over <i>assumed</i>	Mixed sex <i>assumed</i>	Higher education institution  Nursery school	Evaluation: Researcher-manipulated	Formative assessment	Assessed	Under-graduate trainees <i>assumed</i> Post-graduate trainees <i>assumed</i>		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
<i>ABEI 015, Lyle S (1996) The education of reflective teachers? A view of a teacher educator</i>	Details UK: England	Assessment Teaching and learning	Learners Teaching staff	17-20 assumed 21 and over assumed	Female only	Higher education institution  Primary school	Evaluation: Naturally occurring		Assessed		University tutor  Supervising teacher	School-based experience
<i>ABEI 017, Fitzgibbon A (1994) Self-evaluative exercises in Initial Teacher Education</i>	Details Ireland	Assessment	Learners	21 and over	Mixed sex	Higher education institution  Secondary school	Evaluation: Researcher-manipulated	Formative assessment  Self-assessment	Assessment  Assessed	Post-graduate trainees		School-based experience
<i>ABEI 018, Preece PFW (1993) The assessment of teaching practice performance</i>	Details UK: England	Assessment	Learners	21 and over	Mixed sex	Secondary school	Exploration of relationships	Formative assessment	Assessment	Post-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
ABEI 019, Vaughan G (1992) <i>Profiling: a mechanism for professional development of students?</i>	Details UK	Other topic focus <i>ICT in ITT</i>	Learners	17-20 <i>deduced</i>  21 and over <i>deduced</i>	Mixed sex <i>deduced</i>	Higher education institution	Evaluation: Researcher-manipulated			Under-graduate trainees  Post-graduate trainees		School-based experience
ABEI 020, Calderhead J, James C (1992) <i>Recording student teachers' learning experiences</i>	Details UK: England	Teaching and learning	Learners	17-20 <i>deduced</i> 21 and over <i>deduced</i>	Mixed sex <i>deduced</i>	Higher education institution	Description			Under-graduate trainees <i>trainee teacher u/g or p/g not specified</i> Post-graduate trainees <i>trainee teacher u/g or p/g not specified</i>		School-based experience

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
ABEI 021, McLaughlin HJ (1991) <i>The reflection on the blackboard: student teacher self-evaluation</i>	Details USA	Assessment	Learners	21 and over	Female only	Higher education institution Secondary school	Evaluation: Researcher-manipulated	Formative assessment Self-assessment	Assessed	Post-graduate trainees <i>assumed</i>		School-based experience
ABEI 022, Simmons C, Wild P (1992) <i>New forms of student teacher learning</i>	Details UK	Assessment	Learners	17-20 21 and over <i>assumed</i>	Mixed sex <i>assumed</i>	Higher education institution Primary school Secondary school	Evaluation: Naturally occurring	Formative assessment Summative assessment		Post-graduate trainees		School-based experience
ABEI 024, Simbo FK (1989) <i>The effects of microteaching on student teachers' performance in the actual teaching practice classroom</i>	Details Nigeria	Assessment	Learners	17-20 <i>deduced</i>	Mixed sex <i>assumed</i>		Evaluation: Researcher-manipulated		Assessed	Under-graduate trainees <i>deduced</i>		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/ countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 002, Dutt KM (1997) <i>Assessing student teachers: the promise of developmental portfolios</i>	Details USA	Assessment	Learners Teaching staff Other population focus HEI supervisor	17-20 21 and over	Mixed sex <i>assumed</i>	Higher education institution	Description	Formative assessment Self-assessment	Assessment Assessed	Under-graduate trainees <i>assumed</i>		School-based experience
AERIC 005, Dutt-Doner K, Gilman DA (1998) <i>Students react to portfolio assessment</i>	Details USA	Assessment		17-20 <i>assumed</i>	Mixed sex <i>assumed</i>	Higher education institution	Description	Formative assessment	Assessment Assessed	Under-graduate trainees		School-based experience
AERIC 010, Naizer GL (1997) <i>Validity and reliability Issues of performance portfolio assessment</i>	Details USA	Assessment Curriculum	Learners	17-20 21 and over	Female only	Higher education institution Primary school	Evaluation Evaluation: Researcher-manipulated	Formative assessment Self-assessment Portfolio assessment	Assessment	Under-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 013, Reis NK, Villaume SK (2002) <i>The benefits, tensions, and visions of portfolios as a wide-scale assessment for teacher education</i>	Details USA		Learners Teaching staff	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution  Home  Primary school  Special needs school	Evaluation: Researcher-manipulated	Formative assessment  Summative assessment  Ipsative assessment  Self-assessment  Portfolio assessment	Assessment	Under-graduate trainees  Post-graduate trainees	University tutor  Supervising teacher	School-based experience  Exit practices
AERIC 014, Fahey PA, Fingon JC (1997) <i>Assessing oral presentations of student-teacher showcase portfolios</i>	Details USA	Assessment	Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution	Evaluation: Naturally occurring	Formative assessment	Assessment  Assessed	Under-graduate trainees assumed Post-graduate trainees assumed		

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 025, Stroh L (1991) <i>High School student evaluation of Student Teachers: how do they compare with professionals?</i>	Details USA	Assessment Teacher careers	Learners Teaching staff	17-20 assumed 21 and over assumed	Mixed sex	Higher education institution assumed Primary school assumed Secondary school assumed	Evaluation: Researcher-manipulated	Formative assessment	Assessors	Under-graduate trainees  Post-graduate trainees assumed	University tutor  Supervising teacher	School-based experience
AERIC 062, Buchanan D, Jackson S (1998) <i>Supporting self-evaluation in Initial Teacher Education</i>	Details UK: Scotland	Assessment Curriculum	Learners Other population focus ITT students	21 and over	Mixed sex assumed	Secondary school	Evaluation  Evaluation: Researcher-manipulated	Formative assessment	Assessed	Post-graduate trainees		School-based experience
AERIC 067, Brucklacher B (1998) <i>Cooperating teachers' evaluations of student teachers: all 'A's'?</i>	Details USA	Assessment	Learners Teaching staff	17-20 21 and over	Mixed sex	Primary school  Secondary school  Special needs school	Evaluation: Naturally occurring	Formative assessment	Assessors  Assessment  Assessed	Under-graduate trainees  Post-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
<i>AERIC 075, Ashcroft K, Tann S (1988) Beyond building checklists: staff development in a school experience programme</i>	Details UK: England	Assessment	Learners Teaching staff	17-20 21 and over	Mixed sex <i>assumed</i>	Higher education institution  Primary school	Evaluation: Naturally occurring		Assessed			
<i>AERIC 083, Di X, Lee SJ (2000) The impact of an alternative evaluation for group work in teacher education on students' professional development</i>	Details USA- <i>assumed</i>	Assessment	Learners	17-20 21 and over	Mixed sex	Higher education institution	Evaluation: Researcher-manipulated	Formative assessment	Assessment Assessed	Under-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 092, Fishman AR, Raver EJ (1989) 'Maybe I'm just NOT teacher material': dialogue journals in the student teaching experience	Details USA	Assessment	Learners Other population focus <i>The Researcher also a Teacher</i>	17-20 <i>Assumed</i> 21 and over <i>Assumed</i>	Mixed sex <i>Assumed</i>	Higher education institution  Secondary school	Evaluation: Naturally occurring	Formative assessment	Assessment	Post-graduate trainees <i>Assumed</i>		School-based experience
AERIC 101, Baillie LE (1994) <i>Paradigms lost: the role of reason in reflection</i>	Details Canada (east)	Assessment	Learners Teaching staff	17-20 <i>assumed</i> 21 and over <i>assumed</i>	Mixed sex <i>assumed</i>	Higher education institution  Secondary school	Exploration of relationships	Summative assessment  Self-assessment	Assessors Assessment Assessed			School-based experience
AERIC 104, Fueyo V (1998) <i>Pre-professional accomplished practice indicators: a metric for learning to teach</i>	Details USA	Assessment	Learners Teaching staff	17-20 <i>Assumed</i> 21 and over <i>Assumed</i>	Mixed sex <i>Assumed</i>	Higher education institution  Primary school  Secondary school	Evaluation: Naturally occurring	Self-assessment	Assessment Assessed	Post-graduate trainees <i>Assumed</i>		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 105, Edwards A (1997) <i>Guests bearing gifts: the position of student teachers in primary school classrooms</i>	Details UK: England	Assessment Classroom management	Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Primary school	Evaluation: Naturally occurring	Formative assessment	Assessed	Under-graduate trainees		School-based experience
AERIC 112, Berg M, Curry J (1997) <i>Portfolios: what can they tell us about student teacher performance?</i>	Details USA	Assessment Teacher careers Teaching and learning	Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution Primary school Secondary school	Evaluation: Researcher-manipulated	Formative assessment Ipsative assessment Self-assessment Portfolio assessment	Assessment	Under-graduate trainees Post-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
<i>AERIC 115, Jones M (2000) Becoming a secondary teacher in Germany: a trainee perspective on recent developments in initial teacher training in Germany</i>	Details England; Germany	Assessment  Organisation and management  Teacher careers  Teaching and learning	Learners	17-20 <i>assumed</i> 21 and over <i>assumed</i>	Mixed sex <i>assumed</i>	Higher education institution  Primary school  Secondary school	Evaluation: Researcher-manipulated	Summative assessment	Assessed	Under-graduate trainees  Post-graduate trainees	University tutor  Supervising teacher	School-based experience
<i>AERIC 117, Scruggs TE, Mastropieri MA (1993) The effects of prior field experience on student teacher competence</i>	Details USA	Assessment  Curriculum  Teacher careers  Teaching and learning	Learners	17-20  21 and over	Mixed sex	Special needs school	Exploration of relationships	Formative assessment  Summative assessment	Assessed	Under-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 121, Basom M (1994) <i>Pre-service identification of talented teachers through non-traditional measures: a study of the role of affective variables as predictors of success in student teaching</i>	Details USA	Assessment Teacher careers Teaching and learning	Learners	17-20 21 and over	Mixed sex	Higher education institution Home	Development of methodology	Formative assessment Summative assessment	Assessed	Under-graduate trainees		School-based experience Exit practices Statutory Induction
AERIC 122, Smith PL (2001) <i>Using multimedia portfolios to assess preservice teacher and P-12 student learning</i>	Details USA	Assessment	Learners Teaching staff	17-20 <i>Assumed</i> 21 and over <i>Assumed</i>	Female only <i>Assumed</i>	Nursery school Secondary school	Evaluation: Naturally occurring	Formative assessment Self-assessment Portfolio assessment	Assessment	Under-graduate trainees <i>Assumed</i> Post-graduate trainees <i>Assumed</i>	University tutor Supervising teacher	School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 128, Wardlow G (1987) <i>The teaching performance of graduates of teacher education programs in vocational and technical education</i>	Details USA	Teacher careers	Learners Teaching staff	17-20 21 and over	Mixed sex	Higher education institution Secondary school	Description			Under-graduate trainees Post-graduate trainees		School-based experience Exit practices Statutory Induction
AERIC 131, Cassidy J (1993) <i>A comparison between students' self-observation and instructor observation of teacher intensity behaviours</i>	Details USA		Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Primary school	Evaluation: Researcher-manipulated	Formative assessment	Assessed	Under-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 132, Arroyo AA, Sugawara AI (1993) <i>A scale of student teaching concerns (SSTC)</i>	Details USA	Assessment Teacher careers Teaching and learning	Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution	Evaluation: Researcher-manipulated		Assessed			School-based experience
AERIC 136, Reyes JR, Isele F Jr (1990) <i>What do we expect from elementary student teachers? A national analysis of rating forms</i>	Details USA	Assessment Teacher careers		17-20 assumed 21 and over assumed	Mixed sex assumed	Primary school	Development of methodology	Formative assessment Summative assessment	Assessment	Under-graduate trainees Post-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 140, Wile JM (1999) <i>Professional portfolios: the 'talk' of the student teaching experience</i>	Details USA, Luxembourg Germany and Austria	Assessment	Learners Other population focus <i>mentors</i>	17-20 <i>deduced</i> 21 and over <i>deduced</i>	Mixed sex <i>deduced</i>	Higher education institution	Description	Formative assessment  Summative assessment  Self-assessment  Portfolio assessment	Assessment	Under-graduate trainees  Post-graduate trainees	Supervising teacher	School-based experience
AERIC 141, Riner PS, Jones WP (1993) <i>The reality of failure: two case studies in student teaching</i>	Details USA	Assessment  Teacher careers  Teaching and learning	Learners	17-20 <i>inferred</i> 21 and over <i>inferred</i>	Female only <i>inferred</i>		Exploration of relationships	Formative assessment  Summative assessment	Assessed	Under-graduate trainees  Post-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 143, Tse KS, Chung CM (1995) <i>The relationship between personality-environment congruency and teaching performance in student teachers</i>	Details China	Assessment Curriculum Teacher careers Teaching and learning	Learners	17-20 <i>inferred</i> 21 and over <i>inferred</i>	Mixed sex <i>inferred</i>	Primary school	Exploration of relationships	Formative assessment Summative assessment Ipsative assessment	Assessed	Under-graduate trainees Post-graduate trainees		School-based experience
AERIC 145, Isele F (1992) <i>The role of research in evaluating the effectiveness of an elementary teacher's performance: a national study of evaluation criteria</i>	Details USA	Assessment	Learners	21 and over	Mixed sex <i>assumed</i>	Higher education institution Primary school	Description	Formative assessment	Assessment	Under-graduate trainees <i>assumed</i> Post-graduate trainees <i>assumed</i>	University tutor Supervising teacher	School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
<i>AERIC 146, Unrau NJ, McCallum RD (1996) Evaluating with KARE: the assessment of student teacher performance</i>	Details USA	Assessment Teacher careers Teaching and learning	Learners	17-20 21 and over	Mixed sex	Secondary school	Description	Formative assessment Summative assessment	Assessment	Under-graduate trainees Post-graduate trainees		School-based experience
<i>AERIC 148, Ingersoll GM, Kinman D (2002) Development of a teacher candidate performance self-assessment instrument</i>	Details USA	Assessment Teacher careers Teaching and learning	Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution	Development of methodology	Formative assessment	Assessment	Under-graduate trainees Post-graduate trainees		School-based experience Statutory Induction
<i>AERIC 153, Anderson RS, DeMeulle L (1998) Portfolio use in twenty-four teacher education programs</i>	Details USA	Assessment	Learners	17-20 21 and over	Mixed sex	Higher education institution	Description	Formative assessment Self-assessment Portfolio assessment	Assessment	Under-graduate trainees Post-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 154, Kain DL (1999) <i>On exhibit: assessing future teachers' preparedness</i>	Details USA	Assessment	Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Secondary school	Description	Summative assessment	Assessment	Under-graduate trainees  Post-graduate trainees		
AERIC 170, Willis EM, Davies MA (2002) <i>Promise and practice of professional portfolios</i>	Details USA	Assessment	Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution  Primary school  Secondary school	Evaluation: Researcher-manipulated	Formative assessment  Portfolio assessment	Assessment	Under-graduate trainees  Post-graduate trainees		School-based experience
AERIC 172, Rakow SJ (1999) <i>Involving classroom teachers in the assessment of preservice intern portfolios</i>	Details USA	Assessment	Learners  Teaching staff	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution  Primary school  Secondary school	Development of methodology	Formative assessment  Self-assessment  Portfolio assessment	Assessment	Under-graduate trainees  Post-graduate trainees	Supervising teacher	School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 175, Costigan AT III (2000) <i>Teaching the culture of high stakes testing: listening to new teachers</i>	Details USA	Assessment Teacher careers	Teaching staff			Primary school	Evaluation: Naturally occurring	Formative assessment of children Summative assessment of children	Assessors			Statutory Induction
AERIC 178, Wepner SB (1997) <i>'You never run out of Stamps': electronic communication in field experiences</i>	Details USA	Other topic focus <i>Using electronic communications in field experiences</i>	Learners Teaching staff	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution	Evaluation: Researcher-manipulated				University tutor Supervising teacher	School-based experience
AERIC 179, Ryan L, Krajewski JJ (2002) <i>The journey toward becoming a standards driven and performance based teacher preparation program: one college's story</i>	Details USA	Assessment	Learners Teaching staff	21 and over assumed	Mixed sex assumed	Higher education institution	Description	Formative assessment	Assessment	Undergraduate trainees Post-graduate trainees	University tutor	School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 180, Gerlach GJ, Milward RE (1989) <i>A new perspective for strengthening teaching skills: pre-teacher assessment</i>							Description	Formative assessment Ipsative assessment	Assessment	Under-graduate trainees Post-graduate trainees	University tutor	School-based experience
AERIC 185, Hartsough CS (1998) <i>Development and scaling of a preservice teacher rating instrument</i>	Details USA	Assessment Teacher careers Teaching and learning	Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution	Development of methodology	Formative assessment	Assessment	Under-graduate trainees Post-graduate trainees		School-based experience

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 188, Johnson CJ, Shewan CM (1988) <i>A new perspective in evaluation clinical effectiveness: the UWO clinical grading system</i>	Details Canada	Assessment	Learners	17-20 <i>Assumed</i> 21 and over <i>Assumed</i>	Mixed sex <i>Assumed</i>	Higher education institution	Development of methodology	Formative assessment	Assessment	Undergraduate trainees <i>Assumed</i> Post-graduate trainees <i>Assumed</i>	University tutor	
AERIC 190, Ramanathan H, Wilkins-Canter EA (2000) <i>Preparation of cooperating teachers as evaluators in early field experiences</i>	Details USA	Assessment	Learners	21 and over <i>assumed</i>	Mixed sex <i>assumed</i>	Higher education institution  Primary school  Secondary school	Description	Formative assessment	Assessors		University tutor  Supervising teacher	

Appendix 3.1: Keywords of studies included in the in-depth review

Item	In which country/countries was the study carried out? (Please specify.)	What is/are the topic focus/foci of the study?	What is/are the population focus/foci of the study?	Age of learners (years)	Sex of learners	What is/are the educational setting(s) of the study?	Which type(s) of study does this report describe?	Type(s) of assessment on which the study focuses	Main agent within assessment on which the study focuses	Type of learner	Type of teaching staff	Phase of initial teacher training that is the focus of the study
AERIC 195, Johnson J (1999) <i>Professional teaching portfolio: a catalyst for rethinking teacher education</i>	Details USA	Assessment	Learners	17-20 assumed 21 and over assumed	Mixed sex assumed	Higher education institution	Evaluation: Naturally occurring	Formative assessment Portfolio assessment	Assessment	Under-graduate trainees assumed Post-graduate trainees assumed		
AERIC 196, Shannon DM, Boll M (1998) <i>State-mandated assessment of preservice teachers in Alabama</i>	Details	Assessment	Senior management			Higher education institution	Evaluation: Naturally occurring	Summative assessment Portfolio assessment	Assessment		University tutor	Exit practices

## APPENDIX 4.1: Studies excluded from data-extraction

Study
Brucklacher B (1998) Cooperating teachers' evaluations of student teachers: all 'A's'? <i>Journal of Instructional Psychology</i> <b>25</b> (1): 67-72.
Aim and rationale
The aim of the report is to 'describe a study of cooperating teachers' evaluations that resulted in few below-average grades' awarded to trainees (p 67). The rationale for this is that 'researchers have found problems with the ratings made by cooperating teachers', namely the 'halo effect in cooperating teacher's evaluations' (p 67).
Study design
'From the fall of 1994 through the fall of 1996, 662 elementary and high school teachers served as cooperating teachers for the university' (p 68). The university was American, but anonymous in the report. 'Of these, 465 (70%) completed evaluations of their student teachers and returned them to the university' (p 68). Evaluations were based on a 20-item assessment tool, a copy of which is included in the report.
Findings
As predicted in previous studies, few below-average grades were given.
Conclusions
'The cooperating teachers in this study may have assigned higher ratings than were warranted by their student teacher's behaviours (leniency), or they may have attended to global impressions of the student teachers rather than to the specific criteria of the rating scale' (p 69). 'A key element in determining success in student teaching is the relationship between the student teacher and the cooperating teacher' (p 69). Sympathy, empathy, and disparity between cooperating teachers and teacher educators may also load the evaluation. 'The above average ratings in this study may also be the result of a flawed instrument. Making items more behaviourally specific, or training cooperating teachers to use the instrument more objectively might lessen rating errors' (p 69). Other reasons for consistently high grading might stem from the 'need to foster students' self-esteem regardless of academic achievement' and the cultural impact the progressivist and egalitarian doctrines have had on American mainstream culture (p 70). The final conclusion is, 'rater bias, problems with the evaluation instrument, and a progressivist educational paradigm that encouraged high grading regardless of performance are all factors that might have influenced the evaluations examined in this study' (p 70).
Reasons for exclusion
This study, which does not focus on a model or instrument of assessment but focuses instead on the shortcomings of an assessment system already in operation, was keyworded ASSESSORS, ASSESSMENT and ASSESSED in section 6c. ASSESSORS or ASSESSED instead of ASSESSMENT would have been more accurate and avoided recall via the descriptive map. The review question also requires positive valuations to be made of the assessment model or instrument, not negative valuations.

<b>Study</b>
Fishman AR and EJ Raver (1989) 'Maybe I'm just NOT teacher material': dialogue journals In the student teaching experience. <i>English Education</i> <b>21</b> (2): 92-102.
<b>Aim and rationale</b>
No aim is explicitly stated. Through inference, it is possible to identify the aim: to chronicle the experience of dialogue journals between the researcher (a cooperating teacher) and a student teacher. The rationale (again not made explicit) is that their recent experience of these journals was enriching enough to warrant retelling.
<b>Study design</b>
The paper reports, retrospectively, the journal entries of both the student teacher and the cooperating teacher, presenting them as expostulations and replies, so that the two parties engage in ongoing dialogue through their journals. The study does not use samples, data-collection methods, analysis methods or quantitative data.
<b>Findings</b>
The dialogue journals enriched the experience of both the student teacher and the cooperating teacher in a number of stated ways.
<b>Conclusions</b>
'Because journal dialogues allow the student, her cooperating teacher, and/or her field supervisor to highlight, review, analyse, and synthesise what's being learned from a variety of perspectives, they can be used as important instruments for evaluation, offering a multidimensional perspective few other instruments provide' (p 106). It continues, 'they help students assess and appreciate what they have learned, what they are learning, and what they have yet to learn, without the constraints of more generic, pre-packaged instruments. For summative evaluation, dialogue journals can describe and assess what student teachers have achieved and still need to achieve along more dimensions than any behavioural performance, or product checklist can reflect' (p 106).
<b>Reasons for exclusion</b>
The study is more unreliable than the other studies. Firstly, it does not aim either to test an assessment instrument or to determine the benefits and shortcomings of some assessment tool. Instead, it concludes that dialogue journals could be used to assess student teachers, but focuses ultimately on their diagnostic rather than assessment function. The selection of data is unreliable because one of the purposes of the selection is cohesiveness of narrative. No quantitative data are presented. No detail is given in the description of what is taking place in the study. The sample group (excluding the researcher) consists of one individual and, for the purpose of reducing subjectivity (and achieving consensus results), the studies to be data-extracted should have a larger group of participants. The review-specific keywording may have been more accurately keyworded ASSESSOR and/or ASSESSED.

<b>Study</b>
Fitzgibbon A (1994) Self-evaluative exercises in Initial Teacher Education. <i>Irish Educational Studies</i> 13: 145-164.
<b>Aim and rationale</b>
'This paper seeks to discuss the teaching of self-evaluation within one element – that of keeping a journal – of a Higher Diploma in Education course' (p 145). It continues, 'in the present paper, two basic issues are explored. The first is the possible usefulness of the different activities offered in connection with the journal to achieve the objective of self-evaluation. The second is the relationship between the students' responses to these activities and their personality types' (p 146). The rationale is, 'student teachers must be given the necessary opportunities and skills to become self evaluators' (p 145). This is a study inspired by Schon's notion of the 'reflective practitioner' (Schon, 1987).
<b>Study: design</b>
Post-graduate secondary schoolteachers completed an autobiographical journal, performed exercises and answered questionnaires.
<b>Findings</b>
In summary, two-thirds (66.6%) of the trainees found the journal useful. Journals were also said to have increased reflection, clarified issues, acted as a forum for ideas, and increased teacher educator empathy.
<b>Conclusions</b>
The journals helped 'a significant number of students in providing exercises for reflection and self-awareness. The level of self-awareness present in the final reflections of many of the journals is impressive' (p 160).
<b>Reasons for exclusion</b>
The study does not focus on developing a formative assessment process or instrument; indeed, 'the journals are ungraded' (p 160). Although the journal could be used as an assessment instrument it is instead used 'for teaching self evaluation and reflection' (p 160). The study was keyworded ASSESSMENT as well as ASSESSED under section 6c. The latter is an accurate descriptor but the former is not.

## APPENDIX 4.2: Details of studies included in the in-depth review

<b>Study one</b>
<p>Berg M, J Curry (1997) Portfolios: what can they tell us about student teacher performance? <i>Social Studies Review</i> <b>36</b>: 78-84.</p>
<b>Study aim(s) and rationale</b>
<p>The word 'aims' is not used in this study. It is stated, however, that 'the evaluation of portfolios has emerged at the centre of the interest in the [assessment] topic' (p 78). Indeed, the researchers 'saw portfolios offering the promise of identifying both a broader and more in-depth picture of an emerging teacher's thinking and behaviour than other, more traditional forms of assessment' (p 84). The broad aim, therefore, could be stated as: to evaluate portfolios in terms of their capability for identifying trainee teachers' growth and evaluation. The study also suggests that portfolios would allow 'the professors, to look beneath the surface of the teaching act itself, and examine the decisions that shaped the student teacher's actions' (p 79).</p> <p>From being used in classrooms with children, portfolios have increasingly been employed in the US to document teachers' knowledge and skills in teaching. The researchers see portfolios as a 'natural outgrowth of this trend to explore the use of portfolios in defining and documenting more clearly what student teachers know and understand and what they are able to do; to create a vehicle for student teachers to tell their personal story of becoming a teacher' (p 78). The sample of trainees on whom the portfolio system was trialed were all involved at that time in the San Diego State University Model Education Center.</p> <p>The study identifies the involvement of five university professors, five supervisors, 30 trainees and 30 cooperating teachers, who worked against a 'backdrop of teaming as an organisational structure in classrooms across the partnership' (p 79).</p> <p>No mention is made of funding sources for the study. As it was part of institutional assessment arrangements, one assumes it was internally funded.</p> <p>As the study suggests that the portfolio data were gathered during one professional year, and the study itself was published in Spring-Summer 1997, it must be assumed that the previous academic year was the focus year of the study (i.e. 1995/96).</p>
<b>Study research question(s) and its policy or practice focus</b>
<p>The study is related to teacher careers insofar as the trainee teachers are new entrants to the profession in the early stages of their training. The 30 trainees are all involved in the San Diego State University Model Education Center as part of their training. As the US involves mainly undergraduate training for teachers, it is probable that these students are all undergraduates. The portfolios are developed by the trainee teachers in collaboration with the tutors and teachers who are involved in developing teaching and learning. The students themselves are also focusing on their own teaching and learning, and the learning of children.</p> <p>No information is given regarding the age and sex of the trainees.</p> <p>The San Diego State University teaching faculty has a Model Education Center which is described within the study as 'a professional practice school... formed as a university-school partnership for the purpose of examining effective teaching and learning practices... of the next generation of teachers' (p 79).</p>

This study is essentially about the instigation of a portfolio system for assessing the growth and development of 30 trainee teachers. The research team developed a set of eleven goals which 'would form the overarching umbrella under which the team developed the structure of the portfolio experience... [which] became the basis for student teacher portfolio entries' (p.79). The research team also devised a series of questions for student teachers to reflect on under the headings of:

- classroom management
- mathematics
- reading/language arts
- videotaped lessons

The students were then assessed broadly on categories determined by the research team, which were:

- ability to reflect clearly on own growth and change
- ability to discuss strengths and weaknesses of performance
- ability to address cross-cultural and language development infusion

The research team also developed assessment criteria which they defined as 'limited', 'developing' and 'strong' abilities of students to reflect through the portfolio. Students were also required to keep at least four examples of the 'best of the best' examples of their own practice. The portfolios were assessed three times: at the beginning of their professional year at the Model Education Center, at mid-point and near the conclusion of the program (p.82).

The research question appears to be: 'Could this [use of portfolios] not be a powerful way for student teachers to document the stories of their growth in an authentic setting?' Other than this, no research question or hypothesis is identified.

#### **Methods: design**

The portfolios are a means by which the researchers evaluated the trainee teachers' abilities to be reflective about teaching and learning practices. They are researcher-manipulated insofar as the researchers designed the specification of the content of the portfolios and assessed trainees' development and growth at three points during the year. We must infer that the concepts involved relate to reflective practices and, through these, the growth and development of trainees' thinking and development as teachers.

#### **Methods: groups**

The study appears to focus on 30 trainee teachers during one practice-based year of their training.

The potential components of trainee portfolios were determined by the project team of professors, supervisors and cooperating teachers. Questions which trainee teachers should ask themselves under a series of four headings were identified (although no theoretical or empirical basis is provided for these).

Three broad rating categories were also determined and then applied to the portfolios at three times during the teaching year. Inter-rater reliability was established by at least two professors independently assessing each portfolio and final grades were given to trainees.

#### **Methods: sampling strategy**

The study reports that 30 trainee teachers were involved but not whether this constituted a whole cohort or a sample from a larger cohort. There is no attempt at identifying a sampling frame. No information is given of any means or ways of selecting the sample.

#### **Methods: recruitment and consent**

It has to be assumed that the 30 students were a captive audience.

**Methods: Actual sample**

Thirty students participated in the study.

Although it could be assumed that all the participants are from the US, at no point is the nationality of the focus group stated explicitly.

We have to assume that the sample represents the pre-service teachers who are registered with the San Diego State University.

There is an assumption that, because the sample group were all trainees on the course over that one-year period, they were all included and all completed.

**Methods: data-collection**

The data collected were those contained in the trainee teachers' own portfolios, kept as part of the requirements of the course during a single year. The evidence within the portfolios was assessed by a pair of professors and rated on a three-point scale.

Trainee teachers completed their own portfolio of reflections on teaching experiences and examples of best practice.

The authors describe portfolios as 'defining and documenting ... what students know and understand ... a vehicle for student teachers to tell their personal story of becoming a teacher. This is also described as 'a container for storing and displaying evidence of the student teacher's knowledge and skills' (p 78)

Student teachers accumulated all the data in their portfolios. This was then analysed and assessed by the researchers.

Reliability of the data collected in portfolios is not mentioned.

The validity of the collection tools is established only insofar as they relate the contents of portfolios to the contents of other portfolios which have previously been the subject of policy decisions.

The portfolio information was gathered by the trainees themselves.

Data were collected as a result of involvement in the Model Education Center and all data requested to be kept emanated from there.

**Methods: data analysis**

Certain items within the portfolio were 'standardised' which would 'serve as benchmarks of formative assessment' (p 79). The 'best of the best' examples of practice would be 'used by the team to evaluate student teachers' ability to reflect upon performance over time' (p 79).

The faculty team developed a set of eleven goals that became the basis for student teacher portfolio entries:

1. Widening the repertoire of communication strategies and skills
2. Creating a collaborative learning community
3. Promoting teaching/learning strategies that align with constructivist theories
4. Promoting higher order thinking skills
5. Developing greater sensitivity to and respect for cultural differences
6. Widening scheme of educational environment
7. Understanding global linkages/interconnections within and among personal, social environments and technological systems

8. Promoting understanding of interconnections among content area through thematic, interdisciplinary instruction
9. Developing English language skills across all content areas while supporting/respecting children's primary language
10. Understanding the theoretical underpinnings of instructional practice
11. Inspiring a lifelong commitment to curiosity and learning

Using these eleven goals as the basis, the team generated a series of 'reflective questions (p 80). The data gathered from these were then analysed against the following criteria:

- *Limited*: [student] unable to respond or cursory, superficial response, no reflection, no specific mention of culture or second language
- *Developing*: specifics identified [by students] with some substantive discussion of underlying principles and/or rationale for decisions, indicators of awareness of and sensitivity to culture and language.
- *Strong*: rich discussion indicating breadth and/or depth of understanding and critical reflection, indicators of concrete application of strategies, focusing on culture and second language acquisition, and a rationale for instructional decisions.

These were later deepened in focus for the final screening.

There was no intervention intended or carried out. The trainees kept their portfolios as part of their course assessment and these were then assessed at the three points in the year previously described.

Neither the validation of the portfolios nor the validity of the analysis of them is mentioned.

The three categories for analysis of the portfolios outlined earlier were later extended to become:

- *Limited*: no connection to or understanding of the goal, or unclear definition of the best [meaning the 'best of the best' examples]; either no mention or a peripheral mention of the student's role in the lesson; nor or vague mention of cultural differences or language development; does not address major weaknesses (by our standards) of the lesson.
- *Developing*: product shows a relation to the goal, but only a pedestrian, unimpressive relationship; sees the importance or mentions the student's role in the lesson (including student affect and/or cognition); explicitly mentions attending to cultural differences and language development in a general, philosophical way; either the lesson has major weaknesses which they mention but do not discuss, or the lesson has no major weaknesses and there is no mention of minor weaknesses or any changes.
- *Strong*: strongly exemplifies goal in an impressive, well-conceived, creative way; focus on student's role, including motivation, affective and conceptual growth and understanding; includes specific ways cultural differences and language development were or might be addressed; discusses weaknesses, if any, and sees options for subsequent instructional delivery.

### **Results and conclusions**

The results are presented mainly in narrative form with examples of statements from student portfolios.

The findings are as follows:

'The results revealed a progressive, professional growth across time in student teacher insights related to the act of teaching, its impact upon student learning in the classroom, and how to work toward making meaning comprehensible for all learners, irrespective of cultural differences or language proficiency... student teachers were focused on selecting representative examples of their practice that they felt were 'fun' or 'motivating' to students in the classroom. Little concern was evidenced with children's learning needs or how to develop instruction. There were no comments made on changes that they would like to make in the activity' (p 82).

The contexts used by the student teachers for presenting the 'best of the best' examples, 'tended to fall into four areas - social studies, science, reading/language arts and math' (p 82).

One unexpected finding is reported in the section 'Conclusions':

'As the planning and implementation evolved, the integration and assimilation of the faculty team's varied pedagogical perspectives and the cohesion that developed forged an unexpectedly strong bond. It was this joining together as collaborators in learning with student teachers that perhaps became one of the most powerful elements of the Model Education Center student teacher portfolio project' (p 84).

These 'findings' are only minimally presented in the reviewers' views. Whilst we are given some information about analysis, this analysis is, in itself, very subjective and we are not told how the issue of subjectivity was handled. This would make replicability very difficult indeed. Similarly, as we have no theoretical or conceptual base for portfolios given in this study, it is difficult to understand quite what students were intended to do.

Before the section 'Conclusions', there is a longer discussion on the issues related to student teacher development as identified in the findings.

The later 'Conclusions' section states:

'We could conclude that student teachers' reflection on practice showed measurable growth over time as evidenced by their writing and professional judgement. They were able to integrate instruction both within content areas and draw across disciplines to meet the learning needs of the increasingly diverse student population in their classrooms... the examples [best of the best] allowed the student teachers to share their own perspectives on teaching and encouraged introspection through which began the process of informed self-assessment' (p 84).

### **Quality of the study: reporting**

There is no explanation of the Model Education Center or of the deeper concepts related to portfolios (e.g. reflective practice as a theoretical construct). No information is given about the trainee teacher group or the ages of the children who were the focus of the portfolio teaching.

Whilst it is easy to assume that portfolios must be a good thing if they encourage reflection amongst trainee teachers which then has an impact on their ability to teach, this is implicit within the aims of this study rather than actually made explicit.

Whilst considerable information is given about the basis on which the portfolio was assessed, this is all very subjective and further information was needed on how this subjectivity was handled by the research team.

More information would be needed on the sample, the actual portfolio and on the underpinning basis for its development.

On the whole, the authors stick to describing what they set out to investigate but, as we are not given sufficient information about sample and portfolio content, then it is hard to judge whether they have been selective in their reporting.

#### **Quality of the study: methods and data**

It is hard to see how the researchers could have designed what they did differently. If one wishes to research the use of portfolios in student teacher development and growth, then getting the trainees to keep a portfolio and assessing it is the obvious choice of method.

No information is given about why the particular contents of the portfolios were appropriate or used.

Without significantly more information about the subjective nature of the analysis, bias is inherent in this study. The results are not generalisable without much more contextual information.

In US and San Diego terms, one feels certain that there is the basis of a good study. However, the evidence provided is scant in both conceptual and empirical terms, therefore, making it low in trustworthiness.

Results, discussion, conclusions and implications all flow into one another and have to be interpreted.

There is an element within this study that the 'desired' outcomes (only implicitly stated) were achieved. There is no argument presented as to whether or why some students were less effective in their portfolios, for example. It is hard to agree or disagree with the findings and conclusions without further information or evidence.

<b>Study two</b>
Willis EM, MA Davies (2002) Promise and practice of professional portfolios. <i>Action in Teacher Education</i> 23: 18-27.
<b>Study aim(s) and rationale</b>
<p>The study was conducted in the US but this is not made explicit in the report itself.</p> <p>The broad aim of this study was to report 'the impact of portfolios on undergraduate teacher education students' professional development and the broader implications for teacher training programs' (p.18). It also explored 'students' perceptions regarding professional portfolios and presentations' and examined 'whether students considered the process of creating a professional portfolio and sharing it in a presentation of value as they moved from the educational to career environment' (p 20).</p> <p>The report starts with an introduction that refers to previous papers written about portfolios, how they have been introduced into teacher training programmes, and how they have been beneficial. These studies range from 1993 to 2001. The introduction also starts with a quotation (Sanders, 2000; p 11) that touches on 'the current educational reform movement' towards 'authentic assessment' and the consequent rise of portfolios. In the last paragraph of the introduction, the portfolio component of the teacher training programme at a 'southwestern institution', on which the study is based, is said to be a 'response to these multiple uses of professional portfolios for assessment' (p.18). Implicit in the report, therefore, is that the study examines a portfolio component introduced because of the current climate favouring portfolios as a means to 'authentic assessment'.</p> <p>One previous study, which 'investigated preservice teachers' knowledge of portfolio assessment and attitudes toward using portfolios as an alternative to conventional assessment practice' (p 246) was used to inform the questionnaire that was handed out to the 93 student teachers. This other study, therefore, had an impact on the data collected but could not be described as a 'linked report', neither was this report building on data or theory from an earlier report.</p> <p>There is no evidence of consultations with any interest groups at stage in this study or report. There is certainly no evidence of any consultation when considering the broad aims of the study and the issues to be addressed.</p> <p>There is no statement or suggestion about the funding of this study. It could perhaps be inferred that the study was internally funded, as only one teacher training programme was investigated.</p> <p>The time and date of the actual study is not made explicit, but it is inferable from the opening citation (Sanders, 2000; p 11) and the publication date (2002) that the study was conducted between the winter of 2000 and the summer of 2001. Data were collected over 'a three-semester period'.</p>
<b>Study research question(s) and its policy or practice focus</b>
<p>Assessment is the main focus of the study: the report is about a study that examines the use of portfolios as possible instruments for authentically assessing teacher trainees. It is also about teacher careers; the study focuses on initial teacher training, the pre-career stage of teaching, but at the same time makes frequent references to exit practices and the 'job search', and the benefit portfolios have regarding these two things. It is also about teaching and learning, for the study is loosely concerned with certain issues in teaching and learning, such as reflective practice and self-evaluation. 'Undergraduate teacher education students' are the population focus of the study. The programme is an undergraduate one, so minimum age can be confidently inferred (i.e.</p>

post-18), but maximum age cannot be inferred with any certainty. Of the 93 participants in the study '74 were female and 19 were male' (p 19).

The educational settings of the study are (i) An HEI : the physical setting of the study is the 'southwestern teacher training institution'. It is here that the presentation of the students' 'hard copy notebooks and electronic portfolios' are presented and the later questionnaires about their experiences completed. (ii) Primary schools: although primary schools are not explicitly featured in the report, the course specialises in 'elementary education'. (iii) Secondary schools: although the study does not take the school-based experience of the participants as its main focus, the course is 'elementary education'. The American system is not in perfect alignment with the English one. 'Elementary' is not quite 'primary', so it may also cover grades 7 and 8, the first and second years of English 'secondary' education.

The research questions are implicit. They could be qualified as: what impact do portfolios have on undergraduate teacher education students' professional development and what broader implications are there? (p 18). And, what are students' perceptions of portfolio professional development and presentations? (p 19).

#### **Methods: designs**

The study evaluates a recently implemented portfolio component into a teacher education programme. It assess whether it works well, concludes it does, but also draws on suggestions for future improvement. It is researcher-manipulated because the portfolio element has changed the experience of the students that year and 'researchers collected survey data from 93 students enrolled in the third semester of the teacher education cohort program', so the collection of data was an ongoing, intervention-based process, not retrospective.

The study is retrospective, however, insofar as it required participants to look back on their portfolio presentation via a questionnaire.

#### **Methods: groups**

A study design summary could be expressed as follows: 'over a three-semester period, researchers collected survey data from 93 students enrolled in the third semester of the teacher education cohort program. Of the 93 participants, 74 were female and 19 male... The questionnaire consisted of 23 Likert-type questions and five incomplete stems' (p 19).

#### **Methods: sampling strategy**

The gender ratio (74 females to 19 males) suggests that the study is attempting to represent a 'typical' teacher education course regarding the general make-up of initial teacher training courses. It does not state explicitly nor imply, however, that it is representative of a 'typical' or a specific population. Nevertheless, it does draw conclusions that are generalised. This suggests that the researchers believe that conclusions and implications drawn from this study would naturally be applicable elsewhere.

Selection for the study appears to have been all-inclusive. There are, however, three 'delivery options for coursework' (i.e. courses). One of them, the one that the study focuses on, is 'a three-semester on campus cohort which provides all professional preparation coursework'. This particular course, it is stated, 'focuses on applying theory to practice through practicum experiences'. It is inferable that this style of course is most suited to portfolio assessment, and that this is why students from it have been chosen. People within this particular course are not further identified or classified.

No incentives for recruitment onto the study are stated but it is highly unlikely that any were offered because the portfolio component was obligatory.

There was no sampling frame as such, in that there was no distinction drawn between possible participants, ideal participants, and actual participants.

#### **Methods: data-collection**

Data were both quantitative and qualitative. The questionnaire included '23 Likert-type questions and five incomplete sentence stems'. The data from the Likert-type questions were therefore quantitative, whilst the data from the open-ended, incomplete questions were qualitative. The data were used to define the sample. The data appear to have been collected by hand and stored on a software package, The Data Collector (Turner and Handler, 1992).

The report only states who determined the categories for recording the quantitative data; it does not reveal who circulated the questionnaires or who collected the research.

A number of reliability and validity measures were undertaken. The report states that, regarding the collection and collation of qualitative data, 'a team of two, one university faculty and a graduate student, independently identified the response categories and then mutually reached consensus on category labelling' (p 20, paragraph three). There was, therefore, an inter-rater reliability measure for the collection of qualitative data. The questionnaire used to collect data incorporated 'several questions... from an earlier study' (p 19, paragraph seven).

#### **Methods: data analysis**

A couple of software packages are mentioned: the Statistical Packages for the Social Sciences (SPSS) is mentioned and so is the Data Collector. Exactly how these packages were used is unclear. From Table 1 on page 21, it appears that answers to the quantitative questions were simply tallied and the mean average deduced. A separate method was used for providing a 'framework for reporting results' from the qualitative questions.

#### **Results and conclusions**

The results are presented in prose in a section called 'Survey Results', and in part via Table 1 on page 21. In the section 'Survey Results', the Likert-type answers have been classified under five headings. In the section 'Incomplete Sentence Stems', the open-ended questions are also classified under five headings.

The findings of the study mostly relate to the portfolio as a tool for assessment. The findings are drawn from the students' retrospective answers to the questionnaire. They 'perceived the portfolio and presentation assessment a worthwhile experience' (p 20). It gave them 'greater opportunity to express themselves creatively' but they desired 'personal ownership and decision making about the portfolio categories' (p 20). The 'evaluation form promoted reflection' and 'encouraged them to think about their preparation for student teaching and a teaching career', and this preparation 'made them more aware of their growth as a teacher' (p 20). Also, 'the process of selecting portfolio pieces clearly encouraged reflection' (p 20). The presentation of their portfolio 'increased their self-confidence' and students also preferred that the portfolios were not given a letter grade, just merely passed or failed. Their portfolios 'highlighted their professional growth and skills' and 'prepared them for a job search' (p 22). It also made them think about the standards and how they should be incorporated into the portfolio (p 22). Furthermore, students thought that the portfolios 'uniquely represented themselves' (p 22).

An electronic element was part of the portfolio component and this 'supported the application of technology skills in meaningful contexts to develop technological literacy' and this electronic portfolio, they believed, 'would help them to find a job' (p 22). There were also difficulties with the portfolio.

Students had difficulty 'deciding upon both the content to present and designing a format' (p 22). They 'expressed concern over the portfolio page limitation' and found it difficult to determine how best to represent themselves (p.22). The presentation, as part of an integrated portfolio assessment system, assisted students 'in reflecting on and organising their work' (p 23). The reflective process that was a corollary of the portfolio system 'helped to develop confidence through greater self-awareness and heightened their perceptions of being life-long learners' (p 24). The 'exhibition' of the portfolio also 'improved their speaking and interviewing skills' (p 24).

The numerical data (percentages, etc.), on which assertions are made, were not consistently presented. Imprecision clouds many of the assertions. For example, 'over one fourth of the sample...' (p 24). The presentation of the findings is consistently bad, using vague percentages. The 23 Likert-type questions are not represented in their entirety. The stage involving the metamorphosis of these questions into the five headings is poorly related. The quantitative data presented in Table 1 only includes questions 1, 4, 6, 7, 10, 11, 19, 22 and 23. There is no sign of the other questions. Under the section 'Incomplete Sentence Stems', there is no presentation of the qualitative data on which the statements and assertions are based, save for a few particular statements.

The conclusions of this study mostly synthesise the findings related in the previous section. According to the researchers, the portfolio 'programmatically encourages ongoing professional development through reflective practice' (p 24). It also offers 'programs a tool for assessing effectiveness in promoting growth and its related skills and dispositions' (p 25). It 'allows for ongoing opportunities' for students 'to practice reflection and communication skills, model their values, and expose them to a wide variety of self-evaluation criteria' (p 25). It also helps students to 'develop self-confidence in their verbal communication skills' and 'reflection skills' and ultimately, to 'assist in the job search' (p 25). A further summary of this is offered on page 25. The study 'confirms a number of perceived benefits: increased reflective practice, improved communication skills, emphasis on life-long learning and growth, and greater self-confidence in making the transition from school to work' (p 25).

#### Quality of the study: reporting

It is inferable that the study took place when it did and where it did, but this is interpretative and conjectural. The study, so it seems, was a reflex action to a heightened nationwide interest in, and use of, portfolios, but how this was funded, the exact location, when it exactly took place, and whether the study was primarily an internally funded review of a new system or a piece of externally funded research which was supposed to have wider implications is uncertain.

It is difficult to determine whether the introduction states the actual aims of the study or what it actually succeeded in doing. If the aims were to 'report the impact', then these are satisfied (p 18).

Apart from the number and the gender, the sample used in the study is not adequately described. What description there is can be found on page 19, paragraph six.

Apart from a brief description of the questionnaire and a reference to software packages involved in the process, data-collection is not described in any detail.

Although software packages are mentioned and the means by which the qualitative research is classified is described, the process through which the data went from this raw stage to the stage in which they are presented is unclear. Not all the data are presented, so the screening process is not discernible.

The portfolio guidelines are not included, neither is a copy of the original questionnaire. The exact methods used to collect, collate and analyse the data are not made explicit. Table 1 is not explained. Percentages are not made explicit.

**Quality of the study: methods and data**

There is a loose connection between the climate described in the introduction and the actual study, but no sound justification, reasoning or apologetic is discernible.

Decisions are rarely made explicit, let alone given a rationale.

The questionnaire seems to have been the best (perhaps only way) to have drawn data retrospectively on the portfolio component of the teacher training course. Interviews would have been an alternative, for better or worse.

There is no deliberation on the possible shortcomings and flaws of the questionnaire and the collection of it. One major shortcoming may be that the Likert scale is four-pronged, instead of the more common five-pronged scale. One school of thought advocates that there is no room for indifference on this model as everything is either agreeable or disagreeable, to different degrees, and there is no middle ground. If this is perceived to be the case, this study is certainly flawed. Another school of thought, however, advocates that the four-pronged Likert scale is better than the five-pronged one, because it forces participants off the fence. With middle ground, as well, there is the potential for drawing conclusions either way, and seeing them as positive or negative.

The data-extractor (the questionnaire) is based mainly on a set of former survey questions. To some degree, it builds on a tried and tested base. This is a minor validity assurance measure.

Although the findings of the study are for the most part reported in the section 'Survey Results', there are obvious omissions (i.e. Table 1 does not present all the questions and therefore omits a lot of the data). Since the entire process is not adequately described, the research method and design is not easy to follow; therefore the chance of bias and error distorting the findings is quite high.

There is nothing to suggest that the sample would seriously differ from any other taken from an average teacher training course. The findings, therefore, could safely be applied to other courses, especially within the US.

Although the research design and method is not made explicit, and this naturally raises questions about the reliability and validity of the findings, the data and the assertions made from the data were fairly trustworthy, insofar as the study is a simple one, without a huge margin of error.

Quantitative data without complex statistical analysis cannot really be interpreted in too many ways, so this is fairly trustworthy. The qualitative data is the untrustworthy component of the report. As we cannot see the original questionnaire in its entirety, we cannot fairly assess the validity of the questions; however, the report draws a clear distinction between the findings and the conclusions drawn from them. Formulating a conclusion independently from the conclusion in the report is therefore possible.

Because the study is not completely traceable, high trustworthiness cannot be awarded. However, given that it presents most of the data before it interprets it and draws conclusions based on common sense it can be awarded medium trustworthiness.

The concluding section of the report, more than anything, synthesises the findings rather than interprets them. The section on page 25, which offers possible implications and recommendations, is the most interpretative part.

No real justification is offered, however, for the concluding section, but the conclusions are, at the same time, not implausible.

The conclusions drawn from the findings are the same as those the reviewer would make.