

Abstract

This paper outlines a case study of an assessment development process which contrasts with more recent ‘top-down’ trends in assessment which tend to leave assessment development processes opaque to users. This paper describes the processes of a collaborative, multi-agency project which set out to develop an assessment framework which would appropriately recognise the various skills present in the Higher Education Field Academy; a widening participation programme aiming to boost the educational aspirations, enthusiasm and attainment of secondary school students through their engagement in archaeological excavation. The case study describes the stages of building an assessment framework that was sympathetic to the intentions of a learning programme whilst providing robust observation-based outcomes that avoided heavy assessment processes that could corrupt learning relationships. The challenges involved in this development process have important parallels with vocational assessment; not least in providing an example of a localised development of an observation-based assessment model which can be used in diverse settings. The insights gained from this case study are useful for others who are engaged in localised assessment development processes.

Keywords

- assessment,
- access to HE,
- education & training assessment,
- curriculum innovation,
- competence

Introduction: the HEFA course

The higher education field academy (HEFA) is an outreach educational initiative run by the department of archaeology at the University of Cambridge and has been in operation since 2006. One aim of this initiative is to widen access to higher education amongst under-represented groups through allowing young people of secondary school age to develop the skills, confidence and inspiration that they will need in order to aspire to, and ultimately succeed in, higher education or as they move from formal education into the workplace.

Although delivered via an archaeological context, the primary purpose of the programme is not to promote archaeology as a future aspiration for work or study for the young people involved. Instead, the HEFA programme uses the opportunities afforded by this particular context to facilitate the use, development and recognition of a range of ‘soft’ skills that are useful for later success in education or the workplace.

Each HEFA course involves around 40 young people (generally aged between 14 and 17 years of age) from a variety of schools. Courses are currently run around 15 times each year, and in 2010 the programme worked with 537 young people and involved 120 assessors. The course is a 3-day (plus additional writing time) project involving small, below-ground scientific investigations intended to discover, locate, retrieve, record, analyse and interpret new physical evidence of human activity in the past. Participants spend two days working in teams of three or four, excavating one metre square test pits in gardens and other open spaces within local villages and hamlets.

Following an initial introductory and instruction session, teams work independently with support from supervising school and HEFA staff, and following written directions in a programme handbook. Over two days, the team members complete all elements of their project – measuring out, de-turfing, digging and sieving in pits, cleaning and recording finds, section drawing, sampling, backfilling and

making good. During this time, the work of the participants is assessed through observation by each team supervisor. Expertise and guidance, including immediate on-site finds identification and dating, are also provided by HEFA staff.

On the third day of the programme, the participants visit Cambridge for a day synthesising the results from the various excavations and seeing how these results provide new knowledge and understanding of the historic development of the village they were excavating in. After completing the full three days, participants write up their test pit using the records they made. The written assignment is assessed by the University, and returned to participants via their school with written comments. The participants also receive a certificate to mark their achievements.

The skills base that the HEFA programme seeks to impart and recognise is diverse, including a variety of hard (task-associated and specific) and soft (generic skills that are useful in various contexts) skills. In the original assessment, data for assessing these skills were gathered through observations by test pit supervisors and evaluation of participants' written reports by staff from the Department of Archaeology, University of Cambridge.

Aims of the HEFA collaboration with ARD

In 2009, the Assessment Research and Development (ARD) Department of Cambridge Assessment began collaborative work with HEFA to redevelop the programme's assessment procedures. One set of aims of the collaboration related to critically evaluating current assessment arrangements, considering ways to improve them, and exploring the potential for the new assessment to become recognised as a formal qualification. From the perspective of the ARD, another important aim of the collaboration was to help develop its understanding of complex assessment mechanisms. The Cambridge approach (Cambridge Assessment 2009) outlines the principles that the organisation should adhere to in order to maximise the wider social benefits of assessment. Amongst these is a commitment to the use of research-based evidence that can help to augment the corpus of knowledge on assessment, and it was hoped that the experience gained from the collaborative project with the HEFA team could augment this.

In particular, the ARD was interested in collaborating in this project as an example of a development which included relatively non-standard learning environments and which involved the application of applied and academic skills together in group working situations. These complex arrangements contrasted with other research and development work that traditionally tend to operate within more controlled classroom-based situations. It was also hoped that the collaboration would itself give insight into the challenges faced when localised, smaller-scale developments (grounded in, and with sympathetic insight into, specific learning contexts) begin to articulate with large and formal qualification structures, which tend to be more generic in character.

Another issue, which emerged during the early stages of collaboration, was that there appeared to be limited literature which provided a useful theoretical knowledge base about the processes themselves which are involved in developing assessments, and how to take them through to formal qualification status. Although some examples do exist of localised initiatives that have developed into formal qualifications, e.g. Suffolk Science (Dobson 1988), many of these are now outdated and have left a limited literature. The knowledge of how to develop new assessment procedures thus remains mostly in the possession of the now dwindling number of those individuals who have direct experience of such activity: the knowledge surrounding such processes might thus be characterised as a form of 'guild knowledge'. According to Hall and Burke (2004), groups of people build up 'guild knowledge'

over time as a 'community of knowers' who forge communities of practice (e.g. Lave and Wenger 1991) through utilising common practices, employing common tools and communicating in a common language. Sadler (1987, 1989) has used this concept to describe how groups of assessors attain consistency of judgement. At the same time, the concept has a restrictive dimension since it is possible that the knowledge available to those within the guild is not accessible to others outside the guild who might benefit from it. This paper seeks to broaden this limited knowledge base by providing an analytical description of the phases of the research process that informed the collaborative HEFA/ARD assessment development.

HEFA, learner engagement and the recognition and assessment of diverse skills

An integral part of understanding how assessments function, and thus how they should be structured and operated, is to understand the aims of a learning programme and the mechanisms by which these aims are enacted. Without taking these into consideration, it is possible that assessment arrangements can undermine the aims of learning. For example, assessments that prioritise issues of inter-assessor reliability through the use of overly atomistic assessment criteria would jeopardise the potential for such assessments to validly recognise competent performance if true competence required the holistic integration of a number of skills (Johnson 2008a).

A key aim of the HEFA programme is to provide a learning context which promotes positive learner engagement, i.e. when the greatest possible numbers of learners are showing the greatest possible commitment to achieving the task in hand and gaining the greatest possible satisfaction and benefit from doing so. The HEFA learning context seeks to maximise learner engagement through the recognition of a diverse range of skills, including physical/practical skills, documentation/writing skills, academic/cognitive skills and personal/learning skills. Experience on HEFA appears to show that learners are most engaged when they are able to combine their existing skills with an awareness that they have developed new skills, in order to achieve a goal which is recognised as of value by themselves and also by their group. Task completion on the HEFA course requires the use of a combination of different sort of skills. This context allows and requires learners to combine and interweave skills (both individually and between different team members) variously, successively and cumulatively in order to succeed. The progressive nature of the task that forms the central core of the HEFA course provides learners with a very visual index of achievement at every stage, and an appreciation of cumulative achievement generates satisfaction and builds confidence, which in turn reinforces positive engagement.

As an example, the HEFA excavation process requires each 1 m square trench to be excavated in a series of discrete 10 cm layers or spits. Excavating each spit involves careful observation and meticulous recording of the surface of each 10 cm spit before any soil is removed. Once digging begins, different skills are required as sustained hard physical labour is needed to loosen the soil and remove it from the site. This process requires, at the same time, a quite different skill again – careful attention to detail combined with visual acuity and fine motor control in order to ensure that buried remains are noticed and not broken or damaged by the heavy equipment being used. The next stage of the process requires careful attention to detail, involving pattern recognition and creative visualisation as excavated spoil is searched by hand for artefacts. At this stage, learners have no idea what may turn up and need to keep their minds open and alert. This process may be highly productive, yielding large numbers of archaeological artefacts, in which case highly organised systematic working is required as finds are sorted, washed, dried, identified, logged, labelled and

bagged. On the other hand, if very few finds are revealed during sorting, team-building strategies for maintaining concentration and motivation need to be developed by the participants. Whatever has been found, the completion of each 10 cm spit is satisfyingly evident as the trench is visibly and tangibly deeper than it was before, providing a clear and appreciably genuine affirmation of real achievement

The HEFA programme contains many instances of such suites of interlinked activity inclined to develop high levels of learner engagement. This is considered vital since positive engagement helps learners to acquire skills and realise their abilities, helps them to recognise these strengths, validates these strengths through assessment and, as a consequence, helps to boost participants' confidence levels and contribute to their willingness in the future to attempt new challenges in education or the workplace.

Evaluations carried out by HEFA show that through access to the programme participants' 'confidence in their own abilities has improved and their attitudes to their future progression to higher education have become more positive' (Lewis 2011). For example, in 2009–10, 80% of more than 1000 participants reported that they felt more positive about staying on at school after year 11 after completing HEFA than they had done beforehand, while 84% felt more positive about going to university (Lewis 2011, 14). Seventy-six percent of HEFA learners in 2009–2010 considered that HEFA had helped them develop their communication skills; 80% that it had helped develop working to set standards; 78% that it had helped develop skills in applying creative thinking to explain observations; 76% that it had helped develop skills in their skills in self-assessment and reflexive learning; 83% that it had helped develop skills in working with persistence to a high standard; and 87% that it had helped develop skills in team working (ibid., 19–21).

The value of the HEFA learning experience can be seen to be an enduring one which learners remember and appreciate years later. Learner TR, starting a vocational qualification in Engineering in autumn 2010, looked back on their HEFA experience two years before with the comment: 'Once the assignment was completed I felt a real sense of achievement', while learner FD, beginning a vocational qualification in Animal management in Autumn 2010 said that their HEFA experience had showed them: '... that studying at Uni can be fun – not all lectures – and this took a weight off'. Longitudinal tracking of HEFA students in 2009–2010 showed that 88% had continued in school or college to study 'A' levels, while around 10% were starting vocational qualifications (ibid., 30). Post-18 educational aspirations range across academic and vocational subjects, including arts, humanities and sciences, ranging alphabetically from Agriculture to Zoology (ibid., 31). Learner DJ, attending HEFA in 2009, commented: 'I feel the skills and experiences I had will be useful throughout the rest of my life' (ibid., 24), while learner HS said: 'It has given me new skills I will never forget'.

Literature around competence-based learning offers insights into some of the reasons why the HEFA programme has been so successful in achieving its goal of widening participation and improving learner engagement. Some of this engagement might relate to the approach that the HEFA programme takes to recognising holistic performance in a complex and multi-faceted task situation, which contrasts with 'schooled' experiences. The HEFA conceptualisation of competent performance coheres with modern European conceptualisations of competence, which Wesselink notes is perceived to be a holistic process embedded within a context, and which integrates 'knowledge, skills and attitudes that enables a person to perform a certain task in ill-defined and unique environments' (2010, 816).

Some competence-based literature has also taken a closer look at the specific nature of the knowledge found within competent performance and this has consequences for the learning structures that are used in a learning programme. Eraut (2000) has argued that competent performance integrates different types of knowledge, both public and private. Propositional or public knowledge has a codified form, and is open to public scrutiny and 'given status by incorporation into educational programmes, examinations and courses' (Eraut 2000, 114). Within HEFA, such knowledge could, for example, include information recalled from classroom/curriculum learning, such as dates of historical events. On the other hand, performance can also involve personalised knowledge, which can involve codified knowledge filtered through an individual's perception and experience. Again, within HEFA, this could range from knowing how to use tools to move heavy loads most effectively, perhaps gained from a weekend gardening job, to being intuitively able to keep a group motivated through difficult situations, perhaps through humour or leading by example. Such knowledge has tacit elements, and these elements are not easily amenable to public scrutiny, are difficult to represent in formal assessments and are often unrecognised by the self and others (Evans and Kersh 2004). Poikela (2004) argues that this issue has clear implications for assessment arrangements since valid assessment must take into account all forms of knowledge.

Another important consideration is that the formation and recognition of tacit knowledge and its demonstration through applied skills has clear contextual links to the features of a learning environment. For Eraut (2007), tacit knowledge and skills are largely acquired through active participation within particular contexts. Eraut argues that contexts influence behaviours and that the specifics of some contexts can afford skill formation and demonstration. Evans and Kersh (2004) characterise such learning environments as being 'expansive' in nature, often allowing learners to have their tacit skills recognised in a positive way through less formal learning relationships. Soft skills, such as interpersonal skills, a subset of generic skills, can contain tacit elements which challenge their articulation, identification and, ultimately, recognition. Another challenge for assessment relates to the extent to which such knowledge and skills are transferable across contexts. In a review of literature on transfer, Aarkrog (2011) highlights a number of factors that impact on transfer, including factors external to the student, such as the workplace and the methods of instruction, as well as learner-related factors such as personal motivation and self-efficacy. According to theory, a key learner-related factor that supports transfer is their ability to synthesise links across different contexts and to identify elements of fidelity that are shared or similar across those contexts (Aarkrog 2011; Laker and Powell 2011; Shuman, Besterfield-Sacre, and McGourty 2005). Whilst recognising the challenges posed by the concept of transfer, literature suggests that fidelity can be conceptualised into a number of factors which can underpin the similarities or differences across different learning contexts. Aarkrog (2011) identifies one of these factors as 'modality' and represents a key influence on potential transfer. Modality concerns the medium of training and transfer, with 'near transfer' occurring where the medium of learning is very similar across different learning contexts. Aarkrog notes that 'Often VET is characterised by far modality. By reading texts and attending lectures (one kind of medium), the students adopt procedures or principles which they are expected to transfer into practical skills (another kind of medium)' (2011, 7). In essence, modality helps to conceptualise some of the important distinctions that might exist between schooled environments and the vocational context (and the associated barriers to learning transfer) and to enable discussion about why the HEFA learning environment might overcome some of these traditional distinctions.

Another aspect of modality might also relate to the extent to which task completion is seen as an individual or combined group endeavour. Schooled environments arguably cope better with notions of

'the personalised learner' – characterised by the certification of a singular learner's achievements at the end of their schooling. The vocational context more often relies on individual participants applying their skills and knowledge to fulfil roles in complement with each other. The group nature of such applied skills application can also undermine their transferability and recognition. Wood (1986) notes that interpersonal tasks are high in dynamic complexity due to the uncertainty and change present in the task-performance environment. Dynamic complexity complicates knowledge compilation because learners cannot rely on a one-to-one correspondence between the learning and subsequent transfer contexts, making it difficult to implement learned behavioural guidelines. It is suggested that this is a possible explanation for the reportedly low yield of positive learning results attained through soft-skills training programmes (Georges 1996).

Feedback is another factor that can influence the transfer of skills across learning contexts. Laker and Powell (2011) note that feedback has a significant role in developing a skill and that this is characteristically different across different skill domains. Laker and Powell observe that feedback in performances that use hard skills differs from those employing soft skills:

In hard skills the feedback and consequences of one's actions are usually very specific and usually clearly indicate unequivocally whether one has been successful or unsuccessful in demonstrating the skill. With soft skills, this is clearly not the case. The range and complexity of the feedback on and consequences of soft skills is much greater than for hard skills. (2011, 117)

A consequence of this equivocal soft skills feedback is to undermine the potential for these skills to transfer across contexts.

Johnson (2006) suggests that learning relationships are an aspect of schooled contexts that can undermine learning. In a review of research into vocational assessment and issues around access, Johnson reports evidence to suggest that pedagogic style can influence a learning culture, with less formal teacher–learner relationships sometimes improving levels of learner engagement.

There is also literature to suggest that schooled environments which rely on particular modes of assessment find it difficult to recognise young people's holistic performance and inherent aspects of tacit knowledge. Biemans et al. (2004) suggests that the dominant type of learning in schools tends to focus on learners' knowledge deficits rather than taking a developmental perspective. Taking a historical–sociological perspective, Jørgensen (2004) argues that school learning can be narrower and less valid than applied learning because it lacks a focus on the utility of the knowledge gained. Jørgensen suggests that the division of learning into subjects, reinforced by examinations at the level of single subject, reflects this disconnection between schooled and wider learning in its broadest sense. Moreover, this disconnection might represent a disconnection with those skills and aptitudes that can lead to success beyond school. The process of excavation required in the HEFA programme, developed by a professional archaeologist, closely mirrors the processes required in the wider workplace in order to fulfil complex tasks to professional standards, for posterity.

These factors of modality, task completion/learning context, and feedback might help to explain why the HEFA course has attained success in supporting and developing transferable soft skills. The HEFA course shares some important modality links with schooled learning environments, using some academic skills (e.g. report writing, elements of historical knowledge, etc.) as a generic platform for accessing an environment where the tasks differ in modality from school. In HEFA, skills rarely needed or used in the classroom can be highly valued, and it is an environment where physical and academic skills are most effective when used together (such as in the use of physical skills to

excavate while at the same time analysing that which has been revealed) rather than being atomised into separate curriculum subjects such as PE and History. The HEFA programme also contrasts with the types of schooled knowledge that might be more familiar to the participants. Knowledge is not dealt with in terms of subject categories, with the largely observation-based assessment model allowing space for the recognition of wider skills. These include the attitudes which contribute to competence as well as the opportunity to observe and capture instances of tacit knowledge through ongoing assessment mechanisms. In this way, the HEFA environment might be considered to be 'expansive'.

Such issues might be important in understanding why HEFA has been successful in improving learner engagement levels. HEFA provides a context where the focus is on the use of appropriate skills for a clear purpose. In this sense, the skills learned and applied have a direct utility through their connection with purposeful activity. Moreover, the HEFA learning context appears to support the application and recognition of interpersonal skills because any uncertainty and change in the task-performance environment is kept to a relative minimum by the routines of the task which then learner behaviours to be adapted and refined. Similarly, the pedagogic relationships that structure learning in the HEFA programme allow effective feedback to be given where soft, interpersonal skills are demonstrated. The dominant learning model in HEFA is largely based on collaborative participation in shared activity rather than models of didactic knowledge transmission. HEFA shows 'by doing' how school-acquired skills and knowledge can be combined with wider skills in the world of work, and thus provide a foretaste of this for learners, whatever their future might hold. This also affords the opportunity for feedback on skills application to be immediate, contextualised and clearly articulated, helping to counteract concerns that feedback on soft skills tends to be largely equivocal (Laker and Powell 2011).

HEFA assessment challenges

At the outset of the collaborative work between HEFA and ARD, it was clear that any new assessment framework would need to sympathetically reflect the complex learning arrangements and the inclusive aims of the HEFA programme. In order to achieve this, the assessment development process would require a carefully planned research strategy. Oates (2008) outlines the main aspects of such a process, with synchronised evidence-gathering being involved in all of the phases of design, piloting, evaluation and implementation of a new assessment scheme.

Competence-based literature also gives insight into the issues which might need to be considered during the development of a new assessment framework. Reflecting on work-based learning literature, Poikela (2004) suggests that the assessment landscape is changing, with the dominant assessment paradigm shifting from assessment as the method for scientifically measuring outcomes to a more judgementally-based assessment of the processes performed within an activity. This shift essentially involves a move from a position where assessment 'recognises the possession of skills' to the position where assessment 'recognises the application of skills for a purpose', and can take into consideration important situational factors. For Poikela, valid assessment must be sophisticated enough to gather evidence of the existence of both propositional (i.e. encoded/embedded) knowledge and procedural (i.e. tacit/embodied) knowledge. This perspective suggests that there is a need to structure assessment so that it is sympathetic to the contextual features of the assessment, involves expert assessors, and provides robust, trustworthy outcomes.

The development of assessment tools that are capable of matching these requirements can be problematic. Biemans et al. (2004) outlines the pitfalls of such assessments, suggesting that, amongst other things, they tend to be labour and time intensive, difficult to standardise, and require assessment tools of a particularly high quality.

The HEFA assessment development process

The HEFA/ARD collaboration is a localised assessment development which contrasts with much recent assessment development in England. Coles (2004) makes the point that the last generation has witnessed an increasing amount of centralised intervention in assessment development processes. Part of this centralisation reflects the approach that governments have taken to influence the labour market via qualifications development mechanisms. Understandably, these developments have tended to be large scale and a consequence of this seems to be that there is limited literature to inform those who are seeking to develop localised, small-scale and innovative assessment programmes.

Despite this, the need for a research-based approach to the development of such assessments is still imperative. Oates (2004) makes the point that any problems that are intrinsic to the form and content of an assessment will lead to implementation difficulties. Lester (2009) continues on this theme, suggesting that the involvement of an inductive research procedure during assessment development can lead to valid assessment tools being constructed which have high levels of user acceptance and ownership.

The HEFA/ARD collaboration involved three general phases over 18 months; these were:

1. Articulating the aims of the learning programme.
2. Developing the construct base.
3. Evaluating the assessment model.

Phase 1: articulating the aims of the learning programme

The first phase of the project involved explicating the aims of the HEFA learning programme so that the development process could align the assessment model with these aims without compromising them. This phase involved analysing published HEFA documentation (including promotional documents, programme outlines, existing assessment materials, etc.) and focussed discussions involving HEFA/ARD team members to build a shared understanding of the programme aims.

An important part of setting out the aims was to express the values that underpinned them. These elicited values included the belief that knowledge and practical work need to be considered as interconnected, and that there is intrinsic value in allowing young people to develop and demonstrate the application of multiple skills in an integrated way. An important part of the programme was also its focus on 'inclusion', and how this issue of interest in general education could be dealt with through the specific context of working to professional standards as a member of an archaeological team.

One early concern that arose from this articulation of values was that attempts to move the programme into a formal national qualifications framework context could damage the programme by

overburdening the programme with a heavy assessment architecture. Bearing this in mind, the next stages of development involved making explicit the construct bases of the learning programme so that these could be reflected in the new assessment framework.

Phase 2: developing the construct base

Learning programmes (and therefore allied assessment procedures) can be described in terms of three related concepts: *constructs* (the skills and attributes that the programme covers and that the assessment seeks to recognise); *observable behaviours* (the identifiable actions which give inferential evidence about the constructs to be assessed and which might be scalable); and *tasks* (the activities that afford particular behaviours). These concepts are also useful in structuring evaluations of assessment programmes.

The tasks of the HEFA programme are well articulated; being structured around the requirements of professional archaeological activity. For the HEFA/ARD collaboration, the articulation of constructs was a key part of the development as this informed the generation of assessment criteria that would help assess particular observable behaviours. According to Cronbach and Meehl, 'a construct is some postulated attribute of people, assumed to be reflected in ... performance' (1956, 178), and it would be expected that these constructs should relate directly to the aims and values of the programme.

The construct development phase involved two aspects, construct elicitation and construct review. Construct elicitation was a complex task and involved a variety of methods and participants. One important aspect of the process was to bring the experts who developed the HEFA programme to reflect on materials such as curriculum documents and assessment materials from cognate areas designed for young people at the same stage of learning.

The ARD has previously used Kelly's repertory grid procedures (Kelly 1955), or modified versions of this method (e.g. Johnson 2008b), to elicit constructs from experts. These methods generate data through reflective processes by confronting experts with materials which possess similarities and differences from the constructs that they might themselves possess at a tacit, unarticulated level. For the HEFA/ARD project, this introspective process was augmented by additional empirical data-gathering with two ARD researchers attending a HEFA course to collect data about constructs through observations of learners and supervisors at work. Finally, views from stakeholders were gathered through a survey of teachers who had taken part in HEFA courses.

Once the constructs were elicited and codified into a new assessment framework, it was possible to undertake a process of construct review through a series of joint HEFA/ARD meetings. The ultimate aim of this review phase was to be able to establish that the sampling of constructs was adequate to support any validity claims for the assessment. These meetings evaluated the construct base of the new assessment framework through considering the overall balance of construct representation, whether the constructs were considered to be essential or desirable, whether the constructs were binary or scalable, and whether there were any overlaps in constructs.

Once this review was completed, the constructs were broken down into separate criteria during a series of meetings. This process involved the HEFA experts using a variation of an Angoff (1971) standard-setting method which uses experts' conceptualisations of 'minimally competent performance' on a task to establish a reference point. This point can then be used to help articulate performance descriptors that exceed or fall short of this standard. These criteria were then developed into either binary or scaled descriptors (see example in Figure 1).

Figure 1 Extract from HEFA assessment framework.

<i>Structured working</i>			
Plan and carry out scheme of work in structured manner and within required time.	This individual has shown minimal interest in planning or carrying out the required tasks in the correct order or within a reasonable time-scale, taking either far too much or far too little time on tasks.	This individual has used the handbook to plan and follow the required work programme correctly. With some help he/she has identified objectives and defined tasks clearly. Task-flow has been planned and managed; progress has been monitored effectively.	This individual has used the handbook without prompting to pro-actively plan and help drive through an effective programme of work. He/she has identified objectives and defined tasks clearly and has been pro-active in monitoring progress and ensuring tasks are completed on time.

At this stage, the assessment framework consisted of four sections, with two of these containing a number of subsections (Figure 2). In total, the framework included 39 separate assessment criteria, broken down into scaled or binary assessment statements.

Figure 2 The HEFA assessment structure.

Section	Subsection
Record of participation	
Data collection exercise	<ul style="list-style-type: none"> • Setting up and workspace management • Data/sample collection • Measuring and recording • Working to set standards
Personal, learning and thinking skills (PLTS)	<ul style="list-style-type: none"> • Verbal communication • Structured working • Creative thinking • Reflective learning • Effort and persistence • Team working
Written report	<ul style="list-style-type: none"> • Report structuring and data presentation • Research skills • Writing skills • IT skills • Independent working

Phase 3: evaluating the assessment model

The third development phase gathered empirical evidence about the functioning of the assessment framework in two stages. The first evaluation stage involved researchers observing nine team supervisors as they assessed a total of 26 young people taking part in a HEFA course over two days. This evaluation gathered data in a variety of areas. One area was to consider the practicalities of the assessment through looking at the pattern of assessment behaviours: when assessments were taken, how the assessors completed the assessment documents, how the assessors organised their data gathering; and how long these assessments took.

The evaluation also compared the judgements of assessors who were located in one test pit for the whole course, with those of assessors who rotated around different test pits. This is vital, as consistency of assessor judgements is one important measure of assessment validity. A concern with

assessment models which have a large number of assessors working in different contexts is that they might be prone to larger degrees of inter-assessor judgement inconsistency compared with models that use fewer assessors. Part of this phenomenon might be related to the concept of 'assessment density' (Johnson2008b): standardising the judgements of assessors who more frequently see the same sorts of performance evidence in similar contexts is easier than for those assessors who do not.

Researchers also gathered qualitative interview evidence from team supervisors/assessors to investigate whether the language of the assessment framework was clear, what the effect of the assessment was on their relationship with those being assessed, their levels of confidence in making assessment judgements based on the framework, and whether there were any duplicated or missing constructs. Those being assessed were also surveyed to investigate whether they perceived the assessment process to be invasive and how it affected their learning and enjoyment of the HEFA course. Analyses of the assessor evidence suggested that the archaeological supervisors felt confident in the assessment framework and the validity of its outcomes, while learners reported feeling comfortable with the assessment procedure, with some even reporting that they appreciated knowing they were being assessed as it helped them focus.

The final aspect of the first evaluation stage was a statistical analysis of the assessment outcomes. Comparisons between the quantitative mark outcomes for the young people generated by the assessment framework with independently gathered rank orders of their overall performance were used to validate the assessment framework. These data showed that the assessment scoring process correctly ranked 25 of the 30 young people observed during the trial.

These data were then used to make a number of amendments to the wording of the assessment framework, leading to a second evaluation stage which considered issues of inter-assessor agreement. This involved four assessors and 14 young people over two days. In advance of the evaluation, a pre-assessment standardisation training session was arranged, where the assessors could discuss the assessment framework in relation to pieces of video evidence captured during earlier HEFA courses.

The aim of this evaluation stage was to explore whether two independent assessors could reach similar conclusions about the same performance using the assessment framework. Each pair of assessors oversaw two test pits, rotating between these throughout the 2-day period of the course. Analysis then considered the extent of agreement between the two pairs of assessors, whether any assessor variance was systematic or random, whether any variance would affect any notional pass/fail thresholds, and whether there needed to be additional weighting or marks awarded for some constructs. Analysis of the agreement data used measures of the proportion of exact agreement between assessors (P_0 scores) ranging from 0 (complete disagreement) to 1 (complete agreement). The proportion of agreement with a tolerance of ± 1 assessment grade (P_1 scores) was also calculated (for more information on methods to calculate assessor agreement, see Bramley 2007). This evaluation stage showed that P_1 assessor agreement was 1 whilst P_0 scores ranged from 0.5 to 0.8. These data suggested that between assessor agreement was generally good but that there were some apparent systematic patterns with the less experienced assessors being more lenient than their more experienced partners and suggesting the potential need for improved standardisation to help support aligned interpretations of assessment criteria.

These analyses led to further refinements and the final re-drafting of the assessment framework. The final framework covers a diverse skills base, including a variety of 'hard' and 'soft' skills. The 'hard'

skills are mainly represented in the *data collection* and *written report* sections of the assessment (see example in Appendix 1).

Conclusions

This paper has looked to the literature around competence-based learning and assessment to offer insight into the issues that have affected the HEFA/ARD collaborative project, and provided an account of the process whereby a valid assessment framework has been developed from a localised project to one that could be applied much more widely.

A key aim of the project was to develop assessment procedures that were sympathetic to the implicit values that resided in the HEFA programme, namely developing learning and ambition as well as recognising achievement. It was also hoped that the research-informed approach used in this development would help to mitigate any difficulties in assessment implementation (Oates 2004) and result in high levels of user acceptance (Lester 2009). Initial feedback from young people taking part in the pilot stages of the development programme have been positive, suggesting that their experience has not been adversely impacted by the implementation of the new assessment model.

The evaluation phase of the development also suggested that the supervisors who used the assessment scheme felt confident in its application and the validity of its outcomes. Literature suggests a number of reasons why the assessment framework might have worked so well. By focusing on the aims of the learning programme, and the mechanisms by which these aims are enacted, the development process ensured that overly atomistic assessment criteria were not included in the assessment framework. Whilst the adoption of more atomistic criteria might have supported higher levels of inter-assessor agreement, holistic criteria more validly reflect the integrated learning approach characterised by the HEFA programme.

Literature cited earlier suggests that competence integrates different types of knowledge, both codified and tacit (Eraut 2000), with the latter often being unrecognised in formal assessments (Evans and Kersh 2004). Poikela (2004) argues that valid assessment must be sophisticated enough to gather evidence of the existence of both propositional and procedural knowledge for those being assessed. It appears that the assessment model promoted by the HEFA/ARD project coheres with these perspectives on competence and valid assessment. The situated, observation-based assessment model developed in the HEFA/ARD project affords the opportunity to recognise tacit knowledge since this relies on the ability of assessors to integrate judgements within a particular and dynamic context. Importantly, it appears that the HEFA context shared some modalities with schooled environments – particularly with regards to the role of feedback in learning. The immediacy of feedback in the HEFA contexts, supported by the assessment framework, could afford the development and transfer of soft skills in contrast to the commonly observed practices noted in other contexts (Laker and Powell 2011). Another important element in this recognition process might also relate to the way that the assessment does not interfere with the largely informal learning relationships that are central to the HEFA scheme. The supervising assessors can use the assessment framework in ways that do not interfere with their preferred pedagogic style, through separating the observation and recording aspects of assessment if required. The supervisors are able to supervise, advise and encourage during the task and then make formal judgements about observed performances at the end of the sessions in an ongoing manner. In this way, the assessment tool can support iterative assessment, allowing assessors to return to previous judgements about participants to check certainty, as well as to help direct assessor activity towards ensuring that all

participants are given the opportunity to demonstrate performance across all areas of the learning activity.

Establishing an optimal balance between supervision and assessment depends to a great extent on the assessment framework being of a high quality. A major element of this quality relates to the degree to which the assessment criteria are meaningful and relevant to the assessors as they make their observations, and this has direct connexions with the quality of the construct elicitation process.

Through the engagement of research methods in the construct elicitation process, the assessment framework was able to reflect the key aspects of knowledge and skills that the HEFA team felt were crucial to the programme. In this way, the assessment fully supports the learning programme, whose success is itself based on the integration of knowledge which has a high utility value (Jørgensen 2004).

Competence-based literature also recounts the challenges to creating valid assessments in often diverse contexts. Assessment density (Johnson 2008b), labour and time demands, and standardisation difficulties (Biemans et al. 2004) can be some of the pitfalls for those designing assessments of context-based skills assessment. Overcoming these challenges requires that the assessment tools developed need to be of a high quality. It appears that one way of ensuring that any research tools that are developed are fit for purpose is through the engagement of a carefully planned research strategy during the development process (Oates 2008). The HEFA/ARD project was able to embed such a strategy and this shaped the final assessment to the point that the challenges already highlighted were perhaps diminished.

Another issue raised by this development is that the localised, small-scale characteristics of the HEFA/ARD development contrast with dominant trends in the assessment development landscape in England (Coles 2004). One consequence of such a trend is that there appears to be limited available literature to draw on when looking to develop small-scale and innovative assessment developments. This paper seeks to help to fill this gap, providing an account of the developmental phases of a project that might be of use to, or serve as an exemplar for, others engaged in such initiatives.

Appendix 1

HEFA Assessment Framework (Draft)					
Section 1 - Record of participation					
A1	The student has actively participated in an archaeological excavation from start to finish during which he/she has acquired and recorded original primary data for study using appropriate techniques and standards in order to ensure data is valid and health and safety requirements are met.				
A2	The student has experienced of all stages of an archaeological data-gathering practical project				
Section 2 - Data collection					
	Assessment Criteria	Range Descriptor (low)	Range Descriptor (middle)	Range Descriptor (high)	
B1	Use appropriate equipment effectively to excavate a precise square test pit exactly 1m ² , creating and maintaining straight vertical sides at all times and extracting spoil in standard 10cm spits (contexts) leaving horizontal surfaces clean and level on completion of each context.	This individual has rarely used the correct tools (spade, mattock and shovel) in the correct way, and has shown no interest in acquiring or displaying any proficiency in their use. He/she has made hardly any contribution to the task of loosening and removing spoil from the test pit, keeping the test pit the correct shape, size and precision or he/she has repeatedly excavated beyond the required level or required repeated prompting in order to stop at the appropriate level.	This individual has used the correct tools (spade, mattock and shovel) in the correct way more often than not, and has shown interest in acquiring or displaying proficiency in their use. He/she has made a useful contribution to the task of loosening and removing spoil from the test pit and keeping the test pit the correct shape, size and precision. He/she has excavated up to but not beyond the required level most of the time or with some prompting.	This individual has used the correct tools (spade, mattock and shovel) in the correct way at all times, has shown great interest in acquiring or displaying proficiency in their use. He/she has made a very significant contribution to the task of loosening and removing spoil from the test pit and keeping the test pit the correct shape, size and precision. He/she has excavated up to but not beyond the required level at all times without prompting.	
B2	Process a series of extracted 0.1m ² spoil samples using appropriate methods (sieving and/or hand-sorting) in order to identify, recover and retain target data (ie items of archaeological interest).	This individual has taken minimal care to search spoil for finds, has only used the sieves correctly with repeated prompting and has not engaged with identifying when sieving is appropriate and when hand sorting would be more useful. The student has made hardly any contribution to the task of processing spoil.	This individual has taken care to search spoil for finds, has used the sieves correctly with only limited prompting and has been able to identify when sieving is appropriate and when hand sorting would be more useful with only limited prompting. The student has made a useful contribution to the task of processing spoil.	This individual has at all times taken the greatest possible care in searching spoil for finds, has used the sieves correctly with no prompting and has been able to identify when sieving is appropriate and when hand sorting would be more useful. The student has made a very significant contribution to the task of processing spoil.	
B3	Critically observe extraction site while extraction is in progress in order to ensure that contextual information is not overlooked.	This individual has taken hardly any care to look for and identify features such as walls or cuts into the ground in any way, during soil loosening or trowelling, despite repeated prompting.	This individual has mostly taken care, after a little prompting, to look for and identify features such as walls or cuts into the ground by (a) making visual observations while loosening and removing spoil, (b) using a trowel to remove loose spoil from the surface of the pit at least every 10cm and/or at the bottom of each 10cm context and (c) looking for differences in texture and appearance of cleaned surfaces	This individual has taken the greatest possible care, within a reasonable time-frame, to look for and identify features such as walls or cuts into the ground by (a) making visual observations while loosening and removing spoil and (b) using a trowel to remove loose spoil from the surface of the pit at least every 10cm and/or at the bottom of each 10cm context and (c) looking for differences in texture and appearance of cleaned surfaces	
© Access Cambridge Archaeology					

References

- 1. Aarkrog, V. 2011. A taxonomy for teaching transfer skills in the Danish VET system. *Nordic Journal of Vocational Education and Training*, 1(1): 1–13.
- 2. Angoff, W.H. 1971. "Scales, norms and equivalent scores". In *Educational measurement, 2nd ed.*, Edited by: Thorndike, R.L. 508–600. Washington, DC: American Council on Education.
- 3. Biemans, H., Nieuwenhuis, L., Poell, R., Mulder, M. and Wesselink, R. 2004. Competence- based VET in the Netherlands: Backgrounds and pitfalls. *Journal of Vocational Education and Training*, 56(4): 523–538.
- 4. Bramley, T. 2007. Quantifying marker agreement: Terminology, statistics and issues. *Research Matters*, 4: 22–27
- 5. Cambridge Assessment. 2009. *The Cambridge approach: Principles for designing, administering and evaluating assessment*. Cambridge: The University of Cambridge Local Examinations Syndicate. http://www.bulats.org/docs/cambridge_approach.pdf.
- 6. Coles, M. 2004. "Making qualifications fit for the future". In *Identifying skill needs for the future: From research to policy and practice*, Edited by: Schmidt, S.L., Strietska-Ilina, O., Tessaring, M. and Dworschak, B. 87–99. Luxembourg: Office for Official Publications of the European Communities. Cedefop Reference Series 52.
- 7. Cronbach, L.J. and Meehl, P.E. 1956. "Validity in psychological tests". In *The foundations of science and the concepts of psychology and psychoanalysis (I)*, Edited by: Feigl, H. and Scriven, M. 174–204. Minneapolis, MN: University of Minnesota Press.
- 8. Dobson, K. 1988. The Suffolk experience: An introduction and two case studies on the actuality of physics education within 'balanced science'. *Physics Education*, 23: 162–168.
- 9. Eraut, M. 2000. Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology*, 70(1): 113–136.
- 10. Eraut, M. 2007. Learning from other people in the workplace. *Oxford Review of Education*, 33(4): 403–422.
- 11. Evans, K. and Kersh, N. 2004. Recognition of tacit skills and knowledge: Sustaining learning outcomes in workplace environments. *The Journal of Workplace Learning*, 16(1/2): 63–74.
- 12. Georges, J.C. 1996. The myth of soft-skills training. *Training*, 33: 48–54.

- **13.** Hall, K. and Burke, W. 2004. *Making formative assessment work: Effective practice in the primary classroom*, Maidenhead: Open University Press.
- **14.** Johnson, M. 2006. A review of vocational research in the UK 2002–2006: Measurement and accessibility issues. *International Journal of Training Research*, 4(2): 48–71.
- **15.** Johnson, M. 2008a. Exploring assessor consistency in a health and social care qualification using a sociocultural perspective. *Journal of Vocational Education & Training*, 60(2): 173–187.
- **16.** Johnson, M. 2008b. Assessing at the borderline: Judging a vocationally related portfolio holistically. *Issues in Educational Research*, 18(1): 26–43.
- **17.** Jørgensen, C.H. 2004. Connecting work and education: Should learning be useful, correct or meaningful?. *The Journal of Workplace Learning*, 16(8): 455–465.
- **18.** Kelly, G.A. 1955. *The psychology of personal constructs*, New York, NY: Norton.
- **19.** Laker, D.R. and Powell, J.L. 2011. The differences between hard and soft skills and their relative impact on training transfer. *Human Resource Development Quarterly*, 22(1): 111–122.
- **20.** Lave, J. and Wenger, E. 1991. *Situated learning: Legitimate peripheral participation*, Cambridge: Cambridge University Press.
- **21.** Lester, S. 2009. The conservation technician qualification: An employer-led development. *Research in Post-Compulsory Education*, 14(1): 43–55.
- **22.** Lewis, C. 2011. *Higher education field academy. Report 2009 & 2010 monitoring and feedback*. Cambridge: University of Cambridge Department of Archaeology. <http://www.arch.cam.ac.uk/aca/monitoring.html> (accessed March 8, 2011).
- **23.** Oates, T. 2004. The role of outcomes-based national qualifications in the development of an effective vocational education and training system: The case of England and Wales. *Policy Futures in Education*, 2(1): 53–71.
- **24.** Oates, T. 2008. Going round in circles: Temporal discontinuity as a gross impediment to effective innovation in education and training. *Cambridge Journal of Education*, 38(1): 105–120.
- **25.** Poikela, E. 2004. Developing criteria for knowing and learning at work: Towards context-based assessment. *The Journal of Workplace Learning*, 16(5): 267–274.
- **26.** Sadler, D.R. 1987. Specifying and promulgating achievement standards. *Oxford Review of Education*, 13: 191–209.
- **27.** Sadler, D.R. 1989. Formative assessment and the design of instructional systems. *Instructional Science*, 18: 119–144.
- **28.** Shuman, L.J., Besterfield-Sacre, M. and McGourty, J. 2005. The ABET ‘professional skills’ – can they be taught? Can they be assessed?. *Journal of Engineering Education*, 94: 41–55.
- **29.** Wesselink, R., Dekker-Groen, A.M., Biemans, H.J.A. and Mulder, M. 2010. Using an instrument to analyse competence-based study programmes: Experiences of teachers in Dutch vocational education and training. *Journal of Curriculum Studies*, 42(6): 813–829.
- **30.** Wood, R.E. 1986. Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes*, 37: 60–82.