Evidence scope regarding the use of practice observation methods as part of the assessment of social work practice

Lead Author
Gillian Ruch, University of Sussex
Acknowledgements

Thanks are due to colleagues within KPMG, and other members of the consortium, and within the Department for Education for their support, attention to detail and their willingness to engage openly with the findings of this evidence scope.

This piece of work benefited from the invaluable advice of a number of highly skilled academics who shared useful materials, perspectives and links, including Donald Forrester, Liz Beddoe and Marion Bogo. Particular thanks to Alan Murphy, Caroline Humphrey, Cath Holmstrom, Jane Boylan and Kate Leonard for acting as critical readers.

Heartfelt thanks to social work professionals within a number of local authorities who shared their experience of practice observation, including: Bristol, Bromley, Calderdale, Cheshire East, Cornwall, Derby, Essex, Hampshire, Havering, Hillingdon, Kirklees, Leicester, Lincolnshire, Stockport, Tri-Borough, Warrington, West Berkshire, Wigan and York.

The research team are also very grateful to the social work practitioners and managers at East Sussex County Council who gave up their time to take part in the focus group and offer their expertise and knowledge.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Glossary</td>
<td>4</td>
</tr>
<tr>
<td>2. Background</td>
<td>5</td>
</tr>
<tr>
<td>3. Methods of direct observation that have been used to assess social work practice</td>
<td>7</td>
</tr>
<tr>
<td>4. Evidence concerning the reliability and validity of each of these methods as a predictor of actual performance in the field.</td>
<td>18</td>
</tr>
<tr>
<td>5. Lessons for social work which can be learned from the direct observation methods used to assess practice in other related professions</td>
<td>24</td>
</tr>
<tr>
<td>6. Incorporating direct observation most effectively alongside a digital test</td>
<td>38</td>
</tr>
<tr>
<td>7. Methods of direct observation most relevant for each of the three levels of practice</td>
<td>43</td>
</tr>
<tr>
<td>8. The amount of direct observation required for each three levels of practice</td>
<td>45</td>
</tr>
<tr>
<td>9. Issues for wholesale implementation</td>
<td>47</td>
</tr>
<tr>
<td>10. Summary and Key Considerations</td>
<td>52</td>
</tr>
</tbody>
</table>

### List of Appendices:
- Appendix 1: The Knowledge and Skills Statements for the three career levels
- Appendix 2: East Sussex County Council ASYE Direct observation proforma
- Appendix 3: Case studies regarding use of practice observation submitted by local authorities
## 1. Glossary

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASYE</td>
<td>Assessed and Supported Year in Employment (the first year in employment post-qualification).</td>
</tr>
<tr>
<td>Direct observation</td>
<td>Assessment of professional practice involving the assessing person being present during the practice encounter.</td>
</tr>
<tr>
<td>Formative assessment</td>
<td>Occurs during the learning process and provides feedback to the learner for their further learning and future development; corrective feedback for further development.</td>
</tr>
<tr>
<td>HEI</td>
<td>Higher Education Institution.</td>
</tr>
<tr>
<td>Motivational interviewing</td>
<td>A form of collaborative conversation for strengthening a person's own motivation and commitment to change.</td>
</tr>
<tr>
<td>OSCE</td>
<td>Objective Structured Clinical Exam.</td>
</tr>
<tr>
<td>Practice Educator</td>
<td>A qualified practitioner assessing students on qualifying programmes.</td>
</tr>
<tr>
<td>Practice Assessor</td>
<td>A qualified practitioner assessing practitioners in continuing professional development contexts.</td>
</tr>
<tr>
<td>Professional capability</td>
<td>A holistic set of knowledge, skills, and values that are encompassed, but not always immediately visible, in the behaviour of particular professions as they perform roles within their scope of practice.</td>
</tr>
<tr>
<td>Professional competence</td>
<td>A set of knowledge, skills, and values that are evident in the behaviour of particular professions as they perform roles within their scope of practice.</td>
</tr>
<tr>
<td>Reliability</td>
<td>A measure is considered reliable if it authentically accurately captures/measures what is representative of the context of study i.e. is a reliable representation of ‘reality’.</td>
</tr>
<tr>
<td>Simulated observation</td>
<td>Assessment of professional practice using videos of simulated scenarios or actors engaging with practitioners in practice scenarios.</td>
</tr>
<tr>
<td>Social work practitioners</td>
<td>Qualified social workers.</td>
</tr>
<tr>
<td>Summative assessment</td>
<td>Occurs at end of the learning process and assesses how well the learner has done; conclusive evaluations for progression and/or gatekeeping purposes.</td>
</tr>
<tr>
<td>Validity</td>
<td>A valid study is one able to answer the questions it is intended to answer and is using accurate criteria to measure it; validity is associated with a high degree of standardisation and use of standardised measures which can be repeatedly administered in a consistent manner.</td>
</tr>
<tr>
<td>Video/Audio recorded practice</td>
<td>Assessment of professional practice via video or audio recorded material of practice encounters.</td>
</tr>
</tbody>
</table>
2. Background

The social work professional context

Social work is a unique profession. Practitioners working on the frontline are primarily undertaking their work in the community. Unlike other human services professions such as health clinicians in clinical settings or teachers in schools, a high percentage of social work practice happens ‘invisibly’, that is in the privacy of people’s homes, and is statutory in nature. The recent introduction by the Chief Social Worker for Children and Families of Knowledge and Skills Statements for the tripartite practice-based career pathway – Approved Children and Families Practitioner (ACFP), Practice Supervisor and Practice Leader (see Appendix 1) – is a new initiative to assess and accredit the capability of the professional workforce across the career continuum. These new initiatives currently sit alongside the existing Professional Capabilities Framework (PCF) introduced in 2012 by The College of Social Work; at the time of writing, the future of the PCF is under discussion following the announcement that The College of Social Work will close in September 2015. As a mechanism for assessing and accrediting professional practice this proposed new system has prompted significant policy-led activity into how the capability of practitioners and managers across the career trajectory can be assessed.

As an initial response to this challenge the Department for Education and the Chief Social Worker for Children and Families – via a consortium, led by KPMG - are devising a digital solution to assess professional capability. The digital solution will be comprised of a range of simulated practice scenarios and exercises with sophisticated response pathways to assess practitioners’ knowledge and skills. In the context of this development, questions have emerged as to how reliably and comprehensively a single method assessment strategy can capture all aspects of professional social work practice and capability. Interest has been expressed in the extent to which the creation of a multi-faceted assessment strategy, that includes some form of observation of practice, would complement the digital solution. In particular, consideration needs to be given to assessing the complex constituents of social work skills, referred to in the literature as meta-competences (Bogo et al 2011), that a digital approach, used on its own, may be less able to capture.

Professional competency and capability

Within the academic and practice domains of professional practice in general there is wide-ranging debate around the terminology used to describe standards of professional practice (Taylor and Bogo, 2014; Burgess et al, 2014). Professional competence usually refers to a narrow range of knowledge, skills and values that are evident in a particular professional context and which are easily captured and measured. As a term it is often critiqued for implying a reductionist understanding of what professional practice entails. In contrast professional capability is
understood in a more holistic sense as a broader mix of knowledge, skills and values, some of which are quite subtle and difficult to capture in straightforward, visible ways. For the purposes of this scoping review the term professional capability (akin to Bogo’s meta-competence) is being used to emphasise the complexity and diversity of professional social work practice that will be assessed and accredited through the newly devised systems. In addition, Miller’s pyramid of professional confidence - what professionals *know*, what professionals *know how* to do, what professionals *can show* and what professionals *do* (discussed further below) - provides a helpful framework for differentiating practice and how it can be most effectively assessed.

**The scoping review remit, structure and sources of evidence**

It is against this wider professional backdrop that this scoping review seeks to answer the following overarching question:

What do we know about the use and efficacy of practice observation as a method of assessing professional capability in social work and other professional settings, both at qualifying level and beyond?

The review is structured around seven key questions and draws on knowledge from the following sources:

- Literature searches of the role of observation in social work and related professional domains.
- Focus groups: Local Authority practitioners (incl. ASYE), practice educators, managers, workforce development; HEI Practice Learning colleagues; social work doctoral students and researchers.
- Academic colleagues with practice and research expertise in the field.
- Academic colleagues utilising direct observation as an assessment method.
- Social work students perspectives on participating in an Objective Structured Clinical Exam (OSCE) assessment.
- Practice colleagues across the sector who are using innovative approaches to observation for the assessment of professional competence.

Each further section addresses one of the seven questions and begins with a bullet pointed summary of the key messages emerging from the literature.
3. What methods of observation have been used to assess social work practice (including observation of practice used in initial education)?

<table>
<thead>
<tr>
<th>Key messages from the literature:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Direct observations of ‘real practice’ are primary mechanisms for assessing social work capability at qualifying level.</td>
</tr>
<tr>
<td>• Most assessment of the capability of qualified practitioners does not involve direct observation.</td>
</tr>
<tr>
<td>• OSCEs have some potential for assessing practice capability but there are issues regarding their validity and reliability.</td>
</tr>
<tr>
<td>• Alternative observation approaches for assessment purposes are not widely used and as yet have not been reliably tested.</td>
</tr>
<tr>
<td>• Role and skills of the practice educator/assessor are pivotal to effective assessment.</td>
</tr>
<tr>
<td>• To maximise the validity, reliability and credibility of summative assessment strategies, they need to be undertaken in ways which are experienced as developmental and constructive i.e. formative.</td>
</tr>
<tr>
<td>• Importance of any observation method including a reflective dialogical component.</td>
</tr>
</tbody>
</table>

Observation of practice in the social work domain falls into three broad categories (see Glossary), which in some instances overlap:

- Direct observation
- Simulated observation
- Video/audio-recorded observation.

Direct observation is the most widely used with students and some early career qualified practitioners. To date video or audio recorded and simulated practice approaches have not been widely used in the UK social work context, although there are indications that in the longer term they may have some potential for assessing professional capability (Bogo et al, 2014).

**Direct observation**

In its current form, direct observation (first introduced in 1991) is an integral component of qualifying training with students and involves three mandatory direct observations per practice placement undertaken by practice educators and other social work practitioners (Murphy and Laxton, 2014), as required by the Health Care Professions Council (HCPC) who, in 2011, took over responsibility for validating social work qualifying programmes. Direct observations continue into the Assessed and Supported Year in Employment (ASYE) with two such observations, being undertaken by practice assessors in the course of practitioners’ first post-qualification year in practice, although this is not a mandatory requirement and is at the discretion of the employer/HEI partnership to determine. Beyond this point there is no HCPC mandatory requirement for further direct observations to be undertaken, although they may be embedded in some HEI.
continuing professional development modules, for example, practice education qualifications. In an ideal organisational culture, opportunities for informal observation/feedback gathered from, for example, performance in meetings, co-working etc, makes on-going observation feedback a familiar workplace practice. In turn this reduces the ‘pressure’ that can be associated with formal observation practices and mitigates the unreliability that can come from a pressurised one-off observation (see below for further discussion regarding the reliability of assessment methods).

The observations at the qualifying and post-qualifying stages, as stipulated in the HCPC guidance, can be of any aspect of practice, for example a home visit, working with a child in a neutral setting or contributing to a child protection meeting, and involves the student being watched by a practice educator, another social work practitioner or another colleague who is physically in the room with them. Following the observation both parties complete proformas that reflect on how the encounter was conducted and identifies strengths and areas for improvement.

There is a limited body of literature that has evaluated the efficacy of direct observation methods. In a small scale study Murphy & Laxton (2014) evaluated the Assessed Learning in Practice Situations (ALPS) observation tools designed to gather multi-sourced feedback by promoting opportunities for the student to ask for clear evaluative comment from a variety of sources. This process was intended to promote student learning through reflection (Boud, 2000). The model involves:

- direct observations of a student’s practice by practice educators
- the student and practice educator completing reflective accounts of the observed practice
- feedback from colleagues and service users.

Whilst this method was found to promote reflective practice it has not been used to measure enhanced professional capability. Larger scale studies are required to assess its effectiveness in all respects. More broadly within the literature, concerns exist regarding firstly, the accuracy of student/practitioner recall of their practice in general (see Haggerty and Hilsenroth, 2011, in relation to psychotherapy practice) and secondly, the potential for overly generous assessments by practice educators/assessors with pre-existing relationships with students/practitioners (Bogo et al, 2007, 2011; Dorfsman, 2009; Humphrey, 2007). One way to address these concerns regarding the potential for a lack of consistency in the direct observation process would be to develop more standardised assessment recording documents (discussed below).

**Objective Structured Clinical Exams (OSCEs)**

Bogo et al (2011, 2013) have adapted the Objective Structured Clinical Exam (OSCE) model, a well-established simulated assessment mechanism used in medical training, for social work contexts. The OSCE is a performance-based assessment method that measures aspects of professional competency that knowledge-based tests are less suited to capturing. Adapted for social work settings, this version of the OSCE uses quantitative standardised measurement inventories plus qualitative reflective accounts and dialogues focused on the ability to
conceptualise practice. In this way Bogo et al (2011) employ OSCEs more comprehensively than is the case in the medical domain to assess both procedural/technical competences i.e. operational, behavioural techniques and performance skills needed to carry out professional tasks and what they refer to as ‘meta-competences’ i.e. higher order, overarching qualities, abilities and cognitive, affective and reflective capacities (Bogo et al., 2006; 2007; 2011; 2013; 2014).

Participants in the OSCE pilot studies that have been conducted by Marion Bogo and her colleagues in Canada have included current Masters social work students, recent graduates and experienced social workers. An important feature of the Canadian context for practice learning is that it does not have a mandatory requirement to include direct observation of practice at qualifying level as is the case in the UK. Consequently OSCEs are the closest that practice educators can get to ‘real practice’. According to Bogo current competence inventories, including those utilised in the UK (Taylor and Bogo, 2014), can be overly mechanistic, devoid of context, and lose sight of the holistic nature of practice. In their model, social work students take part in a five-scenario OSCE adapted for social work. Each scenario consists of a 15-minute interview with structured questions, followed by a 15-minute reflective dialogue designed to elicit dimensions of meta-competence. Students’ reflections revealed variation in students’ ability to conceptualise practice, to reflect on their professional persona and to learn from reviewing their practice. A post-interview reflection with standardised questions tapping into meta-competencies and rated by trained instructors identifies:

- what and how students are thinking and feeling about their practice
- how students link theory to practice and demonstrate critical thinking, judgment, decision making
- nature of subjective experience
- quality of self-assessment and learning.

Some students described becoming emotionally dysregulated (i.e. being affected by the emotional content of the encounter) and therefore impaired in their use of knowledge in practice. This finding, underpinned by neuroscience knowledge of brain function in stressful situations other than in social work settings (Bogo et al, 2013; Lewis and Todd, 2007), has important implications firstly, for educators and Practice Supervisors assessing practice, and secondly, for the design of assessment methods that can capture the emotional dimensions of professional experience. Crucial to managing this aspect of practice is provision within the assessment method for reflective conversations about incidents of dysregulation, in order to better understand and address them.

Baez (2005) assessed the use of OSCEs on a small scale in the US, finding they have relevance and applicability for assessment of social work skills and capability in more direct ways. Baez, confirmed that in terms of the characteristics of a credible assessment process the OSCEs were deemed to achieve acceptability to all stakeholders, feasibility and educational impact, indicating
that OSCEs had considerable potential in social work education settings. What was not substantively addressed in this study, however, were the reliability and validity (see glossary) of the OSCE method. Further research into this aspect of the OSCE is required which will include determining in particular:

- what type/level of training the OSCE assessors require to maximise the objectivity and standardisation of the method (validity)
- how scenarios can be designed to most accurately simulate the complexity of practice e.g. 10-15 mins does not allow for demonstration of some skills, therefore variable lengths/types of scenario are required (reliability).

Lu et al (2011) used mixed-methods to evaluate use of OSCEs and devised a Clinical Competence-based Behavioural Checklist (CCBC) in US and Taiwan. CCBC seems to be similar to the UK model of assessing for competences. It found that CCBC had potential as an objective, structured and competence-based checklist for clinical assessment. The OSCEs were found to be a viable supplement to traditional methods, but further work is required on assessing the internal reliability of a social work OSCE. In keeping with Baez (2005) and Logie et al (2013), Lu et al (2011) emphasise the importance of developing specific 'fit for purpose' social work scenarios that capture the breadth, depth and diversity and complexity of social work practice.

**Comparing simulated scenarios with direct observation**

Bogo et al (2012) suggest that there are some correlations between OSCE performance and practice in the field, but do so with considerable caution, underlining the need for further carefully designed research to 'test' these findings further. Forrester et al (personal communication, 2015) are engaged in a number of studies assessing social worker communication skills using a variety of research methods, including simulated scenarios and comparing them with direct observations of practice. Whilst they are still involved in analysing their data, early indicators suggest that there may be correlations between the results arising from assessing professional capability by direct observation and simulated means. Alongside these nascent findings, a number of key issues have emerged from these studies:

- the complexity of social work skills and identification of what is empirically measurable e.g. warmth and what can be captured in a research code e.g. humour
- the breadth of dimensions of practice that can be coded for research purposes
- the Motivational Interviewing model used in this research is restricted to one-to-one interaction and does not include the complexity, for example, of working with a family in court contexts or in inter-professional meetings such as child protection case conferences
- the challenges in developing coding frames e.g. some elements appear relatively universal and it is possible to code for them (autonomy, collaboration, empathy, purposefulness) compared with some elements which are not always easy to code for and may need
analysis over longer period – e.g. evocation (requires a behaviour change focus), clarity about concerns (depends on level of concern), focus on child (again, focus of interview shapes this)

- importance of having coding teams to develop valid and reliable tools.

**Video and audio observation**

Evidence of using video or audio recordings as part of assessment strategies beyond the HEI context is negligible in social work. One exception was the audio-recording of simulated phone interviews undertaken as part of a large-scale project to recruit prospective social workers to a new qualifying programme (Forrester, personal communication, 2015). The assessment of practice capability was based on Motivational Interviewing (see glossary) schedules. Whilst it appeared that assessing capability was feasible using this method, one challenge identified by the participants was having in-depth and exploratory discussions over the phone. They felt it was possible to do, but a few said that they found it slightly unrealistic and unauthentic as they would not do this in real life. A plus side of doing it in this way was its flexibility and cost-effectiveness, allowing the assessment interviews to be conducted anywhere. Several features of this observation method suggest its validity/reliability and overall credibility for assessing professional competence is potentially less than other observation methods. Firstly, unless the assessor takes charge of the observation process they rely entirely on the co-operation of the practitioner concerned to determine the representativeness of what is recorded and when it takes place. Secondly, whilst capable of capturing ‘pure observations’ of practice, video and audio recording of practice does have a serious disadvantage insofar as it forces the assessor to rely upon visual and auditory senses at the expense of their other senses. It does not include their attunement to what Ferguson (2010) refers to as ‘atmospheres’ and the immediately visceral, often overwhelming dimensions of practices, for example very smelly households or chaotic domestic settings. A third complicating aspect of using video or audio recordings for assessment purposes are the ethical, confidentiality and Data Protection Act issues that they generate.

In contrast to the limited evidence of video/audio recording being used in social work practice, a number of qualifying social work programmes include students conducting video-recorded interviews with service users (or actors in many instances). Feedback is given by the tutor and the service user/actor and the student reflects at the time. This is followed up by the student writing about the experience having watched their own recording and submitting their write-up for assessment. Educational practices such as these, combined with other digital technology advances, are opening up new possibilities for different forms of assessment of capability in a range of settings. Whilst the concerns about the potentially limited assessment capabilities of digital methods, most notably their capacity to capture the full richness of real practice, remain, it is worth noting that advances in technology such as the use of algorithms and the increasing
capacities for complexity in computer science, might offer useful ways of carrying out assessment and consequently informing practice capability in the not too distant future.

As it currently stands, however, direct observation is the only assessment method that does directly engage with all the complexities and nuances of real practice.

**The role of the practice educator/assessor**

It seems the jury is out when it comes to the role of the practice educator/assessor and the degree to which they should be independent of the student/practitioner. On the one hand, from a formative perspective, if the practice educator/assessor is known to the student/practitioner feedback is more likely to be incorporated into practice, on the other hand, if applied summatively it raises the question of bias and lack of objectivity, which in turn has implications for the validity of the assessment process. Humphrey’s (2007) study of the role of the practice educator/assessor underlines the complexity of observation as an assessment strategy and suggests that, ‘The most accurate and adequate account will be that which is grounded upon detached observation at the time and guided by inter-subjectivity after the event (p.725).’ By this, Humphrey is referring to an observer (practice educator/assessor) who does not have a pre-existing relationship with the person being observed (student or practitioner), but who then engages in a reflective conversation and discussion with them in order to access their understanding of what happened and why they behaved in the way they did. Humphrey highlights the paucity of empirical research into direct observation practices and raises particular concerns regarding the challenges associated with the assessor role and the importance of ensuring their assessments are legitimate and do not ‘contaminate the observation by straying into the realms of participation and intervention, and then to curtail the assessment by downplaying the views of other parties (p.723).’ Humphrey (2007) suggests an observational continuum ranging from a freelance practice educator at one end, privileging neutrality and considered ideal for summative assessments to the abandonment of observation for assessment purposes in the believe that ‘everyday life offers ample opportunities for students’ practice to be observed by their supervisors (p. 735),’ privileging naturalness, ideal for formative assessments, at the other end. In between is a portfolio route incorporating direct observations undertaken by a range of personnel.
Formative versus summative assessment strategies

In highlighting these options Humphrey (2007) raises the important point about the status of the observation - whether it is formative or summative. Most of the research studies focus on formative assessment of students, with a few engaging with it in a summative context with students and/or practitioners. The dilemma revolves around the potential for more formative approaches to yield richer, more reliable data and outcomes, as students/practitioners feel less pressurised to ‘perform’ and, therefore, behave authentically, compared with the tendency of summative approaches to constrain professional performance, which then generates narrower, less reliable data and outcomes. Whilst the focus group respondents engaged in this scoping review understood and accepted the observation tasks being considered for this scope as a summative exercise, they were keen to emphasis the importance of any summative exercise being undertaken in a way which was experienced as developmental and constructive i.e. formative. As Humphrey articulates, the binaries that surround the assessment of professional capability, (i.e. scientific vs humanistic perspectives, formative vs summative purposes) need attending to as they have significant impact on the reliability and validity of different assessment methods (discussed further below) but are not insurmountable.

The importance of reflective dialogue

The literature review and focus group feedback informing this scoping review confirm the pivotal role of the reflective dialogue for ensuring that both the observable knowledge and technical skills, what Bogo et al (2013) refers to as the ‘procedural competencies’, and the more subtle and less visible dimensions of practice - the critical, reflective and analytic knowledge and skills – that Bogo et al (2013) call ‘meta-competencies’ are accurately captured. Observation in and of itself is of limited value if there is no opportunity for the student/practitioner to account for their conduct. When a reflective component, that allows for the rationale lying behind specific behaviours to be
made explicit, is incorporated into an observation method the reliability and credibility of the approach increases significantly. This echoes the important messages contained in the Munro Report (2011) regarding the need for critical, reflective and relationship-based practice and supports the notion that the incorporation of sense-making and self assessment/reflective components into the digital solution would enhance the credibility of the overall assessment package.
### Advantages and disadvantages of the three modes of observation

Reviewing the use of observation as a method for assessing professional capability the advantages and disadvantages can be summarised as follows:

<table>
<thead>
<tr>
<th><strong>Advantages</strong></th>
<th><strong>Comment</strong></th>
<th><strong>Disadvantages</strong></th>
<th><strong>Comments</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoids impact of imprecise memory recall</td>
<td>Videoing/audio-recorded practice could achieve this and have ‘captured’ evidence to refute it</td>
<td>Overly generous grading by practice educator/assessor</td>
<td>Question of familiarity between observed individual and observer can be overcome to an extent by increasing number of observations, using different observers and using standardised tools</td>
</tr>
<tr>
<td>Enables assessor to get a full picture of student/practitioner in context</td>
<td>May not 'see' all interactions in one session; will need to determine how many direct observations and how frequently to get a fuller and accurate measure of capability</td>
<td>Assessor may be overly guided by assessment criteria (risk of using standardised tools)</td>
<td>Depends upon how prescriptive the tools and methods for measuring the assessment are; risks leading to a reductive approach, raises validity-reliability dilemma</td>
</tr>
<tr>
<td>Gives opportunity for valuable feedback and learning; highly developmental component</td>
<td>Feedback being incorporated into practice depends on quality of the assessor; dilemma of balancing summative-formative nature of observation process</td>
<td>Use of specialist time and resources in busy and pressured conditions</td>
<td>Research is needed on the effectiveness of direct observation in order to justify the use of time and resources – though this is arguably true of all methods</td>
</tr>
<tr>
<td>Systems/structures for and expertise in direct observation exist within the sector – allows for a capacity building approach with observers drawn from the sector</td>
<td>New observation initiatives can build on existing infrastructure and expertise, hence reducing additional costs associated with such developments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Additional benefit arising for using direct observation as a summative assessment measure are changes to practice and behaviour with benefits for service user</td>
<td>Feedback and learning change practitioner behaviour and this might have benefits for the service user, though there is conflicting evidence on this and the area needs more research. Moral and ethical reasons to support direct observations, e.g. openness, honesty, transparency of practice and checking for safety and efficacy of the worker?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
<td><strong>Comment</strong></td>
<td><strong>Disadvantages</strong></td>
<td><strong>Comments</strong></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Standardisation: most feasible way to reduce variability in assessing practice; maximises 'objectivity'</td>
<td>Overcomes recognised tendency for direct observation to lead to overly generous assessments</td>
<td>Authenticity: does not fully capture reality and has set pre-determined measurement criteria</td>
<td>Can it fully capture visceral dimension of SW practice? Cannot offer an accurate picture of student/practitioner's reaction due to the false setting and may therefore offer inaccurate evidence</td>
</tr>
<tr>
<td>Gives students/practitioners the opportunity to explore their own practice in a 'safe' environment</td>
<td>Beneficial as a learning and development tool</td>
<td>Question of how 'immediate' feedback might be i.e. during an encounter or post encounter</td>
<td>Risk of delayed responses/feedback if assessor not immediately engaged in assessment task</td>
</tr>
<tr>
<td>Assessors have control and can alter 'scripts' of simulated service users or practice encounters</td>
<td>Can build in complexities and assess reactions and student/practitioner competencies</td>
<td>Use of actors, special venues</td>
<td>A costly approach in terms of additional specialist resources required (a point emphasised by Bogo et al (2013) in terms of organising OSCEs to a good enough standard)</td>
</tr>
<tr>
<td>Advantages</td>
<td>Comment</td>
<td>Disadvantages</td>
<td>Comments</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Assessment material captured in real practice</td>
<td>Authentic: overcoming much of the issues related to simulated practice assessment</td>
<td>Questionable reliability: accuracy/representativeness of what is recorded; cannot fully capture the 'atmospheres' of practice e.g. smelly homes, chaotic visits</td>
<td>To what extent can the 'whole' encounter be captured e.g. assessing students'/practitioners' responses to unsanitised practice realities in these types of tests</td>
</tr>
<tr>
<td>Assessor can be more objective and does not necessarily have a pre-existing or even observational relationship with student/practitioner</td>
<td>Increases validity of the assessment</td>
<td>Reliability/Accuracy of assessment process</td>
<td>Impact of video /audio recording - heighten 'performance anxiety', artificiality for all parties and hence accuracy of practice, and reality of encounter. Familiarity will change extent of this problem, as will depth/time of the event</td>
</tr>
<tr>
<td></td>
<td>Questions of how 'immediate' feedback might be i.e. during an encounter or post encounter</td>
<td>Risk of delayed responses/feedback if assessor not immediately engage in assessment task</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Complexity of data protection and confidentiality</td>
<td>Need to attend to issues of data storage etc</td>
</tr>
</tbody>
</table>
4. What evidence is available about the reliability and validity of each of these observation methods as a predictor of actual performance in the field?

Key messages from the literature

- Measuring the reliability and validity of the different methods of observation for assessing professional capability - direct observation, OSCEs or video/audio-recorded observations - is complex.
- Reliability and validity of observation methods offer no guarantee of predicting future professional capability.
- Careful cross-referencing to the Knowledge and Skills Statements is essential for the reliability and validity of any observation method.
- The creation of standardised assessment and reflective recording templates is essential for the reliability and validity of any observation method.
- The reliability and validity of the different observation methods varies depending on whether it is used for summative or formative assessment purposes.
- Emergent research evidence suggest that simulated observation methods do correlate to some degree with the assessment of real practice.
- The limited evidence base (from all professional contexts) concerning the reliability and validity of different observation methods suggests that more observations per practitioner may be necessary to enhance the credibility of any one method.
- The number of observations assessed and quality/speed of feedback for improvements in practice are significant indicators of the assessment method’s impact and credibility.

Reliability and validity of observation methods for assessing social work practice

Research undertaken in scientific settings has particular understandings of the terms reliability and validity in relation to the research design and hence the credibility (or otherwise) of the research findings. Applying the terms reliability and validity to research conducted in social science settings is not straightforward. None of the cited studies exploring the role of observation in assessing professional capability in practice are based on the findings of a randomised controlled trial, usually regarded in research circles as the ‘gold standard’ type of research in terms of its reliability, validity and credibility. Where the knowledge in this scoping review is informed by empirical research findings most of the studies are based on qualitative methodologies that analyse the accounts of the experiences of key stakeholders involved in the assessment process. To this end, adapted understandings of reliability and validity are used in the context of this scoping review although individual research papers cited do provide their own definitions (e.g. Bogo et al 2011). For the purposes of this review:
**reliability** refers to the ability of an observation assessment method to provide an accurate representation of what is involved in social work practice – how accurately does it replicate and capture/measure what social workers do in everyday social work encounters with individuals and families in their home settings, in meetings and in other professional contexts? How far an observation method is deemed ‘reliable’ is determined by the extent to which what is being observed and assessed is an accurate, ‘reliable’, reflection of ‘real’ everyday practice. In social work contexts the uniqueness, diversity and complexity of practice means the likelihood of needing to observe real practice directly to achieve reliability is high.

**validity** refers to the extent to which the measures used to answer the research question are the right measures and can provide a standardised response i.e. if two different people observed the same piece of practice how far would they replicate their findings. An approach rated highly in terms of its validity is associated, therefore, with a high degree of standardisation and use of standardised measures which can be repeatedly administered in a consistent manner.

Rushforth’s (2007) research into the use of OSCEs to assess capability in nursing, emphasises the high quality ‘gold standard’ of this assessment method but the problem of ‘an inverse relationship between validity and reliability, with the strengthening of one potentially weakening the other,’ meaning ‘achieving good reliability and validity is arguably a considerable challenge (p.483).’ In raising this dilemma Rushforth is outlining how, by developing more standardised measures for assessing capability, validity increases but at the expense of having measures which capture the subtlety and nuance of complex practice scenarios which are considered more reliable. Conversely, more complex and subtle measurement criteria make the assessment artefact potentially more accurate and hence reliable, but simultaneously less valid. The diagram below illustrates the complex inter-relationship between these two important concepts. Designing an observation assessment strategy is therefore a complicated activity. Overcoming this challenge may well mean having more than one observation assessment method and more than one observation at any one point in time, if an observation assessment strategy is to be deemed to be reliable and valid and, therefore by implication, credible.

**The interconnections between reliability and validity of observation methods**

The digital solution can offer higher degree of validity and practice observation can offer higher level of reliability; two combined methods enhance the credibility of assessment and accreditation systems.
Reliability and validity in direct observation

Reviewing the literature highlights that issues of reliability and validity in direct observation approaches to assessing professional capability rest on:

- the use of standardised observation recording templates
- the quality of the practice educator/assessor
- the use of standardised reflective templates.

This suggests, therefore, that if direct observation is the assessment method that is selected to complement the digital solution its use will be enhanced by:

- close cross-referencing to the Knowledge and Skills Statements to inform the creation of explicit assessment schedules and reflective recording templates
- fit-for-purpose training for practice educators and assessors that equips them to know what is being assessed and how best to assess it.

Humphrey (2007) suggests that evidence of the reliability and validity of direct observation of practice is lacking. Having conducted research into the practice educator/assessor role in the direct observation process Humphrey strongly argues that if the reliability and validity of the direct observation assessment process are to be strengthened, then a triangulated model of assessment of capability is essential, that draws together the different stakeholder perspectives, i.e. student/practitioner, practice educator/assessor and service user/other colleagues on the observed practice.

Current direct observation practices do include a reflective component with recording templates devised by the employing body/training agency. No standardised nation-wide reflective template exists to ensure consistency in the approach taken to the reflective task or to facilitate measuring how valid the assessment artefacts are. More general social work articles published in non-peer reviewed outlets, which focus on the inter-personal aspects of direct observations, do nevertheless indicate that rigorous assessment recording templates are available (see Appendix 2) but there is little evidence of the extent to which they are applied consistently. The majority of the academic
literature focuses on ethical aspects and interpersonal factors related to carrying out direct observations including:

- power relations
- the impact of being observed
- the various agendas of those involved in the process i.e. is it assessment or developmental.

By directly observing practice its uniqueness can be captured, offering a high level reliability and when this is combined with the use of a standardised reflective tool, a higher level of validity can also be achieved. A consistent framework and guidance in relation to practitioners, needs to be developed, but can build on and adapt the existing procedures/paperwork designed for practice educators and assessors and there is considerable scope for developing and sustaining good practice through Direct Observer forums and national bodies, for example, Skills for Care etc. Feedback from the scoping review focus group, however, indicated that the tools currently used were acceptable and fit for purpose. The focus of these respondents’ concerns was less on the paperwork and more on the personnel involved in the assessment process and the levels of consistency of practice educators/assessors knowledge and skills to perform the assessment task. This concern echoes those expressed in the literature (Humphrey, 2007).

Direct observation as a form of assessment is argued by some to be reductionist and devaluing and deskilling of practitioner capability and as un-holistic (Gysels et al 2005; Rushforth 2007), while others see it as a useful tool that can be incorporated into existing assessment and training practices.

**Reliability, validity and OSCEs**

In relation to simulated approaches to observation Baez (2005), Bogo et al (2013) and Lu et al (2011) all acknowledge and problematise issues of validity and reliability, recognising that they require further research. Bogo et al (2014) and Logie et al (2013) clearly state that using the OSCE approach requires further carefully designed research to determine its suitability for assessment of professional capability in social work, as its validity and reliability is as yet (as is the case for direct observation) insufficiently proven. In their critical appraisal of eighteen studies, using standardised client simulations in social work education settings, Logie et al (2013) conclude, 'Consistent implementation methods and reliable, valid assessment measures are needed to advance this evaluation method for social work (p.66)'. Furthermore, there is limited research about the effectiveness of this approach in producing improved student practice capability, nor has it been used widely with practitioners. That said, Logie et al (2013) do report high student satisfaction with this method. Students report feeling more confident and knowledgeable as a result of conducting interviews and receiving feedback and guidance about
future learning for practice. As a method it is widely accepted as fair and authentic by students, qualified early career practitioners (ASYEs) and faculty members.

Limitations of the existing studies relate to:

- the voluntary nature of the participants and the way in which a particular sample of volunteer students reflections on their interviews may not be representative of other social work students
- the constraining nature of standardised questions, with questions other than those posed within the standardised assessment template, potentially resulting in different responses and different competencies needing to be assessed
- studies to date having focussed primarily on students and not qualified practitioners, with potential for different responses from practitioners revealing other dimensions of meta-competence that require measurement.

A notable gap in existing knowledge is the suitability of OSCEs for assessing professional capability beyond the qualifying/newly qualified practitioner stage. No evidence currently exists regarding its applicability for assessing professional capability at the Practice Supervisor or Practice Leader levels of the career trajectory.

In order to develop an OSCE-type approach to observing professional competency in social work, it is vital to consider the following:

- Context: implications for the design and number of practice scenarios associated with the diversity of social work settings and the range of practice approaches requiring different knowledge and skills
- Specificity: high level of specificity with regard to the measurement criteria required in order to accurately capture subtlety of all aspects of social workers’ knowledge and skills, including the ‘meta-competencies’
- Skill definition: require definitions and methods that capture the holistic nature of practice; what type of knowledge is used and how it is used (internal cognitive and emotional processes that contribute to decision making and performance; ‘meta-competencies’).

These are important components of any model to maximise its credibility, rather than being specific to the implementation of OSCEs.

Logie et al (2013) make the important point that to date such assessment approaches have been used primarily for educational and formative, as opposed to evaluative and summative, purposes with an over emphasis in the reported outcomes on student satisfaction and an under-emphasis on student performance. They underline the need for more research into the efficacy of these approaches for both ‘improving student learning and in assessing student competence (p.78).’

Added to this is the need for more research into its relevance for the learning and capabilities of Approved Children and Families Practitioners, Practice Supervisors and Practice Leaders.
Reliability, validity and predictability of observation measures

Humphrey (personal communication, 2015) raises important questions regarding the predictive strength of any observation assessment method and emphasises the importance of supportive and development organisational contexts for maximising practitioner potential and for raising practice standards. She cautions against the conflation of the reliability and validity of an assessment method with predictability of performance. According to Humphrey all assessment methods measure actual performance in relation to either a specifically designed task/test or a real-life encounter with relevant people in a natural setting, but however valid and reliable they may be in relation to assessing past progress and current performance, it is questionable whether they could be used to predict future performance in other tasks, tests, encounters or settings. The more experience practitioners have at any given task or test, and the more they work with any given set of service users or in any specific setting, the better their future performance will be in relation to a comparable task/test/encounter/setting. This capability, however, is attributable to their professional experience/expertise, rather than the assessment of the practice, raising the summative-formative dilemma and fundamental question about how practice standards can best be driven upwards – by formal, occasional and relatively artificial assessments of practice or by informal, on-going and natural opportunities for co-working and holistic supervision with collegiate feedback, all of which arise in everyday working contexts? To address this concern and maximise the efficacy, reliability, validity and credibility of observation as a means of assessing professional capability, will require assessment strategies that complement and are complemented by organisational cultures that incorporate observation into everyday practice. Essentially, this means that to maximise the credibility and efficacy of any observation method employed to assess professional capability, it needs to sit alongside increased observation activity more generally within the workplace. Unavoidably, this requirement in the short term will have cost implications and will require significant shifts in some organisational culture and practices. Organisations which have not embraced observation as a formative approach in their everyday practice will need to build the necessary infrastructure to deliver this, for example, employing sufficient assessors and delivering training programmes to educate staff. In the longer term it is anticipated that the benefits of a confident and capable workforce and improved practice will outweigh the initial costs (see Appendix 3 for examples of practice observation).

In relation to the use of OSCEs’ Bogo et al (2013) make some cautiously optimistic statements about the predictive capacity of OSCE assessments. Their caution, relates to the small scale of existing research and the need for further larger scale studies to be conducted and reported before their current tentative claims can be consolidated or contradicted.
5. What lessons for social work can be learned from the direct observation methods used to assess practice in other related professions? How do they assess the reliability and validity of their methods as a predictor of performance, if at all?

Key findings from the literature

- Comparing assessment strategies from other professional domains must be undertaken with caution given the distinctive nature of social work practice.
- Medicine and nursing rely on OSCEs as the primary tool for assessing initial and on-going professional capability, with some OSCE derivative methods also used (eg in transferring to UK having trained and qualified overseas).
- Education involves multiple observations by practice educators/assessors/peers and argues for triangulated observation methods which include immediate, positive, corrective and specific feedback in order to maximise credibility and effectiveness of the observation method.
- Psychotherapy and counselling include some evidence of using video/audio-recording of practice.
- Allied Health (occupational therapists, physiotherapists etc) suggest direct observation is preferable; importance of allowing time for changes to happen.
- Challenge of developing assessment artefacts that are both valid and reliable is an issue across all the professional domains.
- The validity and reliability of all the different observation methods are universally problematic and inversely related.
- Validity and reliability issues impede the ability of observation measures to predict performance.
- Evidence to suggest that some independence of the practice educator/assessor enhances the validity of the assessment process but equally an on-going educator-student/practice assessor-practitioner relationship can add continuity and potentially greater reliability over time.
- Complex inter-relationship between methods used for formative and summative assessments and ways to maximise their accuracy.
- Significance of number of observations assessed and quality/speed of feedback for improvements in practice, and an indicator of assessment impact.
- Potential of summative assessment methods can be enhanced by embedding similar assessment methods in a formative capacity into organisational contexts.

Before outlining assessment practices in other professional domains it is crucial to reiterate the distinctiveness of social work, mentioned in the opening background section, compared with other disciplinary practice. Attention in social work has been turned particularly to practices in the fields of medicine and health care where, as detailed below, OSCEs are the assessment artefact of
choice. Whilst clearly having some merits and relevance for social work contexts the applicability of OSCEs to the assessment of capability of social work practice is not straightforward. It is difficult to think, for example, in the social work arena of anything that is equivalent to the giving of an injection, medication or diagnosis, all tasks that require definitive ‘right’ answers. The problem is that there are few comparable activities in social work that need to be applied to the same standard consistently and have scientific evidence underpinning them such as the use of medication, injections, blood pressure taking etc. So OSCEs tend to be about ‘doing things right’ and are less focussed on the process of doing things, the domain in which most social work activity is located (though OSCEs may also seek to assess, for example, the establishing of rapport between clinical practitioner and patient). Further more, social work evidence is predominantly qualitative in nature. Therefore, even if the starting point in practice is consistent, for example, a medical or social work investigation, as soon as material is gathered in social work settings, it becomes a ‘process’ event that involves weighing up a range of complex factors. The content very quickly becomes dissimilar from that of a medical investigation.

Throughout the literature in other professional domains a recurring theme is the tension between the need to find ways of accurately assessing practice in order to support practice improvements through developing rigorous and valid tools versus reliably capturing the complexity and case-specificity of each encounter. As noted, triangulating, using a range of assessment methods and standardised tools, is one way in the social work domain to overcome these challenges.

**Medicine and Nursing**

Direct observation is extensively used to assess practice in the medical and nursing professions. Although not an exhaustive list, studies included in the review give a good impression of the advantages and disadvantages of direct observation in these fields and most studies find that direct observation is an invaluable means of assessment in the context of the on-going development of students and practice. Most of the approaches are used in the context of formative assessments but are equally applicable to summative settings. That said it is vital for the efficacy and accuracy of summative assessment strategies to be maximised i.e. for them to most accurately capture and measure individual capability they need to be conducted in such a way that the adverse effects of observation e.g. performance anxiety, are minimised. This then challenges the idea that ‘the tougher the test the more accurate result.’ Rather it suggests that the best assessment and accreditation systems should complement and be complemented by workplace practices, i.e. observation should be part of everyday practice and not an unfamiliar experience. Such systems, then, whilst assessing capability, will provide an opportunity for those being assessed to demonstrate practice that most accurately reflects their professional capability and are more likely to be positively experienced as both judging and developmental.
A plethora of tools exist that attempt to capture and measure what goes on in direct observation assessments, for example, the Mini Clinical Evaluation Exercise (Mini-CEX), Clinical Evaluation Exercise (CEX app), Objective Clinical Structured Exam (OSCE), Direct Observation in Clinical Practice (DOPS), Direct Observation in Clinical Settings (DOCS), The Maastricht History-Taking and Advice Checklist, the Direct Observation Clinical Encounter Examination, Clinical Encounter Card, Direct Psychiatric Clinical Examination and Revised Infant Video Questionnaire. The main finding is that when used alongside other forms of assessment, i.e. as a mode of triangulation, direct observation tools can be very useful, particularly through offering protected teaching time in which experienced clinicians can identify problems and feedback to students immediately. As such they are experienced by those being observed as a developmental more than a judgemental activity. Despite the benefits, there remain questions over:

- the ability of particular observation tools to fully capture the diversity and complexity of an encounter e.g. need to do several observations over time
- the validity and reliability and the development of standardised measures
- the potential de-skilling of professionals through the introduction of assessment tools that fail to capture a more holistic form of medical care and the application of critical and competency skills.

These concerns reinforce the need for formative versions of the summative assessment methods to be embedded in organisational contexts to maximise the accuracy of summative measures by reducing performance anxiety and to minimise their potential negative impact on everyday practice, workforce efficiency/wellbeing and ultimately service user wellbeing.

**Medicine and direct observation**

Singh et al (2014) have conducted research into using the Mini Clinical Evaluation Exercise (Mini-CEX) and Direct Observation in Clinical Practice (DOPS) as assessment tools. The Mini-CEX method involves direct observations of different aspects of practice undertaken on 6-8 occasions per year with each encounter observed by a different independent assessor. The frequency of observation and role of the assessor help to increase the reliability and validity of the Mini-CEX which is reported to be much higher than that for the OSCE (Singh, 2012).

Singh et al (2014) also report on the use of Direct Observation of Procedural Skills (DOPS) which take place in real life settings and involve ‘observation and evaluation of a procedural skill performed by a resident on a real patient (p. 715)’. Like Mini-CEX, DOPS serve the twin purpose of assessment as well as enhancing skill learning. Examiner/assessor training is necessary for ‘providing developmental feedback based on direct observation, rather than on historical facts (p. 715).’ A key strength of both approaches is the ease with which they can be integrated into routine work and both provide a reliable assessment if six to eight such encounters are use.
Studies that include a theoretical component drawing on ecological, organisational and learning theories help to inform how the assessment process is experienced. For Dorfsman (2009) direct observation provides medics with specific and immediate feedback and ‘allows them to immediately incorporate that feedback into practice (p.350).’ The study also suggests that direct observation is ‘inherently subjective and subject to the biases of each observer (p. 350)’ in terms of education and clinical practice, hence the importance of undertaking more than one observation event and engaging more than one person to do it, and also why the person being observed has an important ‘checking/ confirmation’ role. The study reported by Cydulka et al (1996) highlights the use of direct observation in identifying problems and the opportunity this affords for establishing remedial action. Kogan et al (2009) acknowledge the multiple tools for direct observation that are available but recognise too that ‘their characteristics and outcomes have not been compared systematically (p.1316)’. Referring to a study of the Mini-CEX method implemented repeatedly with medical students, residents and fellows across specialties they conclude that ‘few [research study participants] demonstrated improved clinical skills or patient care quality with tool implementation in an educational program (p.1324).’ Kogan et al (2009) go on to recommend that ‘randomized study designs should follow whenever possible to assess whether the tool affects educational outcomes (p.1324)’. In a second paper Kogan et al (2012) cite the theoretical knowledge drawn from cognition and ecological psychology in an acknowledgement of the complex interactions that take place in both the observation process and the consequent giving of feedback. They see these as being vital to the learning and development of trainees. They suggest that ‘faculty staff need to learn techniques for delivering feedback in situations in which accurate self-assessment is lacking’ and that ‘feedback is more effective when the source is perceived to be credible...’ and ‘...is maximised when there is a sequential process that includes repeated instances of feedback rather than a series of unrelated events (p. 212)’, thus highlighting the importance of longer-term professional relationships and multiple observations, for high quality assessment; a view endorsed by Mazor et al (2011). In considering the question of how frequently observations might be undertaken, Madan et al’s (2012) exploration of enablers and barriers to direct observation reported that:

‘The most common response to the question about the ideal of direct observation was weekly for both residents and faculty. This was not supported as strongly by senior residents who reported less desire for more frequent observation (p.270-271).’

By inference this might imply that more senior social work professionals may be more averse to observation than frontline practitioners, perhaps due to time pressures – something that will need to be constructively explored and managed in terms of the Practice Leader assessment process.

Kang et al (2009) highlight the challenges associated with direct observation when using an approach called Direct Observation of Clinical Skills (DOCS). The DOCS rate interviewing, physical
examination and oral presentation and include an overall global rating on a 1–5 Likert scale, additional data collected on the forms includes information about the observer (attending/resident), clinical setting in which the DOCS session was conducted, student and observer satisfaction with the DOCS form, and observation and feedback time. The study found a statistically high rate of satisfaction with the programme and recommends future research on the benefits of the use of direct observation and most particularly its appropriateness as a summative assessment tool. Despite this positive finding, Kang et al (2009) acknowledge the inherent tension within the DOCS between ‘the ideal of frequent, real-time, validated assessment by faculty skilled in evaluation and feedback, and feasibility constraints of living and working in a busy medical center (p.1)’. Any assessment approach employed in the context of social worker assessment, therefore, needs to minimise the potential for it to negatively impact on everyday practice, workforce efficiency and ultimately service user wellbeing. Given that the majority of local areas report significant pressures on resources (see Association of Directors’ of Children’s Services website for report by Brookes and Brocklehurst on pressures on safeguarding ¹) and rising demand for services, the issue of ensuring assessment activity does not hamper social work teams and services meeting the needs of children is a critically important one.

Craig et al (2011) similarly cite the costs/risks and benefits for practice of direct observation, especially when combined with focused feedback. Their study, however, found no evidence to demonstrate a sustained change in behaviour indicating that ‘this may be due to the difficulty in attributing behaviour change to a single educational intervention (p.65).’ and suggest ‘comparisons of different models of direct observation (checklist-based assessment and feedback, opportunistic teaching, or a combination) are needed, to determine the optimal method of delivery (p.66).’ It also indicates that observation, as far as its impact on practice is currently understood, cannot in its current form be used as a measure of predictability of capability. This concurs with the cautious position that Bogo (2013) takes in this regard in the social work domain.

A number of studies raise the thorny question of how many direct observations should be conducted (Hasnain et al 2004; Madan et al 2012; Ryan et al, 2010). Hasnain et al (2004) argue that the overall quality and meaningfulness of ratings of clinical performance are improved when practice educators/assessors include more direct observations as a data-source and that direct observation be included as one of several data sources. They concluded ‘that when a rating form includes bases for preceptors’ [observers’] ratings, the ratings of those preceptors who use multiple sources, including more direct observation, are more reliable and valid’. What is clear from research across all the profession domains including social work is that there is no ‘magic’ number of observations that will ensure a reliable and valid assessment strategy. The most that

¹ http://www.adcs.org.uk/news/safeguarding-pressures.html
can be said is that the more observations there are the more credible the assessment strategy will be.

**Medicine and OSCEs**

Rushforth (2007) provides a thorough and comprehensive systematic review of the use of OSCEs and their application in the medical and nursing professions. According to her review the strengths and limitations of OSCEs include:

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greater level of objectivity than in most assessment of practice</td>
<td>Student stress and test induced anxiety affect performance and therefore accuracy and representativeness of assessment scores</td>
</tr>
<tr>
<td>Wide range of assessors results in reduced risk of bias</td>
<td>Fragmented attention to specific skills undermines capacity to undertake an holistic assessment</td>
</tr>
<tr>
<td>Positive reviews from students and lecturers</td>
<td>Reviews largely relating to experience of process, not effectiveness of method</td>
</tr>
<tr>
<td>Wide range of skills tested</td>
<td>Checklists are criticised as being insufficient to reveal students 'mastery' of complex skills and undermine examiners' expert critical judgement</td>
</tr>
<tr>
<td>Increased consistency of experience between students</td>
<td>Costly in terms of creating scenarios, employing actors, arranging venue etc</td>
</tr>
<tr>
<td>High level of reliability and validity reported (although not unequivocal)</td>
<td>Some dissent regarding extent of reliability and validity; need large number of different scenarios to ensure reliability and validity of assessment</td>
</tr>
</tbody>
</table>

Rushforth (2007) concludes by advocating for a range of measures for assessing student capability and that ‘used carefully it [an OSCE] can make a helpful and meaningful contribution to health professional education (p. 481)’. As already recognised, however, most of these studies focus on formative as opposed to summative assessment. The main risks arising from using OSCEs for the latter purpose are associated with its ability to accurately reflect real practice (a risk specific to the simulated nature of the OSCE) and its capacity to induce anxiety and impact on practitioners’ confidence, capability and performance (a generic risk in all summative assessment situations).

**Medicine and audio recording**

Audio-recording of practice has not been extensively reported in the medical domain. The Maastricht History-Taking and Advice Checklist (MAAS) has been developed in order to measure medical interviewing skills (Thiel et al, 1991). The MAAS-R consists of a checklist of concrete behavioural items, covering an extensive domain of medical interviewing skills. Twenty four
doctors participated in a consulting session with eight different standardised patients, each portraying a different case. Audio-recordings were obtained from each consultation, totalling 192 consultations. The recordings were independently scored by six general practitioners trained in using the MAAS-R. The results of generalisability analysis, considering the influences of doctors, cases and raters, are encouraging. In order to overcome case-specificity, feasible and reliable measurement can be accomplished with 8-10 cases in 2-2.5 hours of testing time, each case being scored by a different rater. Reliability improves considerably if assessment is restricted to basic interviewing skills. Poor reliability results on the subscale Medical Content, in contrast to better results on the subscale Basic Skills, support the view that there are marked differences in measurement properties between Basic Skills and Medical Content. In particular, doctors trained as raters should more easily rate Medical Content, whereas the complexity of Basic Skills should give more difficulty. More research is needed to explain this phenomenon. Reliability improves considerably if assessment is restricted to basic interviewing skills. In terms of social work practice this study suggests that audio-recordings are less suited to capturing and being used to assess the more nuanced skills and aspects of practice and should be restricted to assessing only ‘procedural competencies’ (Bogo, 2011).

### Implications for social work assessment

In light of the distinctive nature of social work, one specific challenge for using OSCEs is the difficulty of devising suitably representative scenarios that capture the complexity of social work practice, and providing enough OSCE ‘scenario stations’ in an exam for sufficient aspects of practice to be observed.

The proof of concept phase needs to test the number and type of OSCE scenarios needed for a credible assessment strategy, as well who should observe. Attention must also be paid to ensuring that the assessment strategy has minimal adverse effects on everyday practice delivery and service user experiences.

### Allied Health Professions (AHPs) – physiotherapists, occupational therapists, clinical psychologists

Direct observation has arisen in the AHPs in response to its use in the medical profession and also in response to a move towards patient centred-ness, now enshrined in practice with the introduction of the Care Act 2014 and increasingly driven in the policy field by the personalisation agenda. Such patient/service user centred approaches focus on the importance of professional-patient communication, alongside the attention paid to the technical dimensions of practice. Observation appears primarily in the context of supervision as well as training and evaluation and plays an educational (formative) as well as summative role; some studies drawing on educational psychology highlight this (Schuwirth et al 2003, Morris et al 2012). As an educational tool various
studies report its uses, in terms of feedback, and stress the importance of forms of measurement that take into account the amount of time changes in learning take to evolve into practice, suggesting longer periods post intervention as being beneficial. In a study researching physiotherapy students’ professional capabilities, Parry and Brown (2009) found that the study outcomes and assessments focused on students’ ‘ability to describe and reflect on practices, rather than on their practical communication proficiency’. Overall the evidence in this domain suggests that some form of assessment of actual conduct is far preferable to written assessments. The use of experiential learning is rated highly and tools such as role play, use of simulated and non-simulated patients, vignettes and video recording as feedback tools are suggested, as are the setting of personal goals for practitioners in light of this, albeit all of these approaches are being used in a formative capacity.

Milne and Reiser (2011) researching cognitive behavioural therapy supervision in clinical psychology settings, tested direct observation instruments against DIY criteria - design, implementation and yield - and found that ‘there appears to be no suitably rigorous and pragmatic instrument available with which to evaluate competent clinical supervision through direct observation (p.96).’ They proposed that simpler versions were needed and that ‘Future research should seek to develop a brief, supervisory competence rating scale that addresses fidelity issues with improved design, implementation and yield features (p.97)’.

Common themes referred to by Kaslow et al (2004) in relation to assessment of capability in clinical psychology settings included the importance and value of:

- modelling role plays and vignettes
- real-world settings as critical instructional strategies
- facilitation of learning over the professional life course and a learning environment that encourages development alongside assessment
- better methods of assessment and evaluation
- the use of both formative (i.e. corrective feedback for further development) and summative (i.e. conclusive evaluations for progression and gate-keeping purposes) assessment procedures
- evaluating skills and attitudes requires novel training and assessment methods, for example, standardised patients and 360-degree evaluations used by other professions should be considered.

Malik and Bhugra (2011) raise the issue of how many observations are required and the idea of observing continuity with one case for assessment and developmental purposes. They also consider other tools such as the Mini-CEX and the Mini-Assessed Clinical Encounter (ACE) comparing these methods with the long case, which is a more involved engagement with a patient and has wider applicability to psychiatric settings. The authors also cite the piloting of Direct Observation of Non-Clinical Skills (DONCS) as a useful development.
According to Parry (2008), writing in the context of physiotherapy, measuring actual behaviour through direct observations in real life situations is widely regarded as the 'gold standard' and adjusting behaviour to the uniqueness of every encounter is a vital communication skill that is difficult to measure. This is the essence of what social workers need to do and be assessed on too. Studies far more commonly measure and rate the presence, absence, or amount of behaviours, rather than their appropriateness. Empirical research suggests gradual change and evolution of practice can follow an observation and examining practice over several months post-observation seems advisable. Evidence for the effects and predictive capacity of training/observation comes largely from specific areas, and pose significant problems for application to AHP settings highlighting the difficulties of using observation measures as predictors of on-going/future capability. Direct observation research is required in order to identify the specific practices involved in AHP, to yield detailed practice-relevant descriptions of these and to examine their functions and effects. Further substantial, wide-ranging observational research is needed before the predictive capability of observation as an assessment method can be more confidently asserted.

Questions concerning the reliability and validity of observation measures, again, raise their head in this domain. Standardised rating scales are a source of concern for Gysels (2005) who recognises that however valuable the results emerging from these rating scales are for informing practice improvement, the communication categories measured, inevitably exclude behaviours or levels of communication not targeted by the criteria selected to be assessed. Only more subjective methods that offer a critical interpretation of the practice being observed based on contextual factors and a reflective account of why specific actions were taken or responses made, can contribute to finding out about the use of the skills in practice or the possible 'success' of the interaction.

**Implications for social work assessment**

The proof of concept phase could design and pilot a standardised observation template for assessing supervisory practice, being cautious of creating an overly-prescriptive tool.

Careful attention must be paid to the design of any simulated scenarios to try and capture the complexity and diversity of social work practice as fully as possible.

**Education**

Direct observation of classroom teaching has a long history in education. A cautionary approach that borrows observation methods wholesale from education should be taken, however. Firstly, current assessment approaches in education are primarily heavily driven by a ‘top down’ model to assessment, with evidence to suggest that the validity and reliability of different OFSTED inspectors is inconsistent. Secondly, the distinctive features of practice in educational settings (schools, classrooms) as opposed to social work settings (primarily family homes) impedes the
direct transferability of assessment approaches across the two professional domains. The four key issues emerging across the education literature discussed below are:

- mode of direct observation - in person, audio/video
- context - including whether the observation is done by peers, groups, supervisors etc
- how to use it to make it individually useful – formative vs summative requirements
- how it links to other broader purposes - e.g. professional competencies.

O’Leary (2012) reviewed the role of lesson observation across higher, further and school based education and provides a very detailed summary of research findings, models of observation and policy rationales for observation in each sector. Across each element of the education sector, observation is seen as a useful process to stimulate reflection on practice through discussion with colleagues. In schools and FE, observation has been more closely linked with performance management and controversies exist around the graded observation framework. Contestations exist around the stated aims behind its use, the extent to which the outcomes match these aims and the way in which the process of observation is operationalised. O’Leary makes the point that tensions around observations emerge in trying to reconcile developmental needs of staff and the requirements of performance management systems. This is similarly replicated in the evidence gathered from social work professionals involved in the focus group who expressed their concern about the extent to which there can be a developmental/judgmental balance to any observational processes that might be introduced.

A meta-analysis (Fukkink, Trienekens and Kramer, 2011) of studies using video feedback on the interaction skills of professionals in a range of contact professions including teachers, demonstrated the developmental value of video feedback which improves professionals’ receptive, informative and relational skills and, albeit less so, their verbal, non-verbal and para-lingual aspects of communication, whether they are at the beginning of training or have some years of work experience. More broadly Van Houten (1980) categorises attributes of feedback into three categories:

- nature of feedback content (positive, negative, specific, corrective etc)
- timing (frequency and delayed or immediate)
- who gives feedback (peers or supervisors).

Scheeler, Ruhl and McAfee (2004) reviewed empirical literature on the effects of feedback on both newly acquired and ingrained teaching behaviours, and they concluded that teachers as well as other learners acquire and maintain new behaviours when they receive feedback that is immediate, positive, corrective and specific.

Admiraal (2014) explored the use of web-based video in professional preparation programmes in two universities. He found that teacher feedback was rated as more important than peer feedback and that direct video observation was useful for analysing and understanding the performance
itself rather than deeper thoughts about linkages with theoretical notions, learning aims, or professional context. Another more recent development is the use of ‘bug in ear’ (BIE) technology which provides synchronous coaching/feedback to teachers while teaching (see for example, Coninx, Kreijns and Jochems, 2012 and Rock et al, 2009). The use of video technology to assess real practice in real time may be useful to consider in relation to social work, with all the caveats of data protection acknowledged, and possibly, because of its limited ability to fully capture real practice, more for formative than summative assessment purposes.

Currently available measures can give useful information, but their reliability and validity is limited, hence care must be taken not to over-interpret. A range of measures for assessing practice exist in the teaching domain but no one is deemed to be the ‘gold standard’, so triangulating the measures is the safest approach to adopt. The three with the strongest credentials are:

- classroom observations by peers, principals or external evaluators
- ‘value-added’ models (assessing gains in student achievement)
- student ratings.

Three other approaches had limited evidence:

- principal (or head teacher) judgement
- teacher self-reports
- analysis of classroom artefacts and teacher portfolios.

### Implications for social work assessment

The design of an observation strategy for social work must reconcile the competing need for it to be both developmental and judgemental in order for the chosen assessment strategy to be effective.

The proof of concept phase should trial a range of recording templates which incorporate both the developmental and judgemental aspects of assessment, in order to identify an approach that best achieves this balance. The proof of concept phase needs to attend to how, when and by whom feedback is given within the proposed assessment framework.

### Psychotherapy and Counselling

Within psychotherapy and counselling, direct observation tools and processes are normally situated within supervisory contexts, with direct observation taking the form of audio or video recordings of sessions. Although these recordings are then used in training and supervision, this is not in a consistent way nor do we know enough about how much it is used in training and supervision (Amerikaner & Rose, 2012). One recent study of 150 professional psychology supervisees at all levels of training (from beginning practicum through postdoctoral training)
demonstrated that methods permitting direct observation of supervisee work were used very infrequently (Amerikaner & Rose, 2012). Huhra et al. (2008) reviewed the literature on the use of videotaping in supervision within various mental health programs and noted that the bulk of research is focused on its advantages and disadvantages. The advantages of videotaping sessions include:

- means to bring out change in self-perception
- a tool to bring about self-analysis
- means for supervisors to more accurately evaluate trainees
- a way for both trainees and supervisor to re-experience therapy sessions.

Disadvantages centred on potential trainee anxiety and occasional negative reactions by clients. Overall its strengths are more aligned with formative compared with summative approaches to assessment.

Haggerty and Hilsenroth (2011) make a strong case for the use of video in psychotherapy supervision as it reduces the impact of imprecise memory recall in the discussion of therapy sessions. Diener et al. (2007) found that the use of audio or videotaping for supervision has a positive effect on outcome in short-term, affect-focused dynamic therapy. Others have found that the use of video in psychotherapy training for more psychodynamically-oriented psychotherapy is also advantageous (Binder, 1993, 1999; Hilsenroth et al., 2006; Levenson & Strupp, 1999). Aveline (1992) also advocates supervisors showing supervisees videos of their own client work. Within these articles information about how to use video material within supervision is provided – e.g. the Integrated Developmental Model of Supervision that takes into account where a supervisee is in relation to Self/Other awareness, Motivation and Autonomy as well as areas of therapist activity (e.g. treatment goals and plans, ethics, skills competence etc.) (Huhra et al., 2008). Not without strengths, the use of video and audio recorded practice remains more suited to formative than summative assessment purposes.

Interestingly, recent Competence Frameworks for the Delivery and Supervision of Psychological Therapies using cognitive and behavioural therapies, psychoanalytic/psychodynamic therapy, humanistic therapies and systemic therapies, developed at University College London, make mention of the ‘ability to incorporate direct observation into supervision’ as a specific supervisor competency and this includes being able to establish a context for taping, using tapes and an ability to use in-session direct observation.
Implications for social work assessment

There is potential for direct observation to be used both to assess the capability of Practice Supervisors per se, and to assess the ability of the Practice Supervisor to undertake direct observations of those they are responsible for supervising. Being able to ‘do’ direct observation, therefore, is an important skill that needs to be included in the Knowledge and Skills Statement for Practice Supervisors.

The proof of concept phase can be used to develop recording and reflection templates to assess specific skills of Practice Supervisors.

Law

Although it is difficult to find a translation of the practice of direct observation into law settings in the way that it is understood in medical and health professions, it does seem to appear in various guises. It is used as an educational and reflexive tool in the student-supervisorial relationship (Stevens 2006, Fletcher & Weinstein 2002) with the use of video technology, role play and the vignette providing useful forms of feedback (Kovach 1996) and in this way is hypothesised to lead to better outcomes in the practice of law. It is also used by researchers to shed light on the ways that professionals in legal settings make decisions and influence outcomes through certain communication practices. In a paper by Wasoff (1992) four simulated cases were constructed which varied by age, gender, number and ages of children, social class and income, all typical of individuals who divorce. Fifty eight solicitors were interviewed for the study. The authors cite various benefits of this method including:

- access to more ‘elite’ professional groups
- opportunity to study the treatment of specific issues and how professional responses and behaviours vary across professions
- a more controlled (albeit with limitations too in terms of what is not included) approach to practice content than would be possible in a truly natural setting using more conventional ethnographic methods.

Management

Literature relating to using observation with frontline operational and senior managers is very limited and tends to focus on tools such as the 360 degree feedback models in which employees receive confidential, anonymous feedback from the people who work around them. This typically includes the employee's manager, peers and direct reports but is usually framed as a developmental, as opposed to a ‘judgemental’ (summative), exercise. In this professional arena, recruiting suitably qualified assessors is particularly problematic on two fronts: firstly, finding assessors with the expertise to assess managers, particularly senior management personnel, and secondly, the cost implications if independent consultants are required as assessors. One area that
might yield learning in terms of the observation of colleagues in leadership positions is that of the safeguarding peer review process. This process, part of the Local Government Association’s sector-led improvement programme, involves training professionals to undertake peer review of safeguarding and so includes some observation of leadership roles. The programme has been evaluated\(^2\), and offers some positive indications for peer-led direct observation. However, all the caveats of direct observation regarding the need to caution against overly positive, inauthentic feedback that fails to be sufficiently critically constructive will likely still apply.

Direct observation comes up in the management literature but not in relation to individual competencies. Whilst considered to be a valid measure it was not used in a particularly scientific or systematic way e.g. observing steps in a manufacturing process or boardroom conversations, and is treated more as a research method or as one of many methods used to assess how an organisation is working or could work better.

**Implications for social work assessment**

The proof of concept phase needs to explore the strengths and weaknesses of using different sources / sorts of assessor for the assessment of Practice Supervisors and particularly Practice Leaders.

\(^2\) [http://www.local.gov.uk/c/document_library/get_file?uuid=091b5086-d070-4e5e-82e1-2c685e42b4f&groupId=10180](http://www.local.gov.uk/c/document_library/get_file?uuid=091b5086-d070-4e5e-82e1-2c685e42b4f&groupId=10180)
6. How might a method of assessment that incorporates observation most effectively be used alongside a digital test?

Key messages from the literature

- Research suggests that the assessment of professional capability requires multi-faceted assessment strategies that triangulate different sources of assessment data.
- Social work is a complex and challenging profession that involves a depth and breadth of knowledge and skills that need to be assessed using more than one method for accuracy and credibility.
- Social workers need to be differentially assessed on what they know in terms of knowledge, what they can show in terms of applying their knowledge and what they actually do in practice.
- Digital approaches and OSCEs are more suited to assessing professional knowledge.
- Direct observation is the optimal method for assessing professional skills because it is closest to ‘real practice’.
- Given the digital solution involves practitioners engaging with simulated materials it seems appropriate that the observation component of the assessment strategy provides opportunities to assess practitioners in real life practice and not by other simulated (albeit not digital) approaches e.g. OSCEs.
- Pilot permutations of various configurations of digital and observational assessment methods and pathways should be included as part of the proof of concept stage of the overall project.

A consistent message across the professional domains is that the effective assessment of professional capability by direct observation requires a multi-faceted assessment strategy (Humphrey, 2007). Utilising Miller’s (1990) stratification which differentiates between levels of clinical competence and the appropriate way of assessing them, it is possible to suggest the following assessment strategy for social work practice:

**Pyramid of professional competencies**
(adapted from Miller 1990)

- **Doing and being**
- **Showing**
- **Knowing how**
- **Knowing**

The digital solution includes a mixture of knowledge bank and scenarios which will provide coverage of “knowing” “knowing how” and “showing” Observation allows for demonstration of “doing and being”
On this basis of categorising professional competence, observation of different types could be used to complement assessment undertaken online. For example Baez (2005) relates this typology to the use of the OSCE as follows:

‘Introducing the OSCE had vitality and held interest precisely because it gave students and faculty a context that focused on the student showing how, instead of knowing or knowing how, which is the usual context of relatedness between students and teachers in social work education. Furthermore, while it fell short of the complexity and attributes of does in an actual social work practice situation, it had more realism and more skill-building specificity than any classroom role-play (p.17) ’

If this categorisation is applied to direct observation in the context of social worker assessment then practitioners would be demonstrating their capacity to ‘do’ social work i.e. not just what they know but their skills in implementing their knowledge in practice. This would include their capacity to use their personal and professional qualities i.e. their way of being as a social worker and the centrality of the professional relationship in this process (Ruch, 2010). Both the capacity ‘to do’ social work and ‘to be’ a social worker are critical requirements for evidencing proficiency in relation to the Knowledge and Skills Statement for Child and Family Social Work. Digital approaches are more suited to assessing professional knowledge. And given the digital solution involves practitioners engaging with simulated materials it seems appropriate that the observation component of the assessment strategy provides opportunities to assess practitioners in real life practice and not by other, albeit not digital, simulated approaches e.g. OSCEs. Direct observation, therefore, would appear to be the optimal method for assessing professional skills as it is the assessment method that is closest to practice.
Numerous permutations of using different methods alongside the digital solution exist (see table below) but there is no evidence to say which is best, other than that a combination of methods will increase both the validity and reliability of all the different methods employed. The more methods used, however, will have significant implications for the overall costs incurred and logistical complexities of implementing and delivering the assessment strategy. An important consideration in determining which methods to choose concerns how different methods may be more or less suited to a particular aspect of practice. For the purposes of the scoping review and in order to illustrate the full range of methods available for consideration, the option of video and audio recording practice has been included. In our opinion, however, their feasibility as a summative assessment method is limited for a number of reasons which make it an unlikely observation method of choice. Firstly, the evidence for its use as a summative observation method for assessing professional capability is scant. Secondly, compared to other assessment options it is not an appropriate method for assessing certain significant aspects of practice. For example, video recording work done by intake teams that involves first time practice encounters with individuals or families in crisis would be deemed to be inappropriate. Thirdly, video and audio recordings raise serious ethical considerations, including issues relating to data protection and the potential for recordings to be used for other purposes, such as in court proceedings. Conversely, it may be more suited to assessment of Practice Supervisors in supervision settings in particular.

In light of the assessment methods reviewed, and the clear message from all the professional domains that a combination of methods provides the most credible assessment strategy, the tables below illustrate some possible permutations of implementing a combined digital and observational assessment strategy.
Table illustrating possible configurations for using different observation methods

<table>
<thead>
<tr>
<th>Digital Assessment</th>
<th>Direct Observation</th>
<th>OSCE</th>
<th>Video/audio recording</th>
</tr>
</thead>
<tbody>
<tr>
<td>All practitioners</td>
<td>All</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>All</td>
<td>All</td>
<td>All</td>
<td>None</td>
</tr>
<tr>
<td>All</td>
<td>All</td>
<td>None</td>
<td>All</td>
</tr>
<tr>
<td>All</td>
<td>None</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>All</td>
<td>Some practitioners</td>
<td>Some</td>
<td>Some</td>
</tr>
<tr>
<td>All</td>
<td>Some</td>
<td>Some</td>
<td>None</td>
</tr>
<tr>
<td>All</td>
<td>Some</td>
<td>None</td>
<td>Some</td>
</tr>
<tr>
<td>All</td>
<td>None</td>
<td>Some</td>
<td>Some</td>
</tr>
</tbody>
</table>

Illustrative examples

Example threshold pathways

![Example threshold pathways](image)

Everyone completes all modes of assessment

![Everyone completes all modes of assessment](image)
Following any decision regarding the optimum combination of digital and observational assessment methods, further decisions will need to be made to determine how the assessment strategy is then used for the purpose of summative assessments i.e. what are the thresholds for an individual to pass/fail the digital test or an observation and how many times can an individual be re-tested or repeat observed? This raises important questions regarding the number of opportunities at each stage of the assessment process any one individual would have to demonstrate their capability.

How the decision regarding what the assessment pathway looks like is made will depend on where the thresholds of capability are drawn. At the same time the wider implications of these decisions will need to be taken into consideration. Too high a threshold risks seriously exacerbating the existing workforce retention issues, whilst if the threshold is set too low there is the risk of a sustained negative public perception of the profession and arguably a risk of poor practice being perpetuated. Furthermore, the decision needs to recognise which procedures need to be adhered to in order to meet employers’ HR requirements, in particular to minimise the risk of appeals in the face of a fail outcome.

As noted, it is important to be clear what is being tested by each mode. Within the digital test there is an opportunity to test a wider range of knowledge than in one direct observation, but in direct observation you can see the behavioural aspects which are not fully covered in the digital test. Beyond this consideration, with regards to the how the pathway decision is configured, a key variable is the extent to which the digital test is/is not a good predictor of performance compared with an assessment by a direct observation or an OSCE. If the predictability/comparability of approaches is high this would prove the validity of the digital ‘test’ as a measure of professional capability and provide justification for not requiring everyone to be assessed via an observational method. Whilst emergent research evidence (Bogo, 2013; Forrester, personal communication, 2015) suggests that simulated assessments of practice *appear to compare favourably* with assessment of real practice, as yet this has not been conclusively confirmed. Equally if, as suggested, the digital solution is unable to cover all aspects of the Knowledge and Skills Statements, the pathway proposal needs to decide if, regardless of the digital test score, all students/practitioners should be subjected to an observation, or if some very high scoring candidates (e.g. 90%+) would be exempted.

**Implications for social work assessment**

The proof of concept phase needs to consider how direct observation, as the optimal method for assessing professional capability given it is the assessment method that is closest to practice, can most effectively build on existing systems and expertise to maximise the accuracy of its assessment capacity and its cost effectiveness.
7. Which method(s) of observation may be most relevant for each of the three levels of practice?

Key messages from the literature

- Most of the existing measures of professional capability are focused on social work students and frontline practitioners.
- Little evidence of use of observation with the Practice Supervisor level and non-existent evidence base for use with the Practice Leader level.
- The Knowledge and Skills Statements do not contain any information to suggest that one specific observation method is more appropriate for a particular career level.
- The implications regarding the quality of practice assessor are equally applicable across all three career levels, but are more complex at the Practice Leader level.
- Careful consideration must be given to the distinctive training/role/assessing of assessor requirements at all three career levels.
- To maximise the credibility and minimise the disruption of the digital and observational assessment methods in practice, they need to be embedded in everyday practice across the organisational hierarchy.
- The proof of concept phase of the project should test which methods would best suit each of the three career levels.

In producing this scope a short mapping exercise was undertaken, mapping the three Knowledge and Skills Statements against the different observation assessment. This exercise has not been included as the Knowledge and Skills Statements for Practice Supervisors and Practice Leaders are being consulted upon at the time of writing. Beyond this exercise, undertaken for the specific purposes of this scoping brief, there is no evidence to inform which observation methods are most appropriate to use for what purpose or with whom. Most of the existing measures of professional capability are focused on social work students and frontline practitioners. There is little evidence of use of observation with the Practice Supervisor level and non-existent evidence base for use with the Practice Leader level. Nor do the Knowledge and Skills Statements contain any information to suggest that one specific observation method is more appropriate for a particular career level or that any of three observation methods are entirely unsuitable for any of the three career levels (although as stated earlier evidence suggests that the use video/audio recording has limitations warranting its inclusion in a multi-faceted assessment strategy somewhat questionable – though it may be a useful method in other contexts). Whilst a relatively crude measure, the exercise to map the different assessment methods onto the Knowledge and Skills Statements for each career level highlighted how at each level the observation methods were deemed to be more pertinent for assessing professional skills, whilst the digital solution more effectively captured professional knowledge and its application in practice – as discussed previously.
Assessment, organisational culture and everyday practice

Two important considerations for developing the assessment strategies across the professional hierarchy are firstly, the recognition that the accuracy and credibility of the summative assessment strategy will be enhanced if observation is an integral part of the organisational culture. This would imply that observation should be undertaken informally and be an aspect of everyday practice, not just at the practitioner level but also amongst supervisors/managers and senior management. Secondly, in parallel, consideration should be given to embedding equivalences of the digital solution as an aspect of everyday practice, for example, small scale online tests could be introduced across the organisational culture as an on-going means of formative staff development. In so doing, digital tests would become a familiar aspect of the workplace and of an individual’s professional repertoire and some of the limitations of digital assessment methods, for example, test-related anxiety, could be ameliorated.

The role of the practice educator and assessor

From undertaking the broader scoping review it is clear that all the issues concerning the appointment, quality and support of practice educators and assessors for frontline practitioners are replicated across the three career levels. In light of this, careful consideration must be given to the distinctive training/role/assessing of assessor requirements at all of the levels. Indeed as the seniority of the staff being assessed increases, the assessment issues prove to be more problematic as the pool of potential assessors shrinks. Some local authorities have been trialling cross-authority peer review of senior colleagues, for example as part of the Local Government Association-led peer review process of Local Safeguarding Children's Boards 3, which may offer some useful learning regarding peers-assessment in the absence of substantiated research evidence.

3 www.local.gov.uk/documents/10180/6353023/CSPR+What%27s+it+all+about.pdf/aaf230f0-1650-42ba-a911-2ec97a806983
8. How much observation may be required for each level in order that it will be (and will be seen as) a credible and reliable method of assessment?

Key messages from the literature

- Research indicates that in formative contexts with frontline practitioners more observations increase their reliability and credibility.
- No evidence exists to indicate how frequently observations should be taken when used as summative assessment artefact.
- No evidence exists to indicate how frequently observations should be taken with Practice Supervisors or Practice Leaders.
- Number and frequency of observations depends on the summative-formative balance of the observation task.
- There is a need to test as part of the proof of concept phase how many observations are required.
- Consideration to be given to streamlining the assessment and accreditation arrangements, including their frequency, with the existing HCPC re-registration procedures and timeframes.

Number of observations

Research evidence indicates that in formative assessment contexts the more observations that are undertaken at the frontline practice level the more credible are the results (Madan et al, 2012; Singh et al, 2012). No evidence was found, however, to indicate how frequently observations should be taken when used as a summative assessment artefact or when used with Practice Supervisors or Practice Leaders. There is a need, therefore, to test how many occurrences of each form of observation would be required to determine, if for example, a practitioner’s, supervisor’s or leader’s ‘rating’ changed if s/he was observed more than once using one/all of the methods.

Some further information may be gleaned from those local authorities who are using practice observation (see Appendix 3). A call for practice examples has been issued, and responses are starting to come in. Initial feedback is that there is very little data available on cost or using observations with leaders.

Impact of observations on organisational practice

Benefits of more observations for the assessment of professional capability are arguably offset, however, by the time implications associated with them and the subsequent impact on delivery of practice (and therefore potentially service users’ wellbeing), as well as the costs to the organisation of resourcing them. Several studies report the time consuming and resource intensive nature of the direct observation (Humphrey, 2007; Kogan et al, 2012; Milne and Reiser, 2011). If
it is to be carried out by experienced and more senior practitioners, as studies claim is most effective, it means significant time and resources need to be allocated, which some studies suggest may adversely impact on practice delivery (Craig, 2011). Important to note is that, regardless of whether assessors are internal to organisations, part of cross-organisation peer assessment arrangements or totally independent of the organisation, there will be significant cost implications and implications for practice. The development of supposedly effective tools i.e. standardised assessment and reflective templates, with which to capture the nuances of the direct observation encounter are an attempt to limit and make cost effective use of practitioner time and resources. Incorporating a summative direct observation component within existing formative assessment frameworks is suggested by some to be a particularly cost effective intervention, particularly if practitioner skills are improved as a result (Jelovsek, 2013). Crucial too is the need to acknowledge that it takes time for feedback to be incorporated into practice. Studies that fail to take this into account, by carrying out further assessments too soon after an observation assessment and feedback, fail to give an accurate picture of the changes made to practice.

The table below illustrates the advantages and disadvantages for the assessment process associated with practice assessors who come from different positions

<table>
<thead>
<tr>
<th>Table illustrating the advantages and disadvantages of the different statuses of practice assessors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In house assessors</strong></td>
</tr>
<tr>
<td><strong>Advantages</strong></td>
</tr>
<tr>
<td>Familiar with organisation structures and systems</td>
</tr>
<tr>
<td>Cost effective</td>
</tr>
<tr>
<td>Builds skills, developing capacity within the organisation</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
9. For each of the preferred methods of observation what would be the issues to consider for the full implementation of this approach across the total social worker population in England? This should consider scalability, feasibility and cost effectiveness (including, at this stage, a very indicative cost for the set-up and running costs of each method).

<table>
<thead>
<tr>
<th>Key messages from the literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Any proposals must ensure that appropriate quality assurance mechanisms are in place to safeguard the practice educator/assessor role and performance.</td>
</tr>
<tr>
<td>• Important for cost effectiveness purposes and workforce demands to streamline proposals with existing processes e.g. the HCPC requirements/timeframes for re-registration.</td>
</tr>
</tbody>
</table>

A separate exercise has been undertaken by the KPMG consortium to provide estimated costs for each of the observation modes.

The extent to which this review can make statements regarding the total costs involved in the design and delivery of an observation assessment strategy is constrained by key decisions that need to be made, for example which method(s) is/are selected and where thresholds of capability are set. That said to help inform the decision-making process the key constituents of each of the observation approaches are tabled below.
### Table of constituents of the three observation modes

<table>
<thead>
<tr>
<th></th>
<th>Direct observation</th>
<th>OSCE</th>
<th>Video/Audio recording</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Admin Costs</strong></td>
<td>• Programme administration and logistics</td>
<td>• Programme administration and logistics</td>
<td>• Programme administration and logistics</td>
</tr>
<tr>
<td></td>
<td>• Selection of expert observers</td>
<td>• Selection of expert observers</td>
<td>• Selection of expert observers</td>
</tr>
<tr>
<td></td>
<td>• Selection of scenario to observe (if applicable)</td>
<td>• Venue and actor sourcing</td>
<td>• Selection of scenario to observe (if applicable)</td>
</tr>
<tr>
<td></td>
<td>• Pre-approval for direct observation</td>
<td>• Development of scenarios</td>
<td>• Pre-approval for filming/recording</td>
</tr>
<tr>
<td><strong>Staff Costs</strong></td>
<td>• Training of expert observers</td>
<td>• Training of expert observers</td>
<td>• Training of expert observers</td>
</tr>
<tr>
<td></td>
<td>• Time spent observing the event</td>
<td>• Time spent by expert observers at the event including observation, reflective conversation and assessment update</td>
<td>• Time spent observing the event</td>
</tr>
<tr>
<td></td>
<td>• Time spent recording outcome, reflective conversation and assessment update</td>
<td>• Time spent recording outcome, reflective conversation and assessment update</td>
<td>• Time spent recording outcome, reflective conversation and assessment update</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued overleaf
<table>
<thead>
<tr>
<th>Development Costs</th>
<th>Direct observation</th>
<th>OSCE</th>
<th>Video/Audio recording</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution Design</td>
<td>How to perform the observation</td>
<td>How to perform the observation</td>
<td>How to perform the observation</td>
</tr>
<tr>
<td></td>
<td>Against what criteria will the SW be assessed</td>
<td>Against what criteria will the SW be assessed</td>
<td>Against what criteria will the SW be assessed</td>
</tr>
<tr>
<td></td>
<td>Who will perform observation</td>
<td>Who will perform observation</td>
<td>Who will perform observation</td>
</tr>
<tr>
<td></td>
<td>Standardised observation tools</td>
<td>Standardised observation tools</td>
<td>Standardised observation tools</td>
</tr>
<tr>
<td></td>
<td>Development of training package for expert observers</td>
<td>Development of training package for expert observers</td>
<td>Development of training package for expert observers</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment/External costs</td>
<td>N/A</td>
<td>Actors</td>
<td>Video/audio equipment hire</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Venue</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Coordination/management of large scale event</td>
<td></td>
</tr>
</tbody>
</table>
A scoping exercise to outline the practical implications of the different assessment artefacts has involved grading the challenges associated with the scaling up and feasibility of the three observation assessment processes across the three career levels, tabled below.

**Key**

<table>
<thead>
<tr>
<th>Ratings</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credibility</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Scalability</td>
<td>Impossible</td>
<td>Challenging</td>
<td>Possible</td>
</tr>
<tr>
<td>Feasibility</td>
<td>Impossible</td>
<td>Challenging</td>
<td>Possible</td>
</tr>
<tr>
<td>Cost</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
</tbody>
</table>

**Approved Children and Families Practitioner (ACFP)**

<table>
<thead>
<tr>
<th>Observation Method</th>
<th>Credibility (reliability+validity)</th>
<th>Scalability</th>
<th>Feasibility</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct observation</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>OSCE adaptation</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Video/audio recording</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

As an established approach to assessing practice, direct observation is recognised as a credible method for assessing professional capability. There is existing infrastructure and resources that would facilitate the cost effective scaling up of direct observation as a summative observation method and help support feasibility on a large scale – though it is not a cheap option by any means. In contrast, the creation of a nationwide structure for introducing OSCE assessments would require substantial new investment in developing a resource intensive assessment programme e.g. employing actors on an on-going basis, hiring large scale venues, administering the process. Video and audio recording of practice, whilst likely to be the easiest and cheapest option to scale up, is the least credible method and has significant complicating factors e.g. data protection etc, that make it a less feasible option.
### Practice Supervisor (frontline and senior managers/supervisors)

<table>
<thead>
<tr>
<th>Observation Method</th>
<th>Credibility (reliability+ validity)</th>
<th>Scalability</th>
<th>Feasibility</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct observation</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>OSCE adaptation</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Video/audio recording</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

At the Practice Supervisor level the credibility of the three methods remains largely unchanged from that at the ACFP level, with direct observation arguably being the most credible in light of its established method within social work and its proximity to ‘real practice’. An additional challenge which slightly reduces the feasibility of direct observation at this level concerns the identification of suitable assessors. In contrast, the scalability and feasibility of using OSCEs increases slightly as the scale of their implementation would be less than for the ACFP level (though this depends on exactly who is identified as being a Practice Leader) and therefore the resource implications should be slightly less. The considerations, and therefore the ratings, for video and audio recording are the same as at the ACFP level.

### Practice Leader (Assistant Directors Children’s Services)

<table>
<thead>
<tr>
<th>Observation Method</th>
<th>Credibility (reliability+ validity)</th>
<th>Scalability</th>
<th>Feasibility</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct observation</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2/3</td>
</tr>
<tr>
<td>OSCE adaptation</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Video/audio recording</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

At the Practice Leader level direct observations can be scaled up relatively straightforwardly due to the smaller number of individuals included in this career level. The biggest challenge for direct observations is the identification of an appropriate pool of assessors and the potential higher costs this could incur. OSCEs are a reasonably scalable and feasible assessment method for this smaller group of professionals but remain slightly less credible as an observation method when compared with direct observation, due to the simulated nature of the scenario. The ratings for video and audio recording are the same as for the ACFP and Practice Supervisor levels; depending on the nature of the practice being recorded, it may be possible to avoid many of the data protection identified for other levels.
10. Summary and Key Considerations

Key messages from the literature

- Assessment and accreditation systems of professional capability in social work cannot be scientifically evidence-based but can be evidence-informed.
- Validity vs reliability: a digital solution can offer a higher degree of validity and practice observation can offer a higher level of reliability; incorporating the two assessment methods into an assessment and accreditation system will enhance its credibility.
- Assessment and accreditation systems involving practice observation must be embedded in wider organisational cultures of informal observation in everyday practices across the professional hierarchy.
- Importance of attending to the summative and formative dimensions in the design of the assessment and accreditation system to maximise its positive impact on practice quality standards and workforce morale.
- Importance of attending to the summative and formative dimensions in the design of the assessment and accreditation system to minimise its negative impact on the morale of the social work workforce.
- Quality of assessors is a key issue for all three methods and career levels.
- Careful consideration needs to be paid to the frequency and number of observations.
- Determining the capability thresholds for observation and the digital test has serious implications for how the assessment strategy works in practice and for the stability as well as the quality of the workforce.
- Reflective components and manner in which they are conducted are critical constituents of effective observation process.

Any model developed for assessing capability in social work practice needs to be informed by existing research and knowledge from across the professional domains reviewed. At the same time it is imperative to acknowledge that the existing evidence base for observation as a mechanism for assessing professional capability is underdeveloped and further research and scrutiny is required. The primary issues this scoping review has unearthed are:

- the complexity, acknowledged across all the domains reviewed, of assessing capability in professional practice
- the social work specific characteristics that need to be recognised when determining the appropriate observation methods and their specific design
- the need to design a multi-faceted model of assessment to capture the breadth of knowledge and diversity of skills that constitute capable practice
- who assesses practice is as important as how it is assessed; practice educator/assessor competence is crucial for reliable and credible assessments
• the importance of establishing a clear assessment framework and standardised understanding (as far as is possible) of what is being assessed, whilst being alert to the danger of being over-prescriptive
• the importance of balancing the assessment (summative) and developmental (formative) dimensions of the chosen assessment approach to maximise task engagement, accurate outcomes and professional wellbeing, which, in turn, impact on sustaining and improving (or not) practice quality, the ultimate objective of the exercise
• the potential for direct observation, with its roots in real practice, to be the observation method that most effectively complements the simulated nature of the digital solution.

Alongside designing a summative assessment of professional capability there is a strong case to be made for this to be complemented by formative assessment mechanisms embedded in everyday practices, with local capability to resource them. This would involve, for example, embedding regular informal observation opportunities into everyday practice – some examples of which can be seen in Appendix 3. Direct observation processes in student placements specifically and co-working opportunities (when two practitioners work with families in pairs) for all practitioners, more generally, are both established practices that are recognised as having sound but underdeveloped bases for performance evaluation and feedback. Such practices are an important part of developing an organisational observational culture. The combination of a formative and summative assessment strategy will not only influence the credibility of any summative assessment strategy that is introduced but as a consequence could also positively influence the cost effectiveness of the assessment strategy by improving the capability of the workforce and, as a consequence, increasing the number of successful assessments and reducing the number of re-assessments that are required.

The same principle of developing an organisational culture that supports particular modes of assessment might be applied too to the implementation of the digital solution as an assessment method. By introducing digital tests as part of an on-going approach to everyday practice, workforce familiarity with such approaches will increase, the adverse impact on staff of digital testing for summative purposes will decrease and the overall efficacy of this assessment approach will be enhanced. Developing ‘everyday digital tests’ would at the outset involve financial investment and have some impact on staff time in practice, but proponents would argue the longer term benefits would outweigh these initial costs.

In order to inform the design and delivery of a reliable and valid observation strategy for assessing professional capability in social work practice conjunction with a digital solution, this review proposes that the proof of concept phase pays particular attention to and, where possible, pilots in practice the following aspects of the assessment process:
The extent to which two specific observation methods – direct observation and simulated observation - are suited to the Knowledge and Skills Statements of the three career levels.

Different permutations of digital and observational assessment methods and pathways to maximise the overall reliability and validity of a strategy for assessing professional capability.

The creation of standardised assessment and reflective templates for all three career levels which are closely cross-referenced to the three Knowledge and Skills Statements

The number of observations required at each of the three career levels, to ensure the assessments are reliable and valid.

The role of the practice assessor across all three career levels and, in particular, the issue of who takes up this role at the Practice Supervisor and Practice Leader levels

The status of the emergent evidence relating to the comparability of simulated and real practice observation for the assessment of professional capability.

In closing, as is often the case when confronted by gaps in the evidence, it is tempting to call for randomised control trials (RCTs) to be funded. In this instance, we should first think carefully about what the question is that needs answering; though RCTs are widely understood to be the ‘gold standard’ in terms of asking whether something works, here it may be more useful to ask why something works. In order to develop the evidence base underpinning assessment strategies, consideration needs to be given to conducting larger scale qualitative research studies that capture, report and explain how different assessment methods operate and most particularly how those engaged in the process experience it and apply this experience to practice.

The final, currently under recognised, dimension of the assessment process is the contribution that can be made by service users. Service users are an integral component of formative assessment strategies, providing for example feedback to students/practitioners and practice educators/assessors on their experience and views of an observation of practice in which they were the service user being engaged with. To date, however, service user feedback has not been incorporated into summative strategies. Equally the equivalence of service users for the Practice Supervisors (that is, practitioners) and Practice Leaders (frontline/middle operational managers) also need to be included as a key constituent of credible summative assessment strategies. With other professional fields placing increasing emphasis on the ‘expertise by experience’ held by end users (for example, the patient voice being strengthened in NHS reforms) it is important to consider how assessment strategies within social work might meaningfully engage both children and their families in this process.