

**Title approved for a full paper;  
USING PORTABLE TECHNOLOGY FOR ASSESSMENT OF PRACTICAL  
PERFORMANCE**

**Please change it to:  
THE USABILITY OF PDAs FOR ASSESSMENT OF PRACTICAL  
PERFORMANCE**

## **Abstract**

### **Overview**

The administration of an OSCE (Objective Structures Clinical Examination) using paper checklists presents problems such as illegible handwriting, missing student names and/or numbers and worst of all, lost checklists. Calculating and entering results is not only time consuming but also subject to human errors while feedback to students is rarely available. To rectify these problems PDAs (Personal Digital Assistants) and software (HaPerT) were acquired to replace paper checklists and provide automated results and feedback.

### **Aim**

To explore the usability of PDAs for assessment of performance.

### **Methods**

The usability of the PDA system was evaluated according to effectiveness, efficiency and user satisfaction.

- Effectiveness: Results of an OSCE, which is part of the MBChB and BChD II curriculum, conducted in 2003 (paper based method) was compared to the OSCEs conducted in 2004 and 2005 (PDA assessment).
- Efficiency: OSCE logistics of PDA assessments were compared to the paper based methods.
- User satisfaction: Questionnaires were used to obtain feedback on the assessors' experiences.

### **Results**

PDA assessment was found to be effective on grounds of the small difference in the mean scores of the three cohorts of students.

The efficiency measured in terms of time was superior in the PDA assessed OSCEs. Prior computer skills varied from none to very skilled and the training took 20 minutes on average. The checklists were user friendly and the navigation was easier after having assessed 3 candidates. The ease of selecting students and the swift completion of checklists were highly rated. The paperless method was less intrusive for students.

### **Keywords**

*Electronic checklists, PDA assessment*

## **1. Overview**

At the Skills Laboratory of the Faculty of Health Science, University of Pretoria, assessments of student practical performance are done at an average of 100 OSCEs (Objective Structured Clinical Examination) per year. At present approximately 2660 students, both under and postgraduates and representing all the disciplines of the Faculty, are enrolled for courses or practicals in this facility.

The administration of an OSCE using paper checklists for assessments often presents problems such as illegible handwriting, missing student names and/or numbers and worst of all, lost checklists. The calculation of results and entering them manually into a database are not only time consuming activities but are also subject to human errors. Feedback to

students, which should prove an ideal learning opportunity, is rarely given due to the paperwork and workload implied.

In an attempt to rectify these problems experienced at our faculty, reduce the workload and enhance students' learning, portable technology was explored as medium to assess performance electronically. Twenty-five PDAs and a software program, HaPerT, were acquired to replace paper checklists and provide automated results and feedback (Schmidts 2002). In co-operation with the author of the HaPerT software, Dr M Schmidts of the Vienna Medical School, the program was installed and adjusted to meet faculty's needs of performance assessment.

Existing paper checklists were adapted as follows for use on the PDAs:

- Since the PDA's screen is small the following changes were made to create a "branched" design (Schmidts 2002):
  - o Each checklist was arranged according to *specific outcomes* or steps of a procedure, e.g. preparation for a venepuncture. This ensures that each outcome appears on a separate screen, avoiding scrolling up and down a long checklist.
  - o *Assessment criteria* were then arranged under the respective outcomes. A maximum of 11 criteria can be viewed on the screen without scrolling.
- In order to give feedback to students on their mistakes or omissions the assessment criteria were changed from to positive statements to *negative statements* e.g. "Selected a suitable vein" changed to "Didn't select a suitable vein". The items selected on the touch screen will appear on a student's report if selected by an assessor as the mistakes/omissions made.

Optional attributes of the electronic checklist include the following:

- Individually *weighting* of assessment criteria if some are considered to be more important than others
- Identification of *critical error(s)* that, if one is selected, causes the student to fail.
- A range of overall *comments*, including positive ones, can be set up to select for feedback without affecting the score.
- "*Bonding*" of items, which means that only one of two or more bonded items can be selected, e.g. "Tourniquet was too loose" or "Tourniquet was too tight". Such a statement gives a student more specific feedback if (s)he hasn't applied the tourniquet correctly rather than merely stating "Tourniquet applied incorrectly".

The HaPerT system provides for negative scoring which means that only those items not done or done incorrectly are selected on the touch screen. The scores of those items are deducted from a full score for the checklist and the percentage and grading automatically calculated. Assessors can enter free text annotations for any item which will then form part of the report. The performances during all the stations (different skills) in an OSCE are collated and appear as one report per student with or without a final mark and grading.

OSCE assessors were trained to use the PDAs. The setting up of an OSCE on PC was managed by the same person and involves designating checklists to specific PDAs, listing the tutors/assessors, importing a list of students from an excel file and the synchronization of charged PDAs.

The input of the assessor entails the initial selection of his/her name followed by the selection of the candidate's name and his/her mistakes and or omissions. Additional feedback per item can be made by using free text.

On completion of the assessment(s) and synchronizing of data back to the PC, the following reports are automatically and immediately available:

- Feedback (pdf files) to the student.
- Various filtered reports (pdf files) on the results: variance of students' performances, assessors' performance, performance of each item on a checklist
- Results in excel files can be exported as average scores per student of all the stations (skills) or individual station (skills) scores per student.

The usability of this system however needed exploration to ensure that it is in fact effective and efficient and that the users were satisfied with this kind of electronic assessment.

## **2. Literature review**

The proliferation of computers in education has over the last two decades stimulated the development of new tools that assist instructors in teaching, evaluating and directing student learning. (Segal, Doolen & Porter 2004). Electronic assessment can according to Ridgway and McCusker (2005) be justified in a number of ways. It can inter alia help improve the technical quality of tests by improving the reliability of scoring and users can benefit from immediate and perhaps diagnostic feedback.

Due to their convenient size and reasonable computing power PDAs (Personal Digital Assistants), also known as handheld computers or palmtops, have emerged as a platform for computer-based testing and have lately been incorporated by schools and universities in their curricula. (Segal et al 2004).

Computer-based testing offers obvious benefits with respect to test administration such as improved security and accurate immediate scoring. The primary concern is however whether performance as measured by test scores is equivalent to that of paper-and-pen scores. Perkins (1995) Chin and Donn (1991) and Dimock (1991) The compared mean achievement scores between test versions to test whether computer-based tests are equivalent to paper-and-pencil tests. They reported scores obtained through computer-based tests to be equivalent, significantly higher and significantly lower respectively than paper-and pencil tests. It is therefore not certain if there is a relationship between test administration methods and performance. Segal et al (2004) compared the usability of a PDA quiz with the usability of a standard paper-and-pencil quiz and found no difference in the effectiveness (scores) and user satisfaction. The PDA quiz was more efficient since it took less time to complete. Harward, Wilson & Davis (2003) concurs with the time efficiency of PDA assessments. They reported that scanning 3000 assessment forms and correcting the data took 2 days, resulting in a significant delay in providing students with performance assessments. After introducing PDAs the synchronization was instantaneous, saving two days of scanning.

Ridgway and McCusker (2005) stress the importance of feedback following e-assessment. Feedback to lecturers provides information useful for improving quality and identifying topics that have not been learned well. Feedback to students could be diagnostic information about those aspects of performance most in need of improvement. They further recommend that students must understand criteria for assessment.

Harward et al (2003) expressed concern that older assessors (standardized patients) would not like using PDAs due to the small screen and have difficulty using a stylus but the SPs significantly preferred palms pilots to using paper forms. They also report increased self-efficacy as a result of including training in the use of PDAs as part of formal case training protocol for standardized patients.

## **3. Aim of the study**

The aim of this study is to explore the usability of PDAs for assessment of performance pertaining to relevant metrics of effectiveness and efficiency as well as to user satisfaction.

## **4. Method**

The usability of the PDA assessment of performance, using electronic checklists, was evaluated and compared to the traditional OSCE assessment using paper checklists. A product's usability is often evaluated using three metrics: effectiveness, efficiency and user satisfaction. (Segal et al 2004). The usability of the PDA system was evaluated according to the effectiveness (scores) and efficiency (time saving) as well as the user satisfaction of the PDA assessment system when compared to the traditional OSCE assessment.

### **4.1 Research variables.**

**Effectiveness** is defined as how well the assessment of performance is done. Scores obtained with electronic checklist should therefore not differ from the scores obtained by using paper checklists.

**Efficiency** is defined as the “resources consumed to achieve a goal” (Segal et al 2004). In this study the most important resource is time, meaning that a PDA assessment should take less time. Time spent on the OSCE logistics of PDA assessments was compared to that of paper based methods.

**User satisfaction** is the users’ subjective evaluation of the PDA assessment. User feedback was obtained from students as well as assessors by means of questionnaires. A five-point Likert scale was used to obtain users’ views/experiences on positive and negative statements pertaining to training issues (computer literacy, time spent on training), navigation, identification of students, time to complete checklist, negative marking and free text feedback.

## 4.2 Study design

A comparison survey was done using three cohorts of students. The first cohort was assessed using paper checklists and was compared to a second and a third cohort of students assessed by using electronic checklists.

The time spent on OSCE logistics such as preparing or editing checklists, preparation for the examination as well as producing the results was calculated and compared for the two assessment strategies.

Assessors completed a questionnaire after having done their first PDA assessment in an OSCE.

## 4.3 Participants

- Students: The population samples for this study comprised three cohorts of second year medical and dental students enrolled for the course in Generic Procedural Skills.

Cohort 1 (2003); N = 309

Cohort 2 (2004); N = 314

Cohort 3 (2005); N = 270

- Assessors: A convenience sample comprised the first 43 assessors who have used electronic checklists for the first time during any of the OSCEs conducted in the Skills Laboratory.

## 4.4 Procedure

The students were assessed during an OSCE on completion of their course on generic procedural skills as part of the second year curriculum requirements. In order to pass a student needs a minimum score of 60% and earns a “Complies with requirements” on his/her report. A student making a critical error such as inappropriate handling of contaminated needles or not adhering to principles of aseptic technique is heavily penalized to make sure they he/she does not pass the examination.

Their achievement (mean score) in the OSCEs, which comprised the same stations (skills) each year but followed different assessment strategies, were compared:

Cohort 1 was assessed using paper checklists

Cohort 2 was assessed using electronic checklists

Cohort 3 was assessed using electronic checklists and the students were aware of the assessment criteria beforehand

The time spent on OSCE logistics for PDA assessments such as preparation or editing checklists, setting up the examination on a PC, preparation of PDAs (charging and synchronizing data) and producing the results (including synchronizing of data back to PC) was calculated and compared to the time spent on traditional OSCEs using paper checklists, including preparation/editing and photostating checklists, and producing the results.

The assessors were trained before conducting their first PDA assessments. On completion of the assessments they were asked to indicate their views/experiences on a 5-point Likert scale, comprising positive and negative statements on aspects related to using the PDAs. All the assessors were experienced in using paper checklists.

## 5 Results

The **effectivity** of PDA assessment is shown in the comparison between the results of the three cohorts of students (Table 1).

Table 1. Mean OSCE results obtained through different assessment strategies

Student cohort	Year	Assessment strategy	N	Mean score %
1	2003	Paper checklists	309	64.02
2	2004	Electronic checklists	314	63.45
3	2005	Electronic checklists, known to students	270	65.97

The mean score of Cohort 1 assessed using traditional paper checklist (2003) was only 0.57% higher than the mean score obtained by Cohort 2 assessed with electronic checklists (2004) and 1.95% lower than the mean for the Cohort 3 assessed with electronic checklists (2005) that they were familiar with. This indicates a negligible difference in the scores obtained with paper or electronic checklists respectively and that students' pre-knowledge of the assessment criteria made a bigger difference, although still very small. Although pre-knowledge is not an issue raised in this paper the results are included to show the stability of the average mark obtained in the examination over a period of three years.

The **efficiency** of using portable technology for assessment was clearly shown by comparisons made to paper assessment pertaining to the following:

Table 2. Time (minutes) spent on preparing and grading of performance assessments

Year	Preparing checklists	Photostatting	Preparation of PDAs	Producing results	Training	Total minutes
2003	15	40	-	480	-	525
2004	60	-	20	10	30	120
2005	5	-	20	10	-	35

- The time spent on the initial transformation of checklists and copying them to the HaPerT system took on average 15 minutes for each of the three checklists in 2004. About 5 minutes were spent on a few changes made for the OSCE in 2005.
- The preparation of PDAs for an OSCE, involving allocation of checklists to PDAs, importing the students lists, entering assessors' names, synchronizing the data to PDAs and connecting PDAs to battery chargers, took 20 minutes each year. In 2004 Dr Schmidts aided the researcher while being trained in the use of HaPerT and in 2005, after having done several PDA assessments, the researcher was familiar with the process. This process can take longer during the learning curve of the administrator.
- Grading, together with entering and checking the results took 8 hours in 2003. In 2004 and 2005 results were available automatically and immediately after synchronization. In addition to the results feedback to students as well as examination statistics were generated.

In total the 2003 OSCE using paper checklists took over 9 hours (525 minutes) which is 438% more when compared to the 2 hours of the 2004 OSCE using PDA assessment and 1500% more than the 35 minutes of the 2005 OSCE also using PDA assessment. The time difference between the OSCEs of 2004 and 2005 is due to the checklists that only needed minor editing and the training that wasn't needed in 2005.

It has to be kept in mind that a small OSCE (three stations/procedures) with a large number of students was used for this study . The efficiency (time spent) depends a lot on the number of OSCE stations, the number of students and the experience of the administrator.

The **users' satisfaction** with using portable devices for assessment during OSCEs is shown in Table 3. The ratings of 1 and 2 on the Likert scale are grouped as negative responses and ratings 4 and 5 as positive responses.

Table 3. Users' responses (%) to using PDAs for performance assessment (n=43)

Aspect of assessment	1	2	Negative	3	4	5	Positive
Responsible for PDA	0	0	0	0	21	79	100
Computer literate	12	0	12	12	49	28	77
Not anxious	5	0	5	7	16	65	76
Training sufficient	0	0	0	5	33	63	96
Navigation	0	0	0	2	30	67	97
Checklist navigation	0	9	9	7	16	67	83
Got easier	0	0	0	2	12	86	98
Prefer to select name	0	0	0	2	5	93	98
Quicker to complete	0	7	7	0	12	81	93
Negative marking easy	16	2	18	14	16	51	67
Free text enhances feedback	2	5	7	16	28	49	77

The assessors' were willing to take responsibility for the PDA during the OSCE, in spite of the risk involved for these small devices getting lost. The 77% who reported that they were computer literate matched the 76% not feeling too anxious. None of the participants had previous experience with PDAs.

Almost all the participants were satisfied with the training (96%) which on average took 20 minutes. The navigation on the PDA was easy (97%) but the level of satisfaction with the checklist navigation was lower (83%) due to some of the checklists, e.g. for consultation skills, not being linear procedures. Not knowing such a checklist by heart and not having a global view due to the small screen, results in a search for criteria on different screens. Navigation got easier on average after having assessed 3 candidates, which is probably the same when using paper checklists.

The participants enjoyed (98%) the ability to select a student's name from the screen instead of having to write the name and student number. This saves time and the name cannot be left out as in a paper checklist

The PDA checklists was quick to complete (93%) and many participants commented that they had more time to observe the performance of the candidates. A smaller majority (77%) weren't convinced that free text would enhance feedback.

Additional comments were that the use of PDAs is less invasive during assessment and surprisingly nobody mentioned the small screen to be troublesome and even those who were computer illiterate beforehand reported using PDAs "a good idea".

The downside of the system for 18% of the participants was the negative marking, especially when assessing communication skills. Since it is not a "linear procedure" and fairly long it would have been easier to tick off the criteria as they are performed i.e. positive marking.

Another dilemma was experienced in cases where a student was not able to complete the procedure within the limited time. In these instances the remaining items had to be selected as if s(he) has made a mistake/omission.

## 5. Conclusion

The usability of PDA assessment of performance was evaluated according to the metrics most relevant to its use pertaining to effectiveness, efficiency and user satisfaction. Scores are reported to be the most direct measurement of effectiveness (Segal et al 2004). In this study the electronic checklists were found to be effective, producing almost the same median scores in each of two successive OSCEs in Procedural Skills (2004 & 2005) than the paper checklist OSCE in Procedural Skills OSCE of the preceding year (2003).

With regard to efficiency, especially considering time as a scarce resource, the PDA assessments are far more superior to the paper-and-pencil assessments. Variables such as number of OSCE stations, number of students and the efficiency of the administrator have to be taken into account.

Users were overwhelmingly positive about using the PDAs. They reported being sufficiently trained, found the navigation easy and the identification of a student and completion of the checklist to be quicker. A smaller majority found the free text functionality useful and didn't have a problem marking negatively.

Electronic assessment has added values that are not explored in this paper:

- Student feedback, automatic and available in their electronic student academic portfolios, creates a great learning opportunity.
- The sophisticated statistics that are instantly available makes quality control easy and swift.
- The scores would be more accurate
- The running cost should be less, especially where huge groups of students are involved.

The success of the system lies in the growing number of faculty changing to PDA assessment (about 75% of OSCEs in the first year of implementation) but the challenge lies in the fact that the impact of this technology on student learning has yet to be evaluated. Martin Ripley, quoted by Ridgway and McCusker (2005), underlines the need for understanding the impact on learning in stating that "E-assessment must not simply invent new technologies which recycle our current ineffective practices."

## References

Chin C H L & Donn J S (1991) Effects of computer-based tests on the achievement, anxiety, and attitudes of grade 10 science students. *Educational & Psychological Measurement*, 51(3), 735-745

Dimock P H (1991) The effect of format differences and computer experience on a computer-administered test. *Measurement & Evaluation in Counseling and Development*, 24(3), 119-126.

Harward D, Wilson C, Davis A. (2003) Palm Pilots: Cost Effective and Efficient Tool for the Virtual Assessment Center. Conference Proceedings. ASPE 2<sup>nd</sup> Annual Conference, Virginia Beach, Virginia, July 20-23, 2003

Perkins R F (1995) Using hypermedia programs to administer tests: Effects on anxiety and performance. *Journal of Research on Computing in Education*, 28(2), 209-220.

Ridgway J and McCusker S. (2005) *Literature Review of E-assessment*. A Report for NESTA Futurelab At [http://www.nestafuturelab.org/research/lit\\_reviews.htm#lr10t](http://www.nestafuturelab.org/research/lit_reviews.htm#lr10t)

Schmidts M. (2000) OSCE *Logistics* – Handheld Computers Replace Checklists and Provide Automated Feedback, *Medical Education*, 34, 957- 8.

Segal N, Doolen T L, & Porter J D (2004) A usability comparison of PDA-based quizzes and paper-and- pencil quizzes. Article in press.