Harpur Hill, Buxton
Derbyshire, SK17 9JN
T: +44 (0)1298 218000
F: +44 (0)1298 218986
W: www.hsl.gov.uk

# Munro Digital Pendulum R&R Study

**EPSU/FAL/15/028**

| | |
|---|---|
| Report Approved for Issue By: | **Nigel Corlett, PhD, ACSM** |
| Date of Issue: | **02.06.15** |
| Lead Author: | **Mark Liddle, BSc(Hons)** |
| Contributing Author: | **Gillian Frost, BSc(Hons), MRes, MSc** |
| Technical Reviewer: | **Rob Shaw, BSc(Hons)** |
| Editorial Reviewer: | **Steven Williams** |
| HSL Project Number: | **PE15080** |

This report and the work it describes were undertaken by the Health and Safety Laboratory under contract to Munro Instruments Limited. Its contents, including any opinions and/or conclusions expressed or recommendations made, do not necessarily reflect policy or views of the Health and Safety Executive.

# EXECUTIVE SUMMARY

## INTRODUCTION

The pendulum test, as described in BS 7976-2:2002+A1:2013 (BSI, 2002), is the UK Slip Resistance Group's (UKSRG) and the Health and Safety Executive's (HSE) preferred method for assessing the slip resistance of flooring materials. However, it is widely acknowledged that the results generated by the pendulum test can be susceptible to user error. To help to address this, the UKSRG have produced guidelines for its use and recently Stanley Munro have developed a new pendulum instrument, which incorporates a digital display with user prompts and the facility to record test data on a connected computer via Bluetooth.

In March 2015 the Health and Safety Laboratory's (HSL) Falls Prevention Team evaluated the accuracy and usability of two prototypes of the new Stanley Munro digital pendulum. This report details the findings of the evaluation.

## AIMS AND OBJECTIVES

The aim of the study was to provide an independent evaluation of the digital pendulum and to conduct a statistically valid comparison of the digital and analogue instruments.
The objectives were to:

- Study the variability in the analogue pendulum and digital pendulum and provide a confidence interval that can be placed around an individual measurement.
- Study the variability between different operators, between different pendulums and the residual variability, to investigate the sources of variability in the two methods.
- Identify any systematic differences between the two methods.
- Provide a limit of agreement to give an interval within which 95% of differences between measurements by the two methods are expected to lie, i.e. based on our sample, we can be 95% confident that a single result from the analogue pendulum will lie within ±X of a single result from the digital pendulum result.

## CONCLUSIONS

- The digital pendulum was as reproducible as the analogue pendulum.
- The measurement uncertainty for both methods was ±3 (95% confidence interval ±3 to ±4).
- There was an indication that the digital pendulum may be less affected by operator variability, but this would need further investigation.
- There was no quantitative evidence of a benefit in accuracy when recording the result from the digital pendulum using Bluetooth over recording it manually; however several qualitative benefits were seen.
- All operators thought that the digital pendulum was easier to use than the analogue pendulum.

# CONTENTS PAGE

# 1. INTRODUCTION

## 1.1 INTRODUCTION

The pendulum test, as described in BS 7976-2:2002+A1:2013 (BSI, 2002), is the UK Slip Resistance Group's (UKSRG) and the Health and Safety Executive's (HSE) preferred method for assessing the slip resistance of flooring materials. However, it is widely acknowledged that the results generated by the pendulum test can be susceptible to user error. To help to address this, the UKSRG have produced guidelines for its use and recently Stanley Munro have developed a new pendulum instrument, which incorporates a digital display with user prompts and the facility to record test data on a connected computer via Bluetooth.

In March 2015 the Health and Safety Laboratory's (HSL) Falls Prevention Team evaluated the accuracy and usability of two prototypes of the new Stanley Munro digital pendulum. This report details the findings of the evaluation.

## 1.2 AIMS AND OBJECTIVES

The aim of the study was to provide an independent evaluation of the digital pendulum and to conduct a statistically valid comparison of the digital and analogue instruments.
The objectives were to:

- Study the variability in the analogue pendulum and digital pendulum and provide a confidence interval that can be placed around an individual measurement.
- Study the variability between different operators, between different pendulums and the residual variability, to investigate the sources of variability in the two methods.
- Identify any systematic differences between the two methods.
- Provide a limit of agreement to give an interval within which 95% of differences between measurements by the two methods are expected to lie, i.e. based on our sample, we can be 95% confident that a single result from the analogue pendulum will lie within ±X of a single result from the digital pendulum result.

# 2. METHODOLOGY

## 2.1 EXPERIMENTAL METHOD

Pendulum testing was carried out on the three UKSRG validation surfaces (3µm iron oxide Pink Lapping Film (PLF), float glass, and a verified pavigres tile) by four experienced operators, using two digital pendulum prototypes and two standard analogue pendulums. Each pendulum was fitted with Slider 96 rubber and was assigned its own verified pavigres tile and piece of float glass. For the analogue pendulums, pendulum test values (PTV) were recorded manually on a proforma developed specifically for this study and entered into a spreadsheet by a dedicated data input operative. For the digital pendulums data was captured both manually and via a Bluetooth connected laptop computer. This was done to allow a comparison of the variability to be made between manual entry and using the Bluetooth functionality of the digital pendulums.

Each of the four operators was initially assigned to one of the four pendulums. Operators set up the pendulum by checking it was level and properly zeroed, then conditioned the Slider 96 rubber in accordance with UKRG guidelines. Once the pendulum was satisfactorily set up, the three surfaces were tested in accordance with UKSRG guidelines. Each set of tests was undertaken in the order; (i) a new piece of PLF; (ii) the assigned piece of float glass; and (iii) the assigned pavigres tile. Sliders were freshly prepared at the start of the study and were replaced after completing eight sets of tests. This meant that sliders were replaced before they exceeded the 4 mm maximum working edge dimension described in the UKSRG guidelines.
When all four operators had completed a set of tests they moved onto one of the other four pendulums, repeated the set up procedure, and completed another set of tests. Operators continued to swap pendulums according to a pre-specified sequence until they had each completed six sets of tests on each of the four pendulums. This amounted a total of 24 test results for each of the three surfaces on each of the four pendulums.

All the practical work described above was supervised by a fifth experienced pendulum operator to ensure that the study protocol was being followed precisely by each participating operator.
During testing, a minor set up error was identified and rectified on one of the digital pendulums. All the data obtained on this pendulum has been included, inclusion of the data obtained on this pendulum before the correction of the error does not have a significant impact on the statistical analysis. There was a problem when trying to connect one of the digital pendulums to a laptop via Bluetooth at the start of the study, which means that fewer readings were obtained via Bluetooth than manually.

## 2.2 STATISTICAL METHODS

Throughout the analysis, the digital pendulum with manual recording of PTV was treated as a separate pendulum type from the digital pendulum with Bluetooth recording, giving three pendulum types: analogue, digital (manual), and digital (Bluetooth). PTVs for the pavigres surface were corrected for the target value by subtracting the target value from the observed PTV.

Mixed effects models were used to analyse the PTV. These include both fixed effects, which are analogous to standard regression coefficients and are estimated directly, and random effects, which are not directly estimated but are summarised according to their estimated variances and covariances. A random effects model is a mixed effects model that includes only random effects, and can be used to partition observed variability into its constituent parts. Due to the relatively small number of pendulums and operators involved in testing, models were fitted using restricted maximum likelihood estimation. Unless otherwise specified, separate models were used for each surface type and each pendulum type. When an estimate was required across all three surface types combined, surface type was entered as a fixed effect to adjust for systematic differences between surfaces. All analyses were conducted using Stata/MP 13.1 for Windows (StataCorp, 2013).

### 2.2.1 Descriptive Statistics

The recorded PTVs were summarised using the number of observations, means and standard deviations, and box and whisker plots separately for each pendulum type and surface type.

### 2.2.2 Reproducibility and Measurement Uncertainty

Single-level random effects models were used to provide estimates of the reproducibility standard deviations (SDs) plus 95% confidence intervals (CIs). Separate models were used for each surface type, and the residual variability was allowed to vary by pendulum type to provide separate estimates of the reproducibility SD and enable Wald tests to be conducted, which would test if the reproducibility SDs were different across the three pendulum types. The measurement uncertainty was estimated as 1.96 times the reproducibility SD. This would provide the interval within which we can be 95% confident the true PTV would lie based on a single pendulum test.

### 2.2.3 Sources of Variability and Repeatability

The partition of variability in the measured PTV was investigated using two-level random effects models. Due to the small number of pendulums (two for each type) and operators (four) involved in testing, they were not entered as separate random effects. Instead, the combination of pendulum number and operator (so eight combinations for each pendulum type) was entered as a random effect. This would provide estimates of the between pendulum/operator variability, and the within pendulum/operator variability. Note that the within pendulum/operator SD is also known as the repeatability SD, and would show the repeatability of the test method if conducted by the same operator on the same pendulum.

As mentioned above, it was not feasible to decompose the between pendulum/operator SD into separate between pendulum and between operator SDs due to the small numbers of pendulums and operators involved. Therefore, to investigate the between pendulum/operator variability further, mixed effects models were used. These included the pendulum/operator combination as a random effect, but also fixed effects for the pendulum number (1 or 2) and the operator number (1, 2, 3 or 4). This would show whether there were systematic differences between the specific pendulums and operators used in the study. Note that because this analysis was conducted using fixed effects, the results may not be generalisable to other pendulums or operators outside of this study.

### 2.2.4 Systematic Differences between Methods & 95% Limits of Agreement

Mixed effects models were used to investigate if there were systematic differences between the three measurement methods. The pendulum/operator combination was entered as a random effect, with the pendulum type (analogue, digital with manual recording, and digital with Bluetooth recording) entered as a fixed effect. The Wald test was used to compare mean PTVs estimated using each of the pendulum types.

The 95% limits of agreement between two methods provide the interval within which 95% of differences between measurements by the two methods are expected to lie. This takes into account any systematic differences between the methods and the reproducibility of both methods. For a study with repeated measures like this one, the 95% limits of agreement are estimated by

$$\hat{\alpha}_1 - \hat{\alpha}_2 \pm 2 \times \sqrt{2\hat{\tau}^2 + \hat{\sigma}_1^2 + \hat{\sigma}_2^2}$$

where $\hat{\alpha}_1 - \hat{\alpha}_2$ is the estimated mean difference between the two methods, $\hat{\tau}$ is the estimated common between item/method SD, $\hat{\sigma}_1$ is the residual SD for method one, and $\hat{\sigma}_2$ is the residual SD for method two (Carstensen, 2008). In this study, there were three 'items' being measured: the three different surfaces. The inputs to this equation were estimated using a mixed effects model as described in the paper by Carstensen et al (2008). Specifically, this used a two-level mixed effects model with pendulum type and surface type as fixed effects, the surface/pendulum as a random effect, and allowing the residual variability to differ for the pendulum types. When estimating the limits of agreement for each surface type separately, there was no need to include the surface type as a fixed effect or the surface/pendulum as a random effect. Hence, the formula for the limits of agreement simplifies to

$$\hat{\alpha}_1 - \hat{\alpha}_2 \pm 2 \times \sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}$$

Separate models were used to estimate the 95% limits of agreement between the digital pendulum with manual recording compared to the analogue pendulum, and the digital pendulum with Bluetooth recording compared to the analogue pendulum. The delta method was used to estimate 95% confidence intervals for the limits of agreement.

# 3. RESULTS

## 3.1 DESCRIPTIVE STATISTICS

Table 1 shows the number of observations, and the mean and standard deviation in the observed PTVs, and Figure 1Figure 1 shows the corresponding box and whisker plots. The pink lapping film had the greatest mean PTV of 62 as measured by the analogue pendulum, followed by the pavigres tile before correction for the target value with a mean PTV of 35 and glass with a mean PTV of 7 when measured using the analogue pendulum (Table 1). There were fewer observations for the digital pendulum recorded using Bluetooth as discussed in the methods section. Table 2 shows the descriptive statistics further broken down by pendulum number for each pendulum type.

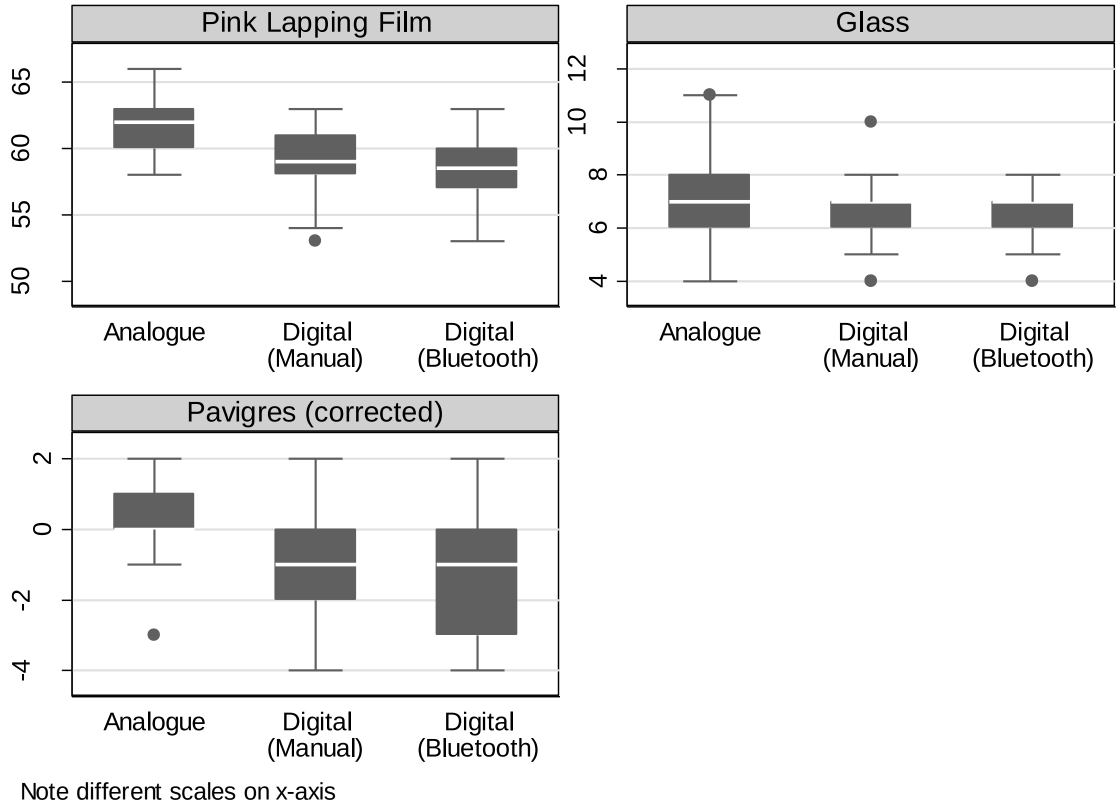


**Figure 1 Box plots of observed Pendulum Test Value, by surface type and pendulum type. The box shows the interquartile range in the PTV, the line in the box shows the median PTV, the whiskers show the upper and lower adjacent values, and the dots show outside values**

**Table 1 Descriptive statistics for the observed Pendulum Test Value, by surface type and pendulum type**

| ***Surface type*** | ***Pendulum type*** | ***Pendulum Test Value*** | | | | |
|---|---|---|---|---|---|---|
| | | ***Number*** | ***Mean*** | ***Standard deviation*** | ***Minimum*** | ***Maximum*** |
| Pink Lapping Film | Analogue | 48 | 61.6 | 1.96 | 58 | 66 |
| | Digital (manual entry) | 48 | 58.9 | 2.25 | 53 | 63 |
| | Digital (Bluetooth entry) | 38 | 58.6 | 2.31 | 53 | 63 |
| Glass | Analogue | 48 | 7.0 | 1.58 | 4 | 11 |
| | Digital (manual entry) | 48 | 6.5 | 1.32 | 4 | 10 |
| | Digital (Bluetooth entry) | 39 | 6.5 | 1.21 | 4 | 8 |
| Pavigres (corrected) | Analogue | 48 | 0.3 | 0.94 | -3 | 2 |
| | Digital (manual entry) | 48 | -1.0 | 1.57 | -4 | 2 |
| | Digital (Bluetooth entry) | 39 | -1.3 | 1.51 | -4 | 2 |
| Pavigres (uncorrected) | Analogue | 48 | 34.8 | 1.01 | 32 | 37 |
| | Digital (manual entry) | 48 | 34.0 | 2.37 | 30 | 38 |
| | Digital (Bluetooth entry) | 39 | 33.4 | 2.25 | 30 | 38 |

**Table 2 Descriptive statistics for the observed Pendulum Test Value, by surface type, pendulum type and pendulum number**

| *Tile type* | *Pendulum type* | *Pendulum number* | *Pendulum Test Value* | | | | |
|---|---|---|---|---|---|---|---|
| | | | *Number* | *Mean* | *Standard deviation* | *Minimum* | *Maximum* |
| Pink lapping film | Analogue | 1 | 24 | 62.8 | 1.67 | 60 | 66 |
| | | 2 | 24 | 60.3 | 1.37 | 58 | 63 |
| | Digital (manual entry) | 1 | 24 | 60.3 | 1.52 | 58 | 63 |
| | | 2 | 24 | 57.5 | 2.00 | 53 | 62 |
| | Digital (Bluetooth entry) | 1 | 15 | 60.3 | 1.62 | 58 | 63 |
| | | 2 | 23 | 57.5 | 2.02 | 53 | 62 |
| Glass | Analogue | 1 | 24 | 7.5 | 1.53 | 5 | 11 |
| | | 2 | 24 | 6.5 | 1.50 | 4 | 11 |
| | Digital (manual entry) | 1 | 24 | 6.6 | 1.41 | 4 | 10 |
| | | 2 | 24 | 6.5 | 1.25 | 4 | 8 |
| | Digital (Bluetooth entry) | 1 | 15 | 6.5 | 1.19 | 4 | 8 |
| | | 2 | 24 | 6.5 | 1.25 | 4 | 8 |
| Pavigres (corrected) | Analogue | 1 | 24 | 0.4 | 0.72 | -1 | 1 |
| | | 2 | 24 | 0.2 | 1.13 | -3 | 2 |
| | Digital (manual entry) | 1 | 24 | 0.0 | 1.12 | -3 | 2 |
| | | 2 | 24 | -2.0 | 1.23 | -4 | 0 |
| | Digital (Bluetooth entry) | 1 | 15 | -0.2 | 1.21 | -3 | 2 |
| | | 2 | 24 | -2.0 | 1.23 | -4 | 0 |
| Pavigres (uncorrected) | Analogue | 1 | 24 | 34.4 | 0.72 | 33 | 35 |
| | | 2 | 24 | 35.2 | 1.13 | 32 | 37 |
| | Digital (manual entry) | 1 | 24 | 36.0 | 1.12 | 33 | 38 |
| | | 2 | 24 | 32.0 | 1.23 | 30 | 34 |
| | Digital (Bluetooth entry) | 1 | 15 | 35.8 | 1.21 | 33 | 38 |
| | | 2 | 24 | 32.0 | 1.23 | 30 | 34 |

## 3.2 REPRODUCIBILITY AND MEASUREMENT UNCERTAINTY

Table 3 shows the reproducibility SD and measurement uncertainty for each of the surface types and pendulum types. The reproducibility SD for the digital pendulum tended to be higher than that for the analogue pendulum, except for the glass tile. The differences in the reproducibility SD between the pendulum types were not statistically significant, apart from the pavigres tile. In this case, the reproducibility SD for the analogue pendulum tended to be smaller than that of the digital pendulum (an SD of 0.94, compared to 1.57 for the digital pendulum with manual recording and 1.51 for the digital pendulum with Bluetooth recording). The reproducibility SD was not statistically significantly different when comparing the two methods of recording data from the digital pendulum, for any of the surface types.

The measurement uncertainty followed the trend observed for the reproducibility SDs. Tiles and pendulums with greater reproducibility SDs had greater measurement uncertainty, but the differences were not substantial. Over all surfaces combined, the measurement uncertainty was the same for all pendulum types at ±3 (95% confidence interval [CI] ±3 to ±4) (Table 3).

**Table 3 Reproducibility and measurement uncertainty of the Pendulum Test Value, by surface type and pendulum type**

| ***Surface type*** | ***Pendulum*** | ***Standard deviation (Reproducibility)*** | ***95% confidence interval*** | ***Wald test*** | ***Measurement uncertainty*** | ***95% confidence interval*** |
|---|---|---|---|---|---|---|
| Pink Lapping Film | Analogue | 1.96 | 1.60-2.39 | *p*=0.490 | ±3.83 | ±3.13, ±4.49 |
| | Digital (manual entry) | 2.25 | 1.84-2.76 | | ±4.42 | ±3.61, ±5.41 |
| | Digital (Bluetooth entry) | 2.31 | 1.84-2.90 | | ±4.53 | ±3.60, ±5.68 |
| Glass | Analogue | 1.58 | 1.29-1.94 | *p*=0.195 | ±3.10 | ±2.54, ±3.80 |
| | Digital (manual entry) | 1.32 | 1.08-1.62 | | ±2.59 | ±2.11, ±3.17 |
| | Digital (Bluetooth entry) | 1.21 | 0.97-1.52 | | ±2.37 | ±1.90, ±2.97 |
| Pavigres (corrected) | Analogue | 0.94 | 0.77-1.16 | *p*=0.001 | ±1.85 | ±1.51, ±2.27 |
| | Digital (manual entry) | 1.57 | 1.28-1.92 | | ±3.08 | ±2.52, ±3.77 |
| | Digital (Bluetooth entry) | 1.51 | 1.21-1.89 | | ±2.96 | ±2.36, ±3.71 |
| All surfaces | Analogue | 1.64 | 1.46-1.86 | *p*=0.617 | ±3.22 | ±2.85, ±3.64 |
| | Digital (manual entry) | 1.78 | 1.58-2.00 | | ±3.49 | ±3.10, ±3.92 |
| | Digital (Bluetooth entry) | 1.78 | 1.56-2.03 | | ±3.48 | ±3.05, ±3.97 |

Note: Separate models were used for each surface type.

### 3.3 SOURCES OF VARIABILITY AND REPEATABILITY

Table 4 shows the variability of the observed PTV decomposed into between pendulum/operator SD and within pendulum/operator SD (also known as repeatability SD). Apart from the analogue pendulum when measuring the pink lapping film, the within pendulum/operator SD tended to be greater than the between pendulum/operator SD for all surface types and pendulum types. However, note that, due to the relatively small number of pendulums and operators used in the study, the confidence intervals surrounding the between pendulum/operator SDs were relatively large. The between pendulum/operator SD for the pavigres surface was estimated to be zero when using the analogue pendulum.

Table 5 summarises the results of the mixed effects models investigating the differences between pendulums of the same type and operator differences. For all surfaces combined, the between pendulum differences were statistically significant for all three pendulum types (all $p<0.001$), and there were statistically significant differences between operators for the analogue pendulum but not the digital pendulums (Table 5). Note, however, that the results were not consistent across all three surface types (Table 5).

**Table 4 Decomposition of the observed variability of the Pendulum Test Value into between pendulum/operator variability and within pendulum/operator variability (aka repeatability SD), by surface type and pendulum**

| *Surface type* | *Pendulum type* | *Between pendulum/operator standard deviation* | *Within pendulum/operator standard deviation (Repeatability)* |
|---|---|---|---|
| Pink Lapping Film | Analogue | 1.71 (0.97-2.99) | 1.10 (0.89-1.38) |
| | Digital (manual entry) | 1.49 (0.79-2.83) | 1.74 (1.40-2.17) |
| | Digital (Bluetooth entry) | 1.43 (0.70-2.94) | 1.87 (1.45-2.41) |
| Glass | Analogue | 0.57 (0.18-1.79) | 1.49 (1.20-1.85) |
| | Digital (manual entry) | 0.55 (0.21-1.47) | 1.21 (0.98-1.51) |
| | Digital (Bluetooth entry) | 0.59 (0.23-1.50) | 1.08 (0.84-1.39) |
| Pavigres (corrected) | Analogue | 0.00 (0.00-0.00) | 0.94 (0.77-1.16) |
| | Digital (manual entry) | 1.09 (0.58-2.03) | 1.17 (0.94-1.46) |
| | Digital (Bluetooth entry) | 0.93 (0.45-1.92) | 1.24 (0.96-1.59) |
| All surfaces | Analogue | 0.75 (0.40-1.40) | 1.38 (1.22-1.56) |
| | Digital (manual entry) | 0.86 (0.47-1.60) | 1.55 (1.38-1.75) |
| | Digital (Bluetooth entry) | 0.79 (0.40-1.55) | 1.57 (1.37-1.80) |

Note: 95% confidence intervals are in parentheses. Separate models for each surface type and pendulum type

**Table 5 Differences between the Pendulum Test Value due to different pendulums being used and different operators, by surface type and pendulum type**

| *Surface type* | *Pendulum* | *Fixed effect* | *Wald test* |
|---|---|---|---|
| Pink Lapping Film | Analogue | Pendulum | $p<0.001$ |
| | | Operator | $p<0.001$ |
| | Digital (manual entry) | Pendulum | $p<0.001$ |
| | | Operator | $p=0.661$ |
| | Digital (Bluetooth entry) | Pendulum | $p<0.001$ |
| | | Operator | $p=0.698$ |
| Glass | Analogue | Pendulum | $p=0.028$ |
| | | Operator | $p=0.273$ |
| | Digital (manual entry) | Pendulum | $p=0.704$ |
| | | Operator | $p=0.003$ |
| | Digital (Bluetooth entry) | Pendulum | $p=0.895$ |
| | | Operator | $p=0.067$ |
| Pavigres (corrected) | Analogue | Pendulum | $p=0.371$ |
| | | Operator | $p=0.806$ |
| | Digital (manual entry) | Pendulum | $p<0.001$ |
| | | Operator | $p=0.898$ |
| | Digital (Bluetooth entry) | Pendulum | $p<0.001$ |
| | | Operator | $p=0.930$ |
| All surfaces | Analogue | Pendulum | $p<0.001$ |
| | | Operator | $p=0.003$ |
| | Digital (manual entry) | Pendulum | $p<0.001$ |
| | | Operator | $p=0.613$ |
| | Digital (Bluetooth entry) | Pendulum | $p<0.001$ |
| | | Operator | $p=0.674$ |

Note: Separate models for each surface type and pendulum type

## 3.4 SYSTEMATIC DIFFERENCES BETWEEN METHODS AND 95% LIMITS OF AGREEMENT

Table 6 shows the estimated systematic differences in the PTV measured by the different pendulum types. For all surface types, except glass, the PTV measured by the digital pendulums was, on average, statistically significantly lower than that measured by the analogue pendulum. For all surface types combined, the PTV measured by the digital pendulum using manual recording was, on average, 1.5 lower than that for the analogue pendulum (95%CI 0.6-2.4). The difference between the PTV as measured by the digital pendulum using Bluetooth recording compared to the analogue pendulum was, on average, 1.6 (95%CI 0.7-2.5). Across all surface types and combined, there was no evidence of a difference in the PTV from the digital pendulum when using manual recording compared to using Bluetooth (Table 6).

Table 7 shows the 95% limits of agreement between the digital pendulums and the analogue pendulum for each surface type and combined over all surfaces.

**Table 6 Differences in the Pendulum Test Value measured by the different pendulum types, by surface type**

| ***Surface type*** | ***Pendulum type*** | ***Estimated difference*** | ***95% confidence interval*** | ***Wald test*** |
|---|---|---|---|---|
| Pink Lapping Film | Analogue | 1.00 | Ref | |
| | Digital (manual entry) | -2.65 | -4.35, -0.95 | *p*=0.002 |
| | Digital (Bluetooth entry) | -2.73 | -4.44, -1.02 | *p*=0.002 |
| Glass | Analogue | 1.00 | Ref | |
| | Digital (manual entry) | -0.50 | -1.31, 0.30 | *p*=0.220 |
| | Digital (Bluetooth entry) | -0.61 | -1.43, 0.22 | *p*=0.151 |
| Pavigres (corrected) | Analogue | 1.00 | Ref | |
| | Digital (manual entry) | -1.31 | -2.19, -0.44 | *p*=0.003 |
| | Digital (Bluetooth entry) | -1.43 | -2.32, -0.54 | *p*=0.002 |
| All surfaces | Analogue | 1.00 | Ref | |
| | Digital (manual entry) | -1.49 | -2.36, -0.62 | *p*=0.001 |
| | Digital (Bluetooth entry) | -1.58 | -2.46, -0.70 | *p*<0.001 |

Note: Separate models were used for each surface type. Ref - reference category

**Table 7 Level of agreement between the digital pendulum and the analogue pendulum, by surface type**

| *Surface type* | *Pendulum type* | *95% limits of agreement with the analogue pendulum* | |
|---|---|---|---|
| | | *Lower limit* | *Upper limit* |
| Pink Lapping Film | Digital (manual entry) | -3.34 (-4.55, -2.14) | 8.59 (7.39, 9.80) |
| | Digital (Bluetooth entry) | -3.07 (-4.39, -1.75) | 9.04 (7.71, 10.36) |
| Glass | Digital (manual entry) | -3.60 (-4.44, -2.77) | 4.65 (3.81, 5.48) |
| | Digital (Bluetooth entry) | -3.41 (-4.25, -2.56) | 4.57 (3.72, 5.41) |
| Pavigres (corrected) | Digital (manual entry) | -2.37 (-3.15, -1.60) | 4.96 (4.18, 5.73) |
| | Digital (Bluetooth entry) | -1.94 (-2.75, -1.12) | 5.19 (4.37, 6.00) |
| All surfaces | Digital (manual entry) | -3.63 (-5.16, -2.10) | 6.59 (5.06, 8.11) |
| | Digital (Bluetooth entry) | -3.46 (-5.25, -1.68) | 6.92 (5.13, 8.71) |

Note: Separate models were used for each surface type. 95% confidence intervals in parentheses

# 4. DISCUSSION AND CONCLUSIONS

## 4.1 DISCUSSION

The variability in the digital pendulum tended to be slightly higher than the analogue pendulum – that is, the reproducibility of the analogue pendulum was slightly better. However, this difference was only statistically significant just for the pavigres tile and not for the other surfaces or overall. Although there were slight differences in the reproducibility, this had little impact on the estimated measurement uncertainty. For all three pendulum types (analogue, digital with manual recording, and digital with Bluetooth recording), the overall estimated measurement uncertainty was ±3 and the confidence intervals surrounding these estimates were narrow, ranging from ±3 to ±4. Hence, for example, if any of the pendulums measured a PTV of 60, then we can be 95% confident that the true PTV for that surface would lie in the range 60±3, so 57 to 63. In order to remain conservative, the upper limit of the confidence interval for the measurement uncertainty could be used, providing a range for the true value of 60±4 or 56 to 64.

The measurement uncertainty estimated in this study is likely to be an overestimate. Each pendulum used a different pavigres tile and the lapping film was destroyed after each test. Therefore, the measurement uncertainty will also include any variability in the manufacturing of the materials. Sheets of PLF were in short supply at the time of the study. Despite placing a single order to supply the study the PLF was supplied as individual sheets which were packaged differently suggesting that the PLF sheets were likely to be from a number of different batches. The operators included in the study were all experienced in using the analogue pendulum, less experienced operators may have greater variability in their PTVs. However, the same was not true for the digital pendulum. Three of the four operators were unfamiliar with the digital pendulum and were trained on its operation on the day of testing. The fact that the digital pendulum achieved a measurement uncertainty that was the same as the analogue pendulum is therefore encouraging.

The method used to estimate the measurement uncertainty assumed that the variability is constant across PTVs. There was an indication that this may not be true, with the pink lapping film having the greatest mean PTV and the greatest variability. However, the pink lapping film was destroyed and replaced after each test, which would add variability. Additional testing involving more surfaces spanning the range of possible PTVs would be required to investigate this further. If it was found that the variability did increase with PTV, then the measurement uncertainty estimated in this study could be an overestimate for small PTVs and an underestimate for high PTVs.

The between pendulum/operator variability tended to be lower than the within pendulum/operator variability for both the analogue pendulum and the digital pendulum. This suggests differences between operators and differences between pendulums of the same type were not the dominant source of variability in the methods. However, these results were based on a small number of pendulums and operators, and so should be interpreted with care.

The estimated between pendulum/operator variability for the pavigres tile when measured using the analogue pendulum was zero. It is unlikely that the true value is zero, but the small numbers of pendulums and operators could have resulted in a zero estimate if the between pendulum/operator variability was low. The PTV for the pavigres tile was corrected for the target value, and so the low between/operator variability for the pavigres tile could suggest that the higher variability observed for the other surfaces was mainly due to different surfaces being used on the different pendulums rather than true differences between the pendulums. However, the same result was not observed for the digital pendulum on the pavigres tile. Another possible explanation could be operator bias due to the target values for the pavigres tiles being written on the tiles themselves. Without realising it, the operators may have been influenced by their awareness of the target value when reading the analogue pendulum and making decisions about where the needle fell. This potential bias comes from the fact that the analogue pendulum scale is only marked in increments of 5 PTV leaving some operator interpretation of the final value obtained. This would not affect the results from the digital pendulum which gives a single clear value for each swing of the pendulum.

When further investigating the between pendulum/operator variability, there was strong evidence of differences between pendulums of the same type for both analogue and digital methods. Only for the analogue pendulum was there evidence of differences between operators. This could suggest that the digital pendulum is less affected by operator variability than the analogue pendulum (possibly for the reasons discussed above). However, this result was not consistent across the three surface types and the results were based on a relatively small number of operators and pendulums, and so further investigation would be needed to be certain of this result.

There was evidence that the PTV measured by the digital pendulum was systematically lower than that measured by the analogue pendulum for both the lapping film and the pavigres tile, but not the glass. Overall, the PTV was, on average, around 1.5 lower when measured using the digital pendulum compared to the analogue pendulum. This could be due to the differences in the control of the pendulums. Both of the analogue pendulums used had been calibrated by BSI in accordance with BS 7976 (BSI, 2002). As prototypes neither of the digital pendulums had been externally calibrated.

Ninety five percent limits of agreement between the digital pendulum and the analogue pendulum were estimated for each of the tile types separately and overall. This takes into account any systematic difference between the two methods, the variability in the digital pendulum and the variability in the analogue pendulum. For all tiles combined, the 95% limits of agreement were -4 (95%CI -5, -2) to +7 (95%CI 5, 8) between the digital pendulum using manual recording compared to the analogue pendulum. For example, if an operator measures a value of 60 on the digital pendulum and recorded it manually, then it would be expected that a single analogue test result on the same flooring (not necessarily by the same operator) would lie between PTVs of 56 and 67, or 55 to 68 if staying on the conservative side and using the confidence limits of the limits of agreement.

## 4.2 OBSERVATIONS

In addition to the statistical analysis of the test results, qualitative observations were made by the operators involved in the testing to evaluate the usability of the digital pendulums. Each operator was asked for their opinion on the design and mode of operation and their comments are summarised here:

- The digital display eliminates the risk of parallax error and operator judgement which can occur when interpreting where the needle rests on an analogue pendulum.
- The on-screen prompts provide users with a helpful reminder of the sequence involved in following UKSRG guidelines. The fact that they can be skipped by experienced operators is a benefit which allows for flexibility in the use of the instrument.
- The Bluetooth feature could be useful (e.g. by reducing the burden on the operator or allowing data capture in wet conditions where traditional handwritten results can be challenging), particularly if it's compatible with mobile phones.
- The regimented way in which data is recorded and the inability to delete readings could provide an excellent audit trail, but can also cause complications in the event of a rouge reading (e.g. when the pendulum arm is accidentally dropped while resetting the arm).
- Setting the zero was often problematic as an error occurred as soon as the needle swung past zero by any amount.
- Setting the footprint could result in rouge readings if the correct sequence wasn't followed precisely on the digital display.
- All operators agreed that the digital pendulum was easier to use than its analogue counterpart.

## 4.3 CONCLUSIONS

- The digital pendulum was as reproducible as the analogue pendulum.
- The measurement uncertainty for both methods was ±3 (95% confidence interval ±3 to ±4).
- There was an indication that the digital pendulum may be less affected by operator variability, but this would need further investigation.
- There was no quantitative evidence of a benefit in accuracy when recording the result from the digital pendulum using Bluetooth over recording it manually; however several qualitative benefits were seen.
- All operators thought that the digital pendulum was easier to use than the analogue pendulum.

# 5. REFERENCES

British Standards Institution. BS 7976-2:2002+A1:2013. "Pendulum testers Part 2: Method of operation."

StataCorp (2013). Stata Statistical Software: Release 13.1. College Station, TX, StataCorp LP.

Carstensen, B., J. Simpson, L. C. Gurrin (2008). "Statistical models for assessing agreement in method comparison studies with replicate measurements." Int J Biostat 4(1): Article 16.