Science & Justice 59 (2019) 544–551

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

The repeatability and reproducibility of the Sheffield Features of Gait Tool

Ivan Birch[a,*], Maria Birch[b], Lucy Rutler[b], Sarah Brown[b], Libertad Rodriguez Burgos[b], Bert Otten[c], Mickey Wiedemeijer[c]

[a] *Sheffield Teaching Hospitals NHS Foundation Trust, Woodhouse Clinic, 3 Skelton Lane, Sheffield S13 7LY, United Kingdom*
[b] *School of Health Sciences, University of Brighton, 49 Darley Road, Eastbourne BN20 7UR, United Kingdom*
[c] *Department of Human Movement Sciences, University of Groningen, Antonius Deusinglaan 1, Groningen 9713 AV, Netherlands*

ABSTRACT

Gait, the pattern or style in which locomotion is undertaken, has kinematic characteristics that may occur in varying proportions of a population and therefore have discriminatory potential. Forensic gait analysis is the analysis, comparison and evaluation of features of gait to assist the investigation of crime. While there have been recent developments in automated gait recognition systems, gait analysis presented in criminal court to assist in identification currently relies on observational analysis by expert witnesses. Observational gait analysis has been the focus of considerable research, and it has been shown that the adoption of a systematic approach to both the observation and recording of features of gait improves the reliability of the analysis. The Sheffield Features of Gait Tool was developed by forensic gait analysis practitioners based on their casework and trial experience, and consists of more than a hundred features of gait and variances. This paper reports the findings of a study undertaken to assess the repeatability and reproducibility of the Sheffield Features of Gait Tool.

Fourteen participants, with experience in observational gait analysis, viewed footage of computer generated avatars walking, and completed the features of gait tool on multiple occasions. The repeatability scores varied between participants from a highest score of 42.59 out of a maximum possible score of 45 (94.65%), to a lowest score of 30.76 (68.35%), with a mean score of 35.79 (79.54%) and a standard deviation of 3.59 (7.98%). The reproducibility scores for the assessment of each avatar varied from a highest score of 137.73 out of the best possible score of 180 (76.52%), to a lowest score of 127.21 (70.67%), with a mean score of 132.21 (73.45) and a standard deviation of 3.82 (2.12%). The results demonstrated that the use of the Sheffield Features of Gait Tool by experienced analysists resulted in what could be considered to be good levels of both repeatability and reproducibility. Some variation was shown to occur both between the results produced by different analysts, and between those produced from the analysis of different avatars. An understanding of the probative value of gait analysis evidence is an important facet of its submission as evidence, and the design and testing of standardized methods of analysis and comparison are an essential element of developing that understanding. This study is the first to test a purpose designed features of gait tool for use in forensic gait analysis.

1. Introduction

Gait, the pattern or style in which locomotion is undertaken [1], has kinematic characteristics that may occur in varying proportions of a population and therefore have discriminatory potential [2]. Forensic gait analysis is the analysis, comparison and evaluation of features of gait to assist the investigation of crime (Draft Forensic Gait Analysis Code of Practice, 2018).[1] Gait was first used as evidence in court in 1839 at the Central Criminal Court London, and was first undertaken using closed circuit television (CCTV) footage in 2000 [3]. Since 2000 there has, in the experience of the authors, been a steady increase in the use of gait analysis as evidence [2,4]. While there have been recent developments in automated gait recognition systems, gait analysis presented in court to assist in identification currently relies on observational analysis by expert witnesses [5].

Observational gait analysis is commonly used in clinical practice by a number of healthcare professions to assess pathological gait [6]. Its use has been the focus of considerable research and it has been shown that the adoption of a systematic approach to both the observation and recording of features of gait improves the reliability of the analysis [7–13]. Key aspects of reliability in this context are repeatability and reproducibility. Repeatability (intra-rater reliability) is the stability of data recorded by one individual across two or more trials. Reproducibility (inter-rater reliability) is the stability of data recorded by two or more individuals for the same trial [14]. Tools specifically designed for use in observational gait analysis have been shown to be repeatable and reproducible and improve objectivity [9,11,15–17]. While clinical observational gait tools inevitably focus on detecting pathological gait,

* Corresponding author.
E-mail addresses: ivanbirch@btinternet.com (I. Birch), M.Birch2@brighton.ac.uk (M. Birch), egbert.otten@umcg.nl (B. Otten).

[1] This document was written by the Chartered Society of Forensic Sciences' Forensic Gait Analysis Working Group in collaboration with the College of Podiatry, who were tasked by the Forensic Science Regulator to write a code of practice for forensic gait analysis capable of being read as a self-contained or standalone document. Once the final code of practice is formally published, the Forensic Science Regulator will require the provider of forensic gait analysis to ensure these services comply with the requirements outlined.

https://doi.org/10.1016/j.scijus.2019.04.001
Received 18 December 2018; Received in revised form 3 April 2019; Accepted 14 April 2019
1355-0306/

some were found to be reliable for assessment of normal gait patterns. The highest mean percentage agreement of both reproducibility and repeatability of the Salford Gait Tool were found in patients exhibiting "normal gait style" [11]. Reliability of the Edinburgh Visual Gait Score [18] also increased in higher functioning patients. This would suggest that the development and use of such a tool could enhance the reliability of gait analysis in the forensic context.

Many factors have been shown to influence the reproducibility and repeatability of observational gait tools in the clinical context. One of the most commonly reported findings is the impact of observer experience [9,11,15,16,19]. Several studies have also reported on the importance of the angle of the camera relative to the subject being recorded [9,11,20,21]. In the clinical context the angle from which gait is filmed can be selected and controlled, and clinical papers generally support the use of a camera angle at right angles to the sagittal plane of the subject [9,11]. Clinical tools such as the Salford Gait Tool [10] are specifically designed for use with a particular camera angle. In the forensic context the analyst has no such control, and the most effective orientation for recording has yet to be established [20], although Birch et al. [21] found that the highest rate of correct identifications was achieved with a camera angle at right angles to the sagittal plane of the suspect, or when the questioned footage (footage of the unknown figure) and reference footage (footage of the known subject) were captured from the same camera angle. Tools intended for use in forensic gait analysis must be suitable for use with footage captured from any angle.

Clinical tools also tend to be designed for use with particular pathologies or demographics, the features of gait included therefore being limited by their relevance. In the forensic context no such predictive assumption can be made as to the type of features of gait that may be demonstrated.

The use of gait analysis in the forensic context is still developing together with the research base underpinning practice, and there remain significant areas requiring further investigation. Birch et al. [21] demonstrated that experienced analysts instructed to focus on features of gait were able to identify successfully individuals from CCTV footage on up to 79% of occasions when the questioned and reference footage were recorded from the same angle. However, this work was undertaken without the use of a tool to structure the participants' observation and recording of the features of gait they saw. To date the published research investigating standardized methods to perform observational gait analysis in the forensic setting is limited. Larsen et al. (2008) published a checklist used in a criminal investigation which appeared not to have been peer-reviewed or tested for reliability. DiMaggio and Vernon (2011) published a list of suggested features of gait heavily based on medical deviations of gait (e.g. hemiplegic, paraplegic and ataxic gait) [20,22]. While these tools could facilitate a general overall observational assessment of gait, in the opinion of the authors, they lack the discriminatory potential of a more extensive and structured tool. Vernon and Dimaggio subsequently published an observational framework based on the early developmental work of the Sheffield Features of Gait Tool [23].

The Sheffield Features of Gait Tool was developed by forensic gait analysis practitioners, based on their extensive theoretical, clinical, research, casework and trial knowledge and experience. The format of the tool was informed by that of observational gait analysis tools developed for use in the clinical context [6,8–13,15–19,24–26]. The content of the tool was initially based on a review of the features of gait noted in 51 forensic gait analysis reports and further informed by subsequent forensic case work. The tool consists of 113 features of gait and variances divided into 14 main sections, containing those features of gait and variances noted most frequently in the gait analysis reports, plus a supplementary 'additional features' section, containing those features which were noted less frequently, but on more than three occasions. A section of the tool is shown in Fig. 1. It was designed for use with any gait, and with footage captured from any angle. Although developed by experienced practitioners, the tool had yet to be tested for repeatability and reproducibility. The knowledge and understanding of the reproducibility and repeatability of strategies and tools used in casework are fundamental requirements of forensic practice [27]. Methodological reliability has been stipulated as a requirement to improve the standing of evidence given by expert witnesses. In the US, reliability it is one of the Daubert criteria governing the admissibility of expert testimony [23,28]. In the UK, the absence of published studies of analyst reliability has recently been highlighted as a key area of concern in a primer providing guidance to UK courts on how gait evidence should be used [29]. Reproducibility and repeatability testing of an observational gait tool could improve the scientific robustness of the use of gait analysis as forensic evidence.

This paper reports the findings of a study undertaken to assess the repeatability and reproducibility of the Sheffield Features of Gait Tool.

2. Method

The study was approved by the School of Health Sciences Research Ethics and Governance Panel at the University of Brighton. Sixteen participants, each with a minimum of 12 months post-graduation experience in observational gait analysis and with a regular caseload of patients requiring observation of gait, were recruited from within the School of Health Sciences community at the University of Brighton to act as gait analysts. None of the participants had any prior experience in forensic gait analysis, nor was any training given in forensic gait analysis in general, or in the use the Sheffield Features of Gait Tool. This approach was adopted as there is currently no requirement for a person wishing to undertake gait analysis in the forensic context to undergo training prior to embarking on casework.

Eighteen pieces of footage, each of approximately six seconds in duration, were produced by the research team at the University of Groningen, The Netherlands, led by Professor Bert Otten as part of a project investigating the use of avatars as a means of testing gait analysis competency. Each piece of footage showed a computer generated avatar walking and wearing the same clothing. These avatars have been specifically designed and developed to replicate human movement, their gait being based on three dimensional motion capture data of a human subject. The avatars were used in order to ensure that the features of gait they demonstrated were known, an important component of the follow up work to this study which will test the validity of the tool in terms of its ability to facilitate the accurate observation and recording of features of gait. Six pieces of footage were used, each showing an avatar walking from a left side frontal oblique perspective, which was selected for use in this investigation as it offered the type of perspective often seen in forensic gait analysis casework and allowed consideration of features of gait occurring in all three body planes (see Fig. 2).

The avatar seen in each piece of footage was programmed with a different combination of variances of features of gait. The selection of the features of gait demonstrated by the avatars was based on forensic casework, ensuring that the combination of features reflected those seen in actual footage as accurately as possible.

The six pieces of footage to be used (one piece of footage of each of the six avatars, captured from a left frontal oblique perspective) were uploaded to a secure folder at Dropbox.com and a link created to each of the six pieces of footage. Each link was made available for a two day period selected by the participant.

Each participant was randomly allocated a pair of avatars (one and two, three and four, or five and six), and was sent the link to each of the two pieces of footages on three occasions in a random order to minimize recall bias. Each participant therefore viewed one of their two pieces of footage on six separate occasions, at one week intervals for a period of six weeks. The participants did not know how many different avatars they had viewed nor did they know whether or not they had viewed the same avatar on multiple occasions. The time period and the

A1	symmetrical gait			
A2	asymmetrical gait			
A3	erratic gait			
	symmetry of gait could not be determined			
B1	no significant rolling of the head and torso			
B2	rolling of the head and torso			
B3	rolling of the head and torso, more to the right on right steps than to the left on left steps			
B4	rolling of the head and torso, more to the left on left steps than to the right on right steps			
	frontal plane motion of the head and torso could not be determined			
B5	no significant yawing of the torso			
B6	yawing of the torso, right side forwards on left steps, left side forwards on right steps			
B7	yawing of the torso, right side forwards on right steps, left side forwards on left steps			
	transverse plane motion of the torso could not be determined			
B8	vertical movement of the head and torso on each step			
C1	head held approximately in line with the midline of the torso in the frontal plane			
C2	head held tilted to the right relative to the midline of the torso in the frontal plane			
C3	head held tilted to the left relative to the midline of the torso in the frontal plane			
	frontal plane alignment of the head could not be determined			

Fig. 1. The first four sections of the Sheffield Features of Gait Tool, showing the features relating to the gross symmetry of gait, the trunk, and the frontal plane alignment of the head.

Fig. 2. Screen shot of an avatar demonstrating the left side frontal oblique perspective used during the study.

number of tasks allocated to each participant were selected so as to gain as much information as possible while minimizing the risk of participant drop out. It was recognized that participation in the study was both potentially challenging and certainly time consuming for the participants. The Sheffield Features of Gait Tool was emailed to each participant each week as a Microsoft Excel document and labeled with a participant identification number and the date of completion. No identifying personal data was included on the completed tools. A supplementary terminology sheet was also provided.

Participants were instructed to view the footage as many times as they wished and to complete the Sheffield Features of Gait Tool for each piece of footage they were sent. For each of the 113 features of gait and variances the participants indicated with a number 1 in the appropriate box of the tool if a feature or variance was seen, leaving the box blank if it was not seen. The completed tool was then returned. Having completed the analysis of each avatar, the participants were instructed not to view the footage again, and not to review their completed and returned tool. The participants were asked not to discuss any aspects of the study with other participants for the duration of the research project, so as reduce the possibility of a participant being influenced by another participant.

The data recorded by each participant on the tool was transferred onto a Microsoft Excel spreadsheet and scored for repeatability and reproducibility. Repeatability scores were calculated by counting the highest number of same responses for each element of the tool, an element being a feature of gait or variant related to a particular part of the body. Eleven sections of the tool contained only mutually exclusive elements,[2] and therefore only positive responses, indicated by a 1 entered by the participant, were counted. The remaining 4 sections of the tool contained non- mutually exclusive elements[3] for which leaving the element without an entry provided equally valuable information regarding the feature of gait as a positive response. The non-mutually exclusive elements were scored by counting the highest number of same responses, whether they were positive responses or blanks. As the sections contained different numbers of elements, section scores were normalized to a maximum score of three to prevent some sections having a greater influence on the final total score than others. Normalized section scores were then added together to give an overall score for each analysis.

Reproducibility scores were calculated by comparing the responses

[2] In a section of the tool containing mutually exclusive elements one, and only one, element must be selected for any one piece of footage. For example:
right knee is flexed at or close to heel strike
right knee reaches a position of full or close to full extension at or close to heel strike
right knee reaches a position of full or hyper extension at or close to heel strike
angulation of the knees at or close to heel strike could not be determined

[3] In a section of the tool containing non- mutually exclusive elements more than one, or no elements can legitimately be selected. For example:
right foot is noticeably inverted prior to heel strike
right foot noticeably everts on weight bearing

of each of the participants for each element of the tool with those of each of the other participants who analyzed the same avatar. Mutually exclusive elements were scored by using only positive responses, while non-mutually exclusive elements were scored by using both positive responses and blanks. Section scores were again normalized, in this case to a maximum of 12, before being added together to give an overall score.

3. Results

Fourteen of the sixteen participants completed all six rounds of the study. Six participants completed analysis of avatars 1 and 2, four participants completed avatars 3 and 4 and four participants completed avatars 5 and 6. As repeatability is a function of each participant's ability to analyze the gait of the avatars on multiple occasions, the results yielded by all fourteen of the participants who completed the study were used to assess the repeatability of the outcomes. However, as reproducibility is a function of multiple participants' ability to analyze the same avatars, only the data yielded by the first four participants to complete the analysis of each of the avatars was used to assess the reproducibility of the outcomes in order to negate a possible skewing effect on the data of a different number of participants analyzing each of the avatars. The results are presented as both actual scores and percentage of best possible scores to facilitate comparison with previous papers.

Table 1 and Graph 1 show the repeatability scores for each of the 14 participants, demonstrating the variation in the ability of the participants to produce repeatable outcomes. The repeatability scores varied between participants from a highest score of 42.59 out of a maximum possible score of 45 (94.65%), to a lowest score of 30.76 (68.35%), with a mean score of 35.79 (79.54%) and a standard deviation of 3.59 (7.98%).

Table 2 shows the mean repeatability scores for each of the 6 avatars, while Graph 2 shows the repeatability scores for each of the six avatars divided into the components contributed by each of the participants. This data demonstrates the effect of the variation in the combinations of features of gait shown by the six avatars on the repeatability scores reported in Table 1. The mean repeatability scores for each of the six avatars showed less variation than those of the participants with a highest mean score of 37.99 out of a maximum possible score of 45 (84.41%), and a lowest of 32.12 (71.38%), with an overall mean score of 35.66 (79.25%) and a standard deviation of 2.38 (5.28%).

Graph 3 shows the total repeatability scores (sum of the scores of the 14 participants) for each of the 15 sections of the Sheffield Features of Gait Tool, demonstrating the variation in repeatability of observing different features of gait. These scores showed a greater degree of variation than those of the participants or of the avatars, with a mean score of 66.81 out of a maximum possible score of 84 (79.54%), a highest score of 75.72 (90.14%), a lowest score of 59.00 (70.24%) and a standard deviation of 4.33 (5.15%).

Table 3 and Graph 4 show the reproducibility scores and the percentage of the best possible score for each of the six avatars, also shown in Graph 4. As was the case with the participant repeatability scores, the avatar reproducibility scores varied with a highest score of 137.73 out of the best possible score of 180 (76.52%), and a lowest score of 127.21 (70.67%), with a mean score of 132.21 (73.45%) and a standard deviation of 3.82 (2.12%).

Graph 5 shows the total reproducibility scores (sum of the scores of the 6 avatars) for each of the 15 sections of the Sheffield Features of Gait Tool. As was the case with the repeatability scores, the section scores varied to a greater degree than did the reproducibility scores for the avatars, with a mean score of 53.88 out of a maximum possible score of 72 (73.45%), a highest score of 61.00 (84.72%), a lowest score of 45.50 (63.19%), and a standard deviation of 4.44 (6.17%).

4. Discussion

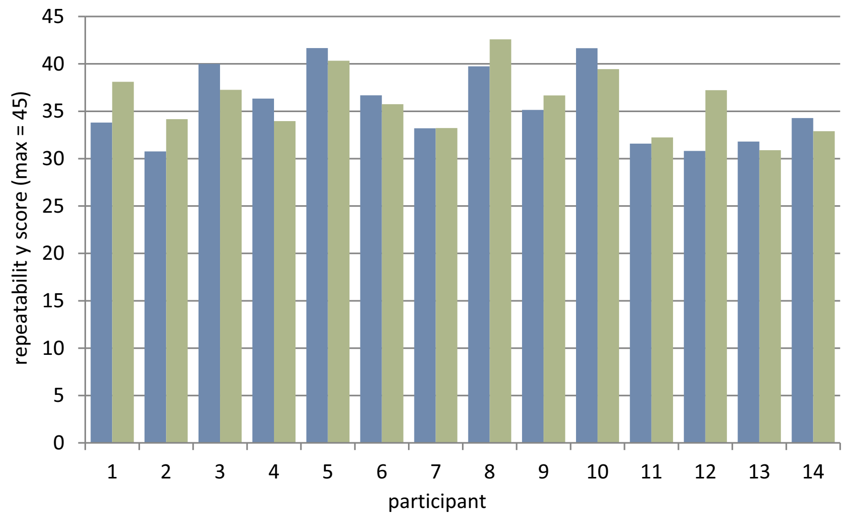
The aim of this study was to test the repeatability and reproducibility of the Sheffield Features of Gait Tool. To date there has been no tested method of observing and noting features of gait in the forensic context. The use of tried and tested methods for which error rates are known is a core element of The Forensic Science Regulator's Forensic science providers: codes of practice and conduct, 2017, and this study is therefore an important facet of ensuring that forensic gait analysis practice meets the same stringent requirements as all other areas of forensic practice.

Participant repeatability was found to vary between observers, ranging from 42.59/45 (94.65%) to 30.76/45 (68.35%). However, the standard deviation of 3.59 suggests a relatively tight grouping of the participant scores. The presence of variation in participant scores is consistent with the findings of investigations into tools to facilitate observational gait analysis in the clinical context [8,11,24], Toro et al. reporting results of 75% for intraobserver repeatability. However, it should be noted that comparison between the results of the papers is less than straight forward in view of the different methodologies and analyses employed. The presentation of the six pieces of footage, showing the two avatars seen by each participant in this study, was randomised, the participants not knowing how many avatars or whether features of gait varied or were constant. This strategy was used to minimize the possible effect of recall bias, the influence of prior knowledge and experience gained in the preceding assessments. Randomising was only explicitly stated in four studies that tested the repeatability of clinical observational gait tools [9,11,15,25].

One explanation for the variation that was found could be the differing levels of experience of the participants, which is widely reported in the clinical literature to be a factor [9,11,15,25]. The number of participants (14 for the assessment of repeatability, and 12, divided into

Table 1
The repeatability scores for each of the 14 participants (maximum possible score 45).

Participant	1		2		3		4		5		6	
Avatar analyzed	1	2	1	2	1	2	1	2	1	2	1	2
Total score	33.81	38.11	30.76	34.17	40.00	37.26	36.34	33.96	41.67	40.33	36.68	35.74
% score	75.12	84.69	68.35	75.92	88.89	82.80	80.75	75.47	92.59	89.63	81.51	79.43
Participant	7		8		9		10		Mean Score		35.79	
Avatar analyzed	3	4	3	4	3	4	3	4	Max		42.59	
Total score	33.20	33.23	39.74	42.59	35.14	36.67	41.65	39.45	Min		30.76	
% score	73.77	73.85	88.32	94.65	78.09	81.48	92.55	87.66	Std		3.59	
									Mean % Score		79.54	
Participant	11		12		13		14		Max %		94.65	
Avatar analyzed	5	6	5	6	5	6	5	6	Min %		68.35	
Total score	31.59	32.24	30.81	37.23	31.80	30.89	34.28	32.90	Std %		7.98	
% Score	70.20	71.64	68.47	82.73	70.67	68.64	76.17	73.11				



Graph 1. Repeatability scores for the two avatars analyzed by each of the 14 participants (maximum possible score 45).

three groups of four, for the assessment of reproducibility) who took part in this study is relatively large compared to many of the studies assessing the reliability of the clinical observational gait tools, some of which were based on just one or two observers [8,16–18]. Three studies reporting specifically on the impact of experience on observational gait analysis found that repeatability increased with experience [8,18,19]. Previous studies in the forensic context have all relied on individuals with experience. Birch et al. [21] compared the results of the experienced analysts in their study who successfully identify individuals from CCTV footage on an overall 71% of occasions to the inexperienced analysts in the previous research of Stevenage et al. [30] showing that they had a correct identification score 21% higher, and suggested that analysts with experience perhaps perform better than those without experience. Although all the participants had been trained in observational gait analysis as part of their pre-registration education, and had at least 12 months post-graduation experience, none of the participants had any experience in forensic gait analysis. Forensic gait analysis requires a more detailed and protracted analysis of gait than is usually undertaken in the clinic, and therefore the participants' experience of gait analysis and that required as part of this study may not have aligned. Experience can be hard to define and it has been suggested that clinical experience does not necessarily equate to experience of observational gait analysis [31], and experience in observational gait analysis may not equate to an ability to apply that experience for a forensic purpose [32]. Future testing of the tool using participants with experience in forensic gait analysis would provide useful data.

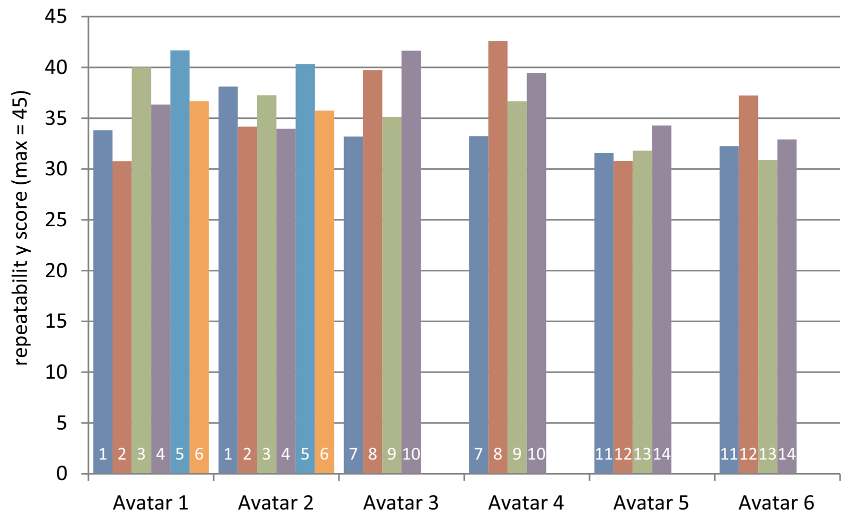
This study relied on participant involvement over a six-week period. One participant withdrew due to time commitments and it is possible that the competing professional commitments of some of the other participants also limited the time available for them to undertake the research task, potentially impacting on their level of engagement. Variations in participant engagement in study designs involving multiple rounds of activity have been previously reported [33]. The tool provides a list of features of gait and variances, and is therefore both extensive and complex. A terminology sheet was provided, but no guidance was given regarding the completion of the tool, the intention being to test the use of the tool under the least favourable conditions. The complex nature of the tool and the lack of experience in its use may also have contributed to demands placed on the participants and the variation in their scores.

Graph 3 shows the variation in repeatability between the 15 sections of the tool, the scores varying from a highest score of 75.72/84 (90.14%) to a lowest score of 59.00/84 (70.24%), with a standard deviation of 4.33, again suggesting a relatively tight grouping of the scores from each of the sections. Each section of the tool deals with either a different part of the body or a different point in the gait cycle. The sections necessarily vary slightly in complexity, and in terms of ease of completion are dependent on aspects of the particular piece of footage being analyzed. The lowest scoring sections were those dealing with the frontal plane movement of the head and torso, the movement occurring at the hip and its effect on the path of motion of the feet, and the angulation of the knee at or close to heel rise. The sections represent a variety of types of observation. Papers reporting the testing of clinical tools generally found that observations of more distally occurring features of gait were more reliable than those of proximal ones [8,9,18,19,25]. This appears not to have been the case in this study. Some studies suggested variation by body segment could be due to a lack of understanding about what constitutes a particular feature of gait and recommended the inclusion of more detailed definitions. This is unlikely to be significant in this case because detailed definitions were provided; indeed the authors of the tool have previously published a paper on the importance of consistent terminology and clear definitions in forensic gait analysis [34]. Also, the great majority of participants were podiatrists and should have been particularly familiar with motion of the foot during the gait cycle and associated terminology. It is therefore possible that although participants knew what to look for they found these subtle movements hard to distinguish from the footage of the avatars.

The section scores also show that the most repeatable section of the tool was that dealing with arm swing. Birch et al. [21] reported that analysts focusing on upper body gait features such as arm swing scored more highly in terms of correctly identifying individuals from their gait

Table 2
The mean repeatability scores for each of the 6 avatars (maximum possible score 45).

Avatar	1	2	3	4	5	6
Participants	1,2,3,4,5,6	1,2,3,4,5,6	7,8,9,10	7,8,9,10	11,12,13,14	11,12,13,14
Mean score	36.54	36.60	37.43	37.99	32.12	33.31
Mean % score	81.20	81.32	83.18	84.41	71.38	74.03



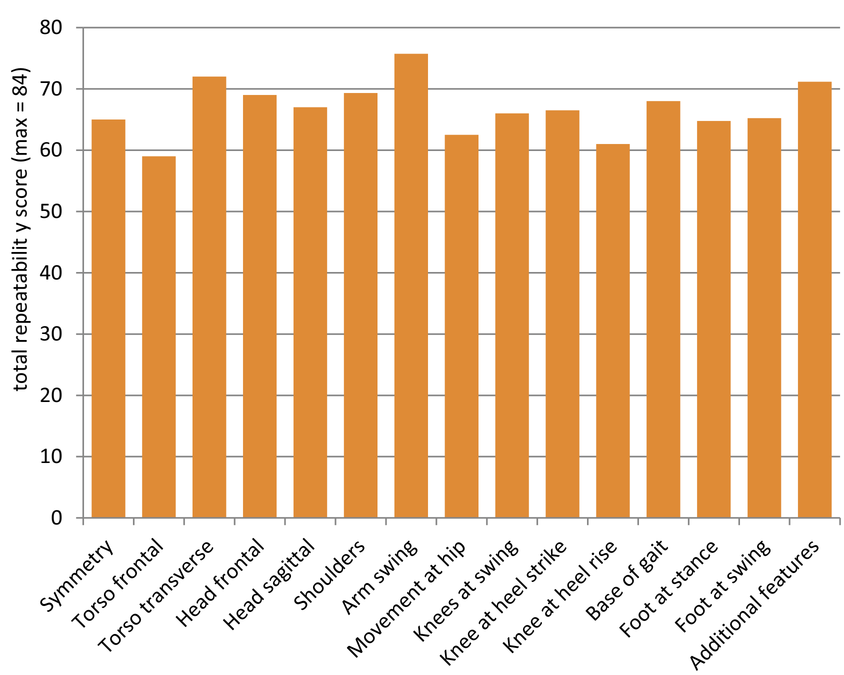
Graph 2. Repeatability scores for each of the six avatars (maximum possible score 45; participant number shown in the base of each column).

features than those focusing on lower body gait features. The high score of the arm swing section would seem to agree with this earlier finding, particularly in view of the high proportions of participants with a podiatric background who were more experienced in assessing lower limb pathology. However, the explanation may not be quite as straight forward. The arm section had a number of detailed subsections, which although the final section scores were normalized, may have played a part in the high score.

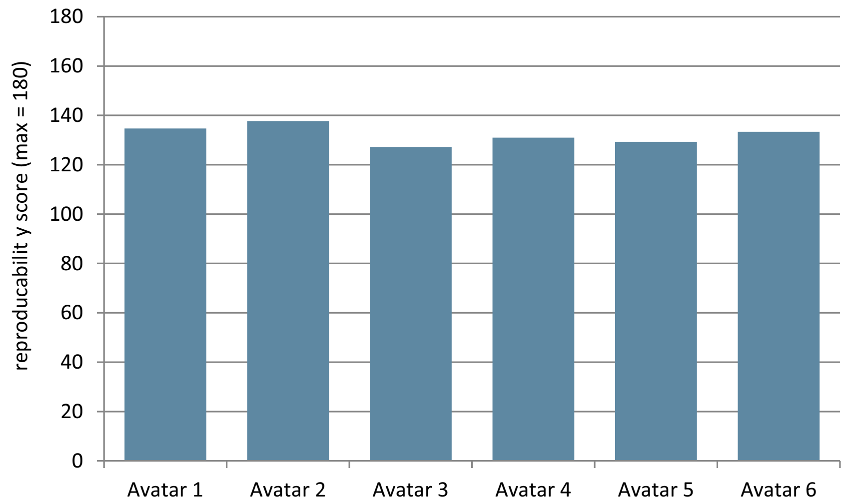
The reproducibility scores for each of the six avatars also varied, with a highest score of 137.73/180 (76.52) and a lowest score of 127.21/180 (70.67), with a standard deviation of 3.82, a proportionately smaller standard deviation than was found for the repeatability scores. The ability of a gait assessment tool to produce reproducible results is important in any context. In the clinical context it allows gait analysis undertaken by one practitioner to be compared, or used in conjunction with, that undertaken by another at a later date. In the forensic context it assures parity between the gait analysis undertaken by different experts working on the same case, allowing the court to compare more easily the findings based on that analysis. The reproducibility scores in this study were calculated using a line by line analysis and section totals rather than a total score or overall assessment. The calculation of reproducibility and repeatability in clinical tools is commonly based on cumulative scores whereby nominal scores for each joint segment are added together to assess overall pathology [8,15–17]. The section scores from this study showed the scores to vary from a highest of 61/72 (base of gait) to a lowest of 45.5/72

Table 3
The reproducibility scores and the percentage of the best possible score for each of the six avatars (maximum possible score 180).

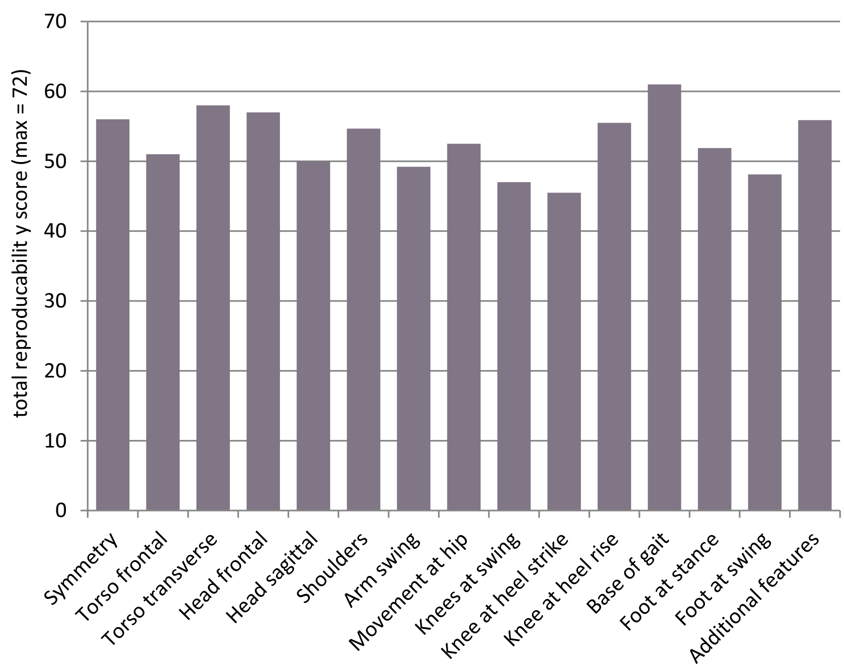
Avatar	1	2	3	4	5	6
Total score	134.70	137.73	127.21	130.98	129.30	133.33
% of best possible score (180)	74.83	76.52	70.67	72.77	71.83	74.07



Graph 3. Total repeatability scores for each of the 15 sections of the Sheffield Features of Gait Tool (maximum possible score 84).



Graph 4. The reproducibility score for each of the six avatars (maximum possible score 180).



Graph 5. Total reproducibility scores for each of the 15 sections of the Sheffield Features of Gait Tool (maximum possible score 72).

(angulation of the knee at heel strike) with a mean score of 52.88/72 and a standard deviation of 4.4. The lower score of the angulation of the knee at heel strike section may have been related to the perspective of the avatar shown in the footage, left side frontal. From this orientation determination of the angulation of the knee would have been challenging for the participants, and in a forensic context the selection of ‘could not be determined’ may have been the most appropriate option.

Whilst comparison with other observational gait tools is limited due to the different statistical measures employed, these results broadly align with tools in the clinical setting shown to have good reproducibility such as the Salford Gait Tool (77% mean agreement) [11] and the Edinburgh Visual Gait Score (70% mean agreement [9], 60–90% agreement [18]). What constitutes good or moderate reproducibility within a particular context is relatively subjective and varies between studies. Toro et al. [11] warn that a tool that is reliable for one patient group may not be reliable for another. It has also been suggested that reliability should be greater when making clinical decisions related to individuals than for clinical decisions affecting groups [14,35]. This raises the question of what acceptable reproducibility and repeatability should be in the forensic context. The 2017 publication ‘Forensic gait analysis: a primer for courts’ noted that identification success rates of 71%, equating to a failure rate of 29%, were unacceptably high [29]. However, the primer mismatches the process used in current forensic gait analysis practice, which is acknowledged as not being capable of identification [2], with research papers that apparently investigated the ability of observers to identify subjects [21,30,36,37], thus conflating two different paradigms. A criminal conviction has a profound and lasting impact on people, however, forensic gait evidence remains class level, providing a likelihood of this being the same person or “congruence of identity” not a definitive match [2,38]. What is currently important is to help courts to understand the probative value of gait analysis evidence, and the more information that can be clearly provided regarding the strategies used during analysis and comparison, the better this objective can be achieved.

The study used computer generated avatars which were specifically designed for use in observational gait analysis competency testing, and were programmed with pre-determined features of gait. Six avatars were used, each programmed with a different set of features of gait. The use of six sets of features introduced the possibility of a variation in the challenge presented to the analyst when analyzing the gait, as would be found in forensic practice. However, two other aspects of this strategy need to be considered. The first is the obvious question of whether using footage of a person rather than an avatar would have made the task easier or more difficult. The object of the study was to test the tool, and the use of the avatar removed the complication of a range of intrinsic and extrinsic factors that can affect gait [4,39–41]. The use of footage of a person would have resulted in the results representing a combination of the tool and that variability. However, the analysts experience was entirely based on analyzing real people not avatars. The second is that the footage used was of a high frame rate, good resolution and good lighting, and therefore footage that would be considered ideal for use in forensic gait analysis. Much of the footage submitted for analysis is suboptimal and the limitations imposed are well established [21,31,42]. The quality of the footage used could have been downgraded, but once again this would have resulted in the results reflecting a combination of determining factors rather than just the ability of the tool to yield appropriate results. Both aspects certainly warrant further consideration.

5. Conclusions

The results demonstrated that the use of the Sheffield Features of Gait tool by experienced analysts resulted in good levels of both repeatability and reproducibility. Some variation was shown to occur both between the results produced by different analysts, and between those produced from different avatars. Reproducibility scores between sections of the tool were shown to vary slightly more than repeatability scores. This is the first observational gait analysis tool specifically designed for use in the forensic context, and the research described in this paper will inform future development and retesting of the tool.

An understanding of the probative value of gait analysis evidence is an important facet of its submission as evidence, and the design and testing of standardized methods of analysis and comparison are an essential element of developing that understanding. This study is the first to test a purpose designed tool for use in forensic gait analysis.

Conflict of interest statement

The authors certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Acknowledgements

The authors would like to thank Sophie Allen and Jonathan Li for their contributions to this research project.

References

[1] D. Levine, J. Richards, M.W. Whittle, Whittle's Gait Analysis, Elsevier Health Sciences, 2012.
[2] I. Birch, C. Gwinnett, J. Walker, Aiding the interpretation of forensic gait analysis: development of a features of gait database, Sci. Justice 56 (2016) 426–430.
[3] M. Nirenberg, W. Vernon, I. Birch, A review of the historical use and criticisms of gait analysis evidence, Sci. Justice 58 (2018) 292–298.
[4] I. Birch, T. Birch, D. Bray, The identification of emotions from gait, Sci. Justice 56 (2016) 351–356.
[5] D. Gibelli, Z. Obertová, S. Ritz-Timme, P. Gabriel, T. Arent, M. Ratnayake, D. De Angelis, C. Cattaneo, The identification of living persons on images: a literature review, Legal Med. 19 (2016) 52–60.
[6] B. Toro, C. Nester, P. Farren, A review of observational gait assessment in clinical practice, Physiother. Theory Pract. 19 (2003) 137–149.
[7] J. Perry, Gait analysis, Los Amigos Research & Education Center, Downey, California, 2002.
[8] S. Lord, P. Halligan, D. Wade, Visual gait analysis: the development of a clinical assessment and scale, Clin. Rehabil. 12 (1998) 107–119.
[9] H.S. Read, M.E. Hazlewood, S.J. Hillman, R.J. Prescott, J.E. Robb, Edinburgh visual gait score for use in cerebral palsy, J. Pediatr. Orthop. 23 (2003) 296–301.
[10] B. Toro, C.J. Nester, P.C. Farren, The development and validity of the Salford Gait Tool: an observation-based clinical gait assessment tool, Arch. Phys. Med. Rehabil. 88 (2007) 321–327.
[11] B. Toro, C.J. Nester, P.C. Farren, Inter-and intraobserver repeatability of the Salford Gait Tool: an observation-based clinical gait assessment tool, Arch. Phys. Med. Rehabil. 88 (2007) 328–332.
[12] C. Rathinam, A. Bateman, J. Peirson, J. Skinner, Observational gait assessment tools in paediatrics - a systematic review, Gait Posture 40 (2014) 279–285.
[13] M.D. Gor-GarcíaFogeda, R. Cano de la Cuerda, M. Carratalá Tejada, I.M. Alguacil-Diego, F. Molina-Rueda, Observational gait assessments in people with neurological disorders: a systematic review, Arch. Phys. Med. Rehabil. 97 (2016) 131–140.
[14] L. Portney, M. Watkins, Foundations of Clinical Research: Applications to Practice, second ed., Prentice-Hall, New Jersey, 2000.
[15] R. Wellmon, A. Degano, J.A. Rubertone, S. Campbell, K.A. Russo, Interrater and intrarater reliability and minimal detectable change of the Wisconsin Gait Scale when used to examine videotaped gait in individuals post-stroke, Arch. Physiother. 5 (2015) 11.
[16] X. Lu, N. Hu, S. Deng, J. Li, S. Qi, S. Bi, The reliability, validity and correlation of two observational gait scales assessed by video tape for Chinese subjects with hemiplegia, J. Phys. Ther. Sci. 27 (2015) 3717–3721.
[17] J. Daly, J. Nethery, J. McCabe, I. Brenner, J. Rogers, J. Gansen, K. Butler, R. Burdsall, K. Roenigk, J. Holcomb, Development and testing of the Gait Assessment and Intervention Tool (GAIT): a measure of coordinated gait components, J. Neurosci. Methods 178 (2009) 334–339.
[18] M.d.P.D. Orozco, O. Abousamra, C. Church, N. Lennon, J. Henley, K.J. Rogers, J.P. Sees, J. Connor, F. Miller, Reliability and validity of Edinburgh visual gait score as an evaluation tool for children with cerebral palsy, Gait posture 49 (2016) 14–18.
[19] C. Brown, S. Hillman, A. Richardson, J. Herman, J. Robb, Reliability and validity of the Visual Gait Assessment Scale for children with hemiplegic cerebral palsy when used by experienced and inexperienced observers, Gait posture 27 (2008) 648–652.
[20] P.K. Larsen, E.B. Simonsen, N. Lynnerup, Gait analysis in forensic medicine, J. Forensic Sci. 53 (2008) 1149–1153.
[21] I. Birch, L. Raymond, A. Christou, M.A. Fernando, N. Harrison, F. Paul, The identification of individuals by observational gait analysis using closed circuit television footage, Sci. Justice 53 (2013) 339–342.
[22] J.A. DiMaggio, W. Vernon, Forensic Podiatry: Principles and Methods, Humana, New York London, 2011.
[23] D.W. Vernon, J.A. DiMaggio, Forensic Podiatry: Principles and Methods, CRC Press, 2017.
[24] A.H. Mackey, G.L. Lobb, S.E. Walt, N.S. Stott, Reliability and validity of the observational Gait Scale in children with spastic diplegia, Dev. Med. Child Neurol. 45 (2003) 4–11.
[25] A. Ong, S. Hillman, J. Robb, Reliability and validity of the Edinburgh Visual Gait Score for cerebral palsy when used by inexperienced observers, Gait Posture 28 (2008) 323–326.
[26] Rancho Los Amigos National Rehabilitation Center, Observational Gait Analysis, Los Amigos Research & Education Institute, Downey, CA, 2001.
[27] Forensic Science Regulator, Codes of Practice and Conduct: For Forensic Science Providers and Practitioners in the Criminal Justice System, The Forensic Science Regulator, Birmingham, 2017, p. 67.
[28] M. Nirenberg, Meeting a forensic podiatry admissibility challenge: a Daubert case study, J. Forensic Sci. 61 (2016) 833–841.
[29] The Royal Society and the Royal Society of Edinburgh, Forensic Gait Analysis: A Primer for Courts, The Royal Society, London, 2017.
[30] S.V. Stevenage, M.S. Nixon, K. Vince, Visual analysis of gait as a cue to identity, Appl. Cogn. Psychol. 13 (1999) 513–526.
[31] I. Birch, W. Vernon, J. Walker, J. Saxelby, The development of a tool for assessing the quality of closed circuit camera footage for use in forensic gait analysis, J. Forensic Legal Med. 20 (2013) 915–917.
[32] Lundy v The Queen (New Zealand), UKPC 28 (October 2013) [Online]. Available, 2013. http://www.bailii.org/uk/cases/UKPC/2013/28.html.
[33] W. Vernon, The Delphi technique: a review, Int. J. Ther. Rehabil. 16 (2009) 69–76.
[34] I. Birch, W. Vernon, J. Walker, M. Young, Terminology and forensic gait analysis, Sci. Justice 55 (2015) 279–284.
[35] J. Bartlett, C. Frost, Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables, Ultrasound Obst. Gynecol. 31 (2008) 466–475.
[36] J.E. Cutting, L.T. Kozlowski, Recognizing friends by their walk - gait perception without familiarity cues, Bull. Psychon. Soc. 9 (1977) 353–356.
[37] N.F. Troje, C. Westhoff, M. Lavrov, Person identification from biological motion: effects of structural and kinematic cues, Percept. Psychophys. 67 (2005) 667–675.
[38] N. Lynnerup, P.K. Larsen, Gait as evidence, IET Biometrics 3 (2014) 47–54.
[39] O. Ludwig, S. Dillinger, F. Marschall, Intra-individual gait pattern variability in specific situations: implications for forensic gait analysis, Forensic Sci. Int. 264 (2016) 15–23.
[40] S.X. Yang, P.K. Larsen, T. Alkjær, E.B. Simonsen, N. Lynnerup, Variability and similarity of gait as evaluated by joint angles: implications for forensic gait analysis, J. Forensic Sci. 59 (2014) 494–504.
[41] J.C. Menant, J.R. Steele, H.B. Menz, B.J. Munro, S.R. Lord, Effects of walking surfaces and footwear on temporo-spatial gait parameters in young and older people, Gait Posture 29 (2009) 392–397.
[42] I. Birch, W. Vernon, G. Burrow, J. Walker, The effect of frame rate on the ability of experienced gait analysts to identify characteristics of gait from closed circuit television footage, Sci. Justice 54 (2014) 159–163.