Published in final edited form as:

Physiother Theory Pract. 2019 January ; 35(1): 94–99. doi:10.1080/09593985.2018.1434708.

Inter-rater and Intra-rater Reliability of a Clinical Protocol for Measuring Turnout in Collegiate Dancers

Amanda Greene, PT, DPT¹, Andrea Lasner, PT, MSPT¹, Rajwinder Deu, MD², Seth Oliphant, MD¹, and Kenneth Johnson, PT¹

¹Department of Physical Medicine and Rehabilitation, Johns Hopkins University, Baltimore, MD

²Department of ^bOrthopaedic Surgery, Johns Hopkins University, Baltimore, MD

Abstract

Reliable methods of measuring turnout in dancers and comparing active turnout (used in class) with functional (uncompensated) turnout are needed. Authors have suggested measurement techniques but there is no clinically useful, easily reproducible technique with established inter-rater and intra-rater reliability. We adapted a technique based on previous research, which is easily reproducible. We hypothesized excellent inter-rater and intra-rater reliability between experienced physical therapists and a briefly trained faculty member from a university's department of dance. Thirty-two participants were recruited from the same dance department. Dancers' active and functional turnout were measured by each rater. We found that our technique for measuring active and functional turnout has excellent inter-rater and intra-rater reliability when performed by two experienced physical therapists and by one briefly trained university-level dance faculty member. For active turnout, inter-rater reliability was 0.78 among all raters and 0.82 among only the PT raters; intra-rater reliability was 0.82 among all raters and 0.85 among only the PT raters. For functional turnout, inter-rater reliability was 0.86 among all raters and 0.88 among only the PT raters; intra-rater reliability was 0.87 among all raters and 0.88 among only the PT raters. The measurement technique described provides a standardized protocol with excellent inter-rater and intra-rater reliability when performed by experienced PTs or by a briefly trained university-level dance faculty member.

Keywords

active turnout; collegiate dancers; functional turnout

INTRODUCTION

Dancers are at high risk for injuries. A recent study showed that professional ballet dancers sustain an average of one new injury per year (Ramkumar et al, 2016). Most dance injuries occur in the lower extremities, and many are overuse injuries (Ekegren, Qvested, and Brodrick, 2014; Liederbach, Dilgen, and Rose, 2008; Ramkumar et al, 2016). Perfect

Corresponding Author: Rajwinder Deu, MD, rdeu1@jhmi.edu.

Declaration of Interest:

The authors report no declarations of interest.

turnout in dance is often described as a position of external rotation of the lower extremities so that the longitudinal axes of the feet are rotated 180 degrees away from each other (Gilbert, Gross, and Klug, 1998). Although most turnout should be produced by external rotation of the hip, studies have suggested that tibial rotation, tibial torsion, and external rotation of the foot also contribute to the total turnout of the leg (Gilbert, Gross, and Klug, 1998; Grossman, 2003). Dancers often perform compensatory movements such as excessive anterior pelvic tilt, external tibial rotation, and subtalar joint pronation to exaggerate their turnout and achieve a more aesthetically pleasing position (Gilbert, Gross, and Klug, 1998). These compensations can increase injury rates among ballet (Coplan, 2002; Negus, Hopper, and Briffa, 2005; Steinberg et al, 2012) and modern (van Merkensteijn and Quin, 2015) dancers. It has been recommended that dance screening tools include measures of both active turnout (the turnout used in class) and functional (uncompensated) turnout for comparison (Liederbach, 1997); however, a consensus has not been reached by the dance medicine community on a standard measurement tool.

Although authors have proposed active and functional turnout measurement techniques (Grossman, 2003; Grossman et al, 2008; Welsh et al, 2008), clinically useful, easily reproducible techniques with established inter-rater and intra-rater reliability have not been described. Gilbert, Gross, and Klug (1998) measured standing turnout by outlining the feet on paper and taking the measurement from the drawings. They reported acceptable intra-rater reliability of this technique but the sample size was lower than the 30 participants recommended to establish good reliability (Portney and Watkins, 2009), and inter-rater reliability was not assessed. Additionally, there was no technique for measuring functional turnout in their study. Welsh et al. (2008) measured standing or “best” turnout, as well as functional turnout on rotational disks by taking photographs from above and measuring the turnout using the photographs. This technique was also found to have good intra-rater reliability, but the authors did not address inter-rater reliability. An interventional study by Pata, Welsh, Bailey, and Range (2014) used Welsh’s technique to assess response to an intervention to improve turnout. Inter-rater reliability of the measurement technique was assessed during this study and was found to be 100%. We believe these two techniques are not easily or quickly reproducible in a screening, classroom, or clinical setting because of their reliance on photographs or foot tracings. A similar technique for measuring active and functional turnout described by Grossman et al. (2008) eliminated the need for paper drawings or photographs; however, the reliability of the technique was not assessed. There has been no investigation of the reliability of these measures when performed by university-level dance faculty members versus healthcare providers. As the field of dance science continues to grow, more university-level dance faculty members are performing research that measures turnout. Therefore, it is imperative to determine if they are reliable when doing so. Previous studies have shown that goniometric measurements are reliable to four to five degrees when taken by the same tester and five to six degrees when taken by different testers (Boone et al, 1978; Norkin and White, 2003). However, reliability of goniometric measurements is influenced by the testing procedure. Studies show that using consistent, well-defined testing procedures improves reliability. Therefore it is important to establish standardized positioning and techniques, including palpation of anatomical landmarks to

align the arms of the goniometer, when performing tests that rely on goniometric measurements (Norkin and White, 2003).

We adapted a technique based on those of Welsh et al. (2008); Grossman (2003); and Grossman et al. (2008), which is easily reproducible in classroom, screening, or clinical environments. Our intent was to establish a standardized position and technique for measuring active and functional (uncompensated) turnout and to test the inter-rater and intra-rater reliability of the technique among performing arts physical therapists (PTs) and a briefly trained university-level dance faculty member.

METHODS

Participants

The study was approved by the Johns Hopkins Medicine Office of Human Subjects Research institutional review board. Written permission was granted from the director of the participating university's dance department, from which participants were recruited. Power analysis was performed for an expected intra-class correlation of 0.5 using PASS 13 Power Analysis and Sample Size software, version 2014, and showed a sample size of 30 subjects with three observations per subject would be required to achieve 98% power at an α level of 0.05. Participants were recruited by flier advertising in a university dance department, as well as through an email to approximately 80 dance major students. Dancers were not excluded on the basis of current or previous injuries. All dancers provided written informed consent to participate.

Procedures

A university dance faculty member received one hour of observational instruction and one hour of hands-on practice performing the turnout measurement technique. She was also provided with a written protocol to review on her own. PTs had a one-hour session to review and practice the measurement technique, which was a technique they had used clinically and during previous dancer screenings.

Dancers arrived three at a time. Each rater performed one trial of active turnout and one trial of functional turnout on one dancer and then tested the next of the three dancers. This nonrandomized sequence continued until each rater had tested each dancer three times. All measurements were recorded by study personnel to blind the raters to the dancers' previous measurements. Dancers were tested in bare feet and were asked not to warm up before testing. A standard script was used to inform them that their turnout was not being assessed, rather the measurement technique was being assessed. They were instructed to give the same amount of effort during each trial and not to try to "outdo" themselves during subsequent trials.

For active turnout, dancers stood on a whiteboard with a line bisecting the board and were asked to stand in their usual first position (bilateral lower extremity external rotation with the knees extended and the heels together) "that you use in class," with their arms in second position (bilateral arms abducted to 90 degrees). The arms were placed in second position because this is a standard dance position and was used in the literature described above.

They were instructed to stand with their heels centered over a fixed line on the whiteboard. No cues were given to correct compensation used by the dancers to increase the apparent turnout. The rater palpated the second metatarsal and used a ruler as a straightedge to draw a line extending from the second metatarsal (Figure 1). The dancer stepped off, and the rater used an extendable goniometer (model 01135, Lafayette Instrument Company, Lafayette, IN) to measure the angle between the fixed center line and the line drawn extending from the second metatarsal (Figure 2). The axis of the goniometer was placed at the bisection of the fixed center line and the second metatarsal line. This was a measure of unilateral turnout. The measurement was taken on each side.

For functional turnout (Figure 3), anti-slip grips were placed on the bottom of a set of Functional Footprints (Balanced Body, Inc., Sacramento, CA) (herein, “footprints”). Footprints were designed as a training tool to help dancers achieve optimum technique. They consist of a resistance-free swivel mechanism that prevents the dancer from using compensation at the foot and ankle to increase turnout. Additionally, they are unstable in the sagittal plane, making it more difficult for dancers to compensate for lack of turnout with lumbopelvic motion. If they attempt anterior tilt to achieve the look of more turnout, the footprints tip forward. The footprints were placed on the whiteboard on pre-marked locations. Dancers used a portable barre for balance while they stepped onto the footprints with the medial malleolus aligned with the horizontal axis line on the footprint, the bisection of the calcaneus aligned with the vertical line at the back of footprint, and the second toe aligned with the bisection line at the front of the footprint. Dancers were asked to place their arms in second position and turnout to first position without tipping the footprints forward. The rater palpated the second metatarsal and used a ruler as a straightedge to draw a line extending from the second metatarsal (Figure 3). The dancer stepped off, and the rater used an extendable goniometer to measure the angle between the fixed center line and the line drawn extending from the second metatarsal (Figure 2). The axis of the goniometer was placed at the bisection of the fixed center line and the second metatarsal line. This was a measure of unilateral turnout. The measurement was taken on each side.

Statistical Analysis

Data were analyzed using Stata, version 13, software (College Station, TX) with an alpha level of 0.05. Intraclass correlations (ICCs) were calculated using multi-level, mixed-effect models, with a correlation of 0.75 being defined as excellent reliability (Cicchetti, 1994). Inter-rater and intra-rater reliability were calculated for active and functional turnout measures using all 3 trials from each rater. Reliability was assessed among all raters (PTs and faculty) and among only the PT raters. ICCs for all raters were derived from a mixed-effects model with fixed effect for type of rater (faculty or PT) and random intercepts for patients, side of patient, and raters. ICCs for PT raters were derived from a mixed-effects model with random intercepts for patients, side of patient, and raters. The difference between the rater's measurements was calculated using the coefficient estimates from the fixed effect for the type of graders in the mixed-effects model. The p values relate to the beta coefficients calculated from the linear mixed-effects regression models.

RESULTS

Thirty-two dance majors aged 18 to 33 years were voluntarily enrolled. One dancer was male and all others were female. They had a mean of 17 years of dance experience (range, 6–29 years). For all situations described above, the intra-class correlations were above the 0.75 threshold for excellent reliability (Table 1). For active turnout, inter-rater reliability was 0.78 among all raters and 0.82 among only the PT raters. The faculty rater's measurements were consistently higher than those of the PT raters by 0.98 degrees ($p = 0.007$). For active turnout, intra-rater reliability was 0.82 among all raters and 0.85 among only the PT raters. For functional turnout, inter-rater reliability was 0.86 among all raters and 0.88 among only the PT raters. There was no significant difference between the faculty rater's measurements and the PT raters' measurements, 0.35 degrees ($p = 0.267$). For functional turnout, intra-rater reliability was 0.87 among all raters and 0.88 among only the PT raters.

DISCUSSION

The technique described for measurement of active and functional turnout has excellent inter-rater and intra-rater reliability when performed by experienced PTs or by a briefly trained university-level dance faculty member. As shown with other goniometric measurements (Boone et al, 1978; Watkins, Riddle, Lamb, and Personius, 1991), intra-rater reliability was slightly higher than inter-rater reliability; therefore, we recommend that measurements be taken by the same rater each time when possible. However, the inter-rater reliability among all raters was also excellent, suggesting that, when necessary, the measurement may be repeated by other trained PTs or university-level dance faculty members. This measurement technique may be used for annual screenings, dance-science research, dance-pedagogy research, and clinical assessment of dancers.

Establishing inter-rater reliability of a measurement technique is important for many reasons. It allows more generalizability of the technique and allows researchers to assume that the measurement is likely to be representative of the subject's true score (Watkins, Riddle, Lamb, and Personius, 1991). Similarly, proving inter-rater reliability of a measurement technique ensures accuracy of the data, which strengthens research conclusions (Watkins, Riddle, Lamb, and Personius, 1991). Ensuring the reliability of turnout measurements is particularly important in dance because of the psychological and emotional effects these measurements may have on dancers (Grossman, 2003). Dancers have been described as perfectionist, achievement-oriented, and frustrated by anatomical deficits such as poor turnout (Hamilton, 1999). Grossman et al. (2008) suggested that reliable, valid turnout measurements enable dancers to learn to use the turnout available to them and avoid the frustration of trying to change something they physically cannot correct. We believe that when providing annual turnout measurements as part of a comprehensive screening program, there is a duty to ensure the reliability of the measurement technique.

In our study, each rater measured a different trial for each dancer. Therefore, there may have been variation in the actual turnout performed with each trial. Variations in actual turnout may have been caused by increased tissue extensibility as the dancers warmed up with repetition, neuromuscular changes caused by a learning effect, fatigue with repetition, or

varying effort given by the dancer for each trial. We attempted to minimize these variations by not allowing dancers to warm up and by providing scripted instruction to the dancers to perform their best during each trial and to avoid trying to outdo themselves during subsequent trials. It should be noted that this instruction may also affect the reliability of the measurement protocol; however, it is likely a negligible effect. The finding that the faculty rater's measurements were higher than those of the PT raters for active turnout but not functional turnout is notable. We suspect that dancers may have put more effort into performing first position in front of their dance instructor than in front of a PT. Another explanation may be that the faculty member had less experience with anatomical palpation than the PT raters, causing error in drawing the line extending from the second metatarsal.

In our study, the dancers were familiar with the measurement technique for functional turnout from past screenings. As a result, they had the motor control to steady the footprints allowing the rater to take an accurate reading. In contrast, those without previous experience with this measurement technique may have difficulty stabilizing the footprints, leading to inconsistent readings. Additionally, the faculty rater was a university-level dance faculty member with some formal training in anatomy. Therefore, the measurement may not have the same reliability when performed by a dance instructor without this knowledge. Additionally, we assessed the reliability of this tool only when used in first position, rather than in all five classical dance positions.

We did not measure dancers' passive or active hip external rotation as other investigators have done (Coplan, 2002; van Merkensteijn and Quin, 2015) because those measures do not take into account the normal contributions of the knee and foot to total turnout. Our focus was to establish a reliable measurement technique to compare the standing first-position turnout that the dancers actually use in class with the turnout they can obtain through the hip, knee, and ankle without using friction from the floor to compensate. Additionally, the unstable footprints (as opposed to stable rotator discs) were used to minimize lumbopelvic compensation. However, it should be noted that this technique likely cannot eliminate all possible compensations.

Future research should examine the difference between active and functional turnout and correlate this difference with injury rates to determine when intervention is warranted. If we know how much of a difference predisposes a dancer to injury, we can provide dancers with valuable information during annual screenings and make better return-to-dance recommendations after an injury. However, noting that a dancer shows less functional turnout than active turnout does not answer the question of why. Does this occur because of a lack of strength to achieve full turnout? Does the dancer lack sufficient passive turnout (from hip, knee, and ankle contributions) to achieve the desired position and, therefore, compensate through the rest of the body? These two scenarios require very different interventions. Grossman (2003) described a method for measuring total passive turnout while the dancer is supine that attempts to account for the contributions of the knee and foot. Using that technique in conjunction with the active and functional turnout measurement techniques described in this study may better answer that question.

CONCLUSION

Standardized measurement tools that are clinically reproducible and reliable will allow for consistency in future dance-medicine and dance-pedagogy studies. Additionally, they are necessary to ensure dancers are receiving accurate information from pre-participation screenings. The techniques described in this paper are easily reproducible and reliable among single or multiple raters. They may be used in future dance research, dancer screenings, and in clinical settings.

ACKNOWLEDGEMENT

We thank the Towson University Dance Department and, specifically, Catherine Horta-Hayden, MFA (Professor of Dance at Towson University), for her participation as the faculty rater. We thank Charlotte Ganuchau (Towson University dance student) and Lauren Friedman (Johns Hopkins Rehabilitation Aide) for their assistance and organization. We thank Jiangxia Wang, MS, MA (Assistant Scientist, Johns Hopkins Biostatistics Center, Johns Hopkins Bloomberg School of Public Health), for assistance with the statistical analysis. We acknowledge support for the statistical analysis from the National Center for Research Resources and the National Center for Advancing Translational Sciences of the National Institutes of Health (grant number 1UL1TR001079).

REFERENCES

- Boone DC, Azen SP, Lin CM, Spence C, Baron C, Lee L 1978 Reliability of goniometric measurements. *Physical Therapy* 58: 1355–1360. [PubMed: 704684]
- Cicchetti DV 1994 Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment* 6: 284–290.
- Coplan JA 2002 Ballet dancer's turnout and its relationship to self-reported injury. *Journal of Orthopaedic and Sports Physical Therapy* 32: 579–584. [PubMed: 12449258]
- Ekegren CL, Quested R, Brodrick A 2014 Injuries in pre-professional ballet dancers: Incidence, characteristics and consequences. *Journal of Science and Medicine in Sport* 17: 271–275. [PubMed: 23988783]
- Gilbert CB, Gross MT, Klug KB 1998 Relationship between hip external rotation and turnout angle for the five classical ballet positions. *Journal of Orthopaedic and Sports Physical Therapy* 27: 339–347. [PubMed: 9580893]
- Grossman G 2003 Measuring dancer's active and passive turnout. *Journal of Dance Medicine Science* 7: 49–55.
- Grossman G, Waninger KN, Voloshin A, Reinus WR, Ross R, Stoltzfus J, Bibalo K 2008 Reliability and validity of goniometric turnout measurements compared with MRI and retro-reflective markers. *Journal of Dance Medicine and Science* 12: 142–152. [PubMed: 19618571]
- Hamilton LH 1999 A psychological profile of the adolescent dancer. *Journal of Dance Medicine and Science* 3: 48–50.
- Liederbach M 1997 Screening for functional capacity in dancers. *Journal of Dance Medicine and Science* 1: 93–106.
- Liederbach M, Dilgen FE, Rose DJ 2008 Incidence of anterior cruciate ligament injuries among elite ballet and modern dancers: A 5-year prospective study. *American Journal of Sports Medicine* 36: 1779–1788. [PubMed: 18753681]
- Negus V, Hopper D, Briffa NK 2005 Associations between turnout and lower extremity injuries in classical ballet dancers. *Journal of Orthopaedic and Sports Physical Therapy* 35: 307–318. [PubMed: 15966542]
- Norkin CC, White DJ 2003 Validity and Reliability. *Measurement of Joint Motion: A Guide to Goniometry*, pp 39–56. Philadelphia, F. A. Davis Company.
- Pata D, Welsh T, Bailey J, Range V 2014 Improving turnout in university dancers. *Journal of Dance Medicine and Science* 18: 169–177. [PubMed: 25474296]

- Portney LG, Watkins MP 2009 Foundations of Clinical Research: Applications to Practice (3rd ed). Upper Saddle River, NJ, Prentice Hall Health.
- Ramkumar PN, Farber J, Arnouk J, Varner KE, McCulloch PC 2016 Injuries in a professional ballet dance company: A 10-year retrospective study. *Journal of Dance Medicine and Science* 20: 30–37. [PubMed: 27025450]
- Steinberg N, Siev-Ner I, Peleg S, Dar G, Masharawi Y, Zeev A, HersHKovitz I 2012 Extrinsic and intrinsic risk factors associated with injuries in young dancers aged 8–16 years. *Journal of Sports Sciences* 30: 485–495. [PubMed: 22288886]
- van Merkensteijn GG, Quin E 2015 Assessment of compensated turnout characteristics and their relationship to injuries in university level modern dancers. *Journal of Dance Medicine and Science* 19: 57–62. [PubMed: 26045396]
- Watkins MA, Riddle DL, Lamb RL, Personius WJ 1991 Reliability of goniometric measurements and visual estimates of knee range of motion obtained in a clinical setting. *Physical Therapy* 71: 90–97. [PubMed: 1989012]
- Welsh TM, Rodriguez M, Beare LW, Barton B, Judge T 2008 Assessing turnout in university dancers. *Journal of Dance Medicine and Science* 12: 136–141. [PubMed: 19618570]

Figure 1.
Active turnout measurement.

Figure 2.
Goniometer placement on board for active and functional turnout measurements.

Figure 3.
Functional turnout measurement.

Table 1.

Inter-rater and intra-rater ICCs of measurements of active and functional turnout of 32 dancers.

Measure by raters ^a	Inter-rater reliability, % (95% CI)	Intra-rater reliability, % (95% CI)	Difference between faculty rater and PT raters	
			Mean degrees (95% CI)	P value
Active turnout				
All raters ^b	0.78 (0.69, 0.85)	0.82 (0.74, 0.88)	0.98 (0.27, 1.7)	0.007
PT raters only ^c	0.82 (0.74, 0.88)	0.85 (0.77, 0.90)		
Functional turnout				
All raters ^b	0.86 (0.79, 0.91)	0.87 (0.81, 0.92)	0.35 (−0.27, 0.96)	0.267
PT raters only ^c	0.88 (0.81, 0.92)	0.88 (0.82, 0.92)		

CI, confidence interval; ICC, intraclass correlation coefficient; PT, physical therapist.

^aRaters were two PTs and one university-level dance faculty member briefly trained in the measurement technique.

^bICCs derived from mixed-effects model with fixed effect for type of rater (faculty or PT) and random intercepts for patients, side of patient, and raters.

^cICCs derived from mixed-effects model with random intercepts for patients, side of patient, and raters.