Montana Tech Library
Digital Commons @ Montana Tech

Safety Health & Industrial Hygiene | Faculty Scholarship

2013

Stairway Step Dimensions: Replication of a Measurement System Study

Christopher Hicks
Montana Tech of the University of Montana

Roger C. Jensen
Montana Tech of the University of Montana

Joselynn M. Adams
Montana Tech of the University of Montana

Follow this and additional works at: http://digitalcommons.mtech.edu/shih

Part of the Occupational Health and Industrial Hygiene Commons

Recommended Citation

Hicks, Christopher; Jensen, Roger C.; and Adams, Joselynn M., "Stairway Step Dimensions: Replication of a Measurement System Study" (2013). *Safety Health & Industrial Hygiene*. 27.
http://digitalcommons.mtech.edu/shih/27

This Conference Proceeding is brought to you for free and open access by the Faculty Scholarship at Digital Commons @ Montana Tech. It has been accepted for inclusion in Safety Health & Industrial Hygiene by an authorized administrator of Digital Commons @ Montana Tech. For more information, please contact sjuskiewicz@mtech.edu.

Stairway Step Dimensions: Replication of a Measurement System Study

Christopher L. Hicks, Roger C. Jensen, Joselynn M. Adams
Montana Tech of the University of Montana, Butte, MT

This paper reports a replication of a prior measurement system study. The earlier study examined the nosing-to-nosing measurement system for measuring steps in a stairway to determine uniformity. In each study, two individuals measured six flights of stairs on two separate occasions. The difference in the first and second study was the different measurers. Step attributes used to define uniformity are riser height and tread depth. The measurers in each study obtained 744 values of riser height and 672 values of tread depth. The ANOVA for each study indicated that less than 4% of the variance in these attributes was due to the measurers; the remainder of variability was due to physical differences in the steps. ANOVA results of this replication led to essentially the same conclusion as the initial study—that the nosing-to-nosing measurement system is acceptable for measuring step dimensions.

INTRODUCTION

Injuries from stairway falls often result in litigation, leading the parties to retain a stairway safety expert. Their investigations include environmental feature, user behavior, and physical characteristics of the stairway. A characteristic regularly examined is step uniformity.

The importance of having uniform step dimensions in flights of stairs has been recognized for quite a while. Summaries of the older studies have been provided by Templer (1992) and Johnson and Pauls (2010). To appreciate why step uniformity is so important, a model of stairway usage is helpful. Archea, Collins, and Stahl (1978) presented a model that helps explain why people tend to misstep on non-uniform steps. According to the model, stair users approach a stairway with an expectation based on their prior experiences using stairs and their visual perception of the stairway ahead. During their first step or two they test that expectation by comparing the kinesthetic, tactile, and visual feedback with their initial expectation. This leads to an adjustment in stepping pattern to match the initial steps. As they proceed, they maintain that stepping pattern while unconsciously assuming that the steps are uniform. If the first step has different dimensions, the user may misstep on the steps next encountered. As they proceed, they do not readily detected steps that differ from the others. When ascending, they can easily catch a toe on the upper edge or nose of a riser. When descending, they can place the ball of their foot too far forward, resulting overstepping or slipping on the nosing. In a paper summarizing findings of in-depth investigations of 80 stairway falls, Cohen, LaRue, and Cohen (2009) concluded that "excessive dimensional variation" within the stairways was a more pervasive factor in stairway falls than individual variables associated with the fall victim.

A system for measuring step dimensions is needed by numerous people. There is the need in the construction industry for a standard method of measuring stairs built in place as well as manufactured stairways installed on site. Building inspectors also need a standard method that yields data suitable for determining compliance with building codes and fire exit codes. Experts in stairway fall injuries also have a need. In civil litigation in the United States, the trial judge is responsible for screening the proposed testimony of experts to ensure it is based on sound science (Daubert v. Merrell Pharmaceuticals, 1993). An expert proposing to offer testimony about their measurements of a stairway needs documentation of the scientific soundness of their measurement system.

The measurement system must be precise because building codes, fire exit codes, and voluntary standards require it. For example, the American National Standards Institute's guidelines for workplace stairs have two types of standards in place (ANSI A1264 Committee, 2007). The first is that for adjacent step risers and tread depths, there should not be a difference greater than 4.8 mm (3/16 inch). The second guideline, for whole flight compliance, specifies that there should not be any difference greater than 9.5 mm (3/8 inch) between any stairs within a flight. Thus, the difference between the shortest riser and the tallest riser should be less than 9.5 mm; and difference between the deepest and shallowest tread should be less than 9.5 mm. Clearly, codes such as these require a precise measurement system.

Measuring dimensional variation in a flight of stairs begins with measuring the riser height and tread depth of each stair. Stairs are traditionally measured using a carpenter square and a ruler. This can be difficult for reasons described by Johnson (2005a). To address this difficulty, Pauls (1998) proposed an alternative method, and Johnson provided a more detailed explanation (Johnson, 2005a, 2005b). These authors called the measurement system the "nosing-to-nosing method." It involves measuring the angle of and the length of the hypotenuse of a right triangle formed by placing a ruler between the leading edge of two adjacent steps. From measurements of the angle (θ) and length of the hypotenuse (H) of the measured right triangle, the lengths of riser height and effective tread depth are calculated using the following trigonometric relationships.

$$\text{Rise} = H \sin \theta$$
$$\text{Depth} = H \cos \theta$$

Copyright 2013 by Human Factors and Ergonomics Society, Inc. All rights reserved. DOI 10.1177/1541931213571123

These computations yield the values for *riser height* and *tread depth* shown in Figure 1. Each dimension should closely match that of the next higher and next lower step within a flight.

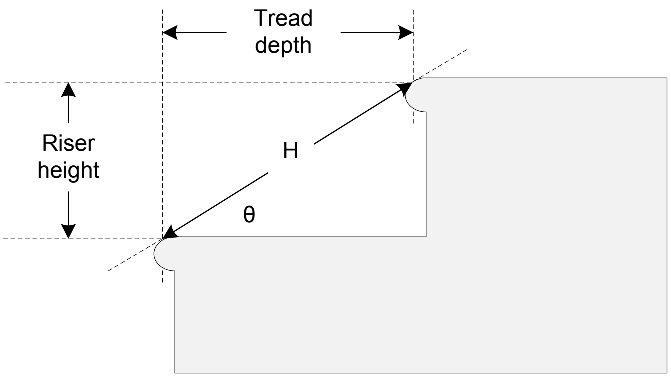

Figure 1. Step dimension and angle determined with the nosing-to-nosing measurement system.

The first of two prior studies of the nosing-to-nosing measurement system had repeated measures by one measurer, of one flight, with one lateral position (Johnson, 2005a). The second study had repeated measures by two measurers, of six flights, with three lateral positions (Jensen, Jensen, & Ross, 2013). Studies with such few participants need follow-up studies to establish the possible generalization of their findings. A replication study is an appropriate way to test the conclusions of a prior study. The purpose of replication studies is to falsify or corroborate the conclusions of an earlier study (Jones, Derby, & Schmidlin, 2010). The particular kind of replication study undertaken for this project was an exact replication in which the same stairways were measured with the same instruments, and in the same order, as the prior study. The difference was different people performed the measurements.

The earlier study by Jensen et al. (2013) examined the contributions to total variability using the measurement system analyses found in Minitab software. Figure 2 depicts how this type of analysis breaks down total variability into finer levels, starting with variation due to step-to-step differences and variation due to the measurement system (the instruments and measurers). The latter consists of an accuracy component and a precision component. The precision component breaks down into repeatability (intra-measurer) and reproducibility (inter-measurer). Minitab and quality control specialists refer to this statistical method as the Gage R&R analysis (Early & Stockhoff, 2010). Some authors use the spelling *gauge* instead of *gage*.

Using this statistical method, the earlier study found that the variability of measurements by two measurers contained contributions from repeatability plus reproducibility (R&R) less than two percent for both step riser height and tread depth. Table 1 provides guidelines from the Automobile Industry Action Group (AIAG) for interpreting results of a Gage R&R experiment (AIAG, 2002).

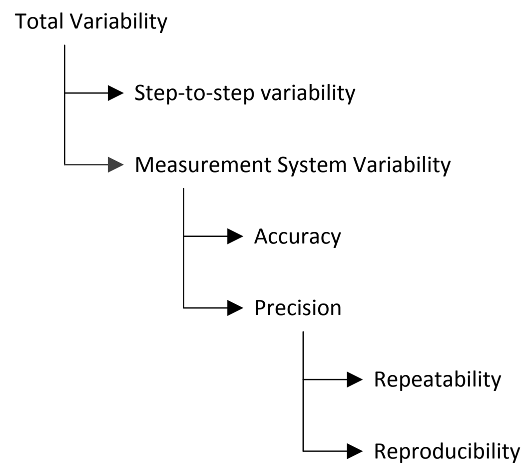

Figure 2. How the Gage R&R method breaks down total variability into components.

Table 1
Guide for Interpreting Results of a Gage R&R Measurement System Study (AIAG, 2002)

R&R Range	Conclusions About Acceptability
0 to 1 %	The measurement system is acceptable.
1 to 9 %	The measurement system is acceptable depending on the application, the cost of the measuring device, cost of repair, or other factors.
> 9 %	The measurement system is unacceptable and should be improved.

This replication study was undertaken for the primary purpose of corroborating or falsifying the findings of the Jensen et al. (2013) study regarding the acceptability of the nosing-to-nosing measurement system.

METHODS

As an exact replication study, we used the same stairways and attempted to use the same methods as the earlier study by Jensen et al. (2013). Because that paper is not easily accessed, and the methods we used were the same, our description of methods is largely identical to that in the earlier paper and included here with their permission.

Experimental Design

The experimental design followed the classic model for a measurement system analysis using Gage R&R ANOVA (Minitab 16, 2012; Hare, 2012). In the quality control environment, two measurers use a gage or other instrument to measure the same batch of parts twice each. This provides data for assessing the consistency of each measurer when repeating a measurement, and the differences in values obtained by one person attempting to reproduce the measurements of the other. In this experiment, the same experimental design was used to measure step dimensions instead of parts. Each measurer measured each step twice, in random order, on two separate days.

Sample of stairways

In the prior study, three older, three-story buildings on campus were selected. From the flights with at least five steps, two flights in each building were randomly selected for study. Each of these flights was measured four times—twice each by two measurers. Table 2 provides basic characteristic of the flights used for the initial and the replication study (Jensen et al., 2013, p. 19).

Table 2
Number of Steps (N) and Basic Characteristics of Sample

Flight	N	Characteristics
1	10	Well-worn terrazzo or granite material
2	11	Steel frame with concrete fill
3	10	Covered with linoleum
4	10	Covered with linoleum
5	13	Painted concrete, very old and worn
6	8	Wood covered with well-worn, thin carpet

Instrumentation

The measurer used a carpenters steel retractable tape measure to measure step width and to determine the lateral points for three measurement locations. A carpenter's chalk line was used to mark three lines from the top to bottom of the flight. A stainless steel ruler with millimeter markings was used to measure the length between the nosings of adjacent steps. A SmartTool^{TM} was used to measure the angle as shown in Figure 1 and it was calibrated before each use according to the owner's manual.

Procedures

Measurements of each flight began by determining the step width. The total width was measured for the narrowest part of the flight. If a handrail was present, the inside surface of the handrail defined the applicable edge. Three lateral points were identified.

- Center point, measured equal distance from the two edges.
- Left point (viewed from bottom of flight) measured 406 mm (16 inches) from the left edge.
- Right point (viewed from bottom of flight) measured 406 mm (16 inches) from the right edge.

The rationale for using 406 mm was that the most worn locations on a flight of steps have somewhat different characteristics than the center location, and the most worn locations occur where pedestrians walk. The following logic was used to estimate these higher-use locations. A pedestrian is forced to walk a path between any handrails or other projections from the sides. The center of that path may be estimated from two parameters: the width of human bodies and spacing between the body and the guardrail, handrail, or wall. Anthropometric data from the U. S. Air Force, as reported by Kroemer and Grandjean (2001) in their Table 4.1, lists the 50 percentile shoulder breadth for males at 491 mm and women at 431 mm. A midpoint of 461 mm was used to represent the mixed population of stair users. The shoulder-to-shoulder distance was halved to approximate the mid-sagittal plane of the body (230 mm). Typically, people keep a distance between themselves and a guardrail, handrail, or wall. That spacing was estimated to be 175 mm. The sum of these two values (406 mm or 16 inches), provided an approximation of the distance of the body center plane from the guardrail, handrail, or wall for a diverse range of pedestrians on the campus.

To make the measurements, a measurer and a recorder were present. The recorder had the list of points to measure, and the random order for the measurements. The recorder informed the measurer which point to measure, and subsequently recorded the measured values of hypotenuse length and angle. Thus, a flight with ten steps required thirty measurements taken in a random order.

Both measurers completed measurements of all six flights. On a later date, each measurer repeated the entire process; including marking the three lateral points and making the measurements. The reason for spacing the two measurements was to avoid memory influencing the second measurement, thereby meeting the ANOVA assumption of independence.

Analyses

From the measured data, the height of each rise and length of each tread was calculated from the trigonometric relationships. Using these values, a Gage R&R ANOVA procedure in the Minitab statistical software suite was used to determine the percentage contribution to total variability of the step dimensions, the measurers, and interaction of the two.

Outputs of the Gage R&R analyses apportion total variability to repeatability, reproducibility, and part-to-part. Repeatability refers to variations attributed to differences in the individual's first and second measurements of the steps. Reproducibility is the variance resulting from the attempts of two measurers to measure the same step. Part-to-part variability in the Gage R&R output means step-to-step variability for this study. It is the physical variations among the dimensions of the stairs measured.

RESULTS

The measurers in each study obtained 744 values of step rise and 672 values of tread depth. These values were analyzed in the same way as the earlier study in order to facilitate comparisons. The Gage R&R ANOVA provided the results displayed in Tables 3 and Table 4. Total variability is apportioned to three factors: total R&R, repeatability, reproducibility, and step-to-step differences.

For riser measurements, the data in Table 3 indicate the measurers in the two studies accounted for 1.42% and 3.82%, respectively. According to AIAG guidelines, both R&R values are in the category "acceptable depending on the application, the cost of the measuring device, cost of repair, or other factors." Of the total R&R in each study, repeatability contributed much more to variation than did reproducibility.

Table 3
Rise Variability from Initial and Replication Study[a]

Source of Variability	Initial Study (%)	Replication Study (%)
Total Gage R&R	1.42	3.82
Repeatability	1.30	3.62
Reproducibility	0.12	0.21
Step-to-Step	98.58	96.18
Total Variation	100	100

[a]Degrees of freedom = 743

For the tread depth measurement, the data in Table 4 indicate the variability contributed by the measurers in the two studies accounted for 0.50% and 1.76%, respectively. According to AIAG guidelines, the R&R contribution to variability of the first study was in the "acceptable" region, and the second study was in the "acceptable depending on the application, the cost of measuring, cost of repair, or other factors" region. Similar to the data in Table 3, repeatability contributed much more than reproducibility.

Table 4
Depth Variability from Initial and Replication Study[a]

Source of Variability	Initial Study (%)	Replication Study (%)
Total Gage R&R	0.50	1.76
Repeatability	0.42	1.49
Reproducibility	0.07	0.26
Step-to-Step	99.50	98.24
Total Variation	100	100

[a]Degrees of freedom = 671

Further analyses provided by the Gage R&R ANOVA indicated the extent to which measured dimensions can be explained by a two factor linear model with interaction. For the riser measurements, the data in Table 5 indicate that the both the step-to-step and the measurer factors contributed significantly to the riser height dimensions (p = .000). The step*measurer interaction terms in the two studies had similar p values (0.102 and 0.051). An inspection of graphs showing measurements for all steps revealed that the interaction occurred primarily on the bottom riser.

Table 5
Contributions to Significance of Variance for Rise (p-values)

Source	Initial Study	Replication Study
Step	0.000	0.000
Measurer	0.000	0.000
Step*Measurer	0.102	0.051

For the tread depth measurements, the data in Table 6 indicate significant contributions to variance from the steps in both studies (p = .000). In the initial study the measurers accounted for a significant amount of variance (p = .018). In contrast, the replication study did not show a significant contribution from measurers (p = .469). The step*measurer interaction did not contribute significantly to the tread depth measurements.

Table 6
Contributions to Significance of Variance for Depth (p-values)

Source	Initial Study	Replication Study
Step	0.000	0.000
Measurer	0.018	0.469
Step*Measurer	0.123	0.194

DISCUSSION

The replication study was compared to the initial study to further analyze the nosing-to-nosing measurement system. The measurements were performed four years apart using the same stairways and instruments. A difference between the studies was that flights 3 and 4 were involved in a remodeling project within the four year span between studies. The old linoleum was replaced with new.

The findings of this experiment corroborate those of the study by Jensen et al. (2013). For both studies, the Gage R&R statistical analysis indicated that less than 4% of variability came from R&R in the rise height and tread depth measurements. Looking into the data further indicates the initial study showed that R&R contributed less than 2% of the variability for both rise and depth. The replication study had larger R&R for both dimensions but still less than 4%. Analyses showed that the total R&R contributions to variance were clearly less than 9% in both studies for rise and depth measurements. Using the AIAG criteria, this indicates that the measurement system is "acceptable depending on the application, the cost of measuring device, cost of repair, or other factors." Considering all these factors, we are of the opinion that the measurement system is acceptable for measuring step dimensions. It yields variance values considerable below the 9% ceiling for the middle acceptable range, and the costs required to use the systems are very low.

The Gage R&R ANOVA outputs also facilitated insight as to what factors affected measured dimensions. For riser height, measurers were significant influences in both studies. For tread depth, the measurer factor was significant in the initial study but not in the replication study. The differences between the two studies are not large. Observers in the replication study may not have been as precise in measurements, which could be a simple explanation for the minor differences.

Selecting the lateral position for measurements is an important decision for three reasons. First, it should be where people commonly walk as they ascend and descend the stairs. Simply picking a point, such as the center point, is arbitrary and not particularly relevant to the matter of stair safety. Second, if the measurements are intended for litigation, reliability of the measurement system is required (Daubert *v.* Merrell, 1993). A measurement that is reliable should be reproducible. Thus, if the plaintiff's expert measures the stairs, and the defendant's expert attempts to reproduce the measurements on a different day, they should obtain close to the same results. Discrepancies in results can easily arise if the two experts do not mark the identical lateral points on the stairway. This is especially true for older, well-worn stairways. We found that some of the older stairs had damaged nosings. Measuring a centimeter to the left or right could produce very different results and conclusions about compliance. Thus, developing a standardize procedure for precisely marking the lateral position of measuring points would improve reproducibility.

A limitation to the studies was that lighting available when performing measurements was poor in some locations. This could have contributed to imprecise reading on the ruler scale. In hindsight, it would have been better to have taken a portable light along with the other equipment for these situations. Another limitation of the studies was that we cannot determine how much variability could be attributed to remarking the three lateral points each time a flight was measured.

Three recommendations for future studies are offered. First, a study of multiple measurers using the same lateral points on selected flights could provide R&R variability percentages free of that factor. Second, studies are recommended directly addressing the related application of this measurement system for determining if adjacent-step differences comply with standards. A third recommendation for future research is to conduct replication studies measuring different flights of stairs using the same measurement system. Like other replication studies, the purpose would be to confirm or falsify the conclusions of a prior study. All these studies would have the potential to extend our understanding the judicially-required scientific soundness of the nosing-to-nosing measurement system.

ACKNOWLEDGMENTS

The authors of this paper were partially supported by Training Grant No. T03/OH008630 from the Centers for Disease Control and Prevention/National Institute for Occupational Safety and Health. The contents are solely the responsibility of the author and do not necessarily represent the official views of the National Institute for Occupational Safety and Health.

REFERENCES

ANSI A1264 Committee. (2007). American national standard: safety requirements for workplace walking/working surfaces and their access: Workplace, floor, wall and roof openings; Stairs and guardrails systems, ANSI A1264.1–2007. Des Plaines, IL: American Society of Safety Engineers.

Automotive Industry Action Group (AIAG). (2002). *Measurement system analysis reference manual*. Chrysler, Ford, General Motors Supplier Quality Requirements Task Force.

Archea, J., Collins, B., & Stahl, F. (1979). *Guidelines for stair safety* (NBS Building Science Series 120). Gaithersburg, MD: National Institute of Standards and Technology.

Cohen, J., LaRue, C. A., & Cohen, H. H. (2009). Stairway falls: An ergonomics analysis of 80 cases. *Professional Safety*, 54(1), 27–32.

Daubert *v*. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Early, J. F., & Stockhoff, B. A. (2010). Accurate and reliable measurement systems and advanced tools. In: K. Juran & J. De Feo (Eds.), *Juran's quality handbook: The complete guide to performance excellence*, (6th ed), (p. 598). New York: McGraw-Hill.

Hare, L. B. (2012). Gage R&R reminders: Running gage repeatability and reproducibility studies properly. *Quality Progress*, 45, 2, 62–64.

Jensen, L. S., Jensen, R. C. & Ross, C.E. (2013, June). A measurement system experiment to evaluate the nosing-to-nosing method for measuring dimensions of steps. *Proceedins: The XXV Annual Occupational Ergonomics and Safety Conference*, Atlanta, GA, pp. 17–22 [CD-ROM]. International Society for Occupational Ergonomics and Safety (www.isoes.info).

Johnson, D. A. (2005a). Error in stair measurements. *Ergonomics in Design*, 13, 2, 18–22.

Johnson, D. A. (2005b). Measurement in pedestrian falls. In: Y. I. Noy & W. Karwowski (Eds.), *Handbook of human factors in litigation* (pp. 20.14–20.18). Boca Raton, FL: CRC.

Johnson, D. A., & Pauls, J. (2010). Systematic stair step geometry defects, increased injuries, and public health plus regulatory responses. In M. Anderson (Ed.), *Contemporary ergonomics and human factors 2010* (pp. 453–461). London: CRC.

Jones, K., Derby, P., & Schmidlin, E. (2010). An investigation of the prevalence of replication research in human factors. *Human Factors*, *52*, 586-595. doi: 10.1177/0018720810384394

Kroemer, K. H. E., & Grandjean, E, (2001). *Fitting the task to the human: A textbook of occupational ergonomics* (5th ed.). Philadelphia: Taylor & Francis.

Minitab 16. Help menu: Statguide, Quality Tools, Gage Study, Gage R&R Study (Crossed). Also see overview at http://en.wikipedia.org/wiki/ANOVA_gauge_R%26R.

Pauls, J. (1998). Techniques for evaluating three key environmental factors in stairway-related falls. In *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (p. 1630). Santa Monica, CA: Human Factors and Ergonomics Society.

Templer, J. (1992). *The staircase: studies of hazards, falls and safer design*. Cambridge, MA: MIT Press.