2013

A Measurement System Experiment to Evaluate the Nosing-to-nosing Method for Measuring Dimensions of Steps

Lea S. Jensen
Liberty Mutual Insurance Group

Roger C. Jensen
Montana Tech of the University of Montana

Craig E. Ross
Montana Tech of the University of Montana

Recommended Citation

Jensen, Lea S.; Jensen, Roger C.; and Ross, Craig E., "A Measurement System Experiment to Evaluate the Nosing-to-nosing Method for Measuring Dimensions of Steps" (2013). *Safety Health & Industrial Hygiene.* Paper 5.
http://digitalcommons.mtech.edu/shih/5

The XXV Annual Occupational Ergonomics and Safety Conference
Atlanta, GA, USA
June 6-7, 2013

A Measurement System Experiment to Evaluate the Nosing-to-nosing Method for Measuring Dimensions of Steps

Lea S. Jensen[1], Roger C. Jensen[2], and Craig E. Ross[3]

[1]Liberty Mutual Insurance Company, Portland, OR

[2]Montana Tech of the University of Montana

[3]Montana Tech Alumnus

Corresponding author's Email: rjensen@mtech.edu

Author Note: This study was part of Lea Jensen's project for her M.S. thesis in Industrial Hygiene. Craig Ross participated and used the project for a senior project report for a B.S. in Occupational Safety and Health. Professor Roger Jensen was the faculty mentor. Partial salary support for Professor Jensen and tuition support for Lea and Craig came through a grant from the National Institute for Occupational Safety and Health. For further information contact either Professor Jensen at the above email address, or Lea at Lea.Jensen@LibertyMutual.com.

Abstract: Non-uniformity of steps within a flight is a major risk factor for falls. Guidelines and requirements for uniformity of step risers and tread depths assume the measurement system provides precise dimensional values. The state-of-the-art measurement system is a relatively new method, known as the nosing-to-nosing method. It involves measuring the distance between the noses of adjacent steps and the angle formed with the horizontal. From these measurements, the effective riser height and tread depth are calculated. This study was undertaken for the purpose of evaluating the measurement system to determine how much of total measurement variability comes from the step variations versus that due to repeatability and reproducibility (R&R) associated with the measurers. Using an experimental design quality control professionals call a measurement system experiment, two measurers measured all steps in six randomly selected flights, and repeated the process on a subsequent day. After marking each step in a flight in three lateral places (left, center, and right), the measurers took their measurement. This process yielded 774 values of riser height and 672 values of tread depth. Results of applying the Gage R&R ANOVA procedure in Minitab software indicated that the R&R contribution to riser height variability was 1.42%; and to tread depth was 0.50%. All remaining variability was attributed to actual step-to-step differences. These results may be compared with guidelines used in the automobile industry for measurement systems that consider R&R less than 1% as an acceptable measurement system; and R&R between 1% and 9% as acceptable depending on the application, the cost of the measuring device, cost of repair, or other factors.

Keywords: Stairs, Safety, Dimensions, Measurement system, Tread depth, Riser height

1. INTRODUCTION

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

Previous research established that non-uniform steps within a flight increase risk of missteps. Summaries of these 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. The person adopts that pattern and proceeds while unconsciously assuming the steps are uniform. As they proceed up or down the flight, they do not readily detected steps that differ from the others, so they do not adjust their stepping to accommodate non-uniform steps. 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. The significance of non-uniform steps relative to other risk factors for stairway falls was summarized in a paper by Cohen, LaRue, and Cohen (2009) in which they summarized findings from in-depth investigations of 80 stairway falls. They concluded that the most pervasive factor in stairway falls was not the individual variables associated with the fall victim; rather, it was the "excessive dimensional variation" within the stairways.

The measurement of step uniformity starts with measuring the rise height and tread depth of each step in a flight. 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 guideline applies to adjacent step risers and treads. It provides that the difference between adjacent step risers should not exceed 4.8 mm (3/16 inch), and the same limit applies to differences in tread depth. The second guideline applies to the whole flight. It provides 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. These relatively small dimensions 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. The traditional method of using a carpenter square and a ruler has several shortcomings (see Johnson, 2005a). To address these shortcomings, 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 two parameters: the angle (θ) and length of the hypotenuse (H) of the right triangle depicted in Figure 1. The lengths of riser height and effective tread depth are calculated using the following trigonometric relationships.

$$\text{Riser Height} = H \sin\theta \qquad (1)$$

$$\text{Tread Depth} = H \cos\theta \qquad (2)$$

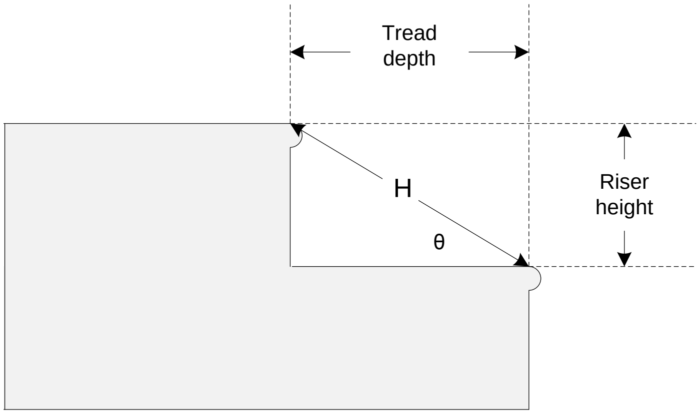

Figure 1. Step dimensions and angle used with the nosing-to-nosing measurement system.

A prior study of the nosing-to-nosing measurement system had repeated measures by one measurer, of one flight, with one lateral position (Johnson 2005a). This study was undertaken in learn more about the measurement system when used by different individuals and applied to a more diverse sample of steps. More specifically, the purpose of the study was to evaluate the measurement system to determine how much of total measurement variability comes from the step variations versus that due to repeatability and reproducibility (R&R) associated with the measurers.

The purpose was accomplished by using a measurement system experimental design often used by quality control professionals (Minitab 16, 2012; Early and Stockhoff, 2010). This experimental design has two or more measurers use a gage or other instrument to measure the same batch of parts at least two times 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. The analysis of these data apportions the total measurement variation into factors as indicated in Figure 2. One factor is the actual physical variation in the items measured, and the other is the variation from the measurement system. Measurement system variation has an accuracy component (how close the measured mean is to a true mean) and a precision component (how small is the variation among measured values). Precision is further apportioned into that due to intra-measurer repeatability and inter-measurer reproducibility.

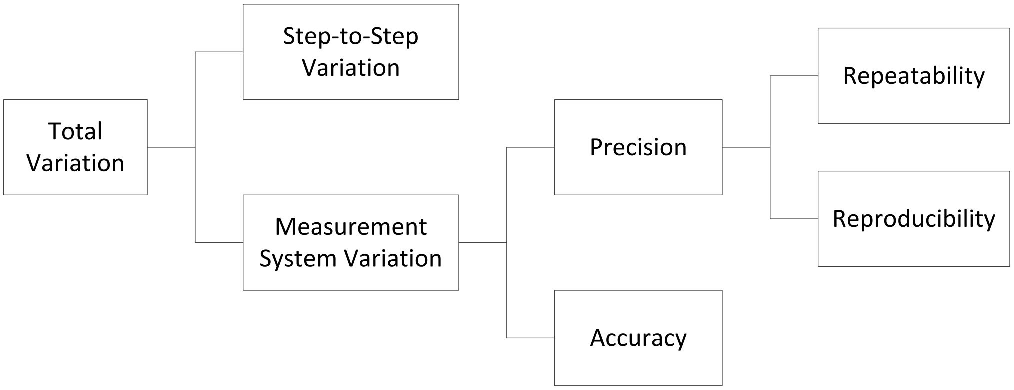

Figure 2. Breakdown of total variance into components. Adapted from Hare, Quality Progress, 2012.

In order to interpret the results, guidelines adopted by the Automobile Industry Action Group (AIAG) were considered (AIAG, 2002). These guidelines apply to the R&R contribution to total variation as follows.

Less than 1% — the measurement system is acceptable.

Between 1% and 9% — the measurement system is acceptable depending on the application, the cost of the measuring device, cost of repair, or other factors.

Greater than 9% — the measurement system is unacceptable and should be improved.

2. METHODS

2.1 Experimental Design

The experimental design followed the classic model for a measurement system analysis using Gage R&R ANOVA. For this experiment, the investigators used this design to measure step dimensions instead of parts. Each measurer measured each step twice, on two separate days.

2.2 Sample of Stairways

The investigators selected three older buildings on the campus. Each was at least three stories. 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 1 provides basic characteristic of the flights used for the initial and the replication study.

Table 1. 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

2.3 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 H, and a SmartTool^{TM} was used to measure θ. The SmartTool™ was calibrated before each use according to the product owner's manual.

2.4 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 H and θ values. Thus, a flight with ten steps required thirty, randomly-ordered measurements. For the bottom step, the depth was set at approximately that same as the typical steps in that flight, but the resulting depth value was only use for the calculations in equations 1 and 2. Thus, one complete measurements of a ten step flight by one measurer yielded for statistical analyses 30 values for riser height and 27 values for tread depth.

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 making it reasonable to assume the two measurements were independent.

2.5 Analyses

From the measured data, the height of each rise and length of each tread depth was calculated from equations 1 and 2. Using these values, a Gage R&R ANOVA procedure (crossed option) in the Minitab statistical software suite was used for analyses. One analysis was to fit the measured values with a two factor linear regression model with interactions. The second analysis apportioned total variability to repeatability, reproducibility, and part-to-part (see Figure 2). Repeatability refers to variations attributed to differences in the individual's first and second measurements of the steps, i.e., intra-individual variability. Reproducibility is the variance component resulting from the attempts of two measurers to measure the same thing, i.e., inter-individual variability. Part-to-part variability in the Gage R&R output means step-to-step variability for this study. It is the physical variations in the dimensions of the stairs measured.

3. RESULTS

The measurers provided for analyses 744 values of riser height and 672 values of tread depth. Two related analyses were used to examine these data sets.

The initial analysis 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. Table 2 indicates the significance level of each term. Both the step and the measurer factors contributed significantly ($p < 0.05$) to the measured riser height dimensions and to the tread depth dimension. The step*measurer interaction terms had similar, non-significant p values (0.123 and 0.102). An inspection of graphs showing measurements for all steps suggested some interaction occurred on the bottom riser of some flights.

Table 2. Factor Significance Levels (*p*-values)

Source	Rise	Depth
Step	0.000	0.000
Measurer	0.018	0.000
Step*Measurer	0.123	0.102

The second analysis examined the contributions to total variance as depicted in Figure 2. The Gage R&R analysis provided the results displayed in Table 3. The first row shows the Total Gage R&R, while the second and third rows show the two components of the Gage R&R. The fourth row shows the percentage of total variability attributed to differences in the actual step dimensions.

Table 3. Variability in Measured Values Apportioned Among Sources[a]

Source of Variability	Rise	Depth
Total Gage R&R	1.42	0.50
Repeatability	1.30	0.42
Reproducibility	0.12	0.07
Step-to-Step	98.58	99.50
Total Variation	100	100

[a]Degrees of freedom = 743 for rise and 671 for depth

For riser height measurements, the data in Table 3 indicate the measurers accounted for 1.42% of the measured values. According to AIAG guidelines, the R&R values are in the category “acceptable depending on the application, the cost of the measuring device, cost of repair, or other factors.” For the tread depth measurements, the data in Table 3 indicate the variability contributed by the measurers accounted for 0.50%. According to AIAG guidelines, the R&R values are in the “acceptable” region.

4. DISCUSSION

The Gage R&R outputs for the initial ANOVA two-factor model indicated the role of measurers to total variability was statistically significant for both riser height and tread depth. The conclusion from this is that the measurement system has a role that should not be ignored; however, that role is much smaller than the actual physical variation among the steps. The second analysis provided a better picture of the importance of the measurer contributions. For riser height, measurers contributed 1.42% of total variability. For tread depth the measurers contributed only 0.5% to the total. Both percentages were much less than 9%. Using the AIAG criteria, this indicates that the measurement system is “acceptable depending on the application, the cost of measuring, cost of repair, or other factors.”

Two limitations need acknowledgment. The first is the process of marking three lateral points on each flight on different days probably contributed to some of the variation. Although specific guidelines were used for each marking, small differences no doubt occurred. The second is the sample studied involved only two measurers and six flights of stairs. Thus, while the authors view the findings as supporting the nosing-to-nosing measurement system, caution about generalizing this conclusion is advised.

Three recommendations for future studies are offered. Replications of this study would help clear up if the results presented here are unique to the two measurers, or if similar results would be obtained by other measurers. There are three types of replication studies (Jones, K., Derby, P., & Schmidlin, 2010). The first type recommended is an exact replication in which two other measurers perform the same experiment using the same sample of stairways. The second recommended type of replication is to have other measurers perform a similar experiment using different stairways. Having two or more individuals use the same lateral points on selected flights could provide R&R variability percentages free of the lateral-point marking factor. Third, studies are recommended directly addressing the related application of this measurement system for determining if adjacent-step differences comply with standards. A fourth recommendation for future research is for others 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 of the scientific soundness of the nosing-to-nosing measurement system.

4. 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.

Automotive Industry Action Group (AIAG). (2002). *Measurement system analysis (3rd ed.). AIAG: Detroit, MI.*

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.

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. and 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.

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

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.