15th Canadian Masonry Symposium
Ottawa, Canada
June 2-5, 2025

Evaluation of Flexural Bond Strength in a Masonry Lab with a Statistical Comparison of Beam versus Bond Wrench Test Methods

Cesar Gerardo Freyre-Pinto[i] and Jennifer E. Tanner[ii]

ABSTRACT

The University of Wyoming masonry design class has a robust laboratory component. Groups of students construct 6-unit stack brick masonry prisms A suite of three specimens were subjected to a 4-point beam test as described in ASTM E518. These tests fail in the center with two 3-unit brick prisms remaining. Both prisms are placed within a bond wrench apparatus in accordance with ASTM C1072. Each prism provides another four measurements of bond strength (two per prism half). Test data indicates that the beam test (ASTM E518) is higher than the average measured by the bond wrench test (ASTM C1072). A statistical analysis was performed to verify the significance of this difference. Although the p-value was higher than the significance level (T(2)=2.05, p-value=0.088), the statistical effect size was found to be of practical importance (d=1.19).

KEYWORDS

Tensile bond strength, brick prism, beam test, bond wrench test, statistical evaluation

[i] Research assistant, University of Wyoming, Laramie, WY, USA, cfreyrep@uwyo.edu
[ii] Professor, University of Wyoming, Laramie, WY, USA, tannerj@uwyo.edu

INTRODUCTION

Flexural bond strength is an important aspect of masonry design. This can be measured by using either a beam test or a bond wrench apparatus. The fall 2024 Reinforced Masonry Design class at the University of University of Wyoming Masonry Design class includes a laboratory session where students build 6-unit stack bond masonry prisms. Specimens are tested using a beam configuration based on ASTM E518. After the beam specimen fails at the center, the remaining two portions are subjected to bond wrench testing in accordance with ASTM C1072. Results are evaluated based on statistics.

Teaching students through hands on activities affords an understanding of behavior in masonry. This process complements a course where students are using design values in the TMS Masonry Code. For example, the flexural bond strength is frequently used to evaluate unreinforced masonry. This paper presents results of a laboratory session where students construct, test brick prisms using two different methods, and compare the results.

The first laboratory session required students to batch, mix, and perform a flow test on masonry prisms. Next, pairs of students used type N mortar to construct prisms. A jig was also used to ensure that the specimens were plumb. Each specimen was wrapped in a plastic bag and tightened down to promote curing during the 28 days before testing. A photo of the specimens immediately after construction are illustrated in Figure 1. Photos of the beam and bond wrench test setups are also shown in Figure 1.

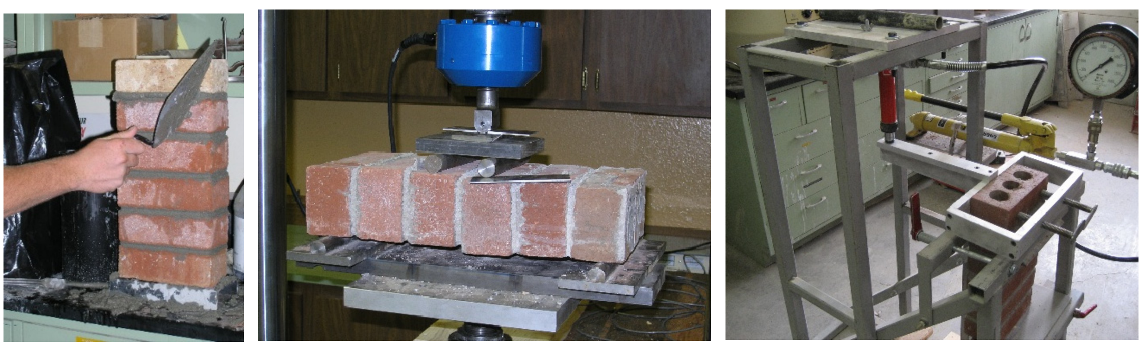

Figure 1: Example of a test specimens built and prior to applying load

Figure 2: Examples of final failures for beam and bond wrench testing procedures

BACKGROUND

Flexural Bond Strength of Masonry

A variety of bond-wrench studies have been conducted over the years. In one study, a suite of bond wrench testing evaluated three different mortars across a total of 15 laboratories. Individual laboratory results indicated that coefficients of variation (COVs) ranged from 15 to 22%; however, repeatability and reproducibility among the different laboratories had higher variability [1]. McGinley reported that stress distributions may not be linear in this testing method. This study recommends clamping as far away from the mortar joint as possible. Sarangapani et al. [2] used a modified loading system to clamp at the base and apply uniform bending to permit failure along a suite of joints in a taller specimen.

Gaggero and Esposito [3] report that COV's can range from 15 to 50% for lower strength mortars. In general, results vary based on surface texture and moisture content of specimen. This is corroborated by Sarangapani et al [2] and Madhada et al. [4]. Kestemont [5] observed that roughly 1/3 of the specimens failed during the process of placing the specimens within the test setup and clamping process.

Nichols et al. [6] investigated a few different modifications to the bond wrench test methods and reports lower bond strengths for the bond wrench testing apparatus versus beam testing.

Statistical Analyses

Flexural bond strength can be thought of as a random variable in the sense that, every time it is measured, different values will be obtained. This stems from the fact that, both brick and mortar, have inherent material variability, there may be variability in testing conditions and, in general, there will always be measurement uncertainties.

The occurrence of random variables can be described through probability distributions such as the widely used normal (gaussian) distribution [7] (Figure 3). When the distribution is used to describe a continuous variable, it is called a probability density function (PDF). Figure 3 shows an example of a normal distribution describing the probability density of the flexural bond strength with a mean value defined as μ and standard deviation σ. For PDF's, the area under the curve in between two value points a and b, is equal to the probability of the variable f_r being between those values $P(a \leq f_r \leq b)$. Similarly, the probability of the variable f_r being greater than a certain value point c $P(f_r \geq c)$ is equal to the area under the right tail of the distribution (Figure 1). It is worth noting that any normal distribution can be transformed into a standard normal distribution, which is a normal distribution with a mean of 0 and standard deviation of 1.

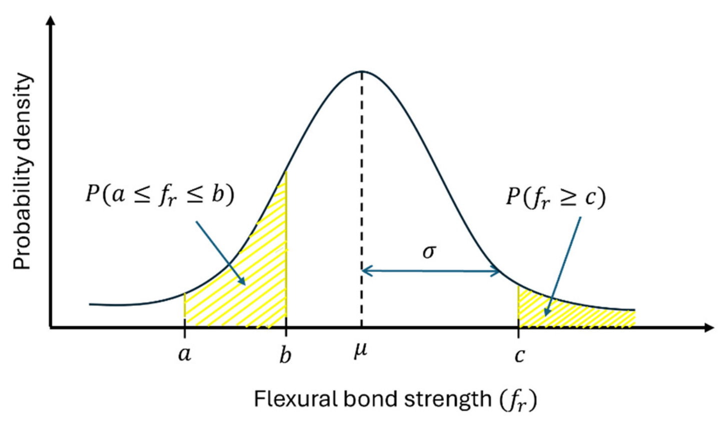

Figure 3: Normal probability distribution for flexural bond strength

When it is desired to compare a sample average with a specific value, a t-statistic is used. The equation for the t-statistic is given by Eq. (1), where $\bar{x}$ is the sample average, s is the sample standard deviation, μ is the true population average, and n is the sample size.

$$(1)\quad T = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

This statistic follows the Student's t-distribution [8] (Figure 4). In the definition of the pdf for the t distribution, there is a parameter called statistical degrees of freedom (denoted St-DOFs), this can be thought of as the available data points to define the value of the statistic. The value of the statistical degrees of freedom (St-DOFs) is one unit less than the sample size ($n-1$). As the statistical degrees of freedom (or the sample size) increases, the t distribution approaches a standard normal distribution. It is good practice to report the t-statistic with its degrees of freedom as T(St-DOFs)=a. This way, any reader can verify the statistical analysis.

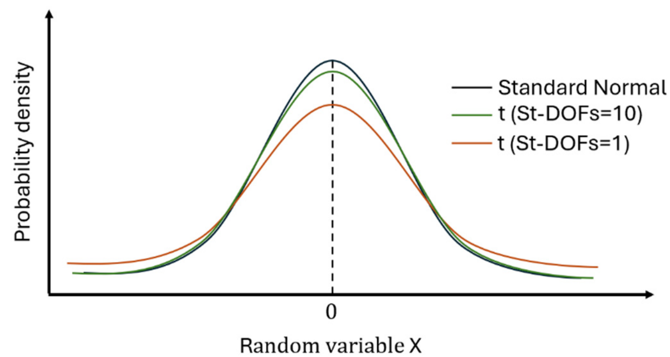

Figure 4: Student's t distribution with different degrees of freedom

Under the classical Null Hypothesis Significance Testing (NHST) the sample average (through the t-statistic) is used to make inferences about the true population average. A null hypothesis is proposed ($\mu = \mu_0$) and this is assumed to be true. This defines a specific distribution. An observed value of the t-statistic can be calculated from sample data and the assumption of the null hypothesis. Then, the probability of obtaining a value of the t-statistic equal or more extreme than the one observed is calculated from the area under the distribution (Figure 5), this value is known as the p-value. It is important to emphasize that this assumes that the null hypothesis is true.

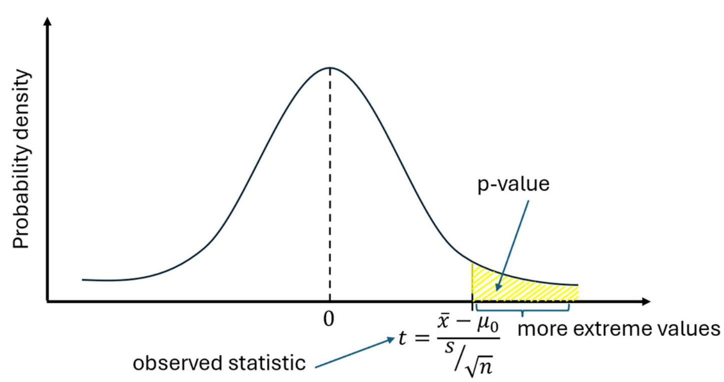

Figure 5: Definition of the p-value

If this probability (p-value) is less than the significance value α (typically $\alpha = 0.05$), this means that the null hypothesis should be rejected (very low probability of observing the data if the null was true). On the other hand, if p-value > α, this means that the null hypothesis cannot be rejected (there is a considerable probability of observing the data if the null hypothesis was true). This process is called a one sample t-test.

There are two choices when comparing samples to a reference value. First, if the two samples are independent, we use an independent samples t-test. Second, if the two samples are not independent, a paired t-test is a better fit. In this type of test, the difference between corresponding data points is calculated and the regular one sample t-test is performed on this difference. The null hypothesis in this case is $\mu = 0$. This means that if the p-value is low enough (<0.005), the null hypothesis is rejected ($\mu \neq 0$) and this is taken as evidence that the average difference in the samples is statistically significant.

Finally, when more than two laboratory groups are present in a study, an analysis of variance (ANOVA) is the appropriate tool to use. This is also an NHST test where the population variance is estimated in two ways. In broad terms, the method calculates the variance of the means (variability between groups $MS_{between}$) and the mean of the variances (variability within groups MS_{within}). The ratio of these two estimates of the variance is called an F-statistic. The F-statistic follows a Fisher's F distribution [9] (Figure 6). This distribution has two parameters named degrees of freedom as well (df1 and df2). The first degree of freedom is the total sample size minus the number of groups ($N - a$), and the second degree of freedom is the number of groups minus one ($a - 1$). Again, it is good practice to report the F-statistic with its statistical degrees of freedom as F(df1, df2)=b. Finally, a p-value is drawn from the area under the F distribution. As usual, a lower p-value is the probability of finding a value of the statistic equal or more extreme than the one observed. If the p-value is below the significance level (α=0.05), the null hypothesis is rejected. For the case of ANOVA, the null hypothesis states that the means of all groups is the same (i.e. they come from the same population).

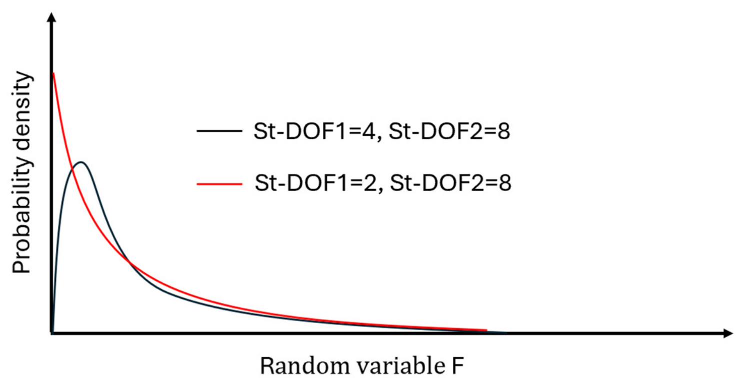

Figure 6: Fisher's F distribution for different degrees of freedom

Parallel to all the NHST framework, there is another important statistical tool that provides insight into data, the effect size. The p-value can answer the question: is there a statistically significant difference? In this case, is there a statistically significant difference between beam and bond wrench test data? On the other hand, effect size gives a measurement of a practical implication that is discipline specific. A widely used measurement of the effect size is Cohen's d (Eq. 2).

$$(2)\quad d = \frac{\bar{x} - \mu}{s}$$

The conventional standards [10] for determining a small, medium or large effect are: d=0.2 small, d=0.5 medium, d≥0.8 large. In this sense, even if the p-value (which depends on sample size) suggests that the difference between averages is not statistically significant, the effect size (independent of sample size) can reflect that there is a difference of practical importance.

A summary of statistical notation is included because the variables are similar to those used in structural engineering applications.

Table 1: Statistical variables

Variables	Interpretation
μ	True population mean
σ	True population standard deviation
s	Sample standard deviation
$\bar{x}$	Sample average (sample mean)
t	T-statistic
d	Statistical effect size

RESULTS

In this section, the results for the flexural tensile strength of masonry are presented. For clarity, this paper differentiates between results of a beam test and those of a bond wrench test.

Bond Wrench

The results for each of the three groups are presented in Table 2.

Table 2: Results of bond wrench tests for the three groups in psi (MPa)

Group 1	Group 2	Group 3
166 (1.14)	140 (0.96)	152 (1.05)
110 (0.77)	102 (0.70)	64.0 (0.44)
71.9 (0.50)	114 (0.78)	111 (0.77)
95.6 (0.66)	-	95.6 (0.66)

Table 3 shows the summary statistics for the results of bond wrench tests. To be able to treat the data coming from the three groups as just a single set of data, an analysis of variance (ANOVA) was conducted to test whether the values were sampled from the same population.

Table 3: Summary statistics for the results of bond wrench tests

Group	Number of specimens in the group (N).	Average flexural tensile strength ($\overline{f_r}$). psi (MPa)	Standard deviation (s). psi (MPa)
1	4	110 (0.76)	40.0 (0.28)
2	3	119 (0.82)	19.4 (0.13)
3	4	106 (0.74)	52.2 (0.36)

Table 4 presents the results of ANOVA. A value of the observed F-statistic of $F(2,8)=0.08$ is obtained, associated with a p-value=0.92, which is greater than the level of significance $\alpha=0.05$. A high p-value leads to a statistical conclusion of failing to reject the null hypothesis. In this context the null hypothesis states that the three laboratory groups are equal, Hence, the conclusion is that there is no significant difference between the three groups.

Table 4: ANOVA results

Source of variation	SS	St-DOF	MS	F	P-value	F crit
Between groups	283	2	142	0.083	0.922	4.46
Within groups	13,700	8	1,720	-	-	-
Total	14,020	239	-	-	-	-

Given the previous results, it is justified to treat the results from all groups as a single sample. In this context Table 5 shows the summary statistics for the overall results of the bond wrench tests.

Table 5: Overall results from bond wrench tests results in psi (MPa)

Average	111 (0.77)
SD	31.35 (0.22)
COV (%)	28.23

Beam tests vs Bond Wrench Tests

Results from beam tests are presented in Table 6. From the test results, it appears that flexural strength from a bond wrench test will be more conservative than that resulting from a 4-point bending test.

Table 6: Results from beam tests

Specimen	f_r. psi (MPa)
1	175 (1.21)
2	130 (0.90)
3	129 (0.89)
Average	144 (1.00)
SD	26.56 (0.18)
COV (%)	18.35

This was verified by using a statistical test (t-test). Because the bond wrench tests came from the remaining pieces of the 4-point beam tests, we cannot assume independence of the two tests within groups. Therefore, a paired sample t-test was performed where each value from the beam test was paired with the average of that group's bond wrench tests.

From the paired sample t-test, a t-statistic of $T(2)=2.05$ associated with a p-value=0.088 is obtained. Because this is strictly larger than a significance level of $\alpha=0.05$. Under the light of classical NHST the fact that the average difference between the beam test strength and bond wrench strength is zero cannot be rejected. For example, there is no statistically significant difference in averages from beam tests or bond wrench tests. On the other hand, if the effect size is calculated using the proposed measure of Cohen's d, a value of d=1.19 is obtained. By conventional standards [10], this suggests that there is a large effect of practical importance. Recall that p-values are dependent on sample size. The sample might just be too small to capture this effect in a NHST.

Table 7: Paired results for beam and bond wrench tests

Specimen	f_r Beam psi (MPa)	$\bar{f_r}$ Bond wrench psi (MPa)	Difference psi (MPa)
1	175 (1.21)	110 (0.76)	65.0 (0.45)
2	130 (0.90)	119 (0.82)	11.0 (0.08)
3	129 (0.89)	106 (0.74)	23.0 (0.16)
		Average	33.0 (0.23)
		SD	26.6 (0.19)

CONCLUSIONS

Masonry design students at the University of Wyoming built and tested 6- unit stack brick masonry prisms. There were three groups of students building one specimen each. The specimens were submitted to a beam test according to ASTM E518. This type of test results in the failure of the middle joint of the specimens (one joint). The resulting two pieces of the specimen (3-unit stack brick prisms) were used to perform bond wrench tests according to ASTM C1072 (four joints in total). The bond wrench strength was considered to be the average of four data points for each laboratory group. That average bond wrench strength was paired with its respective flexural bond strength from the beam test. Statistical analyses were performed to compare the average difference of the flexural bond strength obtained by the two methods.

Although the average difference between flexural bond strength measured from beam tests and that obtained from bond wrench tests was found to be statistically not significant (T(2)=2.05, p-value=0.088), the effect size was found to be of large practical importance (d=1.19).

More testing with a larger sample size is recommended to verify these results. Although more bond wrench data was collected, the analysis required taking the group average. Another recommendation is to have a single professional mason to build the specimens. This would restrict the source variation that comes from different builders.

ACKNOWLEDGEMENTS

The authors are grateful to the students of the fall 2024 University of Wyoming masonry class.

REFERENCES

[1] Holser, P.J., R.E. Klingner, and J.M. Melander. (2002). "Inter-Laboratory Study to Establish the Precision and Bias of Bond-Wrench Testing under ASTM C 1329 and C 1357." In *Masonry: Opportunities for the 21st Century*. ASTM International, 2002.

[2] Sarangapani, G., B.V. Venkatarama Reddy, and K. S. Jagadish. (2005).. "Brick-mortar bond and masonry compressive strength." *Journal of materials in civil engineering* 17, no. 2 : 229-237.

[3] Gaggero, Maria B., and Rita Esposito. (2023). "Experimental characterisation of flexural bond behaviour in brick masonry." *Materials and Structures* 56, no. 3: 62.

[4] Venu Madhava Rao, K., B. V. Venkatarama Reddy, and K. S. Jagadish. (1996). "Flexural bond strength of masonry using various blocks and mortars." *Materials and structures* 29 : 119-124.

[5] Kestemont, Olivier. (2023) "Review and evaluation of the bond wrench test: assessment of its characteristics, applications and limitations." Thesis completed at Université catholique de Louvain.

[6] Nichols, J. M., K. T. Glowacki, and L.H. Feigenbaum. (2018). "A Precision and Bias Study for the Measurement of Flexural Strength of Masonry Prism." *Mason. Int. J. Int. Mason. Soc* 30 : 71-80.
[7] Gauss, C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gauss*. sumtibus Frid. Perthes et IH Besser.
[8] Student. (1908). The probable error of a mean. *Biometrika*, 1-25.
[9] Fisher, R. A. (1922). The goodness of fit of regression formulae, and the distribution of regression coefficients. *Journal of the Royal Statistical Society*, *85*(4), 597-612.
[10] Cohen, J. (1988). The effect size. Statistical power analysis for the behavioral sciences, 77-83.