Measurement uncertainty of field building acoustics testing has typically been measured using a "round robin" inter-laboratory study (ILS) in which the same assembly, subject to defined conditions, is measured by different laboratories. This results in a measurement of the overall uncertainty in the procedure. A gauge repeatability and reproducibility study (GRR) uses analysis of variations (ANOVA) on an appropriately designed experiment to separate and quantify the components of the overall uncertainty [Whitfield and Gibbs, Measurement Uncertainty in Sound Insulation Testing, Proc. ICSV24, London (2017)]. Notably, the uncertainty in the measurement method can be separated from the variation of the specimens. This method offers unique insight into performance of the assembly and can identify the areas that have generate the greatest uncertainty in the measurements. The ability to evaluate the variation due to workmanship and ordinary construction practices separately from the uncertainty in the measurement method is potentially instructive in creating test procedures and methods for evaluation, design, and quality control.

The authors have performed several GRR of impact and airborne testing of floor-ceiling assemblies within multifamily buildings. The components of the overall uncertainty are evaluated and the variation between nominally identical assemblies is discussed.

**Keywords:** repeatability, reproducibility, gauge, impact, field testing

1. **Introduction**

   In the United States, measurement uncertainties for field acoustical testing are determined by inter-laboratory studies (ILS) as described in ASTM E691 [1]. In such a study, a single specimen is tested by the various laboratories or testing agencies, and the assembly may also be tested multiple times by a single operator. The total uncertainty is broken into repeatability, which is the variation when a test is repeated by the same test personnel, and the reproducibility, which is the variation between different test agencies. The total uncertainty is given by the sum of these two terms,

   \[ \sigma^2 = \sigma^2_{\text{repeatability}} + \sigma^2_{\text{reproducibility}}. \]

   A different type of study, a gauge repeatability and reproducibility (“Gauge R&R” or GRR) study, uses analysis of variations (ANOVA) techniques along with a suitable experiment design. This type of study is capable of separately estimating the component of the total uncertainties. Similar to the ILS, repeatability is uncertainty due to the measuring gear itself, i.e., the variation due to repeated measurement of the same part with the same personnel using the same equipment. However, instead of a single combined uncertainty for reproducibility, a GRR returns the uncertainty components associated with the operator, with the part, as well as the interaction terms between the components.

   Gauge R&R studies are commonplace in manufacturing, where they are often used to evaluate the repeatability and reproducibility of a literal gauge used to measure a part during the process. In this context, the repeatability is generally defined as the variation when the same part is re-measured by
the same person, and reproducibility defined as when the same part is re-measured by a different person. Whitfield and Gibbs [2], [3] have applied this concept to building acoustics testing, and here we follow their interpretation. The “part” in this case refers to the assembly under test. Repeatability remains as the variation when the same test personnel re-measure the same assembly, and reproducibility is the variation attributable to different test personnel measuring the same part. That “gauge” in this case refers to the entire process of measuring the sound insulation of an assembly.

Note that in the traditional ILS, the reproducibility is based on measurements by different testing laboratories, while a GRR is traditionally based on measurements by different personnel but within the same company and operating under the same instructions and interpretation of the relevant standards. The GRR methodology could be expanded to include different test companies (with different equipment and practices) by increasing the number of factors in the analysis. This was not investigated here.

Including a large number of assemblies or parts in the GRR further allows an assessment of the uncertainty of the part itself. Traditional ILS either are of a single assembly, or the part variation is included in one of the other terms. Field measurements of a significant number of assemblies will provide an estimate of the overall variation, but does not distinguish between measurement uncertainty and part variation. In principal, the part variation could be estimated by subtracting the measurement uncertainty determined by an ILS. However, this approach is not viable due to the small number of ILS that have been performed for field testing, and thus the low confidence in current estimates of measurement uncertainty.

2. Design of experiment for GRR study

2.1 Assembly (“Part”)

A multifamily project nearing completion was used for the GRR study. The building was four stories and therefore provide three separating floor-ceiling assemblies in each “stack” of units. Two adjacent stacks of units with mirrored floor plans were identified. The living room in each units was selected for the testing. These units therefore had identical dimensions and finishes, and provided six “parts” or nominally identical assemblies for test. The room volume was 1796 cubic feet. The rooms were unfurnished when measured.

The assembly, from top to bottom, was as follows.

- Vinyl plank finish flooring
- Foam sound mat 2 mm thick
- 38 mm (1.5 inch) gypsum concrete
- Tangled mesh sound mat, 10 mm thick
- Plywood subfloor, 18 mm (3/4-inch)
- Nominal 2x10 (50 x 250 mm) solid wood joists
- Fiberglass batt insulation
- Resilient channel
- Damped drywall, 19 mm (3/4 inch)

2.2 Test Personnel (“Operators”)

Testing was performed by Western Electro-Acoustic Laboratories, which is accredited for the test methods by the National Voluntary Laboratory Accreditation Program (NVLAP) of National Institute of Standards and Technology (NIST). WEAL is a subsidiary corporation of Veneklasen Associates, Inc. the authors’ institution. The test personnel had previously been trained and certified for the test method, and had significant experience in performing such tests.

Tests were performed by 4 teams of two people each. Each team used a Bruel & Kjaer type 2250 or 2270 sound level meter running the same building acoustics software for the test. The loudspeakers and noise generators were of identical models, and two models of tapping machines were used. Each
team used the same test equipment as much as possible, although equipment issues forced some teams to switch sound level meters in the middle of the study. This experiment was not designed to separate the effect of personnel from equipment; all variation between teams was assigned to the operator castigatory.

Teams were instructed to perform the measurements as they usually would. The standard requirements allow some choices, such as the loudspeaker location and some of the tapper positions. The locations were noted but no specific instructions were given to the teams, as these decisions would typically be determined by personnel in the field. One member of each team was defined as the leader who made these decisions.

2.3 Design of Experiment

Two GRR’s were performed, one for airborne noise isolation testing per ASTM E336 [4], and the other for impact insulation testing per ASTM E1007 [5]. Each study encompassed 6 assemblies or parts, tested two times each by four teams.

3. Analysis

3.1 Results

There was a data error and it appears that one set of data from one team was lost. Although unfortunate, this is an opportunity to investigate the effects of unbalanced experiments. This analysis will be reported in future work. For simplicity, the analysis was first carried out on the data set of 6 parts tested 2 times each by 3 teams.

A two-factor balanced analysis of variation (ANOVA) with replication was performed on the data set. In this model, the factors are the operators and the assemblies (parts), and the results can be written as

\[ X_{ijk} = \mu + O_i + P_j + (OP)_{ij} + r_{ijk} \]

where \( \mu \) is the overall mean, \( O \) is the random variation associated with operator \( i \), \( P \) is the random variation associated with assembly or part \( j \), and \( OP \) is the variation associated with the interaction of operator \( i \) with part \( j \), and \( r \) is the variation associated with the repeat \( k \). The standard deviation (square root of the variance) of each term is labelled in the same manner as \( \sigma_O, \sigma_P, \sigma_{OP} \) and \( \sigma_r \).

In addition to the variation associated with individual factors, we compute

\[ \sigma_{gauge}^2 = \sigma_O^2 + \sigma_P^2 + \sigma_{OP}^2 + \sigma_r^2, \]

which is the total variance due to the entire measurement process. This is conceptually similar to Eq. 1, with the additional detail of potential operator-part interaction. The total variance of the measurement is

\[ \sigma^2 = \sigma_P^2 + \sigma_{gauge}^2, \]

where the variation attributable to the part or assembly can be separate from the variation due to the measurement.

The results for the single number ratings are shown below. Airborne ratings are Normalized Noise Isolation Class (NNIC), which is similar to \( D_{nT,w} \). Impact ratings are Normalized Impact Sound Rating (NISR), which is similar to \( L'_{nT,w} \) excepted inverted so that higher ratings represent lower levels; the approximate relationship is \( 110 - L'_{nT,w} \). The analysis was repeated for each third-octave band. The variation due to the terms in Eq. 4 are graphed in Fig. 1 and 2. The standard deviations of the three components of the gauge uncertainty are graphed in Fig. 3.
Table 1: GRR Results (standard deviations) for Single Number Ratings

<table>
<thead>
<tr>
<th>Rating</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>$\sigma_p$</th>
<th>$\sigma_{\text{gauge}}$</th>
<th>$\sigma_Q$</th>
<th>$\sigma_{OP}$</th>
<th>$\sigma_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Airborne NNIC</td>
<td>57.5</td>
<td>1.01</td>
<td>0.84</td>
<td>0.54</td>
<td>0.27</td>
<td>0.00</td>
<td>0.47</td>
</tr>
<tr>
<td>Impact NISR</td>
<td>57.9</td>
<td>1.78</td>
<td>1.41</td>
<td>1.09</td>
<td>0.48</td>
<td>0.69</td>
<td>0.69</td>
</tr>
</tbody>
</table>

3.2 Part Uncertainty

The part uncertainty was quite good for the airborne testing, at less about 1 dB or less with only a slight increase at the low frequencies. For impact testing, the part variation was somewhat higher, about 1.5 dB, increasing towards the high frequencies with the most variation at 2-4kHz.

3.3 Measurement Uncertainty

For the mid-frequency range, the gauge standard deviation representing the measurement variation was near or below 1 dB. This is good overall performance, but there were some areas that require further review. Most obviously, there were larger variations than anticipated at high frequencies. The cause for this appears to be background noise measurements. The receiving room levels and background noise levels varied widely both between trials and between teams. These measurements were performed on separate days and at different times. Although all attempts were made to avoid construction noise, exterior noise sources may not have been the same for each test. Further, measuring low noise levels requires care in avoiding self-noise, which may have been implemented to various
degrees of success. This clearly indicates that modifications to the test method should be considered to reduce this variation.

The gauge standard deviation increased at low frequencies. This is unavoidable to some extent as the increased modal character of the room means that small changes in measurement technique can potentially result in large differences in measured level. For impact testing in particular, there was large variations between teams in the low frequency bands. This requires further investigation and training since these are the frequencies shown to be important to human annoyance from footfall and are included in low-frequency impact ratings such as LIR [6].

A ILS on these test methods was performed by ASTM in 2008, in which 6 testing laboratories measured one assembly. Since only one trial was performed, the results do not include any information on the repeatability or the operator-part interaction, and therefore might be best compared to the operator variation. The standard deviation for the single number ratings was 1.34 for NNIC and 0.84 for NISR. The comparison of the third octaves are shown in Figure 3. Note that the high-frequencies are truncated compared to the previous figures.

The operator uncertainty measured in this GRR study was between employees of the same company, while the ILS was between different companies. Therefore, the lower variation is expected.
However, the operator variation at the extremes of the frequency range approach or exceed that of the earlier study, which is not ideal. Further study and review of measurement technique is required to lower variation.

4. Conclusion

This paper continues the work of Whitfield and Gibbs [2], [3] to apply the concepts of a gauge repeatability and reproducibility study to building acoustics. Because of the multiple teams and because each team has to test multiple times, considerably greater effort is required compared to conventional program without repeats. However the additional information gained is valuable. The program indicated that the measurement uncertainties in certain frequency ranges were larger than desired, and pointed towards some factors that may allow this to be reduced. Further, the variation between nominally-identical assemblies was documented, which when implemented over a large number of projects, will provide valuable quality control information to the developer or construction manager.

Fundamentally, knowledge of the actual part performance (field acoustical performance of a floor ceiling assembly in this case) is essential for analysis and understanding of acoustical design and performance. This method provides the means to determine actual part (assembly) performance and removes the guesswork that is typical in the acoustical profession. With part performance defined statistically, the profession has a means to have “the field be the new lab.”

REFERENCES


