The Statistical Methodology Used to Assess the Effect of Limestone Addition to General Purpose Cement

by Dr Tapan Rai

on behalf of accessUTS Pty Ltd

for Cements, Concretes and Aggregates Australia

27 September 2013

Project No: 2013001253

accessUTS Pty Limited

PO Box 123

Broadway NSW 2007 Australia

Tel +61 2 9514 1916 Fax +61 2 9514 1433

ABN 55 098 424 312

accessUTS is a controlled entity of the University of Technology, Sydney

TABLE OF CONTENTS

1. ABSTRACT

This report provides an assessment of the statistical methodology used to analyse the effect of adding up to 12% limestone in general purpose cement.

2. INTRODUCTION

2.1 Background

The cement manufacturing process is associated with a relatively large volume of carbon dioxide (CO_2) emissions. This CO_2 is primarily produced as a result of the use of coal as a fuel, and the calcination of limestone, in the production of clinker. In an effort to reduce the amount of CO_2 released into the atmosphere, during the cement manufacturing process, many jurisdictions have adopted the practice of adding limestone to the cement during the grinding process. As a result a smaller amount of clinker is used in the cement. While the use of a smaller amount of carbon-intensive clinker reduces the CO_2-emissions in the manufacturing process, the result could be a cement or concrete of lower strength, poorer workability, poorer setting times, increased water bleed during the drying process and increased shrinkage. These and other potentially undesirable effects of limestone addition need to be analysed to ensure that the cements and the concretes in which they are mixed, have acceptable properties.

On the global stage, Canada and the United States of America have implemented standards for cements with limestone content up to 15% following extensive Canadian testing; Europe has adopted standards for the addition of limestone in cement following the experience from France and Germany in the 1970s and 1980s, while New Zealand has adopted a standard for 10% limestone addition following extensive testing. In Australia, the cement Standard, AS 3972 was revised in 1991. This revision introduced the concept of “performance based” specification, as well as the practice of allowing up to 5% mineral addition in type GP (general purpose Portland) cement and Type GB (General purpose) blended cement. In 2010, the Standard was reviewed and amended. This revised standard, AS 3972-2010, increased the allowable mineral addition in general purpose cement to 7.5% for Type GP and Type GB cements and established a further cement type, Type GL (general purpose limestone) cement, in which limestone was to be the specific mineral addition at levels from 8 to 20%.

2.2 Standards Australia BD-010 Working Group

The BD-010 Standards Australia Working Group was established in February 2011, to formulate, execute, summarise and report on a program with the objective of recommending to the full BD-010

Cement committee, the maximum level of mineral addition to be permitted in AS 3972 for General Purpose cement (Type GP).

2.2.1 Design of the Study by the BD-010 Working Group

The testing program consisted of two phases:

1. In phase one, the influence of limestone content of manufactured cements on mortar and concrete properties was tested in laboratory conditions.
2. In phase two, the comparative concrete properties of manufactured cements with higher levels of limestone in commercially manufactured and placed concrete, were tested in both the field and laboratory settings.

Stage one was designed to perform tests in several different laboratories, while phase 2 was designed to test in several laboratories, as well as in the field in each state in Australia. The reader is referred to sections 3.3-3.5 (pages 14-18) of the Committee Final Draft (dated 25/08/2013) of the Report of the Standards Australia BD-010 Working Group [BD-010 WG report] for full details of the study design and program changes.

2.2.2 Acceptance Criteria for cement and concrete properties

The cement and concrete properties tested in the two phases of the study are described in sections 3.3 and 3.5 (pages 14-17) of the BD-010 WG draft report. The properties were to be tested for compliance with the standards set out in AS 3972, as well with the NCHRP Report 607 standard of the American Association of State Highway and Transportation Officials (AASHTO) and the acceptance criteria set by the Iowa Department of Transportation, USA (IDOT).

The AS 3972 sets absolute standards for paste setting time, compressive strength, peak temperature rise, sulfate expansion, and drying shrinkage. These standards are outlined in Table 4.2 on page 21 of the BD-010 WG report. On the other hand, the NCHRP Report 607 standard, as well as IDOT, set relative acceptance criteria, as a percent of the measured value for a control cement, with the control cement being current industry standard. These acceptance criteria are outlined on pages 23-24 (section 4.3) of the BD-010 WG report.

2.2.3 Statistical Analysis

The report uses paired sample t-tests to analyse differences between:

1. Cement with 10% added limestone and control cement with a nominal 5% added limestone
2. Cement with 12% added limestone and control cement with a nominal 5%added limestone

In some instances 7.5% limestone added cement is also assessed or used as control cement. However, the primary concern of the current report is with the statistical methodology used for comparison with the control cement and the AS3972, NCHRP Report 607 and IDOT standards.

To compare the test cements with the existing industry standard control, the report pairs each test cement sample with a control cement from the same manufacturing plant or lab. This facilitates the use of paired t-tests, which provide greater power to detect differences than independent sample t-tests.

These tests assess the null hypothesis that there is no difference between the test cement and the control cement against the alternative hypothesis that there is a (statistically) significant difference between the test cement and the control cement. The tests proceed by calculating the probability (p) that the observed data could occur, under the assumption that the null hypothesis is true. If the calculated probability is low, the null hypothesis is rejected in favour of the alternative hypothesis; if the calculated probability is not low, the null hypothesis is not rejected. The standard criterion for rejection of the null hypothesis is $p < 0.05$. i.e. the null hypothesis is rejected if the calculated probability is less than 5% and not rejected if the null hypothesis is 5% or higher. The statistical analysis in the BD-010 WG report complies with this standard.

The reader is referred to section 4.4 (pages 24-25) of the BD-010 WG report for additional details.

2.2.4 Scope of the Current Work

The primary goal of the current work is to assess the appropriateness and adequacy of the statistical analysis performed in the BD-010 WG report, and to recommend additional statistical analysis, if required. This assessment and related recommendations are presented in Section 3 of this report.

Specific comments on other issues in the BD-010 WG report are presented in Appendix A. These comments have been discussed with the authors of the BD-010 WG report and have been adequately addressed, following the discussion.

3. ASSESSMENT OF THE STATISTICAL ANALYSIS (BD-010 WG)

3.1 Summary of the Statistical Analysis Performed in the BD-010 WG Report

The report uses paired sample t-tests to analyse differences between:

1. Cement with 10% added limestone and control cement with a nominal 5% added limestone
2. Cement with 12% added limestone and control cement with a nominal 5% added limestone

In some instances 7.5% limestone added cement is also assessed or used as control cement. However, as mentioned in section 2.2.3, above, this is not of concern in the current report; the primary concern of this report is to assess the adequacy of the statistical methodology used in the BD-010 WG report.

For the analysis reported in the BD-010 WG report, each cement sample was paired with control cement from the same manufacturing plant or lab. Hence the use of a paired/matched design is appropriate. Paired designs are statistically more powerful than independent sample designs. In the current context, this means that a use of paired t-tests provides greater power in the detection of differences in the compared cements than independent sample t-tests would.

These tests assess the null hypothesis that there is no difference between the test cement and the control cement against the alternative hypothesis that there is a (statistically) significant difference between the test cement and the control cement. A standard criterion of $p < 0.05$ for rejection of the null hypothesis is used i.e. the null hypothesis is rejected if the calculated probability is less than 5% and not rejected if the null hypothesis is 5% or higher. This is consistent with the accepted standard for peer-reviewed research.

3.2 Alternative Analyses, Advantages and Disadvantages

Alternatives to the t-test that use a paired/matched sample design include:

1. GLM designs such as repeated measures ANOVA or repeated measures MANOVA
2. Generalised Estimating Equations
3. Linear Mixed Models

If the paired nature of the data is ignored, the data may be analysed by

4. GLM designs such as ANOVA or ANCOVA

One advantage of these alternative designs is that they enable comparison of two or more test samples to a control sample, as part of a single omnibus test. In the context of the current report, this would mean that the test cements with 10% limestone and with 12% limestone addition could be compared to the control cements (with 5% or 7.5% limestone addition) within the same procedure. In addition, some of these tests (e.g. repeated measures ANOVA or MANOVA) may allow the simultaneous testing of several properties (e.g. 1 day, 3 day, 7 day, 28 day and 56 day compressive strengths) within the same procedure. The omnibus test in these procedures would test the null hypothesis that the population mean of all three cements is the same versus the alternative hypothesis that one of the cements is significantly different from the others. It is not designed to determine which specific cement differs from the others. Differences between specific pairs of cements are analysed through "post hoc" tests, if the omnibus test shows that at least one of the cements is different from the others.

This approach of using an omnibus test followed by post hoc tests results in a lower likelihood that the null hypothesis will be rejected, as compared to the use of separate analyses. This more conservative approach is recommended in most situations where multiple testing is required due to comparison of multiple populations or multiple variables. This is due to the fact that most research is designed to show that there are differences between populations rather than to show that there are no differences. In these situations, a more conservative approach in rejecting the null hypothesis is warranted. In addition the p-values of the post hoc tests are usually adjusted to correct for error inflation due to multiple comparisons.

However, in the test program undertaken by the BD-010 WG, the research goal is to show that there is no significant difference between the test cements and the control cement. In this situation, it is appropriate to take the more aggressive approach to finding differences. i.e. to use multiple t-tests rather than to use an omnibus test with post hoc tests with adjusted p-values. Furthermore, if the direction of the expected change is known (e.g. compressive strength is expected to be lower and this is not desirable), the use of a one-tailed t-test may be more appropriate.

3.3 Statistical Theory: Types of Errors in Hypothesis Testing and Error Mitigation

In statistical theory, there are two types of errors associated with hypothesis tests:

1. Type I error refers to the error made by rejecting the null hypothesis when it is true
2. Type II error refers to the error made by failing to reject the null hypothesis when it is false.

Historically, statistical methodology has focussed on minimizing the type I error. This is because most experiments were (and continue to be) designed to show that there are significant differences between populations. In other words, the experiments are set up so that the research hypothesis is accepted when the null hypothesis is rejected. Therefore, a stringent criterion of 5% probability was established to minimize the probability of making a type I error. Most research conducted prior to 1970 simply ignored the existence of type II errors.

However, in the current work, the research goal is to show that there is no significant difference between the test cements and the control cement. In this context, failure to reject the null hypothesis may result in the incorrect conclusion that there is no significant difference between the test cement and the control cement. In view of this it is very important to minimise the type II error.

Type II errors are minimised by increasing the sample size. Ideally, the type II error should be set at 5% to match the criterion for minimising the type I error. In statistical terminology this is equivalent to setting the “power” of the hypothesis test at 95%. The analysis of the balance of type I and type II errors, effect size and sample size is known as power analysis.

3.3.1 Error Mitigation for The Paired-Sample t-tests reported by the BD-010 WG

In the context of the BD-010 WG test program, if the committee chooses to persist with the described paired sample t-tests, power analyses need to be performed and reported for all t-tests that analyse differences between the test cements and the control cement. Some of the questions that need to be answered in this context are:

1. Given that the sample size is fixed, what effect size would the test be able detect with 95% power at 5% significance level? In other words, how large would the difference between the test cement and the control cement have to be, for the null hypothesis to be rejected (assuming that sample size is not changed)?
2. How does the effect size determined by the power analysis compare with the difference stated in the acceptance criterion?

Question 1, above, can be answered by performing a “sensitivity power analysis”. Question 2 requires the comparison of the sensitivity power analysis with the AS3972 acceptance criteria. Sensitivity power analyses generally provide minimum detectable differences as a standardized (unit-less) effect size. The comparison with the AS3972 criterion would require this standardized effect size to be multiplied by the standard deviation of the difference between the test cement and the control cement. This would provide an indication of the minimum difference that would need to be observed before the null hypothesis would be rejected, while setting both Type I error and Type II error to 5%.

If the minimum detectable difference is larger than the acceptable difference determined by the AS3972 criterion, then the hypothesis test may fail to reject the null hypothesis even when the test cements fail to meet the acceptance criterion. In this situation, failure to reject the null hypothesis is not sufficient to conclude that the test cement and the control cement are the same. In this case, additional data may need to be collected to determine whether the acceptance criterion is met.

This is a rather convoluted approach to testing against the acceptance criterion. Moreover it does not easily lend itself to testing against the relative criteria presented in the NCHRP Report 607 or IDOT guidelines. A more elegant formulation of the problem and analysis method is presented in section 3.3.2

3.3.2 Reformulation to Facilitate Testing of Compliance against Acceptance Criteria

Where the acceptance criteria are clearly stated, the problem may be formulated so that the research hypothesis is the alternative hypothesis. In the case of compliance with AS 3972, this may be achieved by a one sample t-test with the following hypotheses:

H_0: The measured value for the test cement is less than or equal to the minimum allowable value stated in the acceptance criterion

H_1: The measured value for the test cement to the control cement is greater than minimum allowable value stated in the acceptance criterion

If comparison is against a ratio of the test cement to the control cement (as in the NCHRP Report 607 or IDOT guidelines), this may be achieved by using the following hypotheses:

H_0: The ratio of the test cement to the control cement is less than or equal to the minimum allowable ratio stated in the acceptance criterion

H_1: The ratio of the test cement to the control cement is greater than the minimum allowable difference stated in the acceptance criterion

Formulating the problem in this way will facilitate direct comparison to the acceptance criteria. In addition, it will facilitate the interpretation of results through a more familiar understanding of the role of the *p*-value in hypothesis testing. The use of these hypotheses might require additional effort on the part of the statistician/data analyst. However, the direct comparison to the acceptance standard makes it the most appropriate approach.

3.4 Recommendations and Conclusion

3.4.1 Recommendations

It is recommended that:

- Compliance with AS 3972 be analysed by using the one-sample t-tests outlined in section 3.3.2, above
- Compliance with NCHRP and IDOT acceptance criteria be analysed using the hypotheses for comparison against a ratio outlined in section 3.3.2 above
- Power analyses be performed and reported for all hypothesis tests.

3.4.2 Limitations of Recommended Approach

The use of t-tests does not permit assessment of:

- differences between cements manufactured in different manufacturing facilities
- differences due to differences in climatic conditions in different states
- differences that may be attributed to conditions (such as water/cement ratio) which may not have been controlled in the testing program

The assessments of these differences will require the appropriate GLM approach to be planned at the experimental design stage. Given the limitation of sample size and data available, testing these differences is not feasible as part of the current program. If these differences need to be tested in the future, it is strongly recommended that the working group employ an experienced statistician to help develop the optimal design for the experiment.

Another limitation is the potential inflation of the family-wise error rate due to multiple comparisons. This can be mitigated by using Bonferroni adjustment to the *p*-values.

3.4.3 Conclusion

The paired t-tests used in the BD-010 WG report are a reasonable approach to assess differences between the test cements and the control cement. However, they fail to assess compliance with the AS 3972 standard or other acceptance criteria. Furthermore, without the related power analysis, they are not adequate to assess whether the test cements are the same (within reason) as the control cement. These goals can be achieved by following the recommendations in section 3.4.1, above.

As a result of consultation during the preparation of the current report, however, the BD-010 WG accepted the recommendations presented in section 3.4.1 report, and analysed compliance of the test cements against the acceptance criteria by methods outlined in section 3.3.2, above. The results were examined by the author of this report and found to be satisfactory (from a statistical perspective), with the test cements appearing to comply with the acceptance criteria on most properties, even after Bonferroni corrections were applied to the *p*-values.

Tapan Rai

Dr Tapan Rai

4. APPENDIX A - SPECIFIC COMMENTS ON OTHER ISSUES IN THE BD-010 WG REPORT

This appendix provides specific comments on scientific (not necessarily statistical) issues noted in the BD-010 WG report. These comments have been discussed with representatives of the BD-010 working group and have been adequately addressed by the authors of the report, as a result of these discussions.

The numbers in the headings and the page numbers in this appendix refer to section numbers and page numbers in the BD-010 WG Report.

2.3 Australian Context – Sustainability Issues:

Figure 2.3.1: Reduction in global CO_2 emissions arising from cement manufacture 1990-2010:

- Graph is misleading since the y-axis begins at 560 rather than at 0; this results in a visual perception of a greater reduction in CO_2 emissions than is actually achieved.
- The text states that reduction between 1990 and 2010 is 14%. A careful look at the graph shows that the graph is consistent with this.
- The issue is only one of visual appearance of the graph – generally caused by poor default settings in most spread sheet software packages.

Figure 2.3.2: Usage of mineral components in Australia cement manufacturing 1990-2012:

- The y-axis shows "Extenders" in 000s of tonnes. It is not clear whether this is for the entire industry in Australia or for a fixed amount of cement.
- If the y-axis refers to extenders for the entire industry, this could be misleading since the amount of cement manufactured over this time period may have grown and the increase in the tonnes of extenders used may be a result of increase in cement production
- It would be more appropriate to show "proportion of extenders used"

Figure 2.3.3: CO_2 emissions per tonne produced - Australia 1990-2012:

- Graph is misleading since the y-axis begins at 500 rather than at 0; this results in a visual perception of a greater reduction in CO_2 emissions than is actually achieved.
- The text states that reduction of 25% has been achieved in the period from 1990 to 2012. The graph is consistent with this figure.
- Once again the issue appears to be one of visual appearance of the graph caused by poor default settings in a spread sheet software program.

2.4 Improved Sustainability:

- P. 9: Addition of limestone can result in increased CO_2 emissions due to calcination. Figure 2.4.1 and Table 2.4.2 address the fact that this is offset by decrease in energy use during clinker production. However, see below (Table 2.4.2).
- On p. 10 there is a claim that "mineral additions reduce the embodied emissions in cement roughly in proportion to the proportion of clinker in the cement". It is not clear whether

this is meant to read "… roughly in proportion to the reduction in proportion of clinker in the cement"; the latter appears to make more sense.

Table 2.4.2: Estimated Annual Reduction in Energy Usage and Emissions Resulting from the use of 10% and 15% limestone in Blended Cement:

- The table does not report baseline energy requirements and baseline emissions. This information would be useful in determining the percent reduction as well as in assessing the magnitude of the problem.

3.1 Design:

Preamble:

- P. 12: It would be useful to report some information about the Canadian study, for comparison with the current Australian study. For example:
    - What was the previous Canadian standard?
    - How many different levels between 10% and 15% were used?
    - How many replicates did each Canadian cement company use?
    - Were their acceptance criteria similar to the ones used in the Australian study?
- P. 12: It would also be useful to see a comparison between Australian and Canadian conditions that would help further justify this study and potential differences in standards

3.2 Manufacture of Australian Test Cement:

- P. 14 mentions the following limitations:
    - Testing is based on "one-off" production trials
    - Cements used in testing have not been optimised; this differs from other test programs

This should be stated more clearly as a limitation

3.3 Stage 1 - Influence of limestone content on mortar and concrete properties:

- Initial Thoughts:
    - 3 levels + Control: Suggests need for One-Way ANOVA with corrections for multiple testing in post hoc analysis
    - Use of different manufacturing plants suggested that data might be clustered and "plant" may need to be included as a random effect
    - One grab sample for sulphate expansion, shrinkage and peak temperature rise may be inadequate
- Subsequent Thoughts:
    - ANOVA designs and inclusion of random effects are used to take a more conservative approach to finding differences – in order to prevent inflation of the type I error rate

- In the current study, the goal is to show that there are no differences. In this case, it may make sense to take the more aggressive approach to find differences (worst case scenario), and work harder of analyse the type II error

3.4 Stage 2 – Comparative concrete properties:

- Initial Thoughts:
    - The use of control/case from the same plant suggests a repeated measures/clustered design
    - Use of different manufacturing plants suggested that data might be clustered and "plant" may need to be included as a random effect
    - Use of control cements with different limestone content, amorphous silica in some cements, different classes of cement and different binders may require a generalised estimating equations (GEE) or linear mixed models (LMM) approach
    - Collection of data from different states suggests the need for an omnibus model that includes "state" as a predictor
- Subsequent Thoughts:
    - GEE and LMM models use maximum likelihood estimation (MLE) and are more likely to find differences than ANOVA approaches. However, they would require large sample sizes. In addition, the effect of controlling for various factors through statistical analysis will result in more conservative findings (no difference) in terms of type I error
    - Given that the goal of the current study is to show that there are no differences, it may make sense to take the most aggressive approach to find differences (worst case scenario), and minimise the type II error

3.5 Program Changes:

- Table 3.5 lists "No Results" for Lab Code D and "No Data Available" under "Field Data" for WA and TAS. It is not clear whether this is because testing was not done at all or whether these locations tired but failed to manufacture the appropriate cements or whether the appropriate cements were manufactured but failed testing. From a statistical perspective, these are different situations, and the reasons for absence of data and/or results should be reported.
- P. 18: Durable Concretes were not tested with 100% Type GP binder. This does not appear to be a serious limitation if 100% Type GP is not commonly used in the market.
- P. 18: Sufficiency of data should be examined through a statistical power analysis (partial power analysis performed as part of this review – see later).

4.1 Uncertainty of Test methods:

<u>Table 4.1: Reported variability associated with given test methods used in this test program</u>:

- Definition of Coefficient of Variation is not clear from context; there are two definitions in the literature:
    - Coeff of Var = (Standard Deviation)/Mean

- Coeff of Variation = (Root Mean Square Error)/(Mean of Dependent Variable)

It is assumed that the former definition is used since the latter assumes the existence of a statistical model. However, this is not a serious issue, as long as the definition is applied consistently in this document and is consistent with industry usage.

- Definitions of Repeatability and Reproducibility are not clear from context. By itself, this is not a serious issue, as long as the definition is applied consistently in this document and is consistent with industry usage. However, the following values are reported:
  - Mortar Peak Temperature Rise (Cement): 8% repeatability; 15% reproducibility
  - Drying Shrinkage (Concrete): 8% repeatability
  - Durability (Concrete):
    - AS 1012.21: 1.5-2% repeatability, 6-6.5% reproducibility
    - NT 492: 14% repeatability, 20% reproducibility

If low reproducibility and repeatability represent poor reliability, (as is the case with most reliability measures), the validity of these tests is questionable, and any statistical analysis performed on data obtained from these tests is not likely to be useful. It is therefore assumed that both repeatability and reproducibility are reported as "variability" in results, with low variability being good and high variability being bad.

- P. 21: It is stated that in this report, "Aggregated Data" refers to an arithmetic average of available data. This may be acceptable as industry standard. However, from a statistical perspective, it would be appropriate to report confidence intervals or standard deviations or standard errors, together with arithmetic averages. This would enable the reader to assess the variability in the available data.

4.3 Acceptance Criteria:

- Acceptance criteria for various measures are listed as a percent of the corresponding measure for the control cement/concrete. These are based on standards developed in the United States of America, and may be considered industry standard at the current time. However, if the control cement/concrete is changed in the future, and future testing is conducted against the new control cement, the results of these future tests would not be comparable to the results of the current tests. For example, suppose that cement with 10% limestone is adopted as industry standard based on the fact that its long term compressive strength exceeds 90% of control cement with 5% limestone addition. Next suppose that in the future there is a move to test cement with 15% limestone addition against existing industry standard of cement with 10% limestone addition. The new cement would be acceptable if its compressive strength exceeds 90% of the compressive strength of the control with 10% limestone addition. However, this would be equivalent to setting an acceptance criterion of 81% of compressive strength of cement with 5% limestone addition.

This may not be a serious issue at the current time, but it would be a limitation for future adoption of higher mineral content. In view of this, it would be appropriate to consider setting a standard for control cements to be used. Even in the current testing program, it is not clear what standard were used to accept cements with 5% limestone addition. As a result, it is not clear how the cements tested as part of this program compare to cement with no limestone addition.

5.1 Cement Mortar Compressive Strength (AS 2350)

Figure 5.1.3: Mortar Compressive Strength for cements with varying levels of limestone mineral addition – Stage 1 – aggregated data:

- The graph shows compressive strengths of mortars with various amounts of limestone added, relative to a control with 5% limestone addition. It is not clear how these compare to mortars with no limestone added. The fact that pure cements/mortars were not used as controls is stated clearly throughout the report. However, if some information was available from the previous studies, as a result of which 5% limestone addition was adopted. It is also not clear whether the American acceptance criteria assumed a pure cement/mortar as the control. The danger in using current market standard as the control is outlined above (under 4.3: Acceptance Criteria).
- It would be useful to display confidence intervals (if available) to show the variability of the results.
- These points are valid for all similar graphs in chapter 5, including 5.6.1, 5.7.1, 5.8.1, 5.9.1

5.2 Cement Paste Setting Time (AS 2350)

Table 5.2.1 Cement Paste Setting Time for cements with varying levels of limestone mineral addition – Stage 1 – aggregated data:

- It would be useful to display confidence intervals (if available) to show the variability of the results.
- Ratios can be misleading in situations where results are rounded off to the nearest 15 minutes – unless the magnitude of the actual time is substantially higher than 15 minutes. In view of this, it would be appropriate to report the raw values, in addition to the percent relative to control.
- These points are valid for all similar tables in chapter 5, including 5.3.1,5.4.1, 5.5.1

5.3 Cement Mortar Drying Shrinkage (AS 2350)

- Not clear how shrinkage is calculated:
    - Lower shrinkage suggests higher final volume if it is calculated as:

$$shrinkage = \frac{final\ vol\ of\ test\ cement - initial\ vol\ of\ test\ cement}{final\ vol\ of\ control - \ initial\ vol\ of\ control\ cement}$$

    - On the other hand, it could mean that final volume was lower, if shrinkage is calculated as:

$$shrinkage = \frac{final\ vol\ of\ test\ cement}{final\ vol\ of\ control\ cement}$$

- Tests in this section are performed in similar experimental conditions (no adjustment to water content etc.) for all mortars. This is appropriate in experimental conditions.
- Please see comments on tables in this chapter (under 5.2)

5.4 Cement Mortar Sulfate Expansion (AS 2350)

- P. 33-34: Literature review indicates inconsistent results for effect of sulphate solution on loss of strength. This could be due to differing sulphate solutions ($NaSO_4$ versus $MgSO_4$ or differing concentrations).
- In view of the above points, it would be appropriate to report the concentration as well as type of sulphate solution used in the Australian Laboratory tests. It is especially important to know that these were the same for all samples of cement tested for the current work.
- Please see comments on tables in this chapter (under 5.2)

6 Stage 2 – Comparative Concrete Properties

Tables 6.2.1.1, 6.2.1.2, 6.2.1.3, 6.2.1.4, 6.2.2.2, 6.2.2.3, 6.3.1.1, 6.3.1.2, 6.3.2.2, 6.3.2.3, 6.4.1.1, 6.4.1.2, 6.4.1.3, 6.4.1.4, 6.4.2.2:

- Ratios can be misleading in situations where results are rounded off – unless the magnitude of the actual measurement is substantially higher than the round off. In view of this, it would be appropriate to report the raw values, in addition to the percent relative to control.
- It would be useful to display confidence intervals (if available) to show the variability of the results.

Figures 6.2.2.1, 6.3.2.1, 6.4.2.1, 6.4.2.1.2, 6.4.2.1.3:

- It would be useful to display confidence intervals (if available) to show the variability of the results.

Tables 6.2.2.5, 6.2.2.6:

- Categorical data is reported in these tables
- Categorical data is notoriously insensitive to statistical analysis. Although, no hypothesis tests were related to these tables, it would be appropriate to report the sample size (number of replicates) used in determining these durability ratings.

Figures 6.3.2.5

- Apparent typo: Limestone content recorded as 4.5%; probably should be 5% or 7.5%

6.4.2.2 Concrete Drying Shrinkage

- P. 75: Please see previous note on potential ambiguity/misinterpretation in shrinkage calculation (under 5.2 Cement Mortar Drying Shrinkage)

7.0 Results Summary

<u>Table 7.0</u>

- P. 76 (Stage 1): Mortar Sulfate Expansion; statement reads: "Type GB cements were inconclusive…" It is not clear if any related tests were performed; unable to identify mortar sulfate expansion data related to type GB cements
- P. 77 (Stage 2): Workability; statement reads "Field concretes showed decreased slump with increased limestone addition." This is not consistent with Table 6.4.1.1 which indicates the following values for field concretes: 100% slump for 5% limestone addition; 100% slump for 7.5% limestone addition; 86% slump for 10% limestone addition and 88% slump for 12% limestone addition

7.1 Compliance to AS 3972 – General purpose and blended cements

- P. 78-79: Table 7.1 provides useful information. However, the compliance should be tested using one-sample t-tests outlined under **4.4 Statistical Analysis of Australian Test Data**

7.2 Compliance to Acceptance Criteria

- P. 81-84: The tables in 7.2.1(NCHRP Acceptance Criteria) and 7.2.2 (Iowa DOT Acceptance Criteria) provide useful information. However, columns showing percents of compliant and non-compliant data points would be useful.
- Compliance to these acceptance criteria should be tested as recommended in Section 3.4.1 of this report

7.2 Statistical Analysis

- Without the related power analysis, it is not appropriate to conclude that the variables not included in this section have equal means.

Differences between control and test cements are not very useful if the size of the difference is not assessed against the acceptance criteria. This is noted on p. 84. Different ways of assessing against the acceptance criteria are outlined in Section 3 of this report. Implementing one or more of these tests will alleviate the problems noted.