| FORM-STAT-04A-01 | Statistical Analysis Plan |
|---|---|
| Version 1.0 | |
| | |

# CPOL

## Improving children's language, literacy and mental health: Evaluating the impact of the Classroom Promotion of Oral Language (CPOL) approach

### Document Version History

| Version Date | Version | Author | Signature | Change Description |
|---|---|---|---|---|
| 18-01-2017 | Draft 1 | Francesca Orsini | | Initial release. |
| 02-11-2017 | Draft 2 | Francesca Orsini | | Incorporate Team's modifications |
| 10-11-2017 | Draft 3 | Francesca Orsini | | Incorporate Team's modifications |
| 16-11-2017 | Draft 4 | Francesca Orsini | | Incorporate Team's modifications |
| 30-11-2017 | Draft 5 | Francesca Orsini | | Incorporate Team's modifications |
| 19-12-2017 | Final Version | Francesca Orsini | *Francesca Orsini* | Final Version |

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AE | Adverse Event |
| ATSI | Aboriginal and Torres Strait Islander |
| CECV | Catholic Education Commission of Victoria |
| CELF | Concepts and Following Directions subscale |
| CI | Confidence Interval |
| CPOL | Classroom Promotion of Oral Language |
| CRF | Case Report Form |
| DEECD | Department of Education and Early Childhood Development |
| DET | Department of Education and Training |
| GCP | Good Clinical Practice |
| GEE | Generalised Estimating Equations |
| ICSEA | Index of Community Socio-Educational Advantage |
| ITT | Intent-To-Treat |
| LBOTE | Language Background Other Than English |
| NAPLAN | National Assessment Program – Literacy And Numeracy |
| NIH | National Institute of Health |
| NPVT | National Institute of Health Peabody Picture Vocabulary Test |
| OLSEL | Oral Language Supports Early Literacy |
| PP | Per Protocol |
| RCT | Randomised Controlled Trial |
| RPT | Reading Progress Test |
| SD | Standard Deviation |
| SDQ | Strengths & Difficulties Questionnaire |
| SE | Standard Error |
| SEHQ | School Entrant Health Questionnaire |
| SEIFA | Socio-Economic Indexes for Areas |
| SES | Socio Economic Status |

# 1. STUDY OBJECTIVES

## 1.1. PRIMARY OBJECTIVE

The Classroom Promotion of Oral Language (CPOL) cluster Randomised Controlled Trial (RCT) aims to determine the effectiveness of a teacher professional learning intervention (based on teacher-led whole-of-class approach to promoting oral language delivered in the first two years of school) on students' literacy, assessed through the NAPLAN (National Assessment Program – Literacy And Numeracy) 'Reading' scores at grade 3 in comparison with students' literacy at grade 3 in the control arm of the study, which will carry out business as usual in the classroom. Teachers in the control arm will not receive the specifically designed professional learning intervention.

## 1.2. SECONDARY OBJECTIVES

Determine the effectiveness of CPOL intervention on students':

- oral language, assessed through the NAPLAN 'Language Conventions' scores at grade 3
- writing, assessed through the NAPLAN 'Writing' scores at grade 3
- numeracy, assessed through the NAPLAN 'Numeracy' scores at grade 3
- mental health, as assessed by the classroom teacher through The Strengths & Difficulties Questionnaire (SDQ) total difficulties score at grade 1
- reading comprehension, as assessed by the Reading Progress Test score at grade 1
- receptive language, as assessed by the Concepts and Following Direction subtest of the Clinical Evaluation of Language Fundamentals® 4th Ed (Aust Adaptation) score at grade 1
- receptive vocabulary, as assessed by the National Institute of Health (NIH) Picture Vocabulary Test score at grade 1
- expressive language, as assessed by narrative production using the Renfrew Language Scales Bus Story script at grade 1

compared to students in the control arm.

# 2. BACKGROUND/INTRODUCTION

## 2.1. STUDY DESIGN

This 2-arm cluster multi randomised controlled community-based trial will involve 72 schools (47 Department of Education and Early Childhood Development (DEECD) schools and 25 Catholic Education Commission of Victoria (CECV) schools), 300 teachers, and 1400 students. It is expected to run for 5 years, from June 2013 to June 2018.

## 2.2. TREATMENT GROUPS

### Randomisation process and blinding

Schools will be randomly assigned to receive the intervention (teacher professional learning days) or to the control group which carries out business as usual in the classroom. Schools will be equally divided between the two arms. A statistician will prepare the randomisation schedule using block randomisation to maintain balance between treatment arms. Randomisation will be stratified by school sector (Catholic Schools and Victorian Government Schools) with variable block sizes. The ordering of schools to be randomised will be alphabetically by school name within each education sector.

Within each school, one Foundation/Grade 1 class will be randomly selected as the 'index class' by the project coordinator. Data will only be collected for the teacher and students of this class to be analysed throughout the CPOL RCT. While every teacher responsible for a Foundation or Grade 1 cohort in the intervention arm of the study will attend the professional learning days, only the data of the index class will be collected and analysed.

The research manager, research assistant, CPOL support workers, randomising statistician and teachers will be aware of the allocation to treatment arm to enable organisation for the teachers to receive the active or control intervention.

Research staff responsible for conducting the Reading Progress Test in term 3, 2015 will be blinded to randomisation allocation. School staff and teachers will be asked not to disclose their randomisation status during this assessment; however, those that do will be recorded in the study database and this 'unblinding' will be examined as a potential confounding variable in the outcome analyses.

### Description of intervention: professional learning days

Teachers whose school is assigned to the intervention arm will attend three face-to-face days of professional learning convened by the research team, and will also engage in a self-directed manner in an online learning network of teachers in like-schools throughout 2014- 2015.

An additional day of face-to-face learning will be held in Term 1 2015. In addition to these formal days of professional learning and access to the online forum, teachers will liaise with CPOL Support Workers via intermittent face-to-face, telephone, and online contact, in order that questions are addressed and program fidelity is enhanced.

The schools in the control arm will conduct business as usual in the classroom. The teachers in the control schools will not attend professional learning days and will not participate in the online professional learning network.

### Subject withdrawals and replacement

A <u>student</u> will be withdrawn from the study if following the initial consent, the student's parents actively choose to no longer give consent for the child's information to be accessed. This may be communicated via the teacher and will be followed up by a member of the study team. In the event that a parent withdraws consent for a student's information to be used in the CPOL RCT, the research team will ask (via communication with the child's Principal) whether they are still happy to take part in the research assessments in an effort to maintain the intention-to-treat analysis. If the participant withdraws fully from the study, they will cease to undergo any further scheduled assessments (SDQ- teacher report, Reading Progress Test and access to NAPLAN results). Data already collected will be used and analysed. Students who discontinue the study (withdrawn by parents) will not be replaced by further recruitment to maintain the required sample size.

<u>Teachers</u> who leave a participating school during 2014 will be replaced by a new teacher for that class. The new teacher will attend any further professional development days and have access to the online professional learning network (forum/resources) if in the treatment arm. In the event that there are changes of teachers during the study period, all teachers of index students (a student in the 2014 randomly selected class for a participating school) will be asked to complete scheduled assessments (e.g., Teacher Survey).

A <u>school</u> will be withdrawn from the study if, following the initial acceptance and participation in the study, the principal of the school advises they will no longer be involved, hence withdrawing the associated index teacher also.

## 2.3. STUDY POPULATION

The study population of this trial is children who are enrolled in schools with the following characteristics:
- >10% children identified as vulnerable in the language and cognition domain of the 2009 and/or 2012 Australian Early Development Index
- Minimum 15 children in a Foundation cohort in 2013

## 2.4. INTERVENTION

Refer to study protocol paper (BMJ Open. 2017 Nov 20;7(11):e016574. doi: 10.1136/bmjopen-2017-016574.).

## 2.5. SAMPLE SIZE

The primary aim of this study is to assess the effectiveness of a teacher professional learning intervention (based on a teacher-led whole-of-class approach to promoting oral language delivered in the first two years of school) on students' literacy compared to standard class teaching, by comparing the NAPLAN Reading scores at grade 3.

The sample size calculation is based on the assumption that 80 points is the average gain in the NAPLAN  Reading score over 2 years, therefore around 40 is the average gain in the Reading score over 1 year. For this study we aim to find a difference between the intervention and control groups in the Reading score at grade 3 of 0.3 SDs (22.47 points, based on a standard deviation of 75). An effect sizes of 0.3 SDs can be meaningful at a population level given the reach of an intervention. In this trial in particular, it equates to a difference of 22.47 points which is approximately a 6 month difference in progress and would represent a "clinically significant" difference in outcomes. Randomisation of 561 children per arm is required to provide 90% power to detect a minimum difference of 0.3 standard deviations on the NAPLAN Reading scores at grade 3, allowing for an average intra-class correlation coefficient of 0.08 and an average cluster size of 17 (based on the average class size). To allow for a potential attrition rate of 20% of children by the time they are in grade 3, 700 children per arm (1400 in total) will be required in the study (approximately 42 clusters based on a cluster size of 17). The study sample size was estimated using the Stata software package.

## 2.6. STUDY PROCEDURE

Figure 1 displays a summary of the study design and timelines.
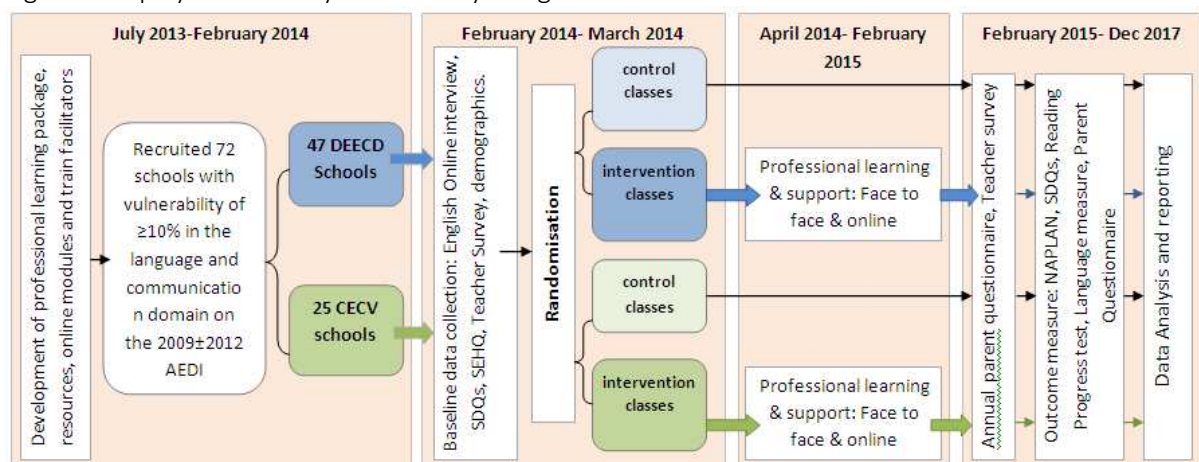
## 3. POPULATIONS OF ANALYSIS

The intention-to-treat (ITT) population will be used in the analyses. Children will be compared according to the group to which they were randomly allocated, regardless of teacher's compliance to intervention, students' exposure and withdrawal from the study. This approach preserves the prognostic balance in the study arms achieved by randomisation.

As a sensitivity analysis, results will also be presented from a per-protocol (PP) analysis. The per-protocol population will include students:

- who complete Foundation in the first year of the intervention and Grade 1 in the second year of the intervention;
- whose NAPLAN Reading score at Grade 3 is available;

- whose index teachers have been exposed to at least three of the four intervention days. If an index teacher leaves during Foundation, his/her replacement must be an intervention teacher who attended at least three of the four intervention days (intervention students only);
- whose Grade 1 teachers have been exposed to at least three of the four intervention days (intervention students only);
- whose school has sent at least one teacher to all four intervention days, i.e. school was represented at each session (intervention students only);

- whose teacher did not work in any of the intervention schools during the two year intervention phase (control students only).

Moreover, intervention students who moved to a control school during Foundation or Grade 1 or control students who moved to an intervention school during Foundation or Grade 1 will be excluded from the PP analysis.

In its original formulation (as stated in the Protocol), the PP population was defined to exclude students:
- who had more than 50 days of absence in Foundation
- who had more than 50 days of absence in Grade 1
- whose school employed a teacher who previously worked in an intervention school during the intervention phase (control students only)

but these information are not available for all the schools randomised, therefore the team has decided to remove these criteria from the definition of PP.

## 4. OUTCOME VARIABLES

### 4.1. PRIMARY OUTCOME:

| Measure | Description | Time point |
|---------|-------------|------------|

| NAPLAN Reading score | When the study cohort is in Grade 3 (2017), we will access and use their Reading score from the National Assessment Program Literacy and Numeracy (NAPLAN) data as the primary outcome measure, in order to investigate the impact of the CPOL intervention on the children's reading ability, in comparison with the children in the control classes.<br>The Reading scaled score **ranges from 0 to 1000**. Students' reading outcomes are mapped onto a ten-band continuum, with each band representing an increasing complexity of skills and understanding. The national minimum standard band for Grade 3 is set at band 2, i.e. scale score >270 and ≤322, with bands 1 to 6 used to report student performance at Grade 3 (however students can achieve scores in any band for each scale). Students with results in the national minimum standard have typically demonstrated the basic elements of literacy/reading for their year level. Students with results in the lowest band, i.e. band 1, have not achieved the national minimum standard ('below national minimum standard') and students with results in the bands higher than the minimum national standard (i.e. band 2) are considered 'above the national minimum standard'. The bands and scale scores line up as follows: | Grade 3 |
|---|---|---|

|  | Band | Year 3 | Scaled scores |
|---|---|---|---|
| Bands 1 to 6 are used to report student performance at year 3 | Band 1 | Below national minimum standard | ≤270 |
|  | Band 2 | National minimum standard | >270 and ≤322 |
|  | Band 3 | Above the national minimum standard' | >322 and ≤374 |
|  | Band 4 |  | >374 and ≤426 |
|  | Band 5 |  | >426 and ≤478 |
|  | Band 6 |  | >478 and ≤530 |
| *Bands 7, 8, 9 , and 10 are used to report student performance at years greater than 3* | *Band 7* |  | *>530 and ≤582* |
|  | *Band 8* |  | *>582 and ≤634* |
|  | *Band 9* |  | *>634 and ≤686* |
|  | *Band 10* |  | *>686* |

A full description of the way in which NAPLAN is administered is available at:
https://www.nap.edu.au/docs/default-source/default-document-library/2015-naplan-technical-report3a1604344b146909a44fff0000c50d63.pdf?sfvrsn=2.
The primary outcome of interest for this study is the raw score for reading. The corresponding standardized score will also be calculated according to the following formula:

$$Z \text{ score} = (\text{raw score} - \text{mean})/SD$$

Mean and SD will be the 2017 mean and standard deviation in Victoria (Mean=444.4 and SD=82.9).

## 4.2. SECONDARY OUTCOMES

| Measure | Description | Time point |
|---|---|---|
| NAPLAN:<br>-Writing<br>-Language Conventions: Spelling<br>-Language Conventions: Grammar and Punctuation<br>-Numeracy scores | In addition to the reading score to be used as the primary outcome measure, the following NAPLAN scores will be analysed as secondary measures:<br>• Writing<br>• Language Conventions – Spelling<br>• Language Conventions – Grammar and Punctuation<br>• Numeracy<br>Writing, language conventions and numeracy scores **range from 0 to 1000**, exactly like the reading score, and are mapped onto a ten-band continuum, with each band representing an increasing complexity of skills and understanding. The bands and scale scores line up as follows: | Grade 3 |

| | Band | Year 3 | Scaled scores | |
|---|---|---|---|---|
| Bands 1 to 6 are used to report student performance at year 3 | Band 1 | Below national minimum standard | ≤270 | |
| | Band 2 | National minimum standard | >270 and ≤322 | |
| | Band 3 | Above the national minimum standard' | >322 and ≤374 | |
| | Band 4 | | >374 and ≤426 | |
| | Band 5 | | >426 and ≤478 | |
| | Band 6 | | >478 and ≤530 | |
| *Bands 7, 8, 9 , and 10 are used to report student performance at years greater than 3* | *Band 7* | | *>530 and ≤582* | |
| | *Band 8* | | *>582 and ≤634* | |
| | *Band 9* | | *>634 and ≤686* | |
| | *Band 10* | | *>686* | |

As with reading, the outcomes of interest are the raw scores for each of these measures. The corresponding standardized scores will also be calculated (see formula provided for NAPLAN Reading Score. In particular, Writing: mean=426.9 SD=57.6; Spelling: mean=422.4 SD=81.2; Grammar and Punctuation: mean=449.7 SD=87.2; Numeracy: mean=420.1 SD=71.6)

| | | |
|---|---|---|
| **The Strengths & Difficulties Questionnaire (SDQ) assessed by then classroom teacher**<br><br><br>**Domain: Mental Health** | The SDQ is a brief behavioural screening questionnaire for 3-16 year olds with 25 questions across five scales (emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and pro-social behaviour). Detailed information on SDQ can be found at http://www.sdqinfo.org/a0.html.<br><br>- A **total difficulties score** is generated from the summed scores of all the scales except the pro-social one, and this score is utilised as the outcome measure. In this study, the SDQ will be administered at the **end of Grade 1 by the classroom teacher**, as close familiarity with the child is needed for valid ratings to be made (Youth in Mind, 2012). Total difficulties **normal scores** will be calculated for males and females separately, using the syntax for Stata provided on the SDQ website (http://www.sdqinfo.org/c3.html ).<br>Teacher SDQ Australian norms (See details in Appendix 1) for sample for ages 7-10 years are:<br>    *Males:    Total difficulties mean = 7.82, SD = 6.87*<br>    *Females: Total difficulties mean = 5.73, SD = 6.0*<br>According to Appendix 1, SDQ Teacher completed are categorized as follow, for males:<br>    *0-13: No concern*<br>    *14-16: Borderline*<br>    *>=17: Abnormal/of concern*<br>And for females:<br>    *0-11: No concerns*<br>    *12-14: Borderline*<br>    *>=15: Abnormal/of concern* | Grade 1 |
| **Reading Progress Test (RPT)**<br><br><br>**Domain: Reading Comprehension** | The RPT assesses pre-reading and early reading skills, including phonological awareness, print concepts, word knowledge, and comprehension through administration of a literacy test, on either a group of children or for an individual child. In this study, the group test will be administered to all the students in the CPOL classes by a blinded CPOL Research Assistant at the **end of Grade 1**, which gives an individual score of reading ability per child. This is a validated tool and has Australian norms.<br><br>Assessments of student achievement will be based on a normative scale, which provides a measure of achievement relative to that of other students of the same age or grade level. Normative measures are provided in the form standardised scores (see below). Australian norms are based on the national sample of students included in the ACER Project on *Curriculum and Organisation in the Early Years of School* (De Lemos, 1996). | Grade 1 |

|  |  |  |
|---|---|---|
|  | - Calculated **standardised scores** range from a minimum of 70 and a maximum of 130 and are derived from raw scores (See details in Appendix 2). Raw scores above and below these limits are recorded as 130 or 70.<br><br>  Standard scores are categorized into five groups:<br>    *<74:      well below average*<br>    *74-88:    below average*<br>    *89-111:   average*<br>    *112-126: above average*<br>    *127+:     well above average* |  |
| **CELF 4- Concepts and Following Directions subscale**<br><br>**Domain: Receptive Language** | Concepts and Following Direction is one subscale of the Clinical Evaluation of Language Fundamentals®–Fourth Edition (CELF®–4). The CELF is an individually administered test for determining if a student (ages 5 through 21 years) has a language disorder or delay. Concepts and Following Direction takes 7-10 minutes to administer and sees the student point to pictured objects in response to oral directions (Pearson, 2013).<br>In this study, CELF-4 will be collected in Grade 1.<br><br>- Derived **scales scores** will be converted from raw scores using CELF manual (see Appendix 3), and have a mean 10 and standard deviation of 3. | Grade 1 |
| **NPVT- NIH Picture Vocab Test**<br><br>**Domain: Receptive Vocabulary** | The NIH Picture Vocab Test is a digitized version of the Peabody Picture Vocab Test which is an individually administered test used to quickly evaluate receptive vocabulary with a test that requires no reading or writing (Pearson, 2015). The digital version of the test was developed by Curve Tomorrow. It takes 3 minutes to administer and sees the student press one of four buttons (pictures) that bests represents the word they hear from the iPad.<br>In this study, NPVT is collected in Grade 1.<br>The NPVT is a measure of general vocabulary knowledge and is considered to be a strong measure of crystallized abilities (those abilities that are more dependent upon past learning experiences and are relatively consistent across the adult life span). To interpret individual performance, we will evaluate the **age-Corrected Standard Scores** which are directly derived from theta scores (see Appendix 4 and 5). The theta score, which is collected for each student, represents the relative overall ability or performance of the participant. A theta score is very similar to a z-score, which is a statistic with a mean of zero and a standard deviation of one.<br><br>- A participant's **Age-Corrected Standard score** at or near 100 indicates vocabulary ability that is average for the age level. Scores around 115 suggest above-average vocabulary ability, while scores around 130 suggest superior ability – in the top 2 percent nationally for age, based on NIH Toolbox normative data. Conversely, a score of 85 suggests below-average vocabulary ability, while a score in the range of 70 or below suggests markedly low language ability (bottom 2 percent nationally), which also is likely to be associated with difficulties in school (for children) or trouble functioning in work environments with a language demand. | Grade 1 |
| **Narrative Analysis**<br><br><br>**Domain: Expressive language (syntax and narrative)** | The Renfrew Bus Story Test is a measure to assess the age level of consecutive speech used from information content, sentence length and grammatical usage in retelling a story (Renfrew, 2015). The test takes 5-8 minutes to administer.<br><br>The Renfrew Language Scales (4th Ed) Bus Story Test will be used to elicit an oral narrative sample from the students. The assessment will be administered by a blinded CPOL research assistant when the students are in Grade 1. It will be administered as per the Bus Story Test protocol, however the student narrative sample will be audio-recorded. The audio files will be transcribed verbatim and coded for narrative macro-structure (story grammar content) and micro-structure (syntax) as per the Oral Language Supporting Early Literacy (OLSEL) pilot RCT. | Grade 1 |

(FORM-STAT-04A-01/Version 1.0)

| | There are 4 continuous variables that we will used for the analysis, higher scores are indicative of better expressive language performances:<br>- Narrative syntactic analysis / expressive grammar – number of t-units (range: 0 to 36)<br>- Narrative analysis – Price et al. (range: 0 to 39)<br>- Story grammar analysis – Snow & Powell (range: 0 to 14)<br>- Syntactic complexity - total conjunctions (range: 0 to 17) | |

## 4.3. OTHER PARAMETERS

### DEMOGRAPHY AND BASELINE
The following demographic and baseline characteristics of the children will be collected:

- Age (source: Census / English Online Interview (EOI) (Department of Education and Training - DET) / School Entrant Health Questionnaire (SEHQ) / Redcap)
- Gender (source: Census / EOI -DET / SEHQ / Redcap)
- General health status (source: (SEHQ)) collected as Very good/excellent, Poor/fair/good,
- Indigenous status (source: SEHQ)
- Language background other than English (LBOTE) (source: Census / SEHQ)
- Nationality (source: Census / SEHQ)
- Previous preschool/kindergarten program attendance (source: SEHQ)
- Presence of developmental delay (source: SEHQ)
- Presence of speech or language difficulties (source: SEHQ)
- Socio-Economic Indexes for Areas (SEIFA) Index of Social Disadvantage Quintile (source: Census / SEHQ)
- Stress level in month prior (source: SEHQ) collected as Highest, High, Middle, Low, Lowest, Invalid response, Not stated
- Mother's highest level of education (source: SEHQ) collected as Completed some high school, Completed high school or equivalent, Completed vocational training (TAFE, trade cert, diploma), Completed a university degree, Other, Invalid response, Not stated
- School demographics - Index of Community Socio-Educational Advantage (ICSEA) value (source: Redcap)
- Parent report SDQ Scale Score (Source: Redcap)
- Teacher report SDQ Scale Score (Source: Redcap)
- EOI Reading scale score (Source: EOI)

In particular, SEHQ is a parent report instrument that records parents' concerns and observations about their child's health and wellbeing as they begin primary school in Victoria. The questionnaire was developed and piloted in 1996–97 as part of the Victorian School Nursing Redevelopment Program and has been distributed to parents and guardians of preparatory (Foundation) grade children in Victorian primary schools since mid-1997 (Department of Education & Training, 2014). The SEHQ includes domains including: general health, medications, immunizations, dental health, speech/language, hearing, vision, disabilities, general development, behaviour and emotional wellbeing, and family stress.

The SEHQ will be collected at baseline by the DE&T school nurses as a part of normal practice.

The EOI is a teacher administered online tool for assessing the English skills of students. The Interview assesses students across the three modes of English in AusVELS (the Australian Curriculum in Victoria) – Reading, Writing and Speaking and Listening. For the CPOL RCT we will use the Reading and Speaking and Listening domains of the Interview. The EOI is a one-to-one interview between a teacher and student, using texts and downloadable resources designed specifically for the Interview. Teachers record each student's responses directly onto the online system. This data are used to generate reports that provide a point-in-time overview of student achievement. The EOI will be administered at baseline by the classroom teachers of the CPOL RCT. It is a mandatory assessment for students in Foundation in Victorian Government Schools. The Catholic sector participating schools/teachers will administer the English Online assessments via hard copy at baseline (Department of Education & Training, 2014). The English Online Interview is not publically available and is the confidential property of the Victorian Department of Education and Training, and can therefore not be made available as an appendix.

# 5.  STATISTICAL METHODOLOGY

## 5.1.  GENERAL METHODOLOGY

Data analysis for the study will be performed by the Clinical Epidemiology and Biostatistics Unit (CEBU) at MCRI. Ms Francesca Orsini, an experienced biostatistician, will conduct the statistical analysis. As indicated above, this study will use a cluster randomization strategy in which each school is randomised to one treatment. Such a design, based on randomization of schools, but with collection of data at the student level, allows an analysis at the student level of randomised treatments. Statistical analysis will follow standard methods for randomised controlled cluster trials and the primary analysis will be by intention to treat (ITT), including all randomised participants where outcome data are available. All analyses will be repeated on using a PP analysis.

## 5.2.  BASELINE DATA

The baseline characteristics of the schools will be will be summarised by group. Categorical variables will be presented as the number and proportion in each category. Continuous variables will be presented as means and standard deviations (SDs), or medians and interquartile ranges (IQR) for skewed data, and the range.

The baseline characteristics of the children will be presented for each group using the mean, SD, median and interquartile ranges for continuous data and proportions for categorical data.

## 5.3.  SENSITIVITY ANALYSES

### Distributional assumptions

Outcomes with a distribution right skewed (as the SDQ is expected to be), will be log-transformed and the mixed-effects linear regression models will be run on the transformed version of the outcome.

### Handling of missing data

Frequency and patterns of missing data will be examined and sensitivity analyses will be performed to compare the results of analyses restricted to students with complete data and analyses where those with missing data are considered using multiple imputation. A single Multiple imputation models will be used to impute all of the missing data in all of the outcomes at

Grade 1 and Grade 3. 50 completed data sets will be imputed by chained equations including all the students initially randomised.

## 5.4. SUBGROUP ANALYSES
The four subgroup analyses are:
1. *Subgroup analysis 1 - Language and literacy vulnerable students.*
   This analysis will examine whether the intervention has differential effects for language and literacy vulnerable students versus not vulnerable students. Language and literacy vulnerable students are defined as those who either reported to have developmental delay at baseline or EOI Reading scale score lower than 22.8 (1.5*SD).
2. *Subgroup analysis 2 - students of 'very high' behavioural concern.*
   This analysis will examine whether the intervention has differential effects for students with very high behavioural concerns versus those without. These students are identified as having a SDQ parent reported total difficulties score reported as "Abnormal/of concern" at baseline (>= 17 for boys, >=15 for girls).
3. *Subgroup analysis 3 - Socio-Educational Disadvantaged students.*
   This analysis will examine whether the intervention has differential effects for students from *Socio-Educational Disadvantaged* families versus those without. These are defined as those students whose mother did not complete high school (or equivalent level of formation) as reported at baseline.
4. *Subgroup analysis 4 - Socio-Educational Disadvantaged schools.*
   This analysis will examine whether the intervention has differential effects in schools with an Index of Community Socio-Educational Advantage 1.5*SD below the mean versus those with an index greater than or equal the cut-off (1.5*SD below the mean).

## 5.5. REGRESSION MODELS

### Unadjusted model
Analysis of all continuous outcome collected either at grade 1 or 3, including the primary outcome, will adopt a <u>two level random effects linear regression model</u>. This mixed-effects linear regression model will include:

- a fixed effect for school sector (Government of Catholic)
- a fixed effect for intervention indicator
- a random intercept for school

The outcome means and standard deviations (SD) will be presented for each group along with the unadjusted mean difference between groups, 95% confidence interval (CI) and p-value from the mixed effects model.

The same mixed model will be run also on the NAPLAN standardized scores (Reading, Writing, Spelling, Grammar and Punctuation, Numeracy Score) to calculate the unadjusted effect sizes and their 95% CIs.

### Adjusted model A
Analysis on all continuous outcome collected at grade 1 and 3, including the primary outcome, will use the same two-level random effects linear regression model as specified for the primary

(FORM-STAT-04A-01/Version 1.0)

analysis but with the addition of a fixed effect for each of the following factors as potentially important confounders:

- Student's age when outcome is collected
- Gender
- LBOTE
- Presence of language or literacy difficulties at baseline (refer to subgroup analysis 1 below)
- Presence of 'very high' behavioural concern at baseline (refer to subgroup analysis 2 below)
- Whether or not student's mother did complete high school (refer to subgroup analysis 3 below)
- Whether or not school ICSEA < 1.5*SD below the mean (refer to subgroup analysis 4 below)
- Although underpowered to identify interactions, we will also consider the inclusion of a fixed effect for the interaction between intervention and school, which will be used to determine if there is evidence the intervention effect varies across Government and Catholic schools.

The results will be presented as adjusted mean difference between groups, 95% CI and p-value from the mixed effects models.

The same mixed model will be run also on the NAPLAN standardized scores (Reading, Writing, Spelling, Grammar and Punctuation, Numeracy Score) to calculate the adjusted effect sizes and their 95% CIs.

### Adjusted model B – Subgroup analyses

Four additional adjusted models will be run to explore potential heterogeneity of the treatment effect. They will include the covariates of adjusted model A, and will also include the following interaction terms, with a separate model for each:

- fixed effect for the interaction between intervention and language and literacy vulnerability status
- fixed effect for the interaction between intervention and presence of 'very high' behavioural concern
- fixed effect for the interaction between intervention and socio-educational disadvantaged student status
- fixed effect for the interaction between intervention and socio-educational disadvantaged school status

Through these models we will examine whether there is evidence that the intervention effect varies by subgroups. Should any of these interaction terms reveal evidence that the intervention effect varies between these subgroups, specific subgroup estimates and confidence intervals will be presented, obtained running the unadjusted model and the adjusted model A specified above. As we have not powered the trial to consider subgroups, these analyses are considered exploratory.

## 5.6. PRIMARY DATA ANALYSIS: NAPLAN Reading score at grade 3

*Hypothesis: grade 3 intervention children will demonstrate higher NAPLAN Reading scores compared with the usual care (teaching) group.*

Analyses will be carried out on the ITT and PP.

| Outcome | Time point | Models used for the analysis | Additional analysis | Sensitivity analysis |
|---|---|---|---|---|
| - NAPLAN Reading Score (raw and standardized scored) | Grade 3 | Unadjusted Model (*Primary Analysis*)<br><br>Adjusted Model A<br><br>Adjusted model B | The proportion of students with a raw score below the national minimum standard (band 1), at the national minimum standard (band 2) and above the national minimum standard (bands 3 to 6) will also be calculated and presented by group. | Per Protocol analysis.<br><br>Multiple Imputation analysis. |

## 5.7. SECONDARY DATA ANALYSES

Analyses of the secondary outcomes will be carried out on the ITT and PP.

Adjusted Model with interaction term(s)

| Secondary Outcome | Time point | Models used for the analysis | Additional analysis | Sensitivity analysis |
|---|---|---|---|---|
| - NAPLAN Writing Score<br>- NAPLAN Language Conventions: Spelling<br>- NAPLAN Language Conventions: Grammar and Punctuation<br>- NAPLAN Numeracy Score<br>(raw and standardized scored) | Grade 3 | Unadjusted Model<br><br>Adjusted Model A<br><br>Adjusted model B | The proportion of students with raw scores below the national minimum standard (band 1), at the national minimum standard (band 2) and above the national minimum standard (bands 3 to 6) will also be calculated and presented by group. | Per Protocol analysis.<br><br>Multiple Imputation analysis. |
| - RPT Standardised Score | Grade 1 | Unadjusted Model<br><br>Adjusted Model A<br><br>Adjusted model B | The proportion of students `well below average' and `below average' will also be calculated and presented by group. | Per Protocol analysis.<br><br>Multiple Imputation analysis. |
| - CELF Scales Score<br>- NPVT Age-Corrected Standard Score<br>- Narrative/Expressive Language Number of T-Units<br>- Narrative analysis<br>- Story grammar analysis<br>- Total conjunctions | Grade 1 | Unadjusted Model<br><br>Adjusted Model A<br><br>Adjusted model B | N.A. | Per Protocol analysis.<br><br>Multiple Imputation analysis. |
| - SDQ Teachers Reported Total Difficulties Score | Grade 1 | Unadjusted Model<br><br>Adjusted Model A<br>+<br>SDQ Teachers Reported Total Difficulties Score at baseline<br><br>Adjusted model B<br>+<br>SDQ Teachers Reported Total Difficulties Score at baseline | Mean and SD of female and male normal scores will be calculated and presented by group.<br><br>The proportion of female and male students of 'borderline concern' and 'abnormal/of concern' will also be calculated and presented by group. | Per Protocol analysis.<br><br>Multiple Imputation analysis.<br><br>Log-transformation. |

(FORM-STAT-04A-01/Version 1.0)

## Appendix

| 1 | "Mellor, David 2005,<br>Normative data for the strengths and difficulties questionnaire in Australia, Australian psychologist,<br>vol. 40, no. 3, pp. 215-222." | Appendix_1_mellor-normativedata-2005.pdf |
|---|---|---|
| 2 | "Reading Progress Test - Australian Norms Supplement"<br><br>Raw scores are converted to standard scores (see table 8, page 10 "Norms For the Conversion of Raw Scores to Standardised Scores: Reading Progress Test 1, Pre-Year 1 and Year 1 Levels - using End of Year 1 column) | Appendix_2_RPT -norms-rpt1.pdf |
| 3 | "The CELF Examiner's Manual - Semel,E.M., Wiig, E.H., & Secord, W. (2003).<br><br>The Clinical Evaluation of Language Fundamentals – 4th Edition (Australian Standardisation) – Concepts and Following Directions subtest. Examiner's Manual. Harcourt Assessment: Marrickville, NSW"<br>To convert raw scores to scaled scores, we used the following tables that can be found in Appendix B of the manual:<br>Table B 6:0 – 6:5 (p 276)<br>Table B 6:6 - 6:11 (p.277)<br>Table B 7:0 - 7:11 (p.278)<br>Table B 8:0 - 8:11 (p.279)<br>Table B 9:0 - 9:11 (p.280)<br>Table B 10:0 - 10:11 (p.281) | Appendix_3_CELF.pdf |
| 4 | "The NIH Toolbox® Scoring and Interpretation Guide for the iPad"<br>Pages 2-3: scoring generally.<br>Page 5: Picture Vocabulary section. | Appendix_4_Toolbox_Scoring_and_Interpretation_Guide_for_iPad_v1.7_original.pdf |
| 5 | "Published Appendices"<br><br>To convert our raw scores to a scaled score, we used the second table in the document, called "Raw to Scaled Score (ss) for use in Age-corrected Formulas: Children (3-17 years old)". To then convert to Age-corrected standard scores, we need to use the table further down called "Appendix 2.2 Scaled Scores to Age-Corrected Standard Scores". There is a specific formula for "Picture Vocabulary":<br><br>$$100 + \cfrac{15 * \cfrac{\text{Vocabulary scaled score} - \left[11.41 - 0.57 * \left(\frac{age}{100}\right)^{-1}\right]}{1.86 + 1.12 * \frac{age}{100}}}{1.24}$$ | Appendix_5_Published Appendices.docx |

SIGNATURES PAGE

Signature of Principal
Investigator:
Print Name               A/Prof Sharon Goldfeld          Date 8 - 1 - 2018

Signature of Trial Statistician:              Date 19-12-2017
Print Name               Francesca Orsini