Glasses in Classes
Statistical Analysis Plan

Evaluator (institution): University of Nottingham
Principal investigator(s): Professor Roisin P. Corcoran

SAP last updated: January 2020

PROJECT TITLE	Glasses in classes: A cluster-randomised controlled trial to evaluate the effects of a school-based intervention to improve academic achievement, visual acuity, and adherence to glasses wear in young children in a disadvantaged multi-ethnic community
DEVELOPER (INSTITUTION)	NHS Bradford Teaching Hospitals Foundation Trust
EVALUATOR (INSTITUTION)	University of Nottingham
PRINCIPAL INVESTIGATOR(S)	Professor Roisin P. Corcoran
PROTOCOL AUTHOR(S)	Professor Roisin P. Corcoran, Dr Michael Adkins
TRIAL DESIGN	Two-arm cluster randomised controlled trial with random allocation at school level
TRIAL TYPE	Efficacy
PUPIL AGE RANGE AND KEY STAGE	Reception (4–5year olds), Early Years Foundation Stage
NUMBER OF SCHOOLS	100
NUMBER OF PUPILS	700
PRIMARY OUTCOME MEASURE AND SOURCE	Reading achievement – Woodcock Johnson IV Letter-Word Identification Subscale
SECONDARY OUTCOME MEASURE AND SOURCE	Reading achievement – Woodcock Johnson IV Word Attack; mathematics achievement – Woodcock Johnson IV Applied Problems; And visual acuity – logMAR

SAP version history

VERSION	DATE	REASON FOR REVISION
1.0 [*original*]	20/4/20	*Published SAP version*

Glasses in Classes

Statistical Analysis Plan

Evaluator (institution): University of Nottingham
Principal investigator(s): Professor Roisin P. Corcoran

SAP last updated: January 2020

Contents

Glasses in Classes
Statistical Analysis Plan

Evaluator (institution): University of Nottingham
Principal investigator(s): Professor Roisin P. Corcoran

SAP last updated: January 2020

Introduction

Eyesight development in children occurs within the first 7–8 years of life, with reduced VA in children indicating potential conditions including *refractive* error, amblyopia and/or strabismus (Bruce et al., 2018a; Daw, 1998; Dobson, 1993). There is growing consensus that vision problems may be a potentially treatable component of mathematics and reading difficulty (Collins et al., 2016; Granet, 2011; Kiely, Crewther, & Crewther, 2001; Levine, 1984; Lubkin, 1968; Solan et al., 2004). As part of the Child Health Promotion programme (Committee, 2009), the UK NSC recommends visual screening for children during their first year of school entry, with glasses wear being the principal treatment recommended for reduced vision. Children who fail to attend follow-up ophthalmic examinations and those who fail to adhere to glasses wear are unlikely to improve their level of VA, affecting their early reading and mathematics (Bruce et al., 2018a).

The purpose of this study is to examine the impact of a school-based intervention to improve glasses wear in children (which involves sharing vision screening results with school and provision of additional glasses to be kept in school) on their reading and mathematics achievement. The causal mechanisms of this effect such as attendance for eye appointments, adherence to glasses wear in young children following vision screening, and improvement in VA will also be examined. The effect of the intervention on academic achievement and VA in the child's first year (reception class) of school will be measured. This cluster randomised study will consist of two groups. The treatment group (50 schools), with approximately 350 pupils in need of glasses, will be randomised to receive the intervention over the academic year; the control group (50 schools) will receive business-as-usual care.

Ophthalmic treatment for the children participating in the trial will not change. However, the intervention schools will receive additional school-based support to promote glasses wear. This intervention has not been tested in the UK using a rigorous RCT approach although elements of the intervention have been studied previously within the Bradford setting (Bruce & Outwaite, 2013; Bruce et al., 2018a; Bruce, Sanders, & Sheldon, 2018b; Cassetti, Sanders, & Bruce 2019).

This statistical analysis plan outlines the analysis planned, discusses the study design and provides sample size calculations. It also addresses the primary and secondary outcome analyses, sub-group analysis, the additional mediation and longitudinal follow-up analyses, the handling of missing data and noncompliance issues, and finally effect size calculations.

Glasses in Classes
Statistical Analysis Plan

Evaluator (institution): University of Nottingham
Principal investigator(s): Professor Roisin P. Corcoran

SAP last updated: January 2020

Design overview

Table 1: GiC Design Parameters

Trial design, including number of arms		Two-arm, cluster randomised
Unit of randomisation		School
Stratification variables (if applicable)		None
Primary outcome	variable	Reading achievement
	measure (instrument, scale, source)	Woodcock-Johnson IV Letter-Word Identification (Continuous)
Secondary outcome(s)	variable(s)	Reading achievement, mathematics achievement and visual acuity
	measure(s) (instrument, scale, source)	Woodcock-Johnson IV Word Attack (continuous); Woodcock-Johnson IV Applied Problems (continuous); logMAR (continuous)
Baseline for primary outcome	**variable**	Reading achievement
	measure (instrument, scale, source)	Woodcock-Johnson IV Letter-Word Identification (Continuous)
Baseline for secondary outcome	**variable**	Reading achievement, mathematics achievement and visual acuity
	measure (instrument, scale, source)	Woodcock-Johnson IV Word Attack (continuous); Woodcock-Johnson IV Applied Problems (continuous); logMAR (continuous)

Sample size calculations overview

The trial has been designed to maximise the possibility of detecting an effect size of 0.2, within a very small geographical area of the Bradford Metropolitan area. The power analysis involved a number of sensitivity analyses conducted with a range of assumptions (varying ICCs, pre-post correlation, number of schools etc) and software – in particular MLPowSim (Browne, Golalizadeh Lahi, & Parker, 2009), PowerUpR (Bulus et al., 2018), and Optimal Design (Raudenbush, 2011). The analyses discussed below use PowerUpR which provides a more precise estimate than the more limited options available in Optimal Design. Power calculations are detailed in **Error! Reference source not found.** below.

Table 2: GiC Sample Size Estimates

		Protocol		Randomisation	
		OVERALL	FSM	OVERALL	FSM
Minimum Detectable Effect Size (MDES)		0.195	0.22	0.194	
Pre-test/ post-test correlations	level 1 (pupil)	0.88[1]	0.88	0.88	0.88
	level 2 (class)	-	-	-	-
	level 3 (school)	0.62	0.62	0.62	0.62
Intracluster correlations (ICCs)	level 2 (class)	-	-	-	-
	level 3 (school)	0.15[2]	0.15	0.15	0.15
Alpha		0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		7	3	8	
Number of schools	intervention	50	50	50	50
	control	50	50	49	49
	total	100	100	99	99
	intervention	350	150	406	

[1] Villarreal (2015) undertook a test review of the WJIV standard battery of tests and found correlations in the range of .83-.95. In the Woodcock Johnson IV manual, test-retest correlations were between 0.83-0.95 for the age 7-11 group (p.94). However, these sort of test-re-test reliability analyses tend to be over very short periods (e.g. one day).

[2] We have selected an ICC of 0.15 as this represents a trade-off in that the schools are centred on a specific small geographical region, but on the other hand recognises that some EEF trials amongst Early Years have been higher at 0.17-0.19.

		Protocol		Randomisation	
		OVERALL	FSM	OVERALL	FSM
Number of pupils	control	350	150	387	
	total	700	300	793	

With the above parameters, at the protocol stage, we estimated that it is possible to achieve an MDES of 0.195 for the main effect and 0.22 for FSM subgroup analysis. During recruitment we found that the number of children failing their initial vision screening was slightly higher than expected giving an average cluster size of 8 children in the study. Despite recruiting one school less than planned, the MDES remained effectively stationary at 0.194. The FSM estimate at randomisation will be updated once we receive the matched data from the DfE containing the FSM indicator.

Analysis

The analysis will examine the impact of the Glasses in Classes intervention in comparison to the business-as-usual control condition on the basis of intention to treat (ITT) using a two-level multilevel model using Maximum Likelihood estimation.

Error terms in linear regression are assumed to be independent. However, pupils are clustered within schools and are considered more similar than those pupils attending other schools, violating this key assumption. Multilevel modelling allows us to relax the independence assumption and represents a compromise approach between complete pooling approaches where all schools are treated the same within the model, and no pooling approaches where a separate model is fitted for each school. Multilevel modelling allows partial pooling where the effect of schools with little data are pooled towards the mean (Gelman & Hill, 2007, p252-259). Practically speaking multilevel models tend to provide the similar point estimates to complete pooling models, but have larger more appropriate standard errors and confidence intervals.

Primary outcome analysis

As discussed in table 1 above we will investigate the main outcome – the impact on reading achievement as measured by Woodcock Johnson IV Letter-Word Identification subscale. We will use the standardised score for both post and pre-test. As randomisation took place at the school-level, the post-test score will be modelled using a varying intercept model which is presented below follows:

$$y_{ij} = \beta_{0j} + \beta_1 Treatment_{ij} + \beta_2 Pre-Test_{ij} + u_{0j} + \varepsilon_{ij}$$

$$u_{0j} \sim N\,(0, \sigma^2_{school})$$

$$\varepsilon_{ij} \sim N\,(0, \sigma^2_{y})$$

This can be understood as follows. The post-test score for the i^{th} student in the j^{th} school is equal to the grand mean score (β_{0j}), the impact of a binary indicator denoting treatment received (β_1) which is coded as 0 or 1, the impact of the mean-centred normally distributed pre-test (β_2), the school-level error term (u_{0j}), and finally the student-level error term (ϵ_{ij}). The two error terms each receive their own probability distribution which are assumed to be normally distributed and centred on 0, with the two variance parameters estimated from the data (σ^2_{school} and σ^2_{y}).

Secondary outcome analysis

The three secondary outcomes: Woodcock-Johnson IV Word Attack subscale standardised score, Woodcock-Johnson IV Applied Problems subscale standardised score, and logMAR (to measure VA) will be modelled in the same manner as the primary analysis on the basis of intention to treat and estimate effect sizes using the same formula presented below. As per the EEF guidelines (2020), regardless of whether there is a significant effect on both the Woodcock-Johnson and VA measures, we will conduct a follow-on analysis investigating the mediating impact of VA on academic outcomes.

Subgroup analyses

An additional multilevel interaction model following the form presented below to examine the impact of the intervention on Free School Meal recipients. This will use the Free School Meals flag in the NPD rather than FSMever as the children involved are in the reception year of primary school.

$$y_{ij} = \beta_{0j} + \beta_1 Treatment_{ij} + \beta_2 Pre\text{-}Test_{ij} + \beta_3 FSM_{ij} + \beta_4 Treatment*FSM + u_{0j} + \varepsilon_{ij}$$

$$u_{0j} \sim N\,(0, \sigma^2_{school})$$

$$\varepsilon_{ij} \sim N\,(0, \sigma^2_{y})$$

In addition, a subgroup analysis of those pupils who receive and do not receive FSM will be analysed separately, regardless of statistical significance of the interaction effect.

Additional analyses

Should there be a significant treatment effect on both the Woodcock-Johnson and VA measures, we will conduct a follow-on analysis investigating the mediating impact of VA on academic outcomes. This three-stage process for the each of primary and secondary academic outcomes will take the following form, where y_1 refers to the Woodcock-Johnson subscale and y_2 refers to the logMAR visual acuity measure:

$$y_{1ij} = \beta_{0j} + \beta_1 Treatment_{ij} + \beta_2 Pre\text{-}Test_{ij} + u_{0j} + \varepsilon_{ij}$$
$$y_{2ij} = \beta_{0j} + \beta_1 Treatment_{ij} + \beta_2 Pre\text{-}Test_{ij} + u_{0j} + \varepsilon_{ij}$$

$$y_{1ij} = \beta_{0j} + \beta_1 Treatment_{ij} + \beta_2 Pre\text{-}Test_{ij} + \beta_3 VisualAcuity_{ij} + u_{0j} + \varepsilon_{ij}$$

$$u_{0j} \sim N\,(0, \sigma^2_{school})$$

$$\varepsilon_{ij} \sim N\,(0, \sigma^2_{y})$$

As part of a sensitivity analysis we will fit adapt the primary model to include the pupil-level pre-test score, but centred around school means, and include an additional school-level average pre-test score at level 2. This follows the advice of Hedges and Hedberg (2013, p450) and Bafumi and Gelman (2006) to reduce the correlation between the pre-test and the school-level units which can cause poor estimates of uncertainty in the parameters.

Longitudinal follow-up analyses[3]

There has been some discussion of conducting a further study on the longitudinal outcomes of Glasses in Classes at the stage of the KS1 examinations. The results of these tests could be collected from schools or requested from the NPD. At this stage it remains an option and

[3] Please see the longitudinal analysis guidance.

if approved, an addendum would be added to the analysis plan and the final report. The model will remain the same format as the primary analysis.

Imbalance at baseline

While we will focus on the imbalance for the primary measure (the Woodcock Johnson IV Letter-Word ID) in terms of pupil (baseline pre-test and FSM status) and school characteristics (Ofsted rating, rural-urban classification, percentage of FSM pupils), we will provide supplementary imbalance characteristics for the three further secondary measures – the Woodcock Johnson IV Applied Problems and Word Attack subscales, and the visual acuity logMAR measure. This will be presented as a cross-tabulation with means, standard deviations and effect sizes for continuous variables and counts and percentages for categorical variables.

Missing data

As per EEF standards, headline figures will be presented throughout. In addition, as set out in Corcoran (2016, p.69) we will carry out an analysis to examine the overall and differential attrition rates that may bias the estimated impact of the Glasses in Classes treatment condition. However, there is an expectation that overall attrition would be less than 10% and the difference between the treatment and control schools less than 5%. These thresholds being within acceptable standards for WWC guidelines (WWC Standards and Procedures Handbook, Version 4.1). This expectation for low attrition and potential for bias is due to random factors and the primary reason for pupil attrition being independent of the treatment, such as families moving. However, attrition at the pupil level will be monitored and the structure of the missing data will be assessed using the MCAR test (Little, 1988). If the structure of the missing data is not missing completely at random (MCAR) or the combination of a non-MCAR test results and pupil overall attrition that is above a 5% threshold, then multiple imputation methods will be considered for a sensitivity analysis (Little and Rubin, 2002).

Compliance

The comparison group variable will be used as the instrumental variable to predict compliance which is a dichotomously measured indicator that distinguishes between the randomly assigned treatment and comparison students. The use of the comparison group is standard per WWC guidelines (WWC Standards Handbook, Version 4.1, p. 46). The variable that will be used as the endogenous independent variable that estimates compliance will be a dichotomous variable that indicates full fidelity vs incomplete fidelity of implementation. The fidelity variable could be derived from two fidelity sub-domains: child attendance with optometrist with prescription glasses and whether the pupil had a full set (two pairs of glasses) compared to a single set or no glasses. These variables could be combined into an index and then dichotomised based on full fidelity. Full fidelity and optimal compliance would consist of those pupils that had greater than 80% attendance (percent of total optometrist sessions attended) with the optometrist and had a full set of prescription glasses. The use of the compliance variable as a dichotomous endogenous independent variable is consistent with WWC guidelines (WWC Standards Handbook, Version 4.1, p. 46). The level at which compliance is measured is at the pupil level.

To assess treatment effects in the presence of non-compliance, CACE analysis will be conducted using a Two-stage least squares (2SLS) analysis that utilizes asymptotic standard errors. The first-stage of the analysis will be specified as:

$$Compliance\ y_i = \beta_{0i} + \beta_1 Treatment_i + \beta_2 Pretest\ Outcome_i + \varepsilon_i$$

Where, i is notation for all individuals in the sample, Compliance y is the endogenous independent variable that measures compliance with the treatment measured as a dichotomous variable, Treatment is a dichotomous variable that captures the random assignment of students to treatment or control groups by school, the Pretest outcome is the pretest measure of the outcome variable, and ε is the error term. This first stage equation is

estimated using OLS and predicted values of the compliance variable will be generated from these estimates and denoted as Compliance $\hat{y}_i$ for the final stage of analysis. The predicted compliance indicator will measure the contrast between those intervention pupils with optimal compliance to the intervention and those intervention pupils with minimum compliance to the intervention in addition to control pupils.

The structural equation with the final stage will be specified as:

$$Posttest\ Outcome\ y_i = \beta_{0i} + \beta_1 Compliance\ \hat{y}_i + \beta_2 Pretest\ Outcome_i + \varepsilon_i$$

Where, Outcome y is the posttest outcome of interest, Compliance $\hat{y}$ is the predicted values of the endogenous independent variable that measures compliance from the first stage of analysis, Treatment is a dichotomous variable that captures the random assignment of students to treatment or control groups by school, Pretest outcome is the pretest measure of the outcome variable, and ε is the error term.

In terms of measuring compliance properly, if the theoretical thresholds for both the attendance with optometrists and pairs of glasses received do not exceed adequate thresholds of the intervention group sample, this may suggest that the theoretical thresholds for attendance may be too strict. In that instance, a Jenks natural breaks optimisation will be conducted on the attendance fidelity measure using raw percentage of attendance (Jenks 1967). This will be done to determine an adequate classification of attendance and inclusion for the purposes of compliance with the study intervention.

In sum, this CACE analysis will be used to assess the potential impact of random assignment and other covariates on compliance with the treatment. These estimates in turn will be used along with all covariates to predict the outcome of interest and therefore provide an unbiased estimate of the impact of compliance and non-compliance on the outcome. Hedge's g will be calculated to determine the effect size. Statistically insignificant and non-meaningful effects of the compliance variable will indicate that noncompliance in the study did not have an impact on the outcomes.

Intra-cluster correlations (ICCs)

Intra-cluster correlations (ICCs) will be calculated from a multilevel null model (i.e. one without the treatment or pre-test covariates, but including the school level) for the primary reading achievement outcome. The formula is presented below. As part of the appendices we will also provide ICC calculations for the secondary reading, mathematics and visual acuity outcomes.

$$ICC = \frac{\sigma^2_{school}}{(\sigma^2_{school} + \sigma^2_y)}$$

Effect size calculation

EEF standard practice will be followed in reporting effect sizes (Hedges *g*). The formula is presented below:

$$ES = \frac{\bar{Y}^t - \bar{Y}^c}{\sqrt{(\sigma^2_{school} + \sigma^2_y)}}$$

Effect size quantities will be computed from the lme4 using the sim() function in the Applied Regression Modelling package (arm) in R to generate a posterior distribution from which 95% confidence intervals can be calculated.

References

Abayomi, K., Gelman, A. and Levy, M. (2008), Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 57*, 273-291. https://doi.org/10.1111/j.1467-9876.2007.00613.x

Bafumi, J. & Gelman, A. (2006) Fitting multilevel models when predictors and group effects correlate. *2006 Annual Meeting of the Midwest Political Science Association, Chicago, Il.* http://www.stat.columbia.edu/~gelman/research/unpublished/Bafumi_Gelman_Midwest06.pdf

Bulus, M., Dong, N., Kelcey, B., & Spybrook, J. (2018). *PowerUpR: Power analysis tools for multilevel randomized experiments*. R package version 1.0.2. https://CRAN.R-project.org/package=PowerUpR

Browne, WJ., Golalizadeh Lahi, M., Parker, RMA. "A guide to sample size Calculations for random effect models via simulation and the MLPowSim software package", School of Clinical Vetinary Sciences, University of Bristol. https://seis.bristol.ac.uk/~frwjb/esrc/MLPOWSIMmanual.pdf

Bruce, A., Kelly, B., Chambers, B., et al. (2018). "The effect of adherence to spectacle wear on early developing literacy: a longitudinal study based in a large multiethnic city". *BMJ Open, 8*, 6:e021277. https://doi.org/10.1136/bmjopen-2017-021277

Bruce, A., & Outhwaite, L. (2013). Uptake, referral and attendance: Results from an inner city school based vision screening programme. *Br Ir Orthopt J*, *10*, 42-46. https://doi.org/10.22599/bioj.71

Bruce, A., Sanders, T., & Sheldon, T.A. (2018). "Qualitative study investigating the perceptions of parents of children who failed vision screening at the age of 4–5 years" *BMJ Paediatrics Open, 2*:e000307. https://doi.org/10.1136/bmjpo-2018-000307

Cassetti, V., Sanders, T. and Bruce, A., 2019. Challenges of eye health care in children and strategies to improve treatment uptake: A qualitative study from the perspective of eye care professionals in the UK. *British and Irish Orthoptic Journal, 15*(1), 96–104. https://doi.org/10.22599/bioj.133

Collins, M. E., Mudie, L., Slavin, R. E., Corcoran, R. P., Owoeye, J., Chang, D., Friedman, D. S., Repka, M. X. (2016). Prevalence of eye disease and reading difficulty in an inner city elementary school population—preliminary results of the Baltimore Reading and Eye Disease Study (BREDS). *Journal of American Association for Pediatric Ophthalmology and Strabismus, 20*, 29-30. https://doi.org/10.1016/j.jaapos.2016.07.112

Corcoran, R.P. (2016) "Principals on the path to excellence: Longitudinal, multisite cluster-randomized controlled trials of the National Institute for School Leadership's Executive Development Program" *International Journal of Educational Research, 79*, 65-75. https://10.1016/j.ijer.2016.05.001

Education Endowment Foundation (2020). Statement on statistical significance and uncertainty of impact estimates for EEF evaluations. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/Statement_on_statistical_significance_and_uncertainty_of_impact_estimates_for_EEF_evaluations

Education Endowment Foundation (2018). EEF statistical analysis guidance. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

Granet, D. B. (2011). Learning disabilities, dyslexia, and vision: The role of the pediatric ophthalmologist. *Journal of American Association for Pediatric Ophthalmology and Strabismus, 15,* 119-20. https://doi.org/10.1016/j.jaapos.2011.03.003

Gelman, A., & Hill, J. (2007) *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32* (4), 341–370. https://doi.org/10.3102/1076998606298043

Hedges, L.V. & Hedberg, E.C. (2013) Intraclass correlations and covariate outcome correlations for planning two- and three-level cluster-randomized experiments in

education, *Evaluation Review* 37 (6), 445-489. https://doi.org/10.1177/0193841x14529126
Jenks, George F. 1967. "The data model concept in statistical mapping", *International Yearbook of Cartography 7*, 186–190.
Kiely, P.M., Crewther, S.G., Crewther, D.P. (2001). Is there an association between functional vision loss and learning to read? *Clinical and Experimental Optometry, 84* (6), 346-53. https://10.1111/j.1444-0938.2001.tb06606.x
Levine, M.D. (1984). Reading disability: Do the eyes have it? *Pediatrics*, *73*, 869-71.
Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198–1202. https://doi.org/10.1080/01621459.1988.10478722
Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New York: Wiley.
Lubkin V. (1984). The ophthalmologist and the reading problem. Symposium on reading disability. *Bulletin of New York Academy of Medicine, 44* (4), 459-69.
Raudenbush, S.W. (2011). *Optimal design software for multi-level and longitudinal research* (Version 3.01) [Software]. William T. Grant Foundation.
Solan, H. A., Shelley-Tremblay, J., Hansen, P. C., Silverman, M. E., Larsone, S., & Ficarra, A. (2004). M-cell deficit and reading disability: a preliminary study of the effects of temporal vision-processing therapy. *Optometry-Journal of the American Optometric Association*, *75* (10), 640-650. https://doi.org/10.1177/00222194070400030701
Villarreal, V. (2015) Test Review: Woodcock-Johnson IV Tests of Achievement. *Journal of Psychoeducational Assessment, 33,* 391-398. https://doi.org/10.1177/0734282915569447
What Works Clearinghouse (WWC) (2020). WWC procedures and standards handbook (version 4.1). Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Handbooks