

Article

Design Plan for an Evolution Study of Related Characteristics of a Population

Juan M. Rodríguez-Díaz ^{1,*} , Rosa E. Pruneda ²  and Mercedes Rodríguez-Hernández ³ ¹ Department of Statistics, Universidad de Salamanca, 37008 Salamanca, Spain² Department of Mathematics, University of Castilla-La Mancha, 13071 Ciudad Real, Spain; rosa.pruneda@uclm.es³ Department of Didactic of Mathematics and Experimental Sciences, Universidad de Salamanca, 05003 Ávila, Spain; mercedes.rodriguez@usal.es

* Correspondence: juanmrod@usal.es

Abstract: The objective is to study the evolution of different characteristics of a population through time. These response variables may be related for each experimental unit, and in addition, the observations for each response may as well be correlated with time, producing a complex correlation structure. The number of responses that can be observed is usually limited for budget, resources, or time reasons, and thus the selection of the most informative time points when data must be taken is quite convenient. This will be performed by using the optimal design of experiments techniques. Some analytical results will be shown, and the results will be applied to obtain the most convenient points when tests about two variables related with the capability of the resolution of mathematical problems in primary school students should be performed.

Keywords: covariance structure; evolution study; mental representation; optimal design of experiments; problem solving

MSC: 62K05

Citation: Rodríguez-Díaz, J.M.; Pruneda, R.E.; Rodríguez-Hernández, M. Design Plan for an Evolution Study of Related Characteristics of a Population. *Mathematics* **2022**, *10*, 792. <https://doi.org/10.3390/math10050792>

Academic Editor: Radi Romansky

Received: 9 January 2022

Accepted: 28 February 2022

Published: 2 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

When studying several characteristics from a group of experimental units, one of the main concerns is the independence of the obtained data, since most of the usual techniques assume independence between the observations. However, this assumption may be difficult to maintain when observing different variables of a specific subject through time. In this situation, at least two kind of covariance structures may arise, one between the different types of observations taken on the same subject at every temporal point and another one between the values of a specific type of characteristic observed through time [1,2]. This may be a complex problem even when only one experimental unit is observed. Considering a group of units in the study means the addition of another layer to the structure, a problem that will be addressed in the present work.

The origin was the study of small children's capacity of solving mathematical problems and the evolution of this skill in students through the years they spent in primary school. This variable is very often related with some others that could be observed as well in the same students in order to check the kind of relationship or estimate the model linking the variables. However, obtaining the data has certain costs, especially those coming from the time that teachers and administrative personal will spend in designing and performing the tests, correction, the translation of the outcomes to a computer program, and analyzing the results. It will be very difficult to have these tests performed very often, or even once a year. For this reason, a design plan stating the conditions under which these observations should be taken is quite convenient.

In this work, such designs will be studied for a general case in which some variables will be observed on a cohort of subjects at different time periods, trying to decide the most informative time-points where observations/tests should be collected using optimal experimental design techniques.

In Section 2, the main concepts and latest results of the optimal design of experiments theory applied to models describing several response variables for one subject will be outlined, and new analytical results for a general case with N subjects and k variables of interest will be obtained. In Section 3, the theory will be applied to the joint study of the variables ‘Ability in problem solving’ (PS) and ‘Mental representation capacity’ (MR) of the primary school students [3], and the results will be commented on in Section 4. Finally, the main conclusions and further applications will be addressed in Section 5.

2. Background and New Results

Let x denote the experimental conditions, with possible values in the experimental domain X . An exact design ξ is a collection of points $\{x_1, \dots, x_n\}$ where samples are to be taken. For n observations $Y = \{y_1, \dots, y_n\}^T$ of the one-response linear model $y = f(x)^T \beta + u$, the system can be expressed in matrix notation as

$$Y = X\beta + U, \quad (1)$$

where β is the m -parameter vector, $U = \{u_1, \dots, u_n\}^T$ the vector of error terms, and $X = (f(x_1), \dots, f(x_n))^T$ the design matrix, with $f(x) = (f_1(x), \dots, f_m(x))^T$ and the $f_i(x)$ linearly independent in X . When there exists a correlation structure $\Sigma = \text{Var}(Y)$ between the samples, the estimator of the parameters for normal-distributed errors is $\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} Y$. One of the most important tools in the optimal design of experiments framework is the information matrix of the design ξ ;

$$M(\xi) = X^T \Sigma^{-1} X, \quad (2)$$

which is proportional to the inverse of the variance of $\hat{\beta}$. Usually the objective of practitioners is to find *optimal designs* that produce precise estimators of the model parameters, thus minimizing $\text{Var}(\hat{\beta})$ or equivalently maximizing $M(\xi)$. Different *criterion functions* may be considered, usually convex functions of $M^{-1}(\xi)$. The most popular criterion is D -optimality, which pays attention to the determinant of $M^{-1}(\xi)$, and that will be the criterion used in this work. When a design is D -optimal, it minimizes the volume of the confidence ellipsoid of the estimators of the model parameters. For non-linear models, the information matrix will depend on the unknown parameters. In this case, nominal values are needed for them, and *locally optimal* designs will be obtained. Reference books on the topic are, for instance, [4–6]. To compute optimal designs, from the analytical expression of the model the linearized version is obtained by computing the derivatives with respect to the parameters. When it is not possible to obtain the analytical expression, alternative methods for computing the derivatives can be employed [7].

When the observations are correlated, the size of the design (number of samples to be taken) should be decided in advance. Furthermore, for an evolution study, the design variable is time, and thus it will be assumed that $x_i \neq x_j$ for all i, j because there is no reason to take several observations on the same subject at the same time. When different responses are of interest (*multiresponse models*), the usual assumption was to consider that the k responses observed on the same experimental unit under the conditions x , $y(x) = (y_1(x), \dots, y_k(x))^T$ were correlated, but the measures taken at different points, $y(x)$ and $y(x')$, were independent. However, when the design variable is time, this last assumption is no longer valid [2], and two types of correlation should be considered: a *static* or *intra* covariance structure between different responses observed at the same time and

a *longitudinal* or *inter* correlation between the same type of response obtained at different times.

Thus, let us assume that we observe k characteristics of one subject at different times, and for each time t let $\mathbf{S}(t)$ denote the covariance matrix of the *sample* $\mathbf{y}(t) = (y_1(t), \dots, y_k(t))$, that is, $\mathbf{S}(t) = \text{cov}[y_1(t), \dots, y_k(t)]$. It is usual to assume that the relation between the different variables is similar for every t ; thus, following that, a constant covariance S will be considered (intra correlation). On the other hand, the covariance between the same type of observations taken at different points will be assumed to be dependent only on the distance between points, that is,

$$\text{Cov}[y_i(t_j), y_i(t_l)] = \rho(|t_j - t_l|), \quad i = 1, \dots, k; \quad j, l = 1, \dots, n,$$

where ρ is a known stationary covariance kernel (different proposals of these functions can be found, for instance, in [8,9]; thus, the longitudinal covariance will be the same for the k different responses, $\mathbf{R} = \text{Cov}[y_i(t_1), \dots, y_i(t_n)] \forall i$. The assumption of a known covariance function may be a controversial issue; however, it becomes more acceptable when restricting to a local consideration of the model, thus finally obtaining locally optimal designs.

In previous studies [1,2], the double (inter and intra) covariance structure has been taken into account for measures over a single experimental unit through time. Now, a multisubject scenario will be considered, observing N subjects at different times $t_1, \dots, t_n$, (design ξ). Balanced designs will be considered, that is, for each t_i in ξ several variables, $Y_1, \dots, Y_k$ will be observed for every subject; for unbalanced designs, a procedure similar to the one employed in Example 2 of [1] can be employed. The aim will be to select the most informative design for the models describing the evolution of the variables.

The Kronecker product of the different covariance structures can be used to express the covariance matrix of the $N \times k \times n$ observations. Throughout this work, a non-trivial $\mathbf{R}$ (but constant for every variable k) will be assumed; therefore, for each subject it will be more convenient to use the order $y_1(t_1), \dots, y_1(t_n), \dots, y_k(t_1), \dots, y_k(t_n)$, getting the covariance matrix $\Sigma_0 = \mathbf{S} \otimes \mathbf{R}$. Furthermore, it is quite usual to assume as well that the N subjects are independent, thus the covariance matrix of the $N \times k \times n$ observations can be expressed as

$$\Sigma = \mathbf{I}_N \otimes \mathbf{S} \otimes \mathbf{R} = \mathbf{I}_N \otimes \Sigma_0, \quad (3)$$

where $\mathbf{I}_N$ is the identity matrix of order N . That is, Σ is the block-diagonal (BD) matrix $BD\{\Sigma_0, \dots, \Sigma_0\}$.

- Let us now fix the notation and the models that will be studied for the case of several subjects: let $y_{ij}^{(v)} = y_i^{(v)}(t_j)$ denote the value of the i -th variable observed on the v -th subject at time t_j , $\mathbf{y}_i^{(v)} = (y_{i1}^{(v)}, \dots, y_{in}^{(v)})^T$ the observations of type i taken on the v -th subject, and $\mathbf{y}^{(v)} = (\mathbf{y}_1^{(v)T}, \dots, \mathbf{y}_k^{(v)T})^T$ all the observations taken on the v -th subject, ordered by type. Then Σ in Equation (3) denotes the covariance matrix of the $N \times k \times n$ observations vector $\mathbf{Y} = (\mathbf{y}^{(1)T}, \dots, \mathbf{y}^{(N)T})^T$. It seems convenient to remember here some properties of the Kronecker product of matrices that will be used later; they can be checked, for instance, in [10]:
- $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.
- $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.

In particular, $(A \otimes B)^{-1} = (A^{-1} \otimes B^{-1})$ and $(A \otimes B)^T = (A^T \otimes B^T)$

- If A and B are squared matrices of respective dimensions n_A and n_B , then

$$\det(A \otimes B) = \det(A)^{n_B} \det(B)^{n_A}$$

- $(A \otimes B) = P(B \otimes A)Q$ for certain permutation matrices P and Q .

If both A and B are square matrices, then $Q = P^\top$.
with A, B, C, D, P , and Q having the right dimensions for the product of matrices.

The aim is to obtain D -optimal designs for fitting the models of the k variables involved, assuming m -parameter linear models with the same structure. The setup is somehow similar to that in [11], where the double covariance structure was applied to a compositional-response model. That case was roughly equivalent to a one-subject model, and the samples of the same variable were assumed independent, while here a non-trivial correlation is considered between the different observations of the same variable, and multiple subjects are observed. Two scenarios will be considered:

Model I.—Different models of the variables for each subject (N k models):

$$\hat{y}_i^{(v)}(t) = \mathbf{f}(t)^T \boldsymbol{\beta}_i^{(v)}, \quad (4)$$

with $\boldsymbol{\beta}_i^{(v)} = (\beta_1^{(v)}, \dots, \beta_m^{(v)})^T$ $i = 1, \dots, k, v = 1, \dots, N$.

Model II.—The model of each variable is valid for all the subjects (k models):

$$\hat{y}_i^{(v)}(t) = \mathbf{f}(t)^T \boldsymbol{\beta}_i, \quad (5)$$

with $\boldsymbol{\beta}_i = (\beta_1, \dots, \beta_m)^T$ $i = 1, \dots, k, v = 1, \dots, N$.

In the following, analytical results will be obtained for the two scenarios, beginning with Model I:

Theorem 1. *The D -optimal designs for the individual models of each variable in each subject are also D -optimal for Model I, given by Equation (4).*

Proof. If n observations are taken for each variable and subject at times $t_1, \dots, t_n$, the model can be expressed as (1) with $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)T}, \dots, \boldsymbol{\beta}^{(N)T})^T$, $\boldsymbol{\beta}^{(v)} = (\beta_1^{(v)T}, \dots, \beta_k^{(v)T})^T$, and $\boldsymbol{\beta}_i^{(v)} = (\beta_{i1}^{(v)}, \dots, \beta_{im}^{(v)})^T$, and the design matrix is

$$\mathbf{X} = BD\{\mathbf{X}_0, \dots, \mathbf{X}_0\} = \mathbf{I}_N \otimes \mathbf{X}_0,$$

where $\mathbf{X}_0 = (\mathbf{f}(t_1)^T, \dots, \mathbf{f}(t_n)^T)^T$ is the design matrix for each individual model. Then, the information matrix (2) is given by

$$\begin{aligned} \mathbf{M} &= \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} \\ &= (\mathbf{I}_N \otimes \mathbf{I}_k \otimes \mathbf{X}_0^T) (\mathbf{I}_N \otimes \mathbf{S}^{-1} \otimes \mathbf{R}^{-1}) (\mathbf{I}_N \otimes \mathbf{I}_k \otimes \mathbf{X}_0) \\ &= \mathbf{I}_N \otimes \mathbf{S}^{-1} \otimes \mathbf{M}_0, \end{aligned}$$

where $\mathbf{M}_0 = \mathbf{X}_0^T \mathbf{R}^{-1} \mathbf{X}_0$ is the information matrix of the model of each variable for each subject. Thus $\det[\mathbf{M}] = (\det[\mathbf{S}]^{-m} \det[\mathbf{M}_0]^k)^N$, which finishes the proof since $\mathbf{S}$ does not depend on the design. $\square$

Theorem 2. *The parameters of the individual models (4) can be estimated independently and do not depend on $\mathbf{S}$. However, the variance of the set of the parameter estimators for each subject does depend on $\mathbf{S}$ because $\text{Var}(\boldsymbol{\beta}^{(v)}) = \mathbf{S} \otimes \mathbf{M}_0^{-1}$.*

Proof. The estimator of the parameter vector is:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Y} \\ &= (\mathbf{I}_N \otimes \mathbf{S} \otimes \mathbf{M}_0^{-1}) (\mathbf{I}_N \otimes \mathbf{I}_k \otimes \mathbf{X}_0^T) (\mathbf{I}_N \otimes \mathbf{S}^{-1} \otimes \mathbf{R}^{-1}) \mathbf{Y} \\ &= (\mathbf{I}_N \otimes \mathbf{I}_k \otimes \mathbf{W}) \mathbf{Y}, \end{aligned}$$

where $\mathbf{W} = \mathbf{M}_0^{-1} \mathbf{X}_0^T \mathbf{R}^{-1}$ is such that $\beta_i^{(v)} = \mathbf{W} \mathbf{y}_i^{(v)}$ for all $i = 1, \dots, k$, $v = 1, \dots, N$, and $\text{Var}(\hat{\beta}) = \mathbf{M}^{-1} = \mathbf{I}_N \otimes \mathbf{S} \otimes \mathbf{M}_0^{-1}$. $\square$

Let us now pay attention to Model II given in (5). In this case the parameter vector is $\tilde{\beta} = (\tilde{\beta}_1^T, \dots, \tilde{\beta}_k^T)^T$, with $\tilde{\beta}_i = (\tilde{\beta}_{i1}, \dots, \tilde{\beta}_{im})^T$. There are only kxm parameters since the model of each variable Y_i is the same for all the subjects. For this reason, it will be convenient to place together all the observations of each variable, that is, consider the following observations vector:

$$\tilde{\mathbf{Y}} = \left(\underbrace{\mathbf{y}_1^{(1)T}, \dots, \mathbf{y}_1^{(N)T}}_{\text{obs. variable 1}}, \dots, \underbrace{\mathbf{y}_k^{(1)T}, \dots, \mathbf{y}_k^{(N)T}}_{\text{obs. variable k}} \right)^T$$

The following results can be derived, the first one similar to that of Theorem 1, but now consider the second model:

Theorem 3. *The D-optimal designs for the individual models of each variable in each subject are also D-optimal for Model II, given by Equation (5).*

Proof. The design matrix $\tilde{\mathbf{X}}$ corresponding to the ordering given by $\tilde{\mathbf{Y}}$ will be

$$\tilde{\mathbf{X}} = \begin{pmatrix} \mathbf{X}_0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{X}_0 & 0 & \dots & 0 \\ 0 & \mathbf{X}_0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & \mathbf{X}_0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{X}_0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \mathbf{X}_0 \end{pmatrix} = \mathbf{A} \otimes \mathbf{X}_0,$$

with $\mathbf{A} = \mathbf{I}_k \otimes \mathbf{1}_N$, where $\mathbf{1}_N$ denotes a column vector of '1's of length N .

Let $\mathbf{P}$ be the Nkn permutation matrix that turns $\mathbf{Y}$ into $\tilde{\mathbf{Y}}$, that is, $\mathbf{P}\mathbf{Y} = \tilde{\mathbf{Y}}$. Since $\mathbf{P}$ permutes in fact the vectors of observations $\mathbf{y}_i^{(v)}$, it can be expressed as $\mathbf{P} = \mathbf{Q} \otimes \mathbf{I}_n$, with $\mathbf{Q}$ the Nk permutation matrix expressed by vector $\mathbf{v}_Q$,

$$\mathbf{v}_Q = \begin{pmatrix} (1, k+1, 2k+1, \dots, (N-1)k+1, \\ 2, k+2, 2k+2, \dots, (N-1)k+2, \\ \dots \dots \dots \\ k, 2k, 3k, \dots, Nk) \end{pmatrix},$$

meaning that for every row i of $\mathbf{Q}$ ($i = 1, \dots, Nk$), we have $\mathbf{Q}_{i, \mathbf{v}_Q(i)} = 1$ and $\mathbf{Q}_{i,j} = 0$ if $j \neq \mathbf{v}_Q(i)$.

Now the covariance matrix of $\tilde{\mathbf{Y}}$ is

$$\begin{aligned}\tilde{\Sigma} &= \Sigma_{\tilde{\mathbf{Y}}} = \Sigma_{\mathbf{P} \mathbf{Y}} = \mathbf{P} \Sigma_{\mathbf{Y}} \mathbf{P}^T \\ &= (\mathbf{Q} \otimes \mathbf{I}_n) [(\mathbf{I}_N \otimes \mathbf{S}) \otimes \mathbf{R}] (\mathbf{Q}^T \otimes \mathbf{I}_n) \\ &= [\mathbf{Q}(\mathbf{I}_N \otimes \mathbf{S}) \mathbf{Q}^T] \otimes \mathbf{R} \\ &= \mathbf{S} \otimes \mathbf{I}_N \otimes \mathbf{R},\end{aligned}$$

and the information matrix can be computed as

$$\begin{aligned}\tilde{\mathbf{M}} &= \tilde{\mathbf{X}}^T \tilde{\Sigma}^{-1} \tilde{\mathbf{X}} \\ &= (\mathbf{A}^T \otimes \mathbf{X}_0^T) \left[(\mathbf{S}^{-1} \otimes \mathbf{I}_N) \otimes \mathbf{R}^{-1} \right] (\mathbf{A} \otimes \mathbf{X}_0) \\ &= \left[\mathbf{A}^T (\mathbf{S}^{-1} \otimes \mathbf{I}_N) \mathbf{A} \right] (\mathbf{X}_0^T \mathbf{R}^{-1} \mathbf{X}_0) \\ &= N \mathbf{S}^{-1} \otimes \mathbf{M}_0,\end{aligned}$$

With the last expression obtained using that $\mathbf{A} = \mathbf{I}_k \otimes \mathbf{1}_N$. Then, $\det[\tilde{\mathbf{M}}] = N^{k \cdot m} \det[\mathbf{S}]^{-m} \det[\mathbf{M}_0]^k$. $\square$

Theorem 4. Regarding Model II given by (5), the estimation of the parameters of the i -th response is the average of the corresponding estimations for each subject and does not depend on $\mathbf{S}$ (but their covariance matrix, which is proportional to the inverse of the information matrix, does depend on $\mathbf{S}$).

Proof. The estimator of the parameter vector is

$$\begin{aligned}\hat{\beta} &= (\tilde{\mathbf{X}}^T \tilde{\Sigma}^{-1} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \tilde{\Sigma}^{-1} \tilde{\mathbf{Y}} \\ &= (\mathbf{S}/N \otimes \mathbf{M}_0^{-1}) (\mathbf{A}^T \otimes \mathbf{X}_0) \left[(\mathbf{S}^{-1} \otimes \mathbf{I}_N) \otimes \mathbf{R}^{-1} \right] \tilde{\mathbf{Y}} \\ &= \left[\mathbf{S}/N \mathbf{A}^T (\mathbf{S}^{-1} \otimes \mathbf{I}_N) \right] \otimes (\mathbf{M}_0^{-1} \mathbf{X}_0^T \mathbf{R}^{-1}) \tilde{\mathbf{Y}} \\ &= (\mathbf{I}_k \otimes \mathbf{1}_N^T / N \otimes \mathbf{W}) \tilde{\mathbf{Y}},\end{aligned}$$

where the last equality comes from

$$(\mathbf{S}/N) \mathbf{A}^T = (\mathbf{S} \otimes \mathbf{I}_1 / N) (\mathbf{I}_k \otimes \mathbf{1}_N^T) = (\mathbf{S} \otimes \mathbf{1}_N^T / N).$$

Let us have a close look at this expression. It means that

$$\begin{aligned}\hat{\beta}_i &= (\mathbf{1}_N^T / N \otimes \mathbf{W}) (\mathbf{y}_i^{(1)T}, \dots, \mathbf{y}_i^{(N)T})^T \\ &= \frac{1}{N} (\mathbf{W}, \dots, \mathbf{W}) (\mathbf{y}_i^{(1)T}, \dots, \mathbf{y}_i^{(N)T})^T \\ &= \frac{1}{N} \sum_{v=1}^N \mathbf{W} \mathbf{y}_i^{(v)T}\end{aligned}$$

which, taking into account that $\mathbf{W} \mathbf{y}_i^{(v)T}$ is the estimation of the parameters of the model of the variable Y_i for the v -th subject, finishes the proof. $\square$

3. Optimal Designs for Evolution of MR and PS

There is a general agreement that problem solving should be the main objective in school mathematics instruction [12,13]. It seems clear that this ability may be related with the capability of understanding the semantic structure of the mathematical problems statements and increases with age. Children could be able to solve real world problems at early age [14], but the acquisition of academic language comes later, after they dominate everyday language [15]. The relation between problem solving and linguistic comprehension is

explored in [16]. However, apart from understanding the statement of the problem, the students must be able to construct a mental representation of it [17–19].

All these facts are discussed in [3], where a test on additive problems using expressions close to practical everyday language is proposed. That work studied the influence of mental representation (*MR*) in the ability of the resolution of additive problems (*PS*) for children from 6 to 12 years old and found a model relating the two variables. Once this relationship has been established, the next step should be to find a convenient design plan for a follow-up study of the evolution of both characteristics. To this end, the results of Section 2 will be employed for obtaining an optimal design plan for the observation of these two variables through time. The design points t_i can denote any convenient temporal unit (school level, semester, etc.).

A constant variance σ^2 will be assumed for the observations of the two variables. Should they have different variances σ_1^2 and σ_2^2 , the covariance structure would depend (by a constant term) on the ratio σ_1/σ_2 that has no influence on the computation of the optimal designs (see the discussion at the end of Section 2.1 of [2], and in this situation (constant variance), for the easiness of computations, $\sigma^2 = 1$ will be assumed without loss of generality. Assuming constant covariance between $y_i(t)$, $y_j(t)$ for every $i, j = 1 \dots, k$ may be arguable, mainly because the k variables may refer to quite different characteristics and even use different scales. An alternative that may soften this problem would be using normalized values of the variables. However, the case $k = 2$ of this example is not controversial, since for $k = 2$ the intra-covariance matrix $\mathbf{S}$ will have this shape:

$$\mathbf{S} = \begin{pmatrix} 1 & s \\ s & 1 \end{pmatrix},$$

with $s = \text{Cov}[y_1(t), y_2(t)]$, an assumed constant for all t . Thus, when observing $Y_1 = MR$ and $Y_2 = PS$, the covariance matrix for (the observations taken at) each student will be

$$\Sigma_0 = \mathbf{S} \otimes \mathbf{R} = \begin{pmatrix} \mathbf{R} & s \mathbf{R} \\ s \mathbf{R} & \mathbf{R} \end{pmatrix}. \quad (6)$$

and the previous results can be applied. From Theorems 1 to 4, the D -optimal designs can be computed assuming just one variable and one subject. An inter-correlation structure usually employed when the measurements are taken on the same individual is the exponential covariance [20], decreasing with the increasing distance in time between measurements,

$$\text{Cov}[y(t), y(t+d)] = e^{-\lambda d}, \quad (7)$$

where the parameter λ is characteristic of the individual. When there is no reason to think that the parameter may vary very much between individuals, the same characteristic (reference individual) is used for every of them. In this work, $\lambda = 1$ will be assumed, which is a typical choice.

Two types of evolution models will be considered:

- Linear regression model: $\hat{y} = \beta_0 + \beta_1 t$, that is, $\mathbf{f}(t) = (1, t)^T$. Since there are $m = 2$ parameters, at least two observations are needed
 - If $n = 2$, it will be convenient to express the design as $\xi = \{t, t+d\}$, where t is the first observation and $d > 0$ the distance between the two samples, with $t+d \leq t_{\max}$, the maximum value for performing the tests. Then, assuming the inter-correlation R given by (7),

$$\mathbf{X}_0 = \begin{pmatrix} \mathbf{f}(t)^T \\ \mathbf{f}(t+d)^T \end{pmatrix} = \begin{pmatrix} 1 & t \\ 1 & t+d \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 & e^{-d} \\ e^{-d} & 1 \end{pmatrix} \quad (8)$$

and

$$\begin{aligned}\mathbf{M}_0 &= \mathbf{X}_0^T \mathbf{R}^{-1} \mathbf{X}_0 \\ &= \frac{1}{1-e^{-2d}} \begin{pmatrix} 1 & 1 \\ t & t+d \end{pmatrix} \begin{pmatrix} 1 & -e^{-d} \\ -e^{-d} & 1 \end{pmatrix} \begin{pmatrix} 1 & t \\ 1 & t+d \end{pmatrix} \\ &= \frac{1}{1+e^{-d}} \begin{pmatrix} 2 & 2t+d \\ 2t+d & 2t(t+d) + \frac{d^2}{1-e^{-d}} \end{pmatrix}.\end{aligned}$$

Thus, $\det(\mathbf{M}_0) = d^2 / (1 - e^{-2d})$ is an increasing function of d , and the optimal design will be the one having the maximum distance d between the observations, therefore taking the samples at the beginning and at the end of the period of study.

- When $n = 3$, let us consider the design $\xi = \{t, t + d_1, t + d_1 + d_2\}$, with $d_1, d_2 > 0$ and $t + d_1 + d_2 \leq t_{max}$. Now

$$\mathbf{X}_0 = \begin{pmatrix} 1 & t \\ 1 & t + d_1 \\ 1 & t + d_1 + d_2 \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 1 & e^{-d_1} & e^{-(d_1+d_2)} \\ e^{-d_1} & 1 & e^{-d_2} \\ e^{-(d_1+d_2)} & e^{-d_2} & 1 \end{pmatrix}$$

and

$$\det(\mathbf{M}_0) = \frac{2e^{2d_1+d_2} (e^{d_2} (e^{d_1} - 1) d_1^2 + (-2e^{d_2} + e^{d_1+d_2} + 1) d_2 d_1 + e^{d_2} (e^{d_1} - 1) d_2^2)}{(e^{d_1+d_2} - 1) (e^{d_1} - 2e^{d_2} + e^{2d_1+d_2})}.$$

Assuming a minimum distance d_0 between consecutive sample points, the maximum is attained for $d_1 = t_{max} - d_0$, and $d_2 = d_0$. Thus, the first test should be taken at the beginning and the rest at the end, with the minimum possible distance d_0 between these last ones.

- Exponential regression model: $\hat{y} = \beta_0 e^{\beta_1 t}$.
After linearizing, the model can be expressed as $\mathbf{f}(t) = e^{\beta_1 t} (1, \beta_0 t)^T$, which depends on the unknown values of the parameters. In this case, the D -optimal designs will depend as well on these values; thus, they will be in fact locally optimal, that is, good for (or near to) those nominal values used in the computation. It is well known that the optimal designs will not depend on the parameters that ‘appear linearly’ in the model [21], that is, β_0 in this case, while they will depend on the ‘non-linear parameter’ β_1 . Thus, an initial value will be needed just for this last one. Again, the cases of two and three observations will be studied:

- $n = 2 \Rightarrow \xi = \{t, t + d\}$. The determinant of the information matrix is

$$\det(\mathbf{M}_0) = \frac{\beta_0^2 d^2 e^{2(\beta_1+1)d+4\beta_1 t}}{e^{2d} - 1},$$

and it is clear that the values t and d maximizing this determinant will not depend on β_0 . Figure 1 shows the determinant of the information matrix after removing β_0^2 and using the nominal value $\beta_{10} = 1$ for the non-linear parameter. For primary school levels ($t_{max} = 6$) and assuming a minimum distance $d_0 = 1$ between tests, the recommended design uses $t = 5$ and $d = 1$, that is, making the tests the last two years.

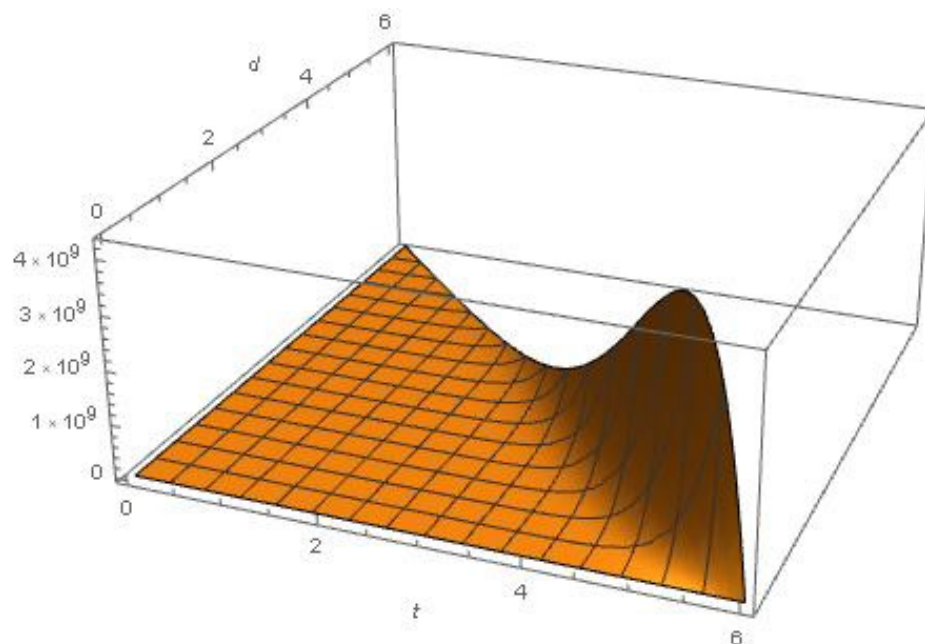


Figure 1. Determinant of the information matrix of $\xi = \{t, t + d\}$ for the exponential model for $\beta_0 = \beta_1 = 1$.

○ $n = 3 \Rightarrow \xi = \{t, t + d_1, t + d_1 + d_2\}$. The determinant of M_0 is

$$\det(\mathbf{M}_0) = \frac{\beta_0^2 e^{2(\beta_1+1)d_1+d_2+4\beta_1 t}}{(e^{d_1+d_2}-1)(e^{d_1}-2e^{d_2}+e^{2d_1+d_2})} \left[e^{d_2} d_1^2 \left(-2e^{\beta_1 d_2} + e^{2\beta_1 d_2+d_1} + e^{d_1} \right) - 2d_2 d_1 e^{\beta_1 d_2} \left(-e^{\beta_1 d_1} - e^{(\beta_1+1)d_2+d_1} + e^{\beta_1 d_1 + (\beta_1+1)d_2} + e^{d_2} \right) + d_2^2 e^{(2\beta_1+1)d_2} \left(-2e^{\beta_1 d_1} + e^{(2\beta_1+1)d_1} + e^{d_1} \right) \right],$$

and when removing β_0^2 and assuming $t_{\max} = 6$ and $d_0 = 1$, the maximum for $\beta_{10} = 1$ is attained when $t = 4$, $d_1 = 1$, and $d_2 = 1$, that is, testing the students the last three courses.

4. Discussion

In the last few years, there has seen a remarkable increase in the relevance of the design of experiments in social science areas [22]. Major issues in the education research area must be treated on a sequential multiscale temporal level. Educational policies such as curricula, reduction in class size, programs for students' support, etc. are developed considering educational theories based on solid experimentation [23]. In addition, large-scale experiments may be unattainable, and some authors alert about extracting generalized conclusions from experiments in education [24]. There is a general agreement in trying to "maximize the scientific benefit using the resources available for an investigation" [25]; thus, the optimal design of experiments becomes a key piece on the research process.

Randomized experiments with multilevel implications have been used by educational researchers in order to determine the effect of some treatment through time. One of the most used techniques in this line is the sequential multiple assignment randomized trials (SMART) [23,26,27]. This approach gives information about the effectiveness of some intervention or treatment through time, but it does not discriminate factor variables [25,28]. As an alternative, factorial designs, with the feature of determining the influence of factors, have also been applied for similar purposes [25,29]. These factorial designs have been used as well to provide a relation between four different mathematical problem representations and the problem solving abilities of elementary school students from grades 1 to 3 [30].

Nevertheless, some limitations regarding to the independence of the variables, the quantity of considered factors, or the presence of more than two levels by factor are addressed [25]. Most of these studies assume independent observations; few consider some kind of relation between samples (for example, the intraclass-correlation between observations within a cluster in [25,31], but not complex correlation structures like the ones derived from the consideration of multiresponse models such as the ones studied in this work.

In educational research, observations are rarely independent, as pointed out in [32]. Mixed-effect regression models cover this and other issues, such as correlated data organized in a multilevel or hierarchical structure, missing data, variability, etc. [33–35]. Applications of these models in educational projects can be found in [32] where the correlation between the results of two tests taken before and after a treatment is studied, controlled for a variable with three levels. The levels correspond to three selected difficult topics to study, and a selection model based on AIC criterion is applied to conclude. In [36], mixed linear regression analysis is applied to a large database to establish the correlation between the student's relationship and their academic performance. Finally, ref. [37] evaluated the impact of sequences of parents' input on children's language outcomes. Although all these applications are longitudinal or level studies, they are focused on selecting the best model for prediction, but no one addresses the issue of optimizing data recollection along time.

Despite being less known in the education area, generalized estimating equations (GEEs) are introduced in [38] as an alternative for the analysis of cross sectional clustered data with repeated measurements along time [39–41]. It shows GEEs as a generalization of linear general models and that its performance is similar to multilevel models (random effects or mixed model) and ordinary least squares. However, some limitations have to be considered, for example the impossibility of applying classical selection model techniques based on likelihood estimation, the lack of non-random data, and the amount of missing data. In addition, the number of clusters should be relatively high, and the observations in different clusters must be independent, although within-cluster observations may be correlated, as it is refereed in [40].

Again, it is necessary to highlight that none of these approaches seem to be fully appropriate to conduct experiments involving multiresponse and multisubject models with repeated observations in time, with potential correlation structures as the ones described in this paper. The novel approach presented here seems the most convenient for studies that could be similar to the one described in the example.

The relation between the variables 'Mental Representation' and 'Mathematical Problem Resolution' in primary school students was extensively studied in [3]. Now, the theory developed in Section 2 has been applied to that case for finding the best design (times when perform the tests to the students) in order to obtain an insight of the evolution of the two variables.

Assuming a sensible hypothesis about the correlation structure, two scenarios have been considered, one with different models of the variables for each subject and another one assuming the same model of each variable for every subject, obtaining interesting theoretical results especially in the second case. Two types of evolution models have been considered as well in order to stress that the optimal allocation of the tests may depend very much on the assumed model. The best design when assuming the linear regression model is making the tests near the extremes of the evaluation period. However, when the evolution is described by an exponential model, the results may be quite different since the optimal designs are very sensitive to the initial value of the parameters. The widely used exponential inter-correlation has been assumed. The preliminary study in [3] discovered that the group of the small (first year) students had a great variability, and thus it was not very informative. For this reason, when a design contains any temporal point belonging to the first school year (for instance when assuming a linear trend evolution), it will be convenient to delay these tests so that the analysis interval starts the second year of the primary school period.

A robustness study has been carried out for the exponential evolution model in order to check whether the optimal designs obtained are sensitive to the choice of the nominal value of β_1 . It has been found that in fact the optimal designs can vary very much; for instance if $n = 2$, the best designs found vary from the ones that take the extremes of the design interval when $\beta_{10} = 0.1$ or the school years 4 and 6 when $\beta_{10} = 0.5$ to those choosing the last temporal points when $\beta_{10} \geq 1$. A similar variability of results is obtained for $n = 3$: when $\beta_{10} = 0.1$, the proposal is to make one test at the beginning of the design interval and the other two at the end, and when $\beta_{10} = 0.5$ assuming $d_0 = 1$, the most convenient school years to perform the tests are 3, 4, and 6.

Optimal designs have been computed for the D -optimality criterion, which is the most used, and two popular 2-parameter evolution models have been considered assuming always exponential inter-correlation. However, the procedure can be immediately extended to other optimality criteria, evolution models, and correlation kernels, provided that these can be assumed similar for all the variables.

5. Conclusions

The need to study the problem solving ability of primary school students, its relation with other variables, and the evolution of all of them was the inspiration of this work, from a previous study of [3]. In order to choose the best temporal points where samples should be taken to maximize the information of the collected data, a general procedure for several variables and subjects has been developed from the point of view of optimal design of experiments theory.

It should be noted that the results obtained in Section 2 are quite general and could be applied in other type of studies involving several subjects, with a $k > 2$ number of variables of interest. Depending on the cost of every type of measure and the available budget, sometimes it might be convenient to select the variables that will be observed at a specific point instead of sampling all of them. The reason may be to save the cost of those measurements that are not considered so important or informative. Cost constraints could be incorporated into the model in a similar way as described in Chapter 5 of [42]. For instance, when two types of variables are to be observed, the problem can be stated as follows: two types of measurements y_1, y_2 can be made for every experimental unit, with costs c_1 and c_2 , respectively. In the general case, every time that a subject is measured entails some cost c . For the total cost there are three possibilities:

- i. Measure y_1 , with cost c_1 . Total cost: $C_1 = c + c_1$;
- ii. Measure y_2 , with cost c_2 . Total cost: $C_2 = c + c_2$;
- iii. Measure both y_1 and y_2 . Now the overall cost is $C_3 = c + c_1 + c_2$.

Possible scenarios are described by $t_x = \{t, x_1, x_2\}$, with $0 \leq t \leq t_{max}$ and x_1, x_2 in $\{0, 1\}$ and $x_1 + x_2 > 0$, where $x_i = 1$ when test y_i is made at a time t , and $x_i = 0$ otherwise. The best designs fitting this budget may be non-balanced. In this case, the covariance matrices could be obtained as described in Example 2 of [1]. Ref. [42] discusses a procedure for studying this issue, but only for one subject and considering intra-correlation and not inter-correlation between samples, which are assumed independent. When there are more subjects and the two types of correlation are non-trivial, the problem becomes far more complex to deal with.

For this study, it has been assumed that the evolution models were similar in the different variables. Should these models be different, the covariance and information matrices would be much more complex, as would the computation of the optimal designs. The assumptions of equal intracorrelation in all the design points and equal intercorrelation for all the variables are quite sensible. In case that any of them (or both) were not true, the covariance matrices could not be expressed by Kronecker products, and the optimal designs for the global model would not be the same as the ones for the individual models of each variable for each subject; in fact, this global optimal design may be extremely difficult to compute.

Author Contributions: Conceptualization, M.R.-H., R.E.P. and J.M.R.-D.; methodology, J.M.R.-D.; software, J.M.R.-D.; validation, M.R.-H., R.E.P. and J.M.R.-D.; formal analysis, J.M.R.-D.; investigation, J.M.R.-D.; resources, M.R.-H., R.E.P. and J.M.R.-D.; data curation, M.R.-H., R.E.P. and J.M.R.-D.; writing—original draft preparation, J.M.R.-D.; writing—review and editing, M.R.-H., R.E.P. and J.M.R.-D.; visualization, M.R.-H., R.E.P. and J.M.R.-D.; supervision, M.R.-H., R.E.P. and J.M.R.-D.; project administration, J.M.R.-D.; funding acquisition, J.M.R.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Spanish Ministerio de Economía y Competitividad and Junta de Castilla y León (Projects ‘MTM2016-80539-C2-2-R’ and ‘SA080P17’, ‘SA105P20’).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Rodríguez-Díaz, J.M.; Sánchez-León, G. Efficient parameter estimation in multiresponse models measuring radioactivity retention. *Radiat. Environ. Biophys.* **2019**, *58*, 167–182. [[CrossRef](#)] [[PubMed](#)]
- Rodríguez-Díaz, J.M.; Sánchez-León, G. Optimal designs for multiresponse models with double covariance structure. *Chemometr. Intell. Lab. Syst.* **2019**, *189*, 1–7. [[CrossRef](#)]
- Rodríguez-Hernández, M.; Pruneda, R.E.; Rodríguez-Díaz, J.M. Statistical Analysis of the Evolutionary Effects of Language Development in the Resolution of Mathematical Problems in Primary School Education. *Mathematics* **2021**, *9*, 1081. [[CrossRef](#)]
- Fedorov, V.V.; Hackl, P. *Model-Oriented Design of Experiments*; Springer: New York, NY, USA, 1997.
- Pukelsheim, F. *Optimal Design of Experiments*; SIAM: Philadelphia, PA, USA, 2006.
- Atkinson, A.C.; Donev, A.N.; Tobias, R.D. *Optimum Experimental Designs, with SAS*; Oxford University Press: Oxford, UK, 2007.
- Rodríguez-Díaz, J.M.; Sánchez-León, G. Design optimality for models defined by a system of ordinary differential equations. *Biometr. J.* **2014**, *56*, 886–900. [[CrossRef](#)]
- Dette, H.; Pepelyshev, A.; Zhigljavsky, A. Design for linear regression models with correlated errors. In *Handbook of Design and Analysis of Experiments*; CRC Press: Boca Raton, FL, USA, 2015; pp. 237–276.
- Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; MIT Press: Cambridge, MA, USA, 2006.
- Harville, D.A. *Matrix Algebra from a Statistician’s Perspective*; Springer: New York, NY, USA, 1997.
- Rodríguez-Díaz, J.M.; Rivas-López, M.J.; Santos-Martín, M.T.; Mariñas Collado, I. Optimal designs for a linear-model compositional response. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 139–148. [[CrossRef](#)]
- NCTM. *Principles to Actions: Ensuring Mathematical Success for All*; National Council of Teachers of Mathematics: Reston, VA, USA, 2014.
- Cockcroft, W.H. *Mathematics Counts: Report of the Committee of Inquiry into the Teaching of Mathematics in Schools Under the Chairmanship of W.H. Cockcroft*; Technical Report; Her Majesty’s Stationery Office and the Queen’s Printer for Scotland: London, UK, 1982.
- Carpenter Thomas, P.; Ansell, E.; Franke, M.L.; Fennema, E.; Weisbeck, L. Models of Problem Solving: A Study of Kindergarten Children’s Problem-Solving Processes. *J. Res. Math. Educ.* **1993**, *24*, 428–441. [[CrossRef](#)]
- Cummins, J. *Language, Power, and Pedagogy: Bilingual Children in the Crossfire*; Multilingual Matters Ltd.: Clevedon, UK, 2000.
- Dixon, J.A.; Moore, C.F. The developmental role of intuitive principles in choosing mathematical strategies. *Dev. Psychol.* **1996**, *32*, 241–253. [[CrossRef](#)]
- De Corte, E.; Verschaffel, L.; De Win, L. Influence of rewording verbal problems on children’s problem representations and solutions. *J. Educ. Psychol.* **1985**, *77*, 460–470. [[CrossRef](#)]
- Hegarty, M.; Mayer, R.E.; Monk, C.A. Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. *J. Educ. Psychol.* **1995**, *87*, 18–32. [[CrossRef](#)]
- Pape, S.J. Compare word problems: Consistency hypothesis revisited. *Contemp. Educ. Psychol.* **2003**, *28*, 396–421. [[CrossRef](#)]
- Cressie, N.; Wikle, C.K. *Statistics for Spatio-Temporal Data*; John Wiley and Sons: Hoboken, NJ, USA, 2011.
- Hill, P.D.H. D-optimal designs for partially nonlinear regression models. *Technometrics* **1980**, *22*, 275–276. [[CrossRef](#)]
- Jackson, M.; Cox, D.R. The principles of experimental design and their application in sociology. *Annu. Rev. Sociol.* **2013**, *39*, 27–49. [[CrossRef](#)]
- Raudenbush, S.W.; Schwartz, D. Randomized experiments in education, with implications for multilevel causal inference. *Annu. Rev. Stat. Its Appl.* **2020**, *7*, 177–208. [[CrossRef](#)]
- Cook, T.D. Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educ. Eval. Policy Anal.* **2002**, *24*, 175–199. [[CrossRef](#)]
- Dziak, J.J.; Nahum-Shani, I.; Collins, L.M. Multilevel factorial experiments for developing behavioral interventions: Power, sample size, and resource considerations. *Psychol. Methods* **2012**, *17*, 153. [[CrossRef](#)]
- Murphy, S.A. An experimental design for the development of adaptive treatment strategies. *Stat. Med.* **2005**, *24*, 1455–1481. [[CrossRef](#)]

27. Almirall, D.; Kasari, C.; McCaffrey, D.F.; Nahum-Shani, I. Developing optimized adaptive interventions in education. *J. Res. Educ. Eff.* **2018**, *11*, 27–34. [[CrossRef](#)]
28. Flay, B.R.; Collins, L.M. Historical review of school-based randomized trials for evaluating problem behavior prevention programs. *Ann. Am. Acad. Political Soc. Sci.* **2005**, *599*, 115–146. [[CrossRef](#)]
29. Collins, L.M.; Dziak, J.J.; Li, R. Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychol. Methods* **2009**, *14*, 202. [[CrossRef](#)]
30. Elia, I.; Gagatsis, A.; Demetriou, A. The effects of different modes of representation on the solution of one-step additive problems. *Learn. Instr.* **2007**, *17*, 658–672. [[CrossRef](#)]
31. Cho, S.-J.; Preacher, K.J.; Bottge, B.A. Detecting Intervention Effects in a Cluster-Randomized Design Using Multilevel Structural Equation Modeling for Binary Responses. *Appl. Psychol. Meas.* **2015**, *39*, 627–642. [[CrossRef](#)]
32. Theobald, E. Students are rarely independent: When, why, and how to use random effects in discipline-based education research. *CBE—Life Sci. Educ.* **2018**, *17*, rm2. [[CrossRef](#)] [[PubMed](#)]
33. Raudenbush, S.W. Educational applications of hierarchical linear models: A review. *J. Educ. Stat.* **1988**, *13*, 85–116. [[CrossRef](#)]
34. McCulloch, C.E.; Searle, S.R. *Generalized, Linear, and Mixed Models*; John Wiley & Sons: Hoboken, NJ, USA, 2004.
35. Gałecki, A.; Burzykowski, T. Linear mixed-effects model. In *Linear Mixed-Effects Models Using R*; Springer: New York, NY, USA, 2013; pp. 245–273.
36. Zhou, H.; Jiang, S.; Liu, X. Regression analysis of intelligent education based on linear mixed effect model. *J. Ambient Intell. Humaniz. Comput.* **2021**, 1–13. [[CrossRef](#)]
37. Silvey, C.; Demir-Lira, Ö.E.; Goldin-Meadow, S.; Raudenbush, S.W. Effects of time-varying parent input on children’s language outcomes differ for vocabulary and syntax. *Psychol. Sci.* **2021**, *32*, 536–548. [[CrossRef](#)] [[PubMed](#)]
38. Huang, F.L. Analyzing cross-Sectionally clustered data using generalized estimating equations. *J. Educ. Behav. Stat.* **2022**, *47*, 101–125. [[CrossRef](#)]
39. Burton, P.; Gurrin, L.; Sly, P. Extending the simple linear regression model to account for correlated responses: An introduction to generalized estimating equations and multi-level mixed modelling. *Stat. Med.* **1998**, *17*, 1261–1291. [[CrossRef](#)]
40. Ghisletta, P.; Spini, D. An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *J. Educ. Behav. Stat.* **2004**, *29*, 421–437. [[CrossRef](#)]
41. Muth, C.; Bales, K.L.; Hinde, K.; Maninger, N.; Mendoza, S.P.; Ferrer, E. Alternative models for small samples in psychological research: Applying linear mixed effects models and generalized estimating equations to repeated measures data. *Educ. Psychol. Meas.* **2016**, *76*, 64–87. [[CrossRef](#)]
42. Fedorov, V.V.; Gagnon, R.; Leonov, S. *Optimal Design for Multiple Responses with Variance Depending on Unknown Parameters*; GSK BDS Technical Report 2001–2003; GlaxoSmithKline Pharmaceuticals: Collegeville, PA, USA, 2003.