Measurement and Research Department Reports 2006-1

Explorations in Recursive Designs

Huub Verstralen

CITO

Measurement and Research Department Reports 2006-1

Explorations in Recursive Designs

Huub Verstralen

Cito
Arnhem, 2006

Abstract

Starting from a set of basic designs, more complex designs are created by recursive application of the basic designs. Properties of these designs, and their effects on the accuracy of Rasch CML-parameter estimates are investigated.

Keys : Designs, Rasch model, CML

1 Introduction

To construct a test from an item bank in a psychometrically sound way, we need data on these items. These data are obtained by administering them to a sample of students, and to register their responses. A response to an item is also called an observation. Because the number of items in an item bank usually is large, each student is administered only a small part of all items in the bank. For these observations to be useful they have to be carefully planned. Therefore, many tests are administered as part of a so-called structurally incomplete design. In such a design the complete set of items is divided into partially overlapping subsets of items called a booklet. The number of items a booklet contains, is such that the intended student can finish it in time. With observations obtained from such a design the item parameters are usually estimated with the method of Conditional Maximum Likelihood (CML). In this report it is investigated for a structured set of design patterns how certain properties of the design affect the quality of the item parameter estimates, as measured for instance by their estimation error. Maximum Likelihood estimation in general accommodates missing data in a natural way. However, the pattern of the observations does affect the quality of the parameter estimates.

First some simple properties of designs that seem relevant for parameter estimate quality are discussed. Then so-called basic designs and their recursive application are introduced, and some properties related to the simple properties are proven. Next, a tractable expression to calculate the estimation error is introduced to quantify the effect of design properties to estimation error. Finally, some numerical examples are discussed that relate some design properties to the quality of the design. For simplicity focus will be on the Rasch model.

2 Design properties

Four design properties will be distinguished:

1. *1. Smallest number of observations per item*

Like the strength of a chain is determined by its weakest link, the quality of a scale can be justly characterized as the largest error of estimation of its item parameters. The estimation error of an item parameter is determined not only by its difficulty for the target population, but to a large extent by the number of observations it shares with other items. Therefore, to optimally distribute observations over items, it is a matter of efficiency to equalize the number of observations for each item, or equivalently, to maximize the minimum amount of observations per item under constraints as cost or total amount of observations. It is evident that under the last constraint this maximum is obtained with an equal amount of observations per item. Usually, we will assume the number of observations per booklet equal over booklets. Then the number of observations of an item equals the number of its booklets times the number of observations per booklet. The number of booklets that contains a certain item is also called item frequency. Under the assumption of an equal number of observations per booklet striving for an equal amount of observations per item is equivalent to striving for equal item frequencies in the design. If all items have equal frequency in a design, the design has the equal item frequency property *(EIF)*.

2. Smallest number of observations per item pair

A pair is called observed if at least one subject responds to both items. With a complete design, the number of observations of a pair of items equals the number of observations of the individual items. With incomplete designs this is not the case. The number of observations for many pairs of items is less than the number of observations of the individual items. However, in IRT-models, only functions of differences of item location parameters are estimable. The minimum number of observations per item pair could, therefore, be an important characteristic of an incomplete design. Because in many incomplete designs, however, the minimum number of observations per pair equals zero, this quality index is not as informative as one might wish. Therefore, in evaluating the quality of a design, other indices using pair frequencies are preferable. One could think of a distribution of pair frequencies, or represent this distribution by a mean and a standard deviation.

3. Largest item pair distance (LPD)

In many incomplete designs, the smallest number of observations for pairs of items equals zero. Given such a design an important characteristic of an item pair is the minimum number of observed item-pairs that connects the item pair. For instance, the distance of an item to itself equals zero, if a pair of different items shares a booklet the pair itself is observed, and the distance equals 1. If the item pairs (a, b), and (b, c) are observed, and (a, c) is not, then the distance of item pair (a,c) equals two, etc.. The item pair in a design with the largest distance determines to a large extent the largest uncertainty about the differences of its item parameters. This property of a design is called largest pair distance *(LPD)*. Of special importance are designs with LPD equal to one or *LPDO,* that is, for each pair of items there is a booklet in the design that contains it. For Conditional Maximum Likelihood (CML) estimation it is crucial that all item pairs are connected whatever the distance. If two items are not connected the difference of their item-parameters cannot be estimated. When all item pairs are connected the design is called connected. The basic designs being explored in this article are connected.

4. Size of common subsets

When the distance of a pair of items equals one, and they are each contained in exactly one booklet, the number of items in the common subset of those two booklets is a major determinant of the accuracy of the estimation of the difference between the two item parameters.

3 Basic patterns

A *basic pattern* is a subset structure on a finite set S of units with certain simplifying limitations. A *unit* may be an item, or a set of items, that always goes together in a booklet. The first two simplifying limitations are the following

1. The subsets all have an equal size of b units

2. The union of all subsets covers S.

To obtain interesting basic patterns, set-theoretical concepts are not sufficient. Graph-theoretical notions do enrich the picture. The subsets in a basic pattern are then identified with *points* in the graph, and if two subsets share a units $(a > 0)$ they are *joined* by an *edge,* also called *directly connected*. The edge connecting two subsets is weighted by a. Two subsets S_1 and S_2 are called *connected* if there is a *path* connecting them. Then there is a set of directly connected pairs of subsets, with the first pair containing S_1, and the last pair containing S_2, and all other pairs not containing S_1 nor S_2.

The number s of edges in a smallest path connecting two subsets S_1 and S_2 defines their *distance* as s. Consequently, the distance between two directly connected subsets equals 1. Two types of chains are discerned. In a *closed path* there exists a path of length > 0 from each point to itself. If in a closed path no edge is traversed twice it is called a *circuit*. If in a closed path no point is traversed twice it is called a *cycle*. The *order* of a point is its number of edges. In a *tree* every pair of points is connected by a unique path. A *linear graph* is a tree where the order of every point is at most two. Now the following types of basic patterns can be distinguished, that all cover S:

1. circular pattern
2. tree pattern
3. linear pattern, is a special tree pattern
4. disconnected pattern, contains a point pair unconnected by a path

It follows immediately from elementary theorems in graph theory that if two points in a tree are connected by an additional edge, it contains a unique cycle, and stops being a tree (Balakrishnan, 1995, theorem A.3, pg 182). The number of units in S, also called the size of S, is denoted by n. On the other hand one should be aware that the proposed Graph-representation in many instances does not capture all properties of a design, because often more essentially different designs are represented by the same graph, as a simple example in Figure 1 shows.

3.1 Circular basic designs

One of the most interesting basic designs is what is here called the circular design. The basic idea can be found in Definition 1, which is presented in Figure 2.

Definition 1: Choose an ordering of S. Start at the left side of S with an uninterrupted subset $B_0 \subset S$, and choose a number s such that $s < b$ and $n = ks$, with k an integer. Next shift B_0 over the length of s consecutive units. This gives the next booklet B_1 in the design. Again shift B_1 over s units. This gives the third booklet B_2. Where the booklet crosses the right border of S it reappears at the left border of S. Repeat this shifting $k - 1$ times. Then the process stops. Shifting one more time would result

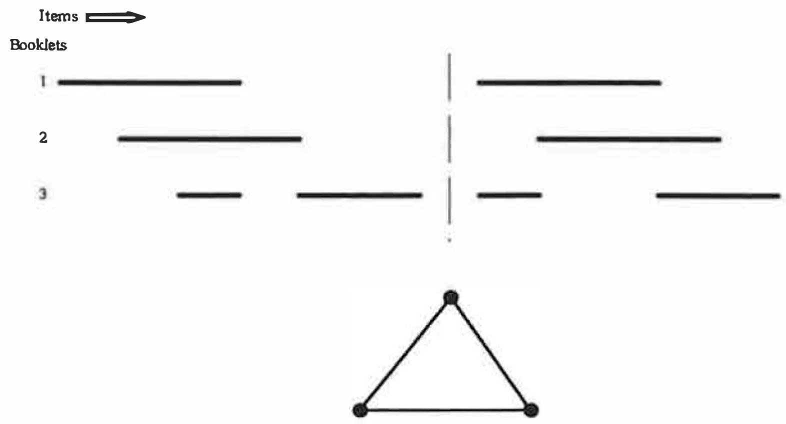

Figure 1: Two different designs represented by one graph. In the left design the first booklet shares part of its intersection with the second and with the third booklet, whereas in the design at the right it does not.

in booklet B_k that coincides with the first booklet B_0, therefore, in Figure 2 B_k has been crossed out, because it already is a part of the design under the name B_0. The design is given by the sequence of booklets $B_0, ..., B_{k-1}$. Note that the shift size s must fit exactly k times into $S : n = ks$, because otherwise B_k would not coincide with B_0. A small real example of a circular design is: S contains items or item units $\{1, ..., 6\}$ and booklets $(1, 2, 3)$, $(3, 4, 5)$, $(1, 5, 6)$, where the shift equals 2 and, therefore, fits 3 times in S.

3.1.1 On the search for the necessary and sufficient conditions of circularity

In this section the essential properties for a basic design to be circular are investigated. Given a subset structure, the question is which properties have at least to be checked in order to be sure that the design is basic and circular according to Definition 1.

The booklets in a circular design as defined by Definition 1 have the property that there is a largest intersection size $c = b - s$ such that for each subset there are **exactly** two other subsets, its c-partners, with which it has an intersection of size c. This property is called 'equal pairwise c–intersection' further abbreviated to '*EPI* '. A necessary condition for this property to hold for all booklets is that the last booklet turns around at the right edge of

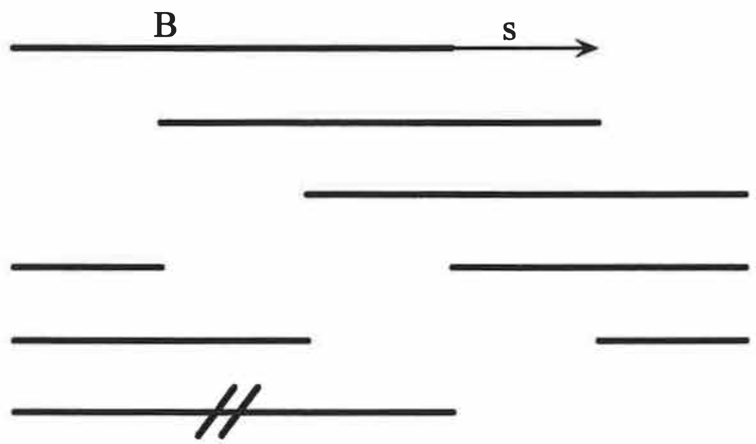

Figure 2: Example of design constructed following definition 1 for circular designs

S to reappear at the left edge with c units to intersect with B_0. This property follows immediately from the requirement that the shift fits an integer amount of times into S. This requirement also implies that the same design is obtained when one shifts from right to left or otherwise. One may also imagine the left and right border of S glued together to form a circle. It is an interesting question whether the property EPI is sufficient to have a circular design. Two simple examples of non-circular designs that prove otherwise are given in Figures (3) and (4). Each horizontal line represents a subset of size b.

Figure 3: A non-circular design with EPI

These counter examples suggest which conditions have to be added to obtain necessary conditions:

Figure 4: EPI with more than one circuit

1. *Minimum Intersection* (MI). For all subsets its two c-partners have minimum intersection size $\max(b - 2\,(b - c)\,, 0)$, to exclude designs as in Figure (3).
2. *One Cycle* (OC). In traversing the booklets by visiting its not yet visited c-partner all booklets are visited, to exclude multiple circuits as in Figure (4).

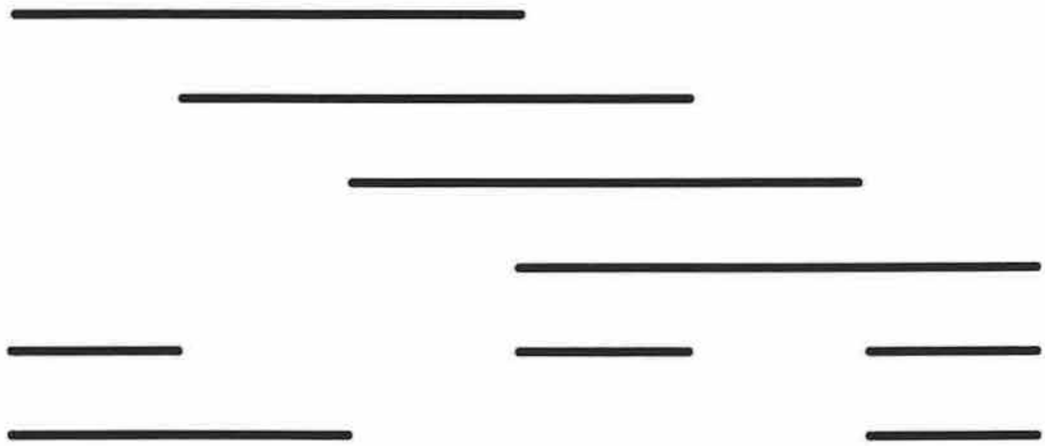

Figure 5: A noncircular design with EPI, MI, and OC

But also these three conditions (EPI, MI, and OC) are not sufficient for a design to be circular according to definition 1 as a counterexample in Figure

5 shows. Note that units 4 and 5 cannot be interchanged without causing a hole in subset 2. Whereas a circular design of the given sizes would have an empty intersection of subsets 2 and 5 in this example these subsets share unit 4. It seems that one has to conclude that the property MI cannot be restricted to only the c-pairs, but must be generalized to all subsets. If in the last counterexample subset 4 would have contained unit 4 instead of 3, the intersection of subsets 1 and 4 would have been empty, without changing the sizes of the intersections with other subsets. Therefore, *EMI* is defined as: every subset has two c-pairs with minimum intersection with each other and all other subsets. By now, one could conjecture that a basic design is circular if and only if it has EPI, EMI, and OC. However, a counter example is still very simple to construct as shown in Figure 6 where EMI and MI are indistinguishable by just having three subsets. The pattern in Figure 6 is not

Figure 6: A basic design that has EPI, MI, and OC, but is not circular. $n = 7$, $b = 5$, $c = 3$

circular, because the subset that results from shifting the last subset over the constant shift size does not result in the first subset, but 'overshoots'.

Two properties of circular designs that are violated in Figure 6 can be discerned. The first is that there exists no integer k such that $ks = n$. Call this property *SI* (shift integer). Apparently, SI is a necessary property to have circularity. The second property concerns the uninterruptedness of the c-intersections. In Figure 6 it is not possible to find an ordering of S where all c-intersections are uninterrupted. In the sequel it will turn out that this property must be replaced by a more encompassing one, and will, therefore, further not be dealt with.

As in Figure 3, Figure 5 also draws attention to another property of circular designs, that is homogeneous unit frequency. If subset 5 in Figure 5 would have contained unit 5 instead of 4, all units would have had frequency 3, whereas now unit frequencies range from 2 to 4. In circular designs frequencies differ 1 at most. Having homogeneous unit frequencies is a favorable property of a design to lower the largest estimation error. The following theorem states exactly the unit frequency in a basic circular design as a function of its determining parameters. From which it immediately follows that in a circular basic design the unit frequencies differ 1 at most.

Theorem 1 *In a circular basic design, the unit frequency is* $b \backslash s$ *or* $(b-1)\backslash s+1$, *where* $\backslash$ *denotes integer division*

Proof. *Choose an arbitrary unit* u *and an arbitrary subset* B *from the design not containing* u. *Consider the design as a series of shifted* B'*s over* s *units. The first subset that hits* u *has passed* u *by at least* 0, *and at most* $s-1$ *units. Therefore, the left most unit* v *of that first hitting* B *has distance at most* $b-1$, *at least* $b-s$ *to* u. *Consequently, it takes* $(b-1)\backslash s+1$ *to* $(b-s)\backslash s+1$ *shifts for* v *to move beyond* u, *resulting in* $(b-1)\backslash s+1$ *to* $(b-s)\backslash s+1 = b\backslash s$ *hits, including the first hit. Clearly, both values are either equal or differ at most 1.* ∎

A basic design with unit frequencies that differ at most 1 is called *homogeneous.* Because homogeneity is easier to check than MI or EMI, it would be interesting if MI or EMI and homogeneity implied each other. However, one can construct a homogenous basic EPI and OC design that is not EMI, as shown in Figure 7, where, for example, the third booklet has an unnecessary nonempty intersection with the first booklet.

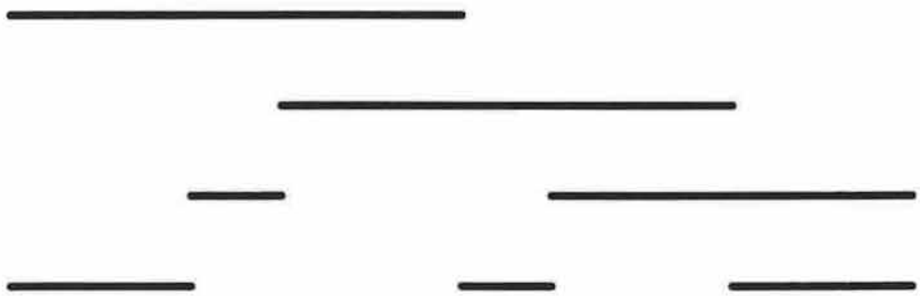

Figure 7: A basic design that has EPI, MI, and OC, but is not circular. $n=7$, $b=5$, $c=3$

However, by violating EMI in these examples another property is also violated. In circular designs, according to Definition 1, there exists an ordering of S where all subsets of the design consist of an uninterrupted series of units *USU*. The edges of S are supposed glued together. In Figures 5 and 7 the USU property is violated. Apparently, also USU is necessary for a circular design. Perhaps USU, EPI, MI, SI, and OC are necessary and sufficient for circular designs. However, in the lemma's below it is shown that USU and EPI are sufficient.

Lemma 2 *MI is implied by EPI and USU*

Proof. Consider a subset B and its two c-partners B' and B'', and assume S to be ordered such that these three subsets are an uninterrupted series of units. For B' to be uninterrupted means that its intersection with B is located and uninterrupted on either side. Suppose B' and B'' share the left part of B, then the units of B' not shared with B must be located at the left of B. The same holds for B'', which must also be on either side of B. However, if it would also be located at the left side of B it would coincide with B' and $Size(B' \cap B'') = b > c$, which violates EPI that demands that c is the largest intersection. Consequently, B' and B'' are located on different sides of B. For instance, the units $B \cap B'$ are all located at the left side of B and $B \cap B''$ at the right side of B, and consequently their intersection is minimal. ■

Corollary 3 *Every subset in an EPI and USU basic design has a left and a right c-partner*

Lemma 4 *OC is implied by EPI and USU*

Proof. Start with an arbitrary subset B. Without loss of generality one can assume an ordering of S such that B is on the left side of S and all subsets are uninterrupted. According to the above Corollary one can select the right c-partner of B, of this subset again its right c-partner and so on, until one finds the left c-partner of B and next reencounters B because the number of subsets is finite. Suppose one does not reencounter B, but a subset at the right of B, then there must be a break in the chain, that can only be result of an interrupted subset, or a nonzero intersection, which both are against the assumptions. If after the first encounter of units common to B, one does not reencounter B itself, but only incomplete intersections, then the reasoning at the end of this proof on a second chain applies, which also results in a violation of assumptions. Moreover, the chain from B to B must cover S, for if it does not, at least one subset must have passed the left side of S without crossing the right side. Consequently, that subset is interrupted by the not covered part of S, against the assumption of USU and the chosen ordering of S. Therefore, the situation depicted in Figure 4, where the third and sixth booklet are interrupted, cannot occur. Now suppose that, with the same ordering of S, there is another chain in the design of uninterrupted subsets of size b with intersections of size c. Because the first chain already covers S, the second chain must have nonzero intersections with every subset of the first chain. Consider the part of this chain that has nonzero intersections with B and take the subset C with the smallest left intersection with B of size a. If $a > c$ then the assumption EPI that c is the largest intersection is violated and we are finished. Therefore, $a < c$. Then the right c-partner of C has an intersection of size $\min(b, a + c)$ with B, which obviously violates EPI. ■

Lemma 5 *SI is implied by EPI and USU*

Proof. In lemma 4 the right c-partner of a subset is shifted c units to the right, and by traversing the chain of right c-partners one closes the circle by reencountering the subset one started with. Now suppose that SI does not hold, then the chain must have crossed the edge of S more than once. Then again consider the subset that while tracing the chain of right c-partners, starting with B, after covering S for the first time crosses the edge of S with smallest left intersection of size a with B, and repeat the reasoning of the previous lemma. ■

Corollary 6 *A basic design with EPI and USU has exactly k subsets with $n = k(b - c) = ks$*

The previous lemma's and the content of their proofs imply the main result of this section:

Theorem 7 *A basic design is circular if and only if it has USU and EPI*

3.2 Linear Basic Designs

In many respects linear basic designs resemble circular designs except that the chain does not go round but has a beginning and an end. There is an ordering of S where the first subset only has a right c-partner, and the last subset only a left c-partner, and the rest of the subsets a left and a right c-partner. Except for the last subset S is partitioned by the first $(k-1)$ shifts. Therefore, in linear designs an adapted property SI does apply for $S - B$, and we have that $(k-1)s = n - b$, again with k the number of subsets. A less favorable property of linear basic designs is that the unit frequency is, in general, not homogeneous. The first and last s units have frequency 1, whereas the frequency of units in the midrange of S is related to the frequency of units in a circular design. Their frequency equals $\min(k, b \backslash s$ or $(b-1) \backslash s + 1)$, because the frequency cannot be larger than the number of subsets.

Linear designs are also less attractive in the present context of the below treated recursive application of basic designs. In a certain variant, where intersection sizes are constant over recursion levels, recursion does not matter. That is the design does not change with the number of recursion levels. With circular designs recursion does matter also with constant intersection sizes. Also linear designs are less homogeneous the more recursion levels are involved. The begin and end units have frequency one whatever the recursion level, whereas the frequency of some of the middle units may double with almost every recursion level.

3.3 Balanced Block Design

The balanced block design is obtained as follows. First one chooses a wanted booklet size. The booklet size is divided by two. This gives the block size d. Next the item bank is partitioned into a number k of blocks of about size d (see Appendix). Then each pair of blocks defines a booklet. And the number of booklets in the design equals $k(k-1)/2$. The balanced block design has some beautiful properties such as EIF and LPDO. A disadvantage is that the number of booklets tends to be high, a property that it shares with designs produced with more than a few levels of recursion of the circular and linear designs.

Although the balanced block design can be recursively applied in an arbitrary number of recursion levels, it is easily shown that the resulting design is equivalent with a design that is directly obtained without a recursive application (see also below the Theorem on LPDO conservation with recursion). Therefore, it is produced directly.

Figure 8: A simple example of a common anchor design

3.4 Common anchor design

The design where all subsets share the same set of units is called the common anchor design, see Figure 8.

An obvious advantage of this type of design is that the distance between two arbitrary units is at most two. An obvious disadvantage is that a small subset of units, the anchor, obtains much more observations than all other units, resulting in very unequal estimation errors of item parameters. Because all subsets are connected, the graph theoretical representation is not very informative for this basic design. Moreover, the recursive application, treated below, is not as straightforward as for the circular and linear basic designs. This is caused by the fact that in a common anchor design the units are not considered equal. The anchor units have, from a design point of view, another status than the other units. By themselves they constitute a complete design, whereas the other units can be considered part of a completely disconnected linear design. The design is connected only by the complete design of the anchor test. Therefore, recursive construction, if at all useful, would have to follow another path than it is conceived here. As such these designs fall outside the scope of this treatment, even more so than the balanced block design. Since common anchor designs are often used, however, it would be nice to compare the results of common anchor designs with the other designs. Therefore, common anchor designs with a range of anchor sizes, booklet sizes, and bank sizes will be studied, and compered with the results for the other design types.

3.5 Other basic designs

Lemma 8 *If a basic pattern of subsets of $b \geq 2$ units is complete, that is, contains all $\binom{n}{b}$ subsets of b units, it is LPDO.*

Proof. Suppose a basic pattern contains all $\binom{n}{b}$ subsets of $b \geq 2$ units. Choose an arbitrary pair P of units, and an arbitrary subset S^P of b units that contains P, then the basic pattern contains S^P, because it contains all subsets of b units, and moreover S^P contains P. ■

The converse is not true in general, only for $b = 2$. For $b > 2$ a basic pattern may be LPDO and yet may not contain all $\binom{n}{b}$ subsets. An example is the complete basic pattern $(n, b) = (4, 3)$ with subset $\{1, 2, 3\}$ left out. More in general there are $\binom{n-2}{b-2}$ subsets that contain a certain pair P of units, so that $\binom{n-2}{b-2} - 1$ of these subsets may be omitted, while retaining a subset that contains P, and a subset for any other pair P', because at most $\binom{n-2}{b-2} - 1$ of subsets that contain P' are omitted. Therefore, the question emerges which subsets a basic pattern must at least contain to be LDPO. For the following situation this can be calculated easily. Not only the number of subsets is found but also a way to construct the desired subset structure.

If $n = mc$, $b = kc$, and $n = r(b - c) + c$ for some integers k, m, r, the minimum number of subsets of S for a basic pattern P to be LPDO can be found as follows. First note that the size of c is immaterial for this problem, therefore, choose $c = 1$, so that $k(= b)$ can also be omitted from the problem. Then the problem can be structured as shown in Table 1.

Table 1: construction of an LPDO design with booklet size b

Used units				
1	$1 + b - 1$	$1 + 2(b - 1)$	...	$1 + r(b - 1)$
#booklets	1	$1 + b - 1$	...	$1 + (r - 1)(b - 1)$
				$= r + (b - 1) \sum_{i=1}^{r-1} i$
				$= r + (b - 1)r(r - 1)/2$

So $n = 1 + r(b - 1)$ is partitioned into one set of one unit and r sets of $b - 1$ units. Take the first unit, and add the first of r sets of $b - 1$ units. This results in 1 booklet. Next add the second set of $b - 1$ units. Each of the $1 + b - 1$ already used units has to be combined with the new set of $b - 1$ units to create an LPDO booklet series of $1 + 1 + b - 1$ booklets. This goes on until all r subsets of $b - 1$ units are used, at which point $r + (b - 1)r(r - 1)/2$ booklets are created.

4 Recursive designs

In this section recursive designs are defined, and some properties are proved. In constructing a design for a set of items, one divides the set into a pattern of subsets. Ideally the pattern exhibits certain desirable properties like equal or homogeneous item frequencies, or largest pairwise distance equal to zero. In a recursive design we start the division of the item set into a basic pattern of subsets, and repeat this process for each of the subsets, either with the same basic pattern, or with some other basic pattern, or even different basic patterns for each of the subsets. This subdivision of subsets can be repeated until subsets (booklets) of a desirable size are obtained. If at each recursive step the same basic pattern is used, the design is called I-Recursive, else D-Recursive. In the sequel, the question will be addressed whether interesting properties of basic patterns are retained with recursive repetition. First, two constancy theorems on equal item frequency (EIF) and zero largest item pair distance (LPDO) are presented.

Theorem 9 *If within each recursive level the same basic pattern with EIF is applied then EIF holds for the entire design. Basic patterns may differ between recursive levels.*

Proof. It is enough to prove that given a pattern of subsets with EIF, if each of the subsets is again divided into the same pattern with EIF, the resulting design retains this property. It is easily seen that item frequency equals the product of the item frequencies over the recursion levels. Because this holds for each item, all items have equal frequency. ■

Theorem 10 *If the basic patterns of all recursion levels conform to LPDO, this holds for the entire design. Basic patterns may differ within and between recursive levels.*

Proof. Again it is enough to prove this for one level to the next. We start with a LPDO pattern. Now choose an arbitrary pair P. LPDO means that for P there is a subset that contains P. This subset is again divided with a LPDO pattern, which means that there is a subset of this subset that contains P. ■

The two Theorems above are easily generalized as follows. Like we have EIF, so we also may have equal pair frequency EPF, and, in general EtF, for subsets of size t. Denote the frequency of subsets of size t in an EtF design with λ_t, then, for instance, an EIF design with item frequency f is E1F with $\lambda_1 = f$, and an EPF design E2F, etc. Then the above theorems are generalized by

Theorem 11 *If a design G is recursively generated by application of D basic EtF patterns with respective frequencies λ_{td}, $(d = 1, .., D)$ then G is an EtF design with $\lambda_{tG} = \prod_{d=1}^{D} \lambda_{td}$.*

Proof. Again proof is only necessary for level 2 recursion. So assume that we have two basic EtF patterns with respective frequencies λ_1, and λ_2, and generate a design D by recursive application of these two basic patterns. Choose a t-subset T. Then there are λ_1 subsets in basic pattern 1 containing T. Each of these λ_1 subsets contains λ_2 subsets that contain T. Therefore D contains $\lambda_1\lambda_2$ subsets that contain T. ■

Because the LPD of a design determines to a large extent the estimation error of pairwise parameter differences, it is important to know how LPD develops with recursive design constructions if basic patterns are used with LPD greater than zero. To derive an inequality we introduce the concept of multilevel graph. A recursive design is in a natural way linked to a multilevel graph. First consider the graph of the highest level design. The blocks are represented as the nodes of the graph, and there is an edge between two nodes if the corresponding blocks overlap. Now apply a design to each of the blocks. On the level of the graph this means that each node is represented by a higher level graph.

The fat triangle in the graph represents the first level of the graph. A fat line between two squares A and B indicates that there are connections between one or more pairs of level 2 points, of which one member of each pair belongs to A and the other to B. The first level connection between the two upper squares is indicated by the fat line, the level two connections between these two squares by thin lines.

With these concepts in mind an upper bound for the distance between two level n points can now be formulated.

Theorem 12 *The LPD in an n-level graph with LPD$_i$ at level i equal to m_i is bounded from above by $\sum_{i=1}^{n} \prod_{j=i}^{n} m_j$*

Proof. Choose a pair of units that have to be connected and for each unit a booklet that contains them. These two booklets can be represented as a

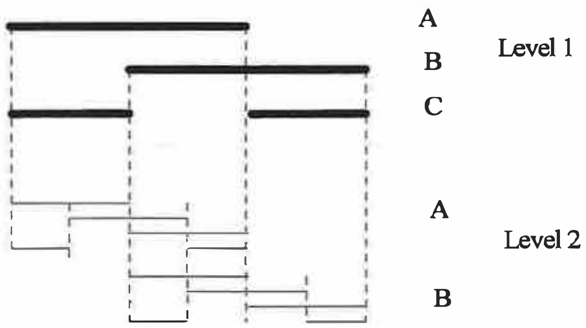

Figure 9: Recursive design with triangle first level design and square second level design (see graph below). To simplify the picture the second level is expanded only for first level blocks A and B.

pair of nodes (a_n, b_n) at level n for which a path has to be found, and start at element a_n. Take the element a_{n-1} from which a_n was produced, and likewise b_{n-1}, and so on up to a_1 and b_1. Then at level 1 there are at most m_1 edges to be crossed to go from a_1 to b_1, visiting level 1 nodes $a_1 = a_{10}...a_{1m_1} = b_1$. However, to cross a level 1 edge like a_{10} to a_{11} one has at level 2 to find a route of at most m_2 edges from a_2 to a level 2 node a_2' that shares a unit with a level 2 node a_2'' produced by a_{11}. One has to process the next level 1 edge in the same way. After processing the last level 1 edge one still has to process at most m_2 level 2 edges to find b_2, giving a total distance of at most $m_1m_2 + m_2$ level 2 edges. Continuing at level 3 one obtains in the same vain at most $(m_1m_2 + m_2)m_3 + m_3 = m_1m_2m_3 + m_2m_3 + m_3$ level 3 edges. Continuing up to level n one obtains a distance of at most $\sum_{i=1}^{n} \prod_{j=i}^{n} m_j$ level n edges from a_n to b_n. Note that after arriving at booklet b_n one has to increase the distance at most by one to reach the wanted unit in b_n. ■

This theorem yields a Corollary that is essential for CML-estimation. Connectedness of a recursive design is guaranteed if the employed basic designs are:

Corollary 13 *Recursive application of connected basic designs results in a connected design.*

Remark 14 *Note that the LPD will in most cases be smaller than this upper bound. As an example see the above figure. For instance, above it was shown*

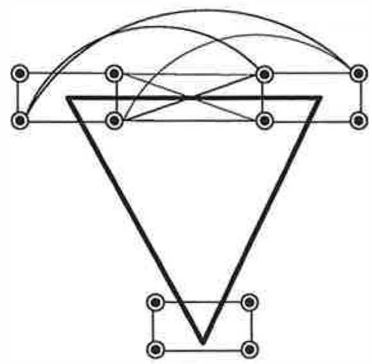

Figure 10: Multilevel graph of the above design with triangle first level design and square second level design. The second level connections between the upper two square designs are depicted.

that LPDO is kept with recursive constructions, whereas this theorem gives n as an LPD upper bound for an n-level design construction of LPDO patterns.

5 Standard errors of item parameter estimates

In this section the effect on standard errors of parameter estimates are evaluated as a function of recursive design characteristics. To this end a number of simplifying assumptions are to be made to eliminate less essential variables. The evaluation is restricted to the Rasch model where all item and person parameters are assumed equal. Moreover, the number of records per booklet is assumed constant. The investigation is restricted to I-recursive designs, so that the basic design is the same at all recursion levels.

The standard errors of item parameters are found as follows. Minus the second derivatives w.r.t. all item pairs in a record are accumulated over records in the so-called Information matrix I. With the above assumptions the information matrix I is proportional to the $n \times n$ matrix where diagonal cell (i, i) holds the count over booklets of the occurrence of item i, and offdiagonal cell (i, j) $i \neq j$ the frequency of items i and j in the same booklet weighted by $-1/(d-1)$, with d the booklet size (see also Verhelst, 1993, Formula (12)). To compare different designs, the information matrix must be made invariant for the number of observations per item. Because we are only interested in the effects of the design structure, and not in the amount of ones that is produced to represent this structure, I is divided by the average number of observations per item.

$$I_s = I \frac{n}{D_{++}} \tag{1}$$

where D_{++} denotes the count of nonblank cells in D. To obtain the covariance matrix C of the parameter estimates, the first row and column of the information matrix I_s is deleted. The resulting matrix I_0 is inverted, and after adding a first zero row and column it is pre- and postmultiplied by a normalization matrix P. As an estimation of the standard errors of the parameter estimates one takes the square root of the diagonal elements of the covariance matrix C.

$$C_{n\times n} = P_{n\times n} \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & I^{-1}_{0,n-1\times n-1} \end{bmatrix} P_{n\times n} \tag{2}$$

For the normalization matrix P we take the normalization with mean item parameter equal to zero.

The quality of a design is difficult to evaluate directly from the contents of I_s or $C_{n\times n}$. Pukelsheim (1993) discusses several real functions on the space of information matrices or covariance matrices to judge design quality. Among them are the n-th root of the determinant of I. See also Eggen, 2004, ch 3, for the meaning of the determinant of I_0 as a measure for the information content of a design. Other measures of design quality, which are clearer to interpret, are derived from the covariance matrix $C_{n\times n}$: the mean and standard deviation of the square root of the variances, the diagonal elements of $C_{n\times n}$, and the square root of the values of the smallest and largest element on this diagonal.

The number of items n will take values 100, 200, 500 and the booklet sizes are 20, 30, 40, 60. The basic designs are linear or circular. The number of recursion levels is related to the subset ratio, the larger the subset ratio the more recursion levels are needed to arrive at the desired booklet size (see also the Appendix). The maximum of the subset ratio is set at 0.75, with a maximum number of booklets 77147 at $n = 500$, $b = 20$, resulting from 11 recursion levels. The intersection ratio c/b will take values .25, and .50. Note that b is the subset size, and only represents the booklet size for subsets of the last recursion level. In the Appendix it is explained why the user set intersection ratio often is not equal to the realized intersection ratio, and indeed may differ quite substantially.

5.1 Results

Our primary interest is in the mean square root of the estimation error per observation as a function of booklet size and subset ratio. Furthermore, it must be emphasized that user control of intersection ratio is sacrificed for constant bank size and booklet size (see Appendix). This loss of control is especially felt for user set intersection ratio 0.25. At intersection ratio set equal to 0.5 its realized values are always close to 0.5 (Figure 12). The dependency of realized intersection ratio on subset ratio for user set intersection ratio = 0.25 is shown in Figure 11. In comparing Figures 11 and 12 note also the scale values on the vertical axis.

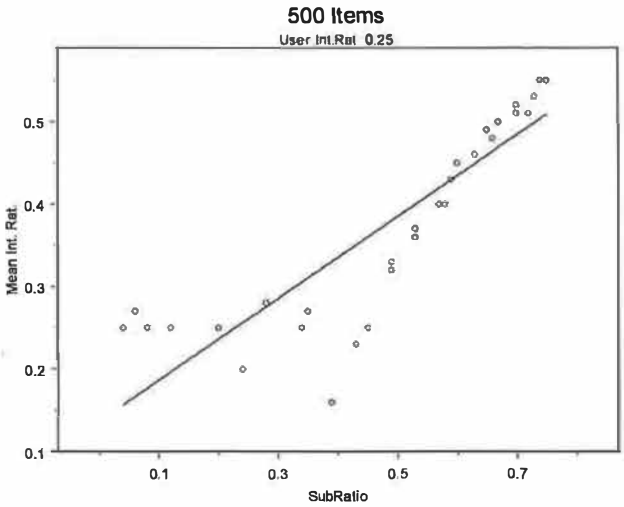

Figure 11: Dependeny between realized intersection ratio and subset ratio for user set intersection ratio = 0.25

Because the results for $n = 100$, and 200 do not essentially differ from $n = 500$ items, discussion is restricted to the latter case.

Figures 13 and 14 show that for linear designs the estimation error hardly depends on recursion depth for the lower and intermediate levels, as was to be expected, but is negatively related for the higher recursion levels. The increasing loss of homogeneity of unit frequencies with recursion level is to be hold responsible for this effect. Moreover, for all combinations of booklet

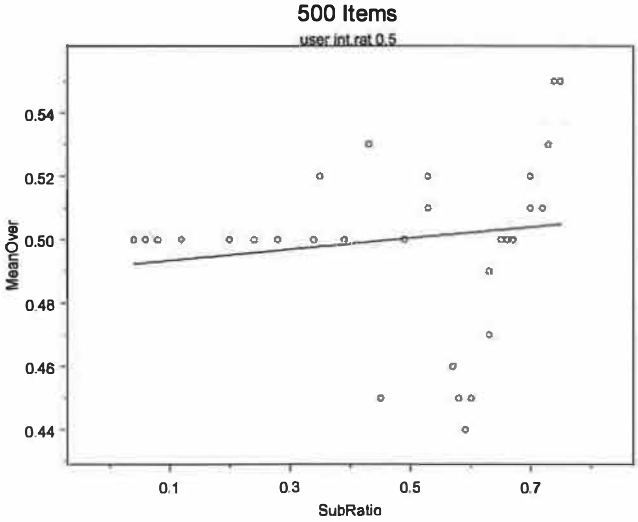

Figure 12: Dependency between realized intersection ratio and subset ratio for user set intersection ratio = 0.5.

size and subset ratio the estimation error is higher than for circular designs. For both design types estimation error decreases with booklet size, although this dependency almost vanishes for the highest recursion levels in circular designs. The user set intersection ratio 0.5 gives a smoother picture of the mentioned relations, because for this value the effective intersection ratio remains about constant with subset ratio. Minimum and Maximum values are shown in Figure 15. Apparently for intermediate values of the subset ratio and for small and large booklet sizes the estimation error for some unit parameters surges. This must, of course, be a consequence of greater variability of unit frequencies. Because of an artifact of the spline smoothing algorithm the maximum surface sometimes runs below the minimum.

The n-th root of the determinant gives an altogether different picture in Figure 16. Note, that a low estimation error is a positive property, while a low determinant is a negative property.

For the lower subset ratios, that is, less recursion levels, linear designs exhibit only a little less quality than circular designs as measured by the $n-th$ root of the determinant of the standardized information matrix. They diverge, however, markedly for the higher subset ratios. Although the esti-

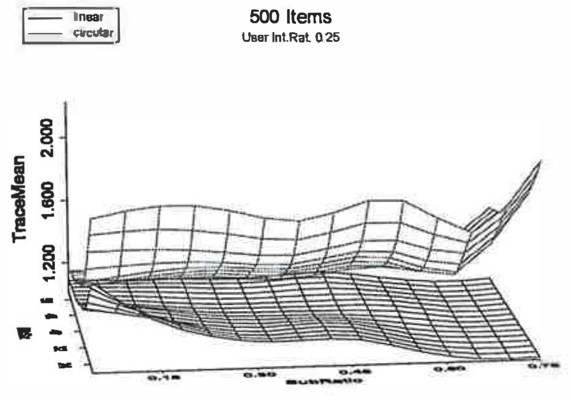

Figure 13: Mean square root of estimation error as a function of subset ratio (x-axis) and booklet size (y-axis). User intersection ratio = 0.25, effective 0.16...0.53. Booklet size and subset ratio decrease toward the left front corner.

mation error also diverges for higher recursion levels, it also shows a large quality difference for the designs with less recursion levels. It is not clear why these two quality measures do not agree for the designs with less recursion levels. For user set intersection ratio=0.25 the dependency between recursion level and realized intersection ratio also interferes. Therefore the two pictures for user set intersection ratio 0.25 and 0.5 differ, and the second in Figure 17 reflects the behavior of the n-th root of the determinant in a less obtrusive way.

In Figure 17 for circular designs the n-th root of the determinant appears almost constant around 0.95, not affected by recursion level nor booklet size, whereas for the linear designs it worsens for higher subset ratios. Especially with decreasing booklet size the quality of linear designs deteriorates. In a complete design the n-th root of the determinant equals 0.96 for $n = 100$, 0.99 for $n = 500$ and its limit for $n \dashrightarrow \infty$ equals 1.0.

Looking at the homogeneity of estimation variances over items one sees a negative dependency for linear designs on the number of recursion levels, whereas, see Figure 18, this property for circular designs is hardly effected. This negative relationship for linear designs also results from increasingly unequal item frequencies with increasing recursion level.

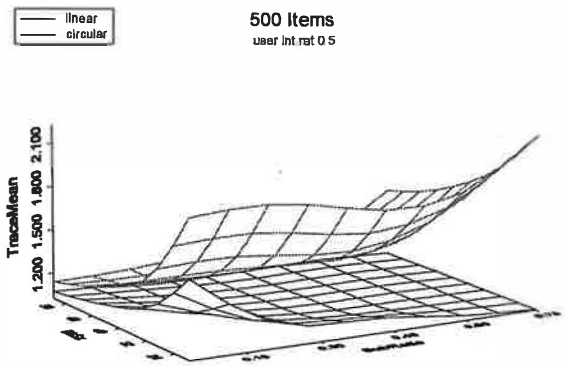

Figure 14: Estimation error as a function of subset ratio (x-axis increasing from right to left) and booklet size (y-axis, increasing from front to back). User intersection ratio = 0.5, effective ±0.5. Booklet size and subset ratio decrease toward the left front corner.

As already mentioned in section 3.3, Balanced Block Designs are produced directly, without the intervention of recursion. Balanced Block Designs are better than any of the other designs, as could be already expected from its properties EIF and LPDO. The results are shown in Table 2.

Table 2. Results of Balanced Block Designs

Bk Size	Realized	#Bklets	$\sqrt[n]{\text{Deter}}$	Mean $\sqrt{\text{EstErr}}$
20	20	1225	0.973	1.0186
30	30	528	0.979	1.0115
40	40	300	0.982	1.0080
60	58	136	0.985	1.0044

On both criteria, the n-th root of the determinant, and the mean square root of the estimation error, Balanced Block Designs score better than the other designs. However, circular designs come close, with appreciably smaller numbers of booklets for the lesser recursion levels. For $n = 500$ recursion level 2, for example, the results are shown in Table 3.

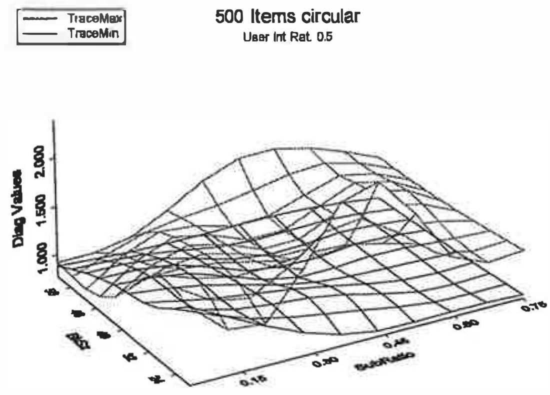

Figure 15: Circular designs min and max of square root values on the diagonal of the covariance matrix

Table 3. Results of circular designs recursion level 2

Bk Size	Realized	#Bklets	$\sqrt[n]{\text{Deter}}$	Mean $\sqrt{\text{EstErr}}$
20	20	100	0.937	1.198
30	30	64	0.949	1.105
40	40	49	0.965	1.053
60	60	36	0.969	1.028

For Common Anchor designs, recursive application is not meaningful. However, interesting relations between booklet size, anchor size and design quality can be obtained. Quality as reflected by the mean square root of the estimation error is shown in Figure 8.

With an item bank of 500 items, the optimum anchor size for small booklets (20 Items) is 4, whereas for 60 items it is 8. With an Item bank of 100 items these figures are respectively 5 and 12. Comparing the means with those of the recursive circular designs it appears that the latter are somewhat more efficient. With an item bank of 500, items, and booklets of 60 items, the lowest value for circular designs is 1.01, the highest 1.06, whereas for common anchor designs these values ar respectively 1.09, and 1.14. The minima and maxima of the square root of the estimation errors are shown in Figure 20. The minima are, of course, for the common anchor units with a

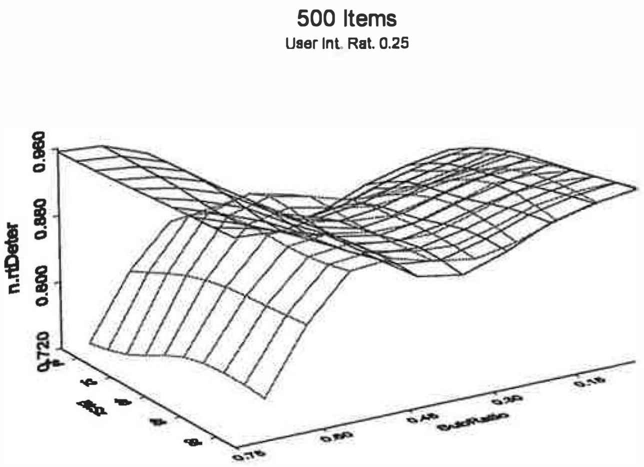

Figure 16: n-th root of the determinant of the standardized Information matrix, user set intersection ratio = 0.25, effective 0.16...0.55

maximum number of observations, whereas the maxima for the other units are large.

However, the n-th root of the determinant gives another relation with quality of anchor size and booklet size. According to this measure, quality increases with booklet size and decreases with anchor size. A strong decrease of quality is found for the small booklets (20 items), and a moderate decrease for the large booklets (60 items). For small anchors the quality increase for booklet size is hardly noticeable.

As already mentioned, it is a disadvantage of common anchor designs that the number of observations for the anchor is in general so much higher than those for the rest of the items, resulting in large differences in estimation error. This is reflected in Figure22. Comparing these values with those in Figure 18 reveals that the standard errors of estimation are here much larger.

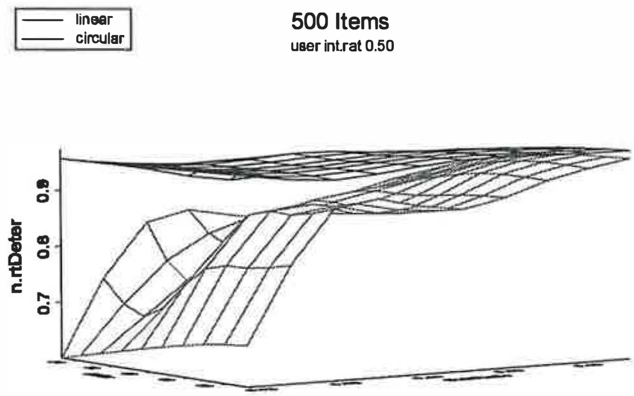

Figure 17: n-th root of the determinant of the standardized Information matrix, user set intersection ratio = 0.5. Booklet size and subset ratio increase toward the left front corner.

6 Discussion

In the above pages we have tried to create and unveil some structure in the overwhelming variety of subset structures that can function as a design of test booklets from a larger set of items. We first studied the basic circular and linear designs, and next unveiled some properties of recursive applications of this type of basic design. The main result of this exploration is that necessary and sufficient properties for a design to be circular are USU and EPI. This means that there is an ordering of S where the booklets consist of an uninterrrupted series of items and each booklet has exactly one pair of booklets with intersection of size c, where c is the largest intersection in the design. We also gave expressions for the frequency of items, in a plain circular design, and upper bounds for recursive designs.

To investigate the accuracy of item parameter estimates we had to introduce some simplifications. Circular designs are, as was to be expected, superiorto linear designs in this respect . Our main indices for design quality per observation were the average estimation error and the determinant of the

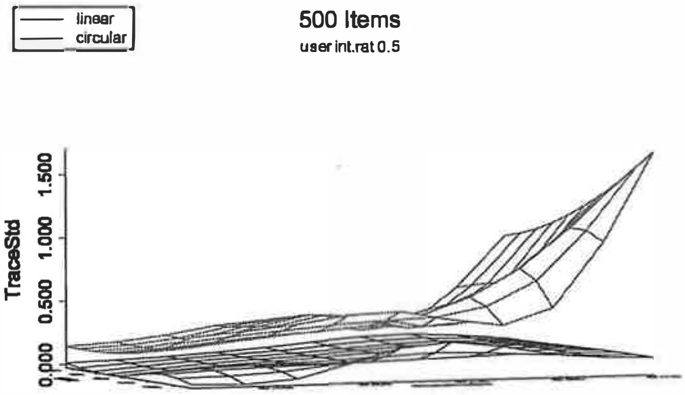

Figure 18: standard deviation of the square root of diagonal elements of the covariance matrix. Booklet size and subset ratio decrease toward the left fron corner.

standardized information matrix. We found that the average estimation error with circular designs for smaller booklets (20 units) increases somewhat with increasing recursion level, especially for the first two or three levels. However, the determinant seems invariant for number of recursion levels. The results on the effect of recursion level on design quality differ. The n-th root of the determinant appears not to be influenced, but the average estimation error shows a positive relationship. Because the latter measure directly represents what we aim for, it seems safe to suppose that quality increases with recursion level.

The number of booklets can grow tremendously with increasing recursion level . In our examples the largest number of booklets in one design amounts to more than 177.000. With paper and pencil tests these numbers are not manageable. In the near future, however, where every student is supposed to interact with a wireless foldable foil screen, these numbers will not pose a real problem. However, the results show, that if a large number of booklets poses no problem, the best results are obtained with a balanced block design. When

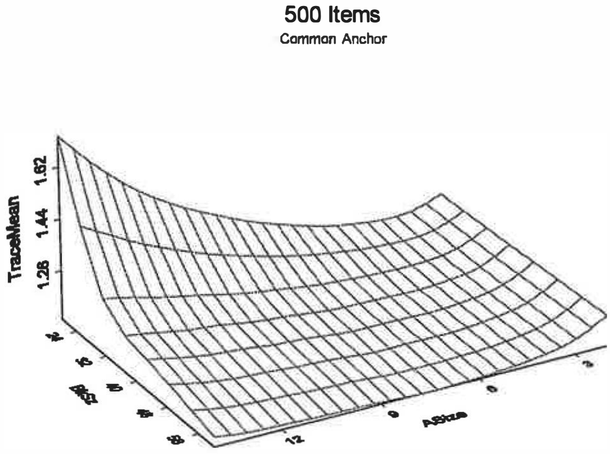

Figure 19: Common Anchor design: mean square root diagonal elements of covariance matrix

the number of booklets is relevant, then one should opt for a circular design with a few recursion levels. With such designs the number of booklets is quite manageable. For the first three levels, for example, with $n = 500$, booklet size 60, they are 17, 36, 64. But even if the variety of designs as presented here is not used in practice, the present study gives some insight in how the construction of a design influences its quality, and how, for instance, these designs compare with the often used common anchor design.

With the common anchor designs it was remarkable that small anchor tests with 4 to 8 items for respectively 20 and 60 item tests were optimal for a bank with 500 items. The common anchor design is less efficient than the circular design, as measured by the estimation error as well as the n-th root of the determinant of the standardized information matrix.

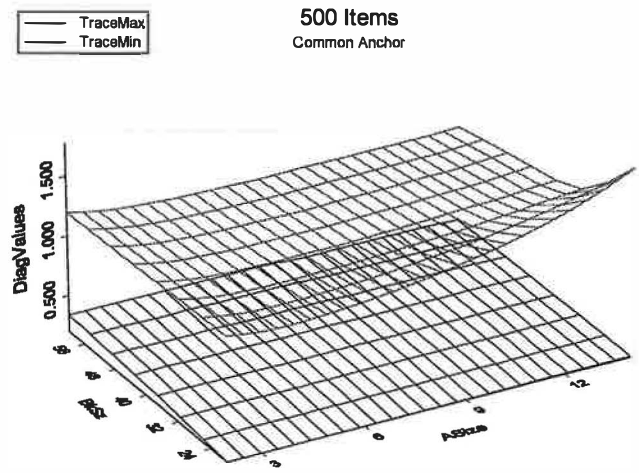

Figure 20: Common Anchor: min and max values of the square root of the estimation errors

7 References

Balakrishnan, V. K. (1995). *Combinatorics, including concepts of graph theory.* Schaum's outline series. New York: McGraw-Hill.

Pukelsheim, F. (1993). Optimal design of experiments. New York, Wiley.

Eggen, Th,J.H.M., (2004). Contributions to the theory and practice of computerized adaptive testing. Dissertation Universiteit Twente. ISBN 90-5834-056-2, Cito, Arnhem, the Netherlands.

Verhelst, N.D., (1993), *On the standard errors of parameter estimators in the Rasch model.* Measurement and Research Department Reports 93-1, Cito, Arnhem, the Netherlands.

8 Appendix : Algorithmic implementation

Circular Designs

In a practical setting some of the determinants of a design, such as total number of items n, booklet size d, and wanted booklet intersection ratio

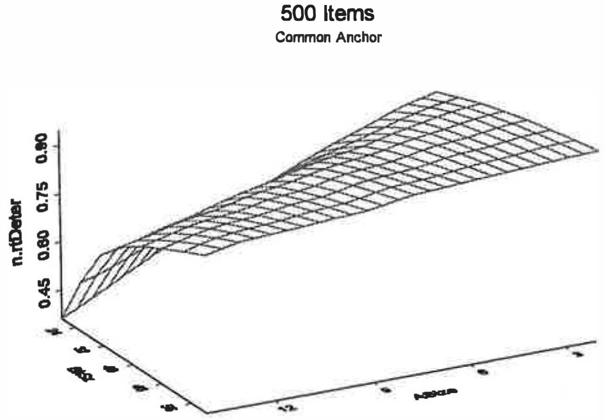

Figure 21: Common Anchor : n-th root of the determinant of the standardized information matrix

which determines the shift size s are fixed. Clearly, within these practical constraints it is, in general, not the case that $n = ks$, for circular designs, or $n - b = (k-1)s$ for linear designs. Moreover, given a certain ratio of subset size and unit count, one does not automatically end up with the wanted booklet size d. Therefore, some precautions are taken to produce a design as close as possible to the wanted criteria within the practical constraints.

In the computer program the user sets the values
n : number of units at recursion level 0
d : booklet size
r_s: subset ratio
r_i : intersection ratio

A subset ratio r_s is specified, that is given a set of size n, the subset size m_v at recursion level v is given by $m_v = r_s^v n$. Now, given a total number of units, not every subset ratio will result in the wanted booklet size d. Therefore, if, after $v = v_d$ recursion levels we want

$$nr_s^{v_d} = m_{v_d} = d \tag{3}$$

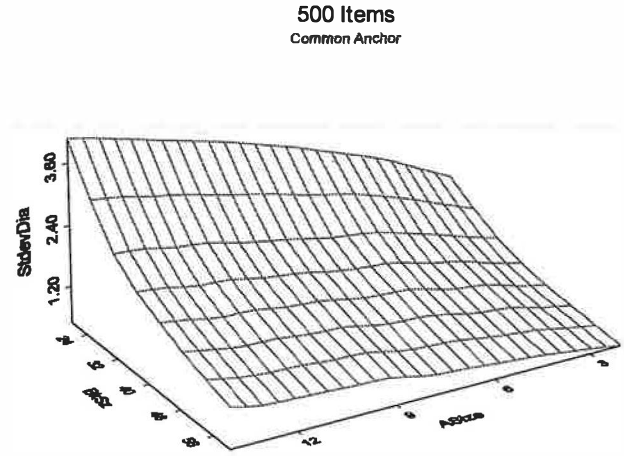

Figure 22: Common Anchor: Standard deviation of the square roots of the estimation variances.

then the subset ratio must be

$$r_s = \left(\frac{d}{n}\right)^{\frac{1}{v_d}} \tag{4}$$

Consequently, if the number of levels v_d is chosen, the subset ratio r_s that results in the wanted booklet size is fixed.

Further, given a wanted intersection ratio r_i, that is the intersection size at recursion level v, is given by

$$c_v = r_i m_v \tag{5}$$

with m_v the subset size, with $m_0 = n$. The shift size then equals $s_v = m_v - c_v$. However, in general, for instance for circular designs, it is not the case that n/s at the zeroth recursion level or m_v/s_v at recursion level v is an integer, as required for a basic design. To accommodate this requirement, the shift size, and, therefore, the intersection size, is adapted for circular designs as follows. First, from one level to the next the set size o_v is passed as a floating point number to be sure that with the right subset ratio one ends with the wished booklet size d, with $n_v = round(o_v)$. Then, at level v the number of subsets $k_v = round(o_v/s_v)$, because for a circular design one needs at least three subsets. Therefore, if $k_v < 3$ its value is changed to 3. This gives a rest number of not used or too few units $p_v = n_v - s_v * k_v$. Note that p_v can be negative. Then the shift size is increased with $p_v \backslash k_v$. This leaves $q_v = p_v \bmod k_v < k_v$ unused units. These are accounted for by increasing the shift size for q_v subsets with one unit if $q_v > 0$ or decreasing the shift size by one for $-q_v$ units if $q_v < 0$. For a linear design a similar strategy is implemented. In this strategy complete control by the user of the intersection ratio is sacrificed for using the entire item bank and producing booklets with the size the user wants.

Real Circularity

There is another pointthat should be stressed in implementing the production of recursive circular designs. In a first implementation, the construction of a basic design given a subset was always started at the item with the lowest index in the subset. Take, for instance, a subset that starts near the end of its superset and 'turns around' the right edge of its superset. The first booklet of the basic design in this subset starts at the first item of its superset instead of its own starting item. This proved to produce an unwanted high dependency of item frequency on position. The frequency of the last item could be 1, against 32 for the first item which resulted in a very high estimation variance of the last item parameter compared to the variance of the first. The dependency of frequency on position was much less by starting the construction of subsets at the start of the superset itself.

A small example can clarify the issue. Attend to the subdivision of the third subset (third thick line in upper part) of the circular design of Figure 23. In the first set of thin lines, the subdivision has started at index 1, in the second set of thin lines the subdivision has started at the start of subset three, that is at one third from the right edge. In shifting the second subset (second thick line) to create the third susbset, its first item is at a distance of one third from the right edge. The item with the lowest index in the third subset, however, has index 1. If one starts the second level recursion at item indexed 1, that is at the very left of the figure, (first set of thin lines) this item has at the second level frequency four, and the last item, at the right keeps having frequency one, whatever the recursion level. In the second set of thin lines we start at the start item of the third subset (implementing real circularity), that is, at one third from the right edge. Then this start item

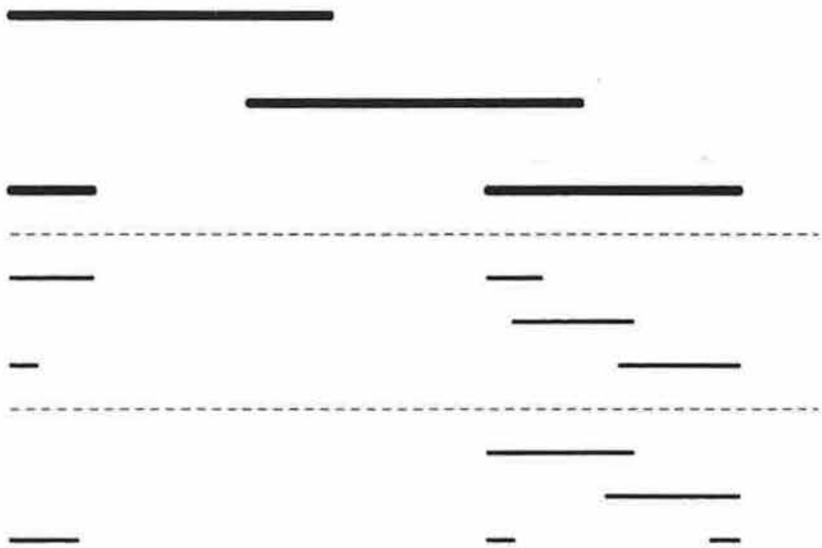

Figure 23: Starting the recursive repetition of the upper basic design for the third subset at index 1 (first subdivision) or at the first item of the third subset, index at 2/3 of the index range (second subdivision)

has frequency three or four at level two, and the last item frequency two. Start items always have a frequency advantage, but with real circularity it is not too often the item with index 1 in the previous level set.

Balanced Block Designs

The balanced block design is constructed as follows: The user defines a wanted block size d, it is assumed that his wanted booklet size equals $2d$. Then

$$\begin{aligned} k &= round(n/d) \\ m &= round(n/k) \\ e &= n - mk \end{aligned} \tag{6}$$

where k denotes the number of blocks, and m is the realized block size. The first $|e| \leq \min(m/2, k/2)$ blocks are diminished by one ($e < 0$), or increased by one ($e > 0$). And every pair of these blocks defines a booklet. Because, in general, $m \neq d$, the realized booklet size can also deviate from the wanted booklet size. Note that booklets can deviate at most two units from the 'realized' book length $2m$.

Common Anchor Designs

In the implementation of common anchor designs, the anchor size a is considered a given. Therefore, no intersection size can be varied to obtain an integer number of booklets of the desired size that fit into the item bank. Given a, and a wanted booklet size, only the booklet size can be varied to obtain a cover of the item bank. The algorithm to obtain a booklet size as close to the wanted booklet size, and have a cover of the bank is as follows. Denote the wanted booklet size with d, realized booklet size with m (except for correction by e, see below)

$$\begin{aligned} k &= round((n-a)/(d-a) \\ m &= round(a+(n-a)/k) \\ e &= n-a-k(m-a) \end{aligned} \tag{7}$$

where $|e| \leq \min((m-a)/2, k/2)$ denotes the number of booklets that is one larger ($e > 0$) or one smaller ($e < 0$) than m. Consequently, the realized booklet size can deviate somewhat from the desired size. Keeping the anchor on a fixed value, and allowing relatively small deviations from the wanted booklet size allow a clearer picture of the effect of anchor size on design quality.