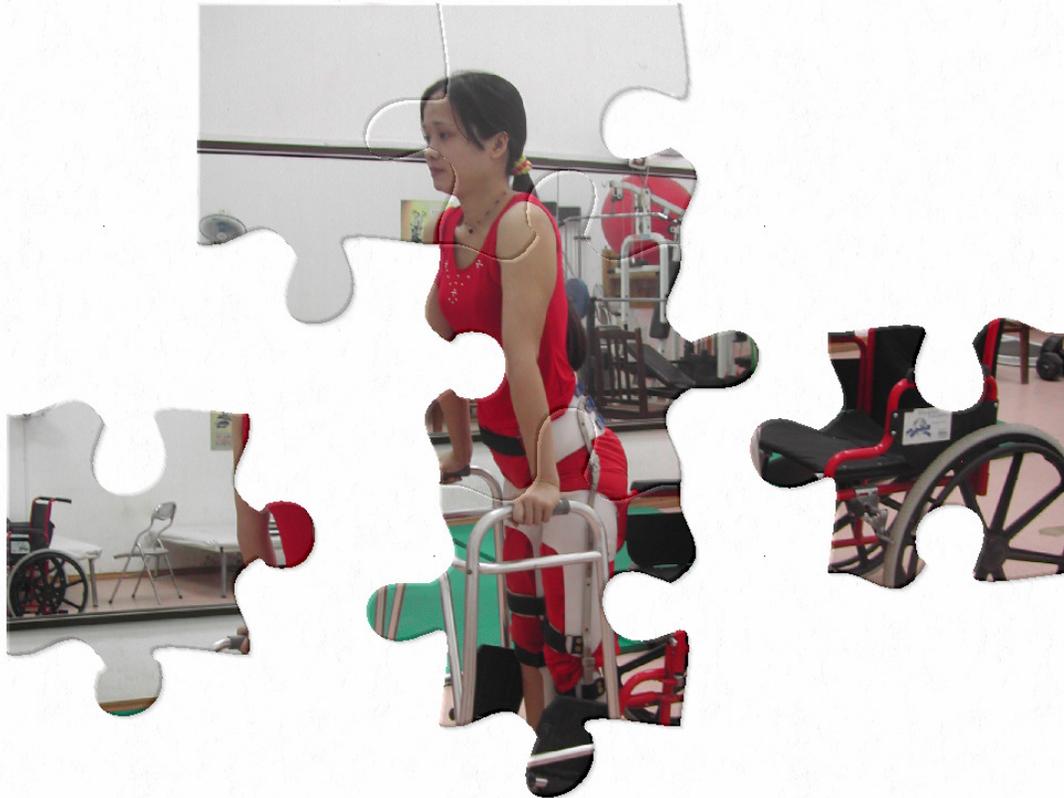


Aldo Benini

Putting the Pieces Together: *Follow-up of persons with disabilities using multiple indicators, Hanoi, Vietnam*



Afterthoughts 2008

Recently, when I was working on trauma patient data in Asia, I was reminded that I had written this note in Vietnam in 2005. It offers some guidance for administrators of similar programs who may want to measure the progress that participants with disabilities make in their everyday coping routines. These observations are of minor value now, given the special constellation in which the program in Vietnam was being monitored; chiefly, the reader may benefit from a number of warnings of how not to do progress measurements.

The note retains a greater interest as a testimony to the precarious conditions in which social programs – both of governments and of NGOs – in poor countries attempt to

observe and report their activities and achievements. Pressures to demonstrate impact have grown, but the design and execution of data collections meant to produce the evidence are all too often poorly supported. They tend to be relegated to persons marginally qualified for the job, without professional supervision worth this name, and under pressure to produce results quickly before they have had time to assimilate the necessary concepts and technical skills. When I look at their work environments, I cannot help but feel that many of these humble monitoring officers discharge their duties in a mood of stoic heroism. They have nowhere to go and ask technical questions. They improvise data structures in undocumented software or in closed database applications that fleeting consultants installed without the ability to modify them, and yet are required to produce output that others will somehow find meaningful. Their products may be accepted mostly because the consumers are too innumerate to question the steps from data to reports.

On the positive side, I have noticed a growing eagerness among program staff (I have had far more exposure to NGO than to government programs) to assess the impact they are making and to develop, on their own terms, systems of measuring it. Instruments can be looked up more readily on the Web, and the ability to do such searches productively may have grown. In the program monitoring units, I often find a great eagerness to improve data management and analysis skills, and the enthusiasm and gratitude shown to someone from outside who bothers to sit down and work with them are infectious. This was the case also in the Bach Mai Hospital in Hanoi, where the concerned staff, and particularly an isolated statistician who had been struggling with patient survey data, asked me to help with data tables and measurement scales.

Washington, December 2008

Summary

At the request of the statistician in the Vietnam Veterans of America Foundation team in Bach Mai Hospital, Hanoi, I analyzed pre-/post-treatment data on over 300 patients that VVAF personnel had followed up. This follow-up was part of the project “*Sustainable Benefits for the Mobility Impaired, 2003-2005*” conducted jointly by the Ministry of Health, Government of Vietnam, and VVAF.

To the extent that the information, particularly the retrospective baseline, is reliable, this sample shows significant improvements in personal hygiene, mobility, cognitive ability and social integration. These findings should be interpreted with caution because of a likely strong “reverse gratitude” effect.

In several meetings at the office and at Bach Mai, the statistician, Ms Nguyen Mai Huong, and I reshaped the data from the inconvenient current format to tractable tables, using statistical applications that she is familiar with. Ms Huong is now capable of arranging and, although in statistically more modest approaches, describing the data and, specifically, the functional gains.

Also, Ms Huong showed me a proposed new assessment tool, the “CIR/RERC Prosthetics Users’ Survey”, from which she said certain items may be inserted into the current instrument. I advise against adopting yet another format; it will make old and new evaluations less comparable and will do little to mitigate the kinds of problems inherent in the current instrument. The emphasis should be on improving reliability and analytic capacity with the existing tool.

An earlier draft of this note was mailed to Ms Kerry Fisher and was discussed with Ms Huong and with the Clinical Supervisor, Mr. Wendell Endley, at a meeting on 27 January. They graciously allowed me to use the case material and the cover photo for this note.

SUMMARY	1
INTRODUCTION.....	5
PROGRESS WHILE IN THE PROGRAM – SOME FINDINGS.....	6
WHO ARE THE PERSONS IN FOCUS HERE?	6
WHAT WAS ASSESSED?	7
HOW CONSIDERABLE ARE THE IMPROVEMENTS?	8
PROGRESS IN MOBILITY FOR DIFFERENT DISABILITIES	9
PROGRESS OVER TIME	10
HOW SHOULD VVAF TALK ABOUT THE FINDINGS?	10
METHODOLOGICAL CONSIDERATIONS.....	13
THE “IMPACT INDICATOR FORM”	13
DOMAINS AND ITEMS	15
ITEM CATEGORY DISTRIBUTIONS BEFORE AND AFTER TREATMENT	15
AN ALTERNATIVE: POLYCHORIC PRINCIPAL COMPONENT ANALYSIS	16
CAPACITY BUILDING: ANALYSIS BY BACH MAI STAFF	18
CONCLUSION	19
REFERENCES.....	20
STATISTICAL APPENDIX	21
DESCRIPTIVE STATISTICS	21
<i>Age and gender of patients</i>	21
<i>Patient diagnoses</i>	21
<i>Time from first to last evaluation</i>	22
POLYCHLORIC PRINCIPAL COMPONENT ANALYSIS OUTPUT.....	22
<i>Personal hygiene</i>	22
<i>Mobility</i>	23
<i>Cognitive ability</i>	24
<i>Social integration</i>	25

Introduction

Mr. X. [traceable through his patient ID BM0908] is a 39-year old man from Hoah Binh with both legs amputated above the knee. A mobile outreach team from Bach Mai registered him in November 2001, and a year later he was fitted with prostheses. During a follow-up interview in May 2003, he responded to a fair number of questions that Bach Mai staff asked him as part of an impact assessment. These questions were asked twice, concerning his contemporary conditions, as well as retrospectively as far as he recalled his condition at the time of registration.

Mr. X., like the other 339 patients who underwent follow-up interviews by 2004, answered a series of questions regarding his ability to take care of his personal hygiene, his mobility, as well as a number of psycho-social areas. These included segments on rehabilitation knowledge, pain, participation in family and community, education and work, self-esteem and communication. All in all, Mr. X. was asked two times 24 questions, each detailing an item of his personal and social functioning, and some of them enhanced with specific technical sub-questions such as to whether he was given a special device to meet some basic function.

For example, we learn about Mr. X. that prior to 2002 he was used to eating independently, but was limited in his ability of wash, dress or undress, or go to the toilet alone. Since he came in contact with the Bach Mai workers, he has learned to take care of his personal hygiene himself, except for some help that he still needs in the toilet.

Mr. X. has seen improvements in some other areas of his life too, but things are still difficult for him. He rarely finds work, and still feels that people sometimes look down upon him.

Mr. X.'s case is easy and quick to present. Making sense of all the data collected through the follow-up interviews is harder. The relevant database "impact" has 759 records [as of 19 January 2005] and 29 fields. The data is in an analytically uncomfortable format. The VVAF Bach Mai statistician has not found a viable approach to evaluating it.

Also, the 340 patients present with evaluation data from more than one point of time – the condition to assess progress - represent only a tiny fraction of the 9,065 patients with records in this database – less than four percent. Working on these select few has an opportunity cost in terms of time left to look after the large majority for normal updating and reporting.

In the remainder of this note, I will give a nutshell description of the "Impact Indicator Form" used by the patient assessment staff and a digest of some findings that the rapid reader may want to glance over. Then the note turns increasingly methodological. This has two elements. I will briefly point out some basic principles of data management that were discussed with the VVAF Bach Mai statistician for making the assessment data analytically accessible. This is in the spirit of capacity building, hoping that simple

analyses by local workers are worth more than exotic algorithms run by a fleeting consultant.

However, in the context of international debates that challenge methods such as the one implied in the current VVAF instrument (see the references), a different method that produces more valid score estimates will be briefly referenced - Polychoric Principal Component Analysis. Results from applying it to the Bach Mai data are used in the first part, but the detailed estimates are mercifully relegated to an appendix. Ms Huong can produce descriptive statistics using simpler methods such as figuring means of item weights and of before/after differences, but it should be kept in mind that an unfavorable reviewer could always assail these as statistically not strictly valid.

I conclude with advice against the proposed adoption of a new assessment tool and for the continuation of the current instrument, whatever its shortcomings, at least until such time when the project management better understands its needs, opportunity costs and limits of its analytic resources. Sticking to the current tool and whatever the Bach Mai staff can make with their statistical resources at hand will at least enlarge the sample of patients with comparable evaluations. It will allow national workers like Ms Huong to produce descriptive accounts of how the persons with disabilities have progressed during their common careers with VVAF.

Progress while in the program – Some findings

Who are the persons in focus here?

VVAF began a mobile outreach program in 1999. Consequently the venue for patient evaluations was changed from home visits to interviews at provincial centers and in Hanoi, and pre-treatment conditions were elicited retrospectively. Six VVAF staff share the follow-up burden. Data began to be computerized in 2002 only when the evaluation form was revised. Its current title “Impact Indicator Form for Village Spot Checks” is a misnomer from a previous era; the interviews no longer take place in the patients’ home villages and towns.

Using part of the database information, we make comparisons between pre-treatment and latest evaluations of 336 patients. These persons are very diverse in terms of their disabilities, demographics, and patient careers including time elapsed between first and last evaluations.

The most frequent diagnosis is cerebral palsy, accounting for over a third of the sample. Other conditions present with sufficient cases to permit before/after comparisons include hemiplegia, polio and amputations.

Background information is given in the appendix. Most of the 336 patients followed the invitation to come to meetings at the provincial centers. These invitations were sent out by provincial hospitals against lists supplied by VVAF Bach Mai. A cash incentive of Dong 50,000 (approx. US-\$ 3) is offered to meet travel and meal expenses. A selection

effect, as over the sample seen in home visits, is to be assumed. Essentially the sampling characteristics are not known.

It is important to note that the information from these follow-up interviews is not the only source of feedback from the patients. In eight provinces, approx. 200 persons with disabilities have formed self-help groups. VVAF workers attend some of their meetings and thus are privy to the patients' self-assessments in this group format.

What was assessed?

Patients were assessed on 24 behavioral and social items. The authors of the assessment tool grouped the items under a number of domains, among which

- Personal hygiene (4 items), and
- Mobility (6 items)

have straightforward meaning. The other fourteen items are less persuasively arranged or are restricted to children, adults or device users. After statistical testing, I have rearranged some of them into

- Cognitive ability (4 items), and
- Social integration (4 items)

domains. These latter two almost certainly have inferior validity and reliability than the hygiene and mobility measurements, but I retain them because we expect progress in the former domains to go hand in hand with, or even be dependent on, progress in the latter two. In fact, their inter-correlations can be demonstrated for this sample too.

An example is in place. Hygiene and mobility item response was rated on five levels, ranging from “Totally dependent” to “Independent”. This table summarizes the before and after-treatment ratings for the first item in the personal hygiene segment, eating:

Personal hygiene: Eating		Latest evaluation					Total
		Totally dependent	Can cooperate	Limited independ	Almost independ	Independ-ent	
Before treatment	Totally dependent	17	5	1	2	0	25
	Can cooperate	0	22	6	3	3	34
	Limited independ	0	0	3	20	6	29
	Almost independ	0	0	0	20	15	35
	Independ-ent	0	0	0	0	216	216
Total		17	27	10	45	240	339

It is obvious from this sample table that about two thirds of the sample patients never had a problem with eating. Improvement in any itemized behavior, therefore, concerns a minority of the patients. The beauty is that averaged over all items in a domain, improvements can be detailed exactly for each patient.

For, within each domain and for each patient, the information on all items was compressed into a single score, with one estimate for the baseline (roughly the condition at time of registration) and one for the time of the latest evaluation. The four scores – hygiene, mobility, cognitive, social – were computed using a statistical procedure that assigns numbers such that a valid before/after difference can be obtained.

Other information was not used. For example, I did not look at changes in reported pain levels. Patient background other than diagnosis and career elements other than time under evaluation were not considered because factoring them in would require controlled samples.

How considerable are the improvements?

To ease readability for the VVAF staff, the estimates are expressed on a similar scale to the one they are using to rate personal hygiene and mobility items. This ranges from -4 for the worst expression, “Total dependence”, to 0 for independence. The mean scores and mean gain then present as follows:

	Baseline	Latest	Gain
Personal hygiene	-1.24	-0.97	0.27
Mobility	-1.33	-0.90	0.44
Cognitive ability	-1.33	-1.02	0.31
Social integration	-2.08	-1.45	0.64

Assuming that the evaluated patient group is a representative sample of all the patients treated in the network of VVAF-supported institutions, the gains are statistically significant in each and all domains. Since we lack information on the sample design, this line of reasoning cannot be pursued any further. Instead, we look at the inter-correlations of gains across domains, as shown in this table of correlation coefficients:

Correlation betw. individual gains	Personal hygiene	Mobility	Cognitive ability	Social integration
Personal hygiene	1.00			
Mobility	0.57	1.00		
Cognitive ability	0.45	0.36	1.00	
Social integration	0.39	0.41	0.35	1.00

Two findings seem noteworthy: The social isolation before treatment was assessed by the administrators of this impact tool as a more serious problem than the limitations in

personal hygiene, mobility or cognitive ability posed. The gain in social integration (family and community) is more significant than the gains in any other domain.

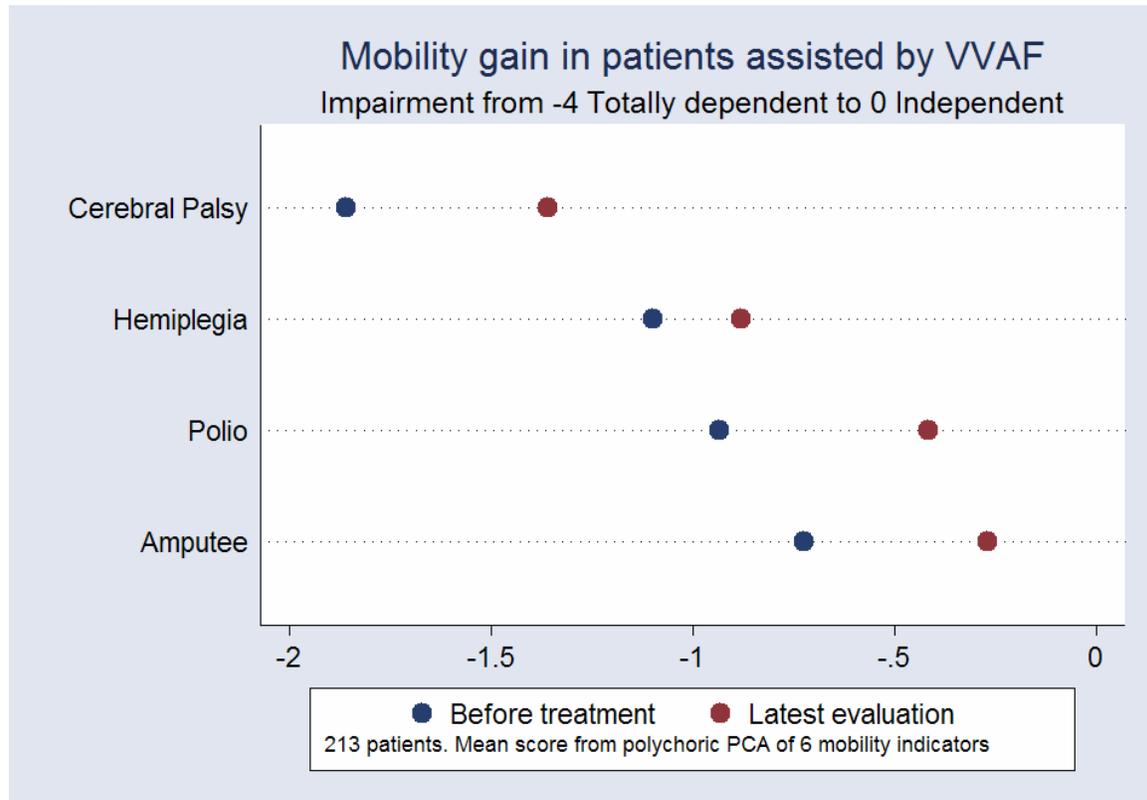
Secondly, while a strong positive correlation between dimensions of personal functioning, such as hygiene and mobility, is to be expected, the correlations between social integration gains and gains also in the physical-behavioral domains are non-trivial (italicized cells).

It is impossible to say, with this data, to what extent this is due to retrospective rationalization among assessment personnel. Alternatively, if the ratings are reliable, one is entitled to assume that functionality gains were additionally facilitated in persons whose social integration too was improved during, if not because of, exposure to the VVAF-supported programs.

Progress in mobility for different disabilities

Typically, mobility impairment differs by diagnosis. Gains are shown for 17 amputee, 120 cerebral palsy, 34 hemiplegia and 42 polio patients for whom before and after score are available.

Note similar gains for all groups except hemiplegia, although from very different baselines.



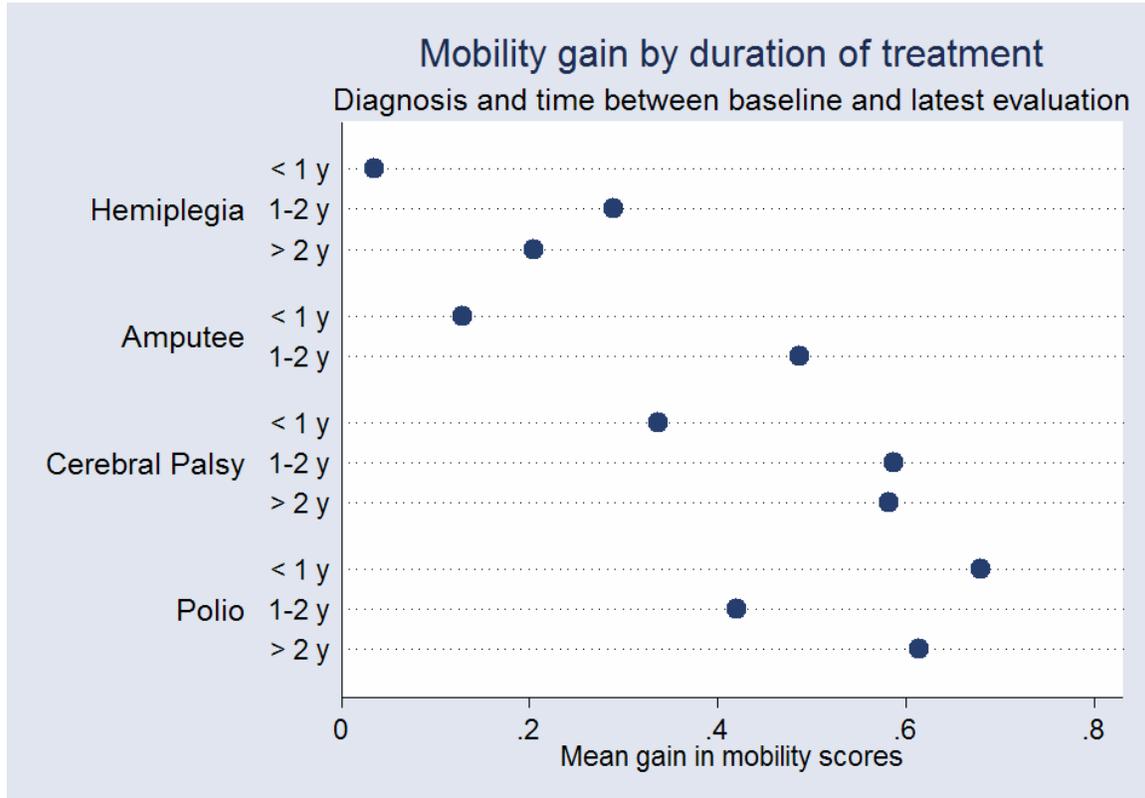
Progress over time

Because the previous example is about mobility gains, I continue with this domain.

Hemiplegia, amputee and cerebral palsy patients that have stayed in the system for more than one year show greater mobility gains than those for whom the latest evaluation took place less than a year after registration. In the small sample (42) of polio patients, this trend is reversed.

Hemiplegia and cerebral palsy patients evaluated after more than two years from baseline show no greater mobility gains than those of duration 1 – 2 years.

This is likely the consequence of selection effects, meaning that more difficult cases stay in the system for longer while rapidly improving patients are not seen again after a short time, or alternatively have not been re-evaluated lately.



[There was only one amputee with a time interval greater than two years.]

How should VVAF talk about the findings?

For this sample, the overall gains in the four domains assessed with the VVAF Bach Mai “Impact Indicator” tool are statistically significant. This confidence is based on the fact

that the statistical method condenses a host of qualitative ratings into scales that make for valid numerically expressed differences.

These gains can and should be mentioned as a general reflection on the work of the institutions supported by VVAF to the extent that

- The retrospectively elicited baseline ratings can be trusted
- The sample of patients evaluated for more than one point in time is typical of the larger caseload
- It is understood that the measurement of “cognitive ability” and “social integration” of patients is more problematic than personal hygiene and mobility measures.

A fascinating finding is the correlation between gains in social integration and gains in the other domains. VVAF communications with partner institutions and donors could suitably elaborate on this, perhaps by way of case studies that highlight the interaction of social work, family support and personal healing. Below the reader finds one of three small case studies written by a former staff member, Mr. Nguyen Van Hanh. I selected this one because the patient happens to figure among those in the follow-up table. I list the claimed improvements in the record, pointing out some of the reliability problems.

Sweeping causal attributions should be avoided. The measured gains are not proof of the effectiveness of any treatment or program – they could come about by any number and combination of unobserved circumstances with which the medico-social interventions simply happen to be in lockstep. This is unlikely, though, and the methods I have used at least demonstrate a happy progress in persons who otherwise are struggling on the hard side of life.

A small case study, written by program staff: *Phuong Thuy*

“**Nguyen Thi Phuong Thuy** was diagnosed with ‘inflammation of the tibia’ and was taken by her father to Huong Sen Rehabilitation Centre in the northern province of Tuyen Quang. She had a very bad gait and walked with great difficulty. Her left leg was thirteen centimeter shorter than the right and she suffered from low back pain as a result of gait pattern. She walked on the lateral border of her foot and had no range of motion at the ankle.

She was assessed by the Bach Mai team during an outreach conducted by the Bach Mai Rehabilitation team and The Vietnam Veterans of America Foundations' Rehabilitation Rights Program in late February, 2004. After another assessment by the Orthotist/Prosthetist it was decided to make her a polypropylene knee, ankle, foot orthosis (K.A.F.O) to control and support the deformed limb, and provide compensation for the leg length discrepancy.

This was manufactured at the workshop at Bach Mai and at the first fitting Phuong Thuy walked with great difficulty. After some practice supervised by her physiotherapist she began to use her K.A.F.O very effectively.

Phuong was really happy with the result and can now walk with a much better gait. Her mother reports that she is now able to walk most of the day. The whole family is delighted with the outcome. ***"I now no longer have to worry about the possibility of my daughter developing a curved spine and other deformities"*** said her mother ***"and things are as good as can be expected."***

Amputation of the deformed limb may have to be considered in the future according to her doctors and surgeons, and prosthesis made for her.

Thuy was born in 1996 in the little town Yen Son, Tuyen Quang. She was a normal healthy child until she was three years old and then suffered from a fever. She was taken to a district hospital by her parents but became weaker and weaker. Unfortunately, her parents were not told that she had an inflammation of the tibia and Thuy did not recover as expected even though she received extensive medication.”

This patient (ID BM3384) was evaluated twice, in February and October 2004, on the 24-item inventory. Items for which improvement was reported include:

Item	Before	After
Self-care	-1	0
Dressing	-2	0
Toileting	-2	0
Lying-to-sitting	-2	0
Sitting	-1	0
Sitting-to-standing	-1	0
Standing	-4	0
Ambulation	-4	0
Device use	-3	0
Participation in family life	-3	0
Playing/socializing in the community	-6	0
Work, play, study	-2	0



The negative numbers stand for the severity of the problem. Zero denotes the unproblematic state, e.g. full personal functioning.

On all of the other items, Thuy was rated 0 both before and after treatment. Also, the fields in her record for questions reserved for adults were filled with zeros.

A number of questions arise immediately that should be asked about the entire scheme. The changes claimed through the across-the-board upgrading to the “no problem” category in all items that had denoted a personal functioning, social or cognitive deficit before treatment seems improbable for a child that had gone through so much suffering. Thus, how much realism does this system observe?

The social integration ratings are particularly dubious. From the medical history, it appears that the parents were caring. On what evidence was it determined that, before, the girl had not participated in family life regularly, and was doing so fully after she received her orthosis?

In Thuy’s case we know a bit more; both Ms Huong and Mr. Endley know her well. Both vouch for her tremendous progress. Thuy’s parents had been very protective and were seen as restricting her activities for fear she might get hurt. It is this observation that prompted the follow-up team to give a -3 for Thuy’s earlier participation in the life of her family. This determination illustrates a problem of item validity rather than of reliable measurement.

There are reliability issues as well. In the database, Thuy is recorded as having used an assistive device “sometimes” [item #11, cat. -3], but the case study suggests she had not had any before she was seen at the Rehabilitation Center.

The use of the zero category in the items reserved for adults does not affect her gain statistics, but it does bias the domain mean if any were calculated for the patient, and the practice upwardly biases the baseline and after-treatment means in the item of the sample.

Methodological considerations

The “Impact Indicator Form”

All the preceding is based on data collected with the help of this form, which was revised in the 4th quarter 2002. Technically, the form lays out a social-behavioral rating inventory for the construction of one or several scales. It carries 24 items in ordered response categories, with typical monotonically-increasing salience or frequency. These are coded zero for the most desirable option, and with variable, albeit always equi-distanced negative weights for the less desirable ones. This is a typical code template used for all the items in the personal hygiene and mobility domains:

- 4 1-Totally dependent
- 3 2-Can cooperate
- 2 3-Limited independence
- 1 4-Almost independent
- 0 5-Independent

Checkboxes for positive responses are in two vertical stacks, one for the “Before treatment” retrospective condition, and one for the contemporary one. The left side of the response nomenclature holds a box for N/A and, for most questions, one for a supplementary question. Unforgivably, the code for N/A is the same as the one for the most desirable option –zero.

What was supposed to be done with the information collected into the form may be guessed by looking at its first page:

The form is divided into two main sections: 'POINTS FOUND AT START' and 'POINTS FOUND NOW'. At the top right, there is a box labeled 'Total points of functioning gain'. Below the main sections, there is a row of numbers 1 through 18, with the instruction 'Circle the activities number, PWD improved'. The 'Personal Hygiene' section is divided into two parts: '1 Eating' and '2 Self Care:'. Each part has a list of activities with checkboxes for 'Before' and 'NOW' and corresponding scores from 0 to -4. For example, under '1 Eating', the activities are 'independently', 'can eat alone, but uncontrolled, (bowl to mouth)', 'limited task', 'can cooperate with caregiver to eat', and 'totally depended'. There are also checkboxes for 'non acured for the case N/A' and 'special device given?'.

[Form slightly squeezed for space reasons]

The two box labels “Points found at start” and “Points found now” suggest that only one scale is to be computed, over all 24 items and as a simple addition of item scores for the time in point. The difference then is to be displayed in the “Functioning gain” box. In addition, by circling the sequential numbers of those items that represent improvements, later users returning to a filled form can quickly form a picture as to where the specific improvements took place.

As is known from the evaluation literature (see Rowe and Rowe 2004, for many), the simple unit-weight addition of indicator variables for scale formation is not statistically valid. Typically, indicators contribute differently to an underlying concept, and pre-set weights cannot capture those differences. However, this practice is wide-spread, and for the diagnostic and reporting purposes of many organizational program alternatives may not be available. This is the case particularly where assessments have to inform program decisions rapidly, such as during the very visit of a medical team that administers the assessment tool.

Even in sophisticated research environments the limitations of ordinal-level data in scale formation are neglected, and factor-analytic approaches are applied that imply that the ordinal variables have meaningful origins and units of measurement. This is the same as saying, in terms of a VVAF impact indicator item, that the difference between “totally dependent” and “can cooperate with the caregiver” is the same as the difference between

the ability to do “things alone, although uncontrolled” and complete independence. Arithmetically, $0 - -1 = -3 - -4 = 1$. Similarly, the numeric codes suggest that “Independence” has the same value across all itemized behaviors.

This criticism is purely academic because in actual fact VVAF Bach Mai has not yet had time to attempt an evaluation of its patient evaluation data, and thus scales, even if considered, have not been practically calculated.

Domains and items

The form lists the 24 items under eight different titles, which I call domains. One domain has as many as 6 items (Mobility), “Pain” has only itself as an item. In the rightmost column, anticipating on the use that I make in scale score computation, I have rearranged items in four domains.

Item #	Domain	Item object	Polychoric PCA score gain scales
1	Personal hygiene	Eating	Personal hygiene
2		Self-care	Personal hygiene
3		Dressing and undressing	Personal hygiene
4		Toileting	Personal hygiene
5	Mobility	Rolling over	Mobility
6		Lying-to-sitting	Mobility
7		Sitting	Mobility
8		Sitting-to-standing	Mobility
9		Standing	Mobility
10		Ambulatory type	Mobility
11	Rehab. knowledge	Devise use	[Not used]
12		Family involvement	Social integration
13		Knowledge	Cognitive ability
14	Pain	Pain	[Not used]
15	Social. and particip.	Family life	Social integration
16		Community immersion	Social integration
17		Community inclusion	[Not used]
18	Education and work	Education	[Not used]
19		Work	[Not used]
20		Skills to work	[Not used]
21	Self-esteem	Can work, play or study well	Cognitive ability
22		Looked down upon	Social integration
23	Communication	Speaking	Cognitive ability
24		Expressing needs	Cognitive ability

Item category distributions before and after treatment

By way of example, the distribution of patients over the five categories of the personal hygiene items is shown in this table. Note that no illegal statistical operations have been committed as far as the distributions within an item, and the comparisons before and after treatment are concerned. The statement that before treatment only 48% of the sample

patients could use the toilet independently, while now this figure has moved up to 58% is statistically correct. The italicized averages over all items in the domain are correct in the limited sense of saying that over all subjects and items used x percent of the ratings fell into y category.

Personal hygiene	Total dependency	Can cooperate	Limited independence	Almost independent	Independent
<i>Before:</i>					
Eating	7%	10%	9%	10%	64%
Self-care	9%	12%	20%	10%	49%
Dressing	13%	12%	13%	13%	49%
Toileting	12%	14%	15%	11%	48%
<i>All items</i>	<i>10%</i>	<i>12%</i>	<i>14%</i>	<i>11%</i>	<i>52%</i>
<i>After:</i>					
Eating	5%	8%	3%	16%	69%
Self-care	7%	10%	10%	15%	58%
Dressing	8%	12%	10%	14%	57%
Toileting	8%	12%	7%	16%	58%
<i>All items</i>	<i>7%</i>	<i>10%</i>	<i>7%</i>	<i>15%</i>	<i>60%</i>

However, summary gain calculations over all the items used within a domain will necessarily require calculating some kind of means, either by comparing the before / after means for a respondent over his/her item weights, and then averaging over individuals, or the other way round. These scores depend on weights that are pre-set, referencing some outside norm or standard or common-sense assumptions of program managers. Items and item categories are thus pre-calibrated by administrative decisions, and only the position of the respondent depends on empirical data. This forfeits the chance to calibrate the position of items and respondents on the intended scale jointly when the impact of an item on the overall performance in the domain should not be pre-judged.

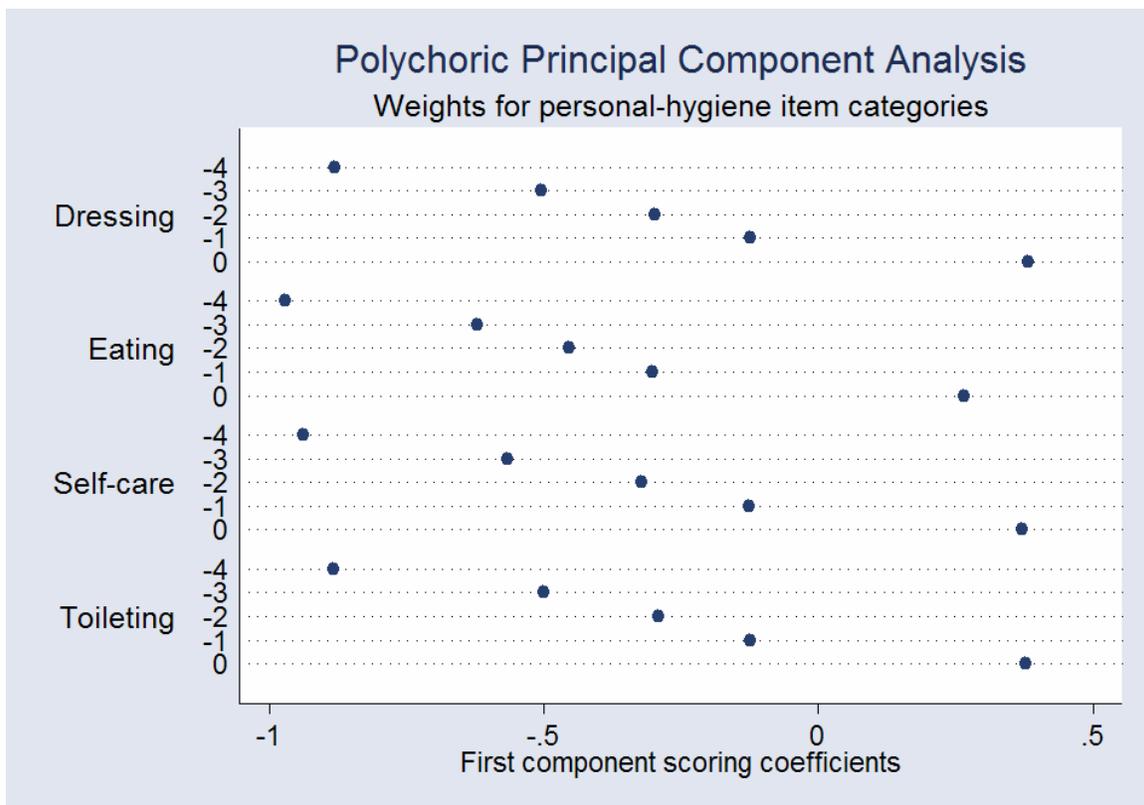
An alternative: Polychoric Principal Component Analysis

All the categories in any item used in the Impact Indicator Form are strictly ordered, from the least desired to the most desired option. For such ordinal variables, polychoric correlation (and some other equally exotic procedures such as Latent Cluster Analysis) assumes that there is a pair of underlying, unobserved, continuous and normally distributed variables with ranges to which the ordinal categories can be assigned such that their joint distribution is best predicted.

The mathematics of calculating the correlations and, from several of them, the principal components (Kolenikov and Angeles 2004, who also wrote the code for STATA that I used) is arcane and irrelevant for this note. What matters is that the procedure, out of a number of ordered-category item responses, distills a scale with interval properties, i.e. between the values on which differences have a valid and comparable meaning. Also, this academic discussion has a distantly political flavor in as much as the authors criticize two Washington DC-based institutions for applying a defective research method in numerous

developing countries - the World Bank and the Macro International-run Demographic and Health Surveys.

The ordinal-vs.-interval problem is again best explained with a graphic illustration. The following graph displays the internal “weights” that the polychoric PCA analysis of the personal hygiene ratings assigned to the categories that are used across the items. By looking at the different positions along the dotted lines of -4 “Total dependence”, -3 “Can cooperate with care-giver”, etc. up to 0 “Independence”, it is obvious that these steps are of very different length, and that the position of a category varies also across items, although not that much as the within-item differences.



The rest of my calculations were based on this procedure, repeated for each set of items in the hygiene, mobility, cognitive and social integration domains under which I gathered the 24 items. The weights were calculated on the joint table of before-treatment and after-treatment observations. This was facilitated by the fact that the first principal component explained a very large portion of the ordinal variance; the other components contributed comparatively little and were discarded.

Subsequently, the estimated scale and item values for the two point of time were lodged in two sets of variables. For ease of visualizing gains, the scores in each scale were rescaled to the [-4, 0] range used in the Impact Indicator Form, with the minimum of the pre-treatment values set to -4, and the maximum to 0.

It should be noted that the explanatory proportion of the first component used for the scales declined from personal hygiene (92%), to mobility (79%), cognitive ability (67%) and finally social integration (52%). The internal validity for the latter two is therefore hardly satisfactory.

A method like Polychoric PCA may not be available to the Bach Mai staff for a while although, if desired, it would be easy to write a set of functions in Excel to transform category weights for use in additional follow-up data. It is therefore time to return to some observations on what Ms Huong and her colleagues can do with the resources at hand.

Capacity building: Analysis by Bach Mai staff

This note demonstrates a way of summarizing changes in the functioning of patients followed-up by VVAF. The statistical method used here was chosen for its validity. It is not easy to replicate in non-statistical environments.

It is more important to enable the staff who continue the follow-up work at Bach Mai to do some of the analysis themselves even if they are limited to methods that may produce slightly less valid results.

The statistician, Ms Huong, is familiar with MS Excel and occasionally works with SPSS. Moving data from patient database tables and queries to Excel, and hence to SPSS should

be easy.

A screenshot from the help function of SPSS. It illustrates a way to solve a difficulty that kept the Bach Mai staff from making the follow-up data from the patient database accessible for analysis. The follow-up patients have undergone assessments between one and four times, and creating a table for the differences between the first and the latest ratings requires a combination of prep-work in Excel and case restructuring in SPSS.

Example of Cases to Variables

In this example, test scores are recorded twice for each subject, before and after a treatment.

	id	scor	time
1	1	1014.00	bef
2	1	864.00	aft
3	2	684.00	bef
4	2	636.00	aft

You want to do a paired samples t test. Your data structure is case groups, but you don't have the **repeated measures** for the paired variables that the procedure requires. Select **Restructure selected cases into variables** in the Restructure Data Wizard, use *id* to identify the row groups in the current data, and use *time* to create the variable group in the new file.

	id	aft	bef
1	1	864.00	1014.00
2	2	636.00	684.00

When you run the paired samples t test, you can now use *bef* and *aft* as the variable pair.

Reformatting the exported tables, given the database table architecture, however, is not straightforward. I created sample tables in Excel that demonstrate formatting options helpful for subsequent analysis steps, and two practice files to show the preparatory work

needed in Excel before the data can be processed further in SPSS. Ms Huong and I did a practice session together in her Bach Mai office.

A key difference discussed is between long-form and wide-form tables. To calculate gains on indicators (from pre-treatment to latest evaluation), long-form tables need transformation to wide-form ones. SPSS has a utility called “Restructure” to achieve the transformation, and Ms Huong will be able to use it for this kind of analysis.

Whatever data-formatting route is taken to studying the impact in the growing number of follow-up evaluations, the methods to summarize gains in particular abilities and for patients with certain diagnoses, treatments and social characteristics should remain simple. Once wide-form tables are made, Pivot tables in Excel meet most needs.

The routines that Ms Huong and her colleagues can muster in Access, Excel and SPSS to produce tables, graphs and occasionally tests (say, of differences of means) can be used to *describe* changes in the lives of the persons served by VVAF and its partner organizations.

Care should be taken to avoid claims of *causality* such as about the efficacy of treatments. At current sample sizes, and with scarce information on patient background variables, it will not be possible in the foreseeable future to isolate causal effects from the numerous selection and confounding effects that intervene from patient background, program policies, and patient careers.

Conclusion

VVAF Bach Mai collected data on the progress of a sample of patients followed up over varying lengths of time. When properly summarized, the data shows marked progress in the sample patients. Some of the correlations among the dimensions of improvement are fascinating and warrant further exploring, perhaps with other methods such as in-depth case studies.

However, as in many aid programs and NGO reporting and monitoring systems, there are a number of problems that need to be addressed in the right order. Apart from problems of cost, relevancy and timeliness, which are extraneous to the research logic, NGO self-monitoring faces these kinds of analysis challenges, frequently in this descending order of importance:

- Reliability of the information
- Data management
- Sampling and inference
- Validity of the models

Some of this may be true also of the Impact Indicator data and its intended uses. The fact that the baseline data is retrospective can mean that it is much less reliable, particularly from “reverse generosity errors”. Reverse generosity is a frequent effect when an assessor

for a service agency and the client conspire to say that his condition prior to treatment was particularly bad, in order to reflect positively on the treatment effect achieved from then to time of interview.

Ms Huong recalled that several years back the program did do baseline assessments at the time of registration, or just after issuing the device. I was shown a form that had boxes for evaluations at regular intervals starting at registration. However, this system fell by the wayside when home visits were abolished in favor of patients meeting at provincial centers, and the “real-time” assessments were replaced with retrospective interviewing for the pre-treatment condition part. From a social science viewpoint, this may have entailed poorer reliability, but may be necessary on organizational grounds.

A problem in the Bach Mai statistics may be limited data management capacity. The fact that no one had helped Ms Huong find ways to recast the impact indicator data into an analyzable form supports this view. Improving this capacity and creating some success stories from their own analysis work on existing data seems more urgent than borrowing from yet another assessment tool. There may be very good technical reasons to look to the CIR/RERC Prosthetics Users’ Survey¹ for additional behavioral items, but every change comes with a cost, including a cost in terms of comparability.

This analysis is hurried, incomplete and a mixture of didactic and analytic intentions. Its findings, however tentative, underline that VVAF is doing great work for persons in need, and documenting this worthy endeavor should not be rocket science.

References

Kolenikov, S. and G. Angeles (2004). The Use of Discrete Data in Principal Component Analysis with Applications to Socio-Economic Indices. CPC/MEASURE Working paper No. WP-04-85.

Rowe, K. J. and K. S. Rowe (2004). Developers, Users and Consumers Beware: Warnings about the design and use of psycho-behavioral rating inventories and analyses of data derived from them. International Test Users’ Conference 2004, Melbourne, 19-20 July 2004.

¹ See <http://www.cirnetwork.org/>

Statistical appendix

Descriptive statistics

Age and gender of patients

Gender	N	Median age
Follow-up sample		
Female	143	18
Male	193	22
Total	336	20
All patients in database		
Female	4,032	18
Male	5,052	17
Total	9,084	17

The main difference concerns the age of male patients. Those in the sample are considerably older than in the overall patients in the database.

Patient diagnoses

Diagnosis	Evaluated for at least two points in time	All recorded patients, as of 31 December 2004
Cerebral Palsy	35.7%	30.1%
Polio	12.5%	8.8%
Hemiplegia	10.1%	7.1%
CVA	6.5%	5.1%
Amputee	5.1%	3.3%
Club Foot	4.2%	5.9%
Post Encephalitis	3.0%	1.8%
Scoliosis	2.1%	3.1%
Spinal problem	2.1%	7.0%
Congenital deformity	1.8%	2.0%
Paraplegia	1.8%	2.1%
Spinal cord injury	1.5%	3.8%
Flatfoot	0.9%	1.3%
Osteomyelitis	0.9%	0.4%
Tetraplegia	0.6%	0.5%
Arthritis	0.3%	0.9%
TB bone / joint	0.3%	0.3%
Spina Bifida	0.0%	0.3%
Other	10.7%	16.0%
No diagnosis	0.0%	0.3%
Total	336	9,088

Time from first to last evaluation

The typical time lapsed between first and last recorded evaluation is 18 months; this interval can be as short as one month, and as long as 13 years. The detailed summary statistics is as follows:

Time first to last evaluation - years				
Percentiles		Smallest		
1%	.166626	.083374		
5%	.416626	.083374		
10%	.416748	.166626	Obs	336
25%	.916748	.166626	Sum of Wgt.	336
50%	1.5		Mean	1.831605
		Largest	Std. Dev.	1.442211
75%	2.583252	7.333374		
90%	3.416748	7.666748	Variance	2.079973
95%	4.083374	8.416626	Skewness	2.52706
99%	7.333374	13.08337	Kurtosis	15.52896

Polychloric principal component analysis output

Personal hygiene

Polychoric correlation matrix

	eating	selfcare	dress	toileting
eating	1			
selfcare	.89056545	1		
dress	.8690407	.93942353	1	
toileting	.87983181	.91150964	.9467186	1

Principal component analysis

k	Eigenvalues	Proportion explained	Cum. explained
1	3.719141	0.929785	0.929785
2	0.150618	0.037654	0.967440
3	0.087432	0.021858	0.989297
4	0.042810	0.010703	1.000000

Scoring coefficients

Variable	Coeff. 1	Coeff. 2	Coeff. 3
eating			
-4	-0.969896	-1.685278	-0.255192
-3	-0.621057	-1.079140	-0.163408
-2	-0.453614	-0.788194	-0.119352
-1	-0.302815	-0.526167	-0.079675
0	0.266102	0.462375	0.070015
selfcare			
-4	-0.936915	0.190973	1.404698
-3	-0.565718	0.115311	0.848170
-2	-0.321625	0.065557	0.482206
-1	-0.127055	0.025898	0.190492
0	0.371359	-0.075695	-0.556771

dress	-4	-0.879673	0.735846	0.023749
	-3	-0.503760	0.421395	0.013600
	-2	-0.297843	0.249145	0.008041
	-1	-0.122848	0.102762	0.003317
	0	0.380889	-0.318613	-0.010283
toileting	-4	-0.882432	0.524727	-1.129750
	-3	-0.499076	0.296770	-0.638952
	-2	-0.291861	0.173552	-0.373661
	-1	-0.124159	0.073830	-0.158957
	0	0.376781	-0.224048	0.482380

Mobility

Polychoric correlation matrix

	rollover	lie2sit	sit	sit2stnd	stand
typeambu					
rollover	1				
lie2sit	.82624841	1			
sit	.77348155	.88881868	1		
sit2stnd	.72358769	.80283156	.82921259	1	
stand	.71534713	.7813516	.80472356	.91916453	1
typeambu	.54404246	.62269324	.59461804	.7485388	.77041733
1					

Principal component analysis

k	Eigenvalues	Proportion explained	Cum. explained
1	4.797465	0.799578	0.799578
2	0.572528	0.095421	0.894999
3	0.262775	0.043796	0.938795
4	0.192566	0.032094	0.970889
5	0.095880	0.015980	0.986869
6	0.078785	0.013131	1.000000

Scoring coefficients

Variable	Coeff. 1	Coeff. 2	Coeff. 3	
rollover	-4	-0.952669	1.112553	-1.601002
	-3	-0.672113	0.784913	-1.129516
	-2	-0.473283	0.552713	-0.795373
	-1	-0.264554	0.308953	-0.444594
	0	0.220062	-0.256995	0.369824
lie2sit	-4	-0.941735	0.727255	-0.072968
	-3	-0.647645	0.500144	-0.050181
	-2	-0.419023	0.323590	-0.032467
	-1	-0.197303	0.152367	-0.015287
	0	0.279171	-0.215590	0.021631
sit	-4	-0.952789	0.668570	0.880722
	-3	-0.669789	0.469989	0.619127
	-2	-0.470653	0.330256	0.435054
	-1	-0.226552	0.158971	0.209417
	0	0.278754	-0.195601	-0.257670
sit2stnd	-4	-0.846443	-0.346709	0.705028

	-3	-0.539638	-0.221039	0.449481
	-2	-0.317316	-0.129975	0.264302
	-1	-0.061439	-0.025166	0.051175
stand	0	0.395581	0.162032	-0.329491
	-4	-0.827529	-0.474019	0.551006
	-3	-0.523686	-0.299974	0.348694
	-2	-0.315471	-0.180705	0.210054
	-1	-0.028371	-0.016252	0.018891
typeambu	0	0.444749	0.254757	-0.296133
	-4	-0.774517	-1.531184	-0.984548
	-3	-0.481467	-0.951838	-0.612030
	-2	-0.224380	-0.443589	-0.285227
	-1	0.027891	0.055139	0.035455
	0	0.393846	0.778616	0.500648

Cognitive ability

Polychoric correlation matrix

	knowledg	canwell	speaking	needexpr
knowledg	1			
canwell	.34010157	1		
speaking	.39216996	.58893781	1	
needexpr	.54668112	.59853629	.84668193	1

Principal component analysis

k	Eigenvalues	Proportion explained	Cum. explained
1	2.691882	0.672970	0.672970
2	0.703044	0.175761	0.848731
3	0.471394	0.117848	0.966580
4	0.133681	0.033420	1.000000

Scoring coefficients

Variable	Coeff. 1	Coeff. 2	Coeff. 3
knowledg			
	-4 -0.623384	-1.351514	-0.310977
	-3 -0.410743	-0.890502	-0.204901
	-2 -0.225586	-0.489077	-0.112534
	0 0.277876	0.602443	0.138619
canwell			
	-4 -0.658757	0.559760	-1.101216
	-2 -0.050008	0.042493	-0.083596
	0 0.567933	-0.482585	0.949388
speaking			
	-8 -1.222935	0.612165	1.059298
	-6 -0.856390	0.428684	0.741799
	-4 -0.639517	0.320123	0.553945
	-2 -0.414491	0.207482	0.359029
	0 0.265690	-0.132996	-0.230139
needexpr			
	-6 -1.901115	0.107508	1.163449
	-4 -1.007375	0.056967	0.616496
	-2 -0.435229	0.024612	0.266352
	0 0.337308	-0.019075	-0.206427

Social integration

Polychoric correlation matrix

	faminvlv	familylife	playsocl	discrim
faminvlv	1			
familylife	.14842751	1		
playsocl	.24959679	.7570734	1	
discrim	.11918255	.36157193	.29340032	1

Principal component analysis

k	Eigenvalues	Proportion explained	Cum. explained
1	2.064313	0.516078	0.516078
2	0.922654	0.230664	0.746742
3	0.781495	0.195374	0.942116
4	0.231537	0.057884	1.000000

Scoring coefficients

Variable		Coeff. 1	Coeff. 2	Coeff. 3
faminvlv	-4	-0.388435	-1.334365	-0.175861
	-3	-0.235046	-0.807437	-0.106415
	-2	-0.067118	-0.230565	-0.030387
	-1	0.085217	0.292740	0.038581
	0	0.277475	0.953192	0.125625
familylife	-6	-0.707160	0.251865	0.318453
	-3	0.022650	-0.008067	-0.010200
	-2	0.367603	-0.130927	-0.165542
	0	0.746311	-0.265810	-0.336084
playsocl	-6	-0.699625	0.062924	0.415112
	-3	0.011845	-0.001065	-0.007028
	-2	0.332941	-0.029945	-0.197545
	0	0.720572	-0.064808	-0.427540
discrim	-6	-0.830072	0.458886	-1.784718
	-4	-0.428556	0.236917	-0.921429
	-2	-0.143553	0.079360	-0.308650
	0	0.307733	-0.170123	0.661649

27 January 2005