A Psychometric Rasch-Type Model for the Viewing Time Assessment Paradigm
Work in Progress

S. Hammond
CDMSU
Broadmoor Hospital

M. Crowder
Department of Mathematics
University of Surrey

Introduction

Deviant sexual arousal is a key element in the assessment of sex offenders. It is known to be a major predictor in risk of re-offending and also provides essential insight into the dynamics of offender motivation and behaviour. However, assessing deviant sexual arousal is a problematic matter. The two most widely used techniques available to the clinician are self-report, either through interview or questionnaire, and penile plethysmography (PPG). The self-report strategy may provide useful information within a standardised and normative context but it is very sensitive to denial and other forms of dissembling on the part of the offender. The PPG approach is less effected by faking but it is highly complex and technical in both administration and interpretation and comes with serious resource and methodological issues (Simon and Scouten 1991)..

A new approach has been put forward by Gene Abel (Abel et al. 1998) which may offer a simple solution to the problem of deviant arousal assessment. Briefly, a series of images involving adults, pubescent and prepubescent children are provided and the offender is asked to browse through them. A surreptitious record is kept of the amount of viewing time that is spent on each image. The theory states that individuals will dwell on images that they find most arousing. The clinician is then able to identify any deviant arousal on the assumption that images considered sexually interesting may be inspected more intensely than those which the viewer does not find arousing. Abel has developed a computer-based test that now being marketed. However, there are a number of unanswered questions that remain before such a technique can be used credibly as a scientific tool in the standard assessment of sex offenders.

Firstly, there needs to be more work carried out to confirm the underlying tenet that viewing time may be used as an index of arousal. A number of studies have confirmed a *correlation* between the two variables (Brown et al. 1973; Wright and Adams 1994) but this is a long way from assuming them to be equivalent in terms of measurement. For example, it is also known that viewing time may be a function of stimuli novelty. Before a correlation can form the basis of a measurement it is necessary to construct and test a proper measurement model (Suen 1990; Barrett 1998). Secondly, the issue of whether the 'scores' derived from the assessment should be ipsative (simply proportioning the amount of time spent on each stimuli group within offender) or normative (comparing the amount of viewing time on each image and stimuli group with norms drawn from a large sample). Although statistically complex, this is an important issue because it defines limitations of use for the procedure both clinically and in a research context (Hammond and Barrett 1994). Thirdly, the kinds of stimuli that can be used needs to be fully explored. Supportive research (Harris and Quinsey 1996) used images of naked people although Abel argues that the procedure precludes the need for nudity or, indeed, any sexually explicit image. This is an important issue as one of the oft-cited drawbacks with PPG assessments is the use of stimuli of dubious legal and/or ethical status.

The first and second points raise technical issues that are all too often overlooked by clinical researchers whose focus is more towards a substantive goal than the means of achieving it. However, there is now a growing awareness in psychological measurement of the paucity of properly developed, technically defensible and appropriately evaluated assessment procedures particularly in the clinical area (Mitchell 1997; Kline 1998; Hammond 1998). In a time of greater concern for clinical governance it is vital that assessment procedures meet clear technical and formal criteria (Barrett 1998). A measurement model developed specifically for this assessment paradigm has been developed and is described here. Before detailing its mathematical derivation it will be necessary to describe in more detail the actual assessment procedure it addresses.

Respondents are presented with images displayed on a computer screen over which they have control. The experimenter remains present but unobtrusive and the participant is left to interact with the computer. Instructions are provided on the computer along the lines of *'On the computer you will find a number of images. Please acquaint yourself with them by scrolling through using the buttons on the mouse to move forward and backward. When you are satisfied that you have looked at each one press the NEXT STAGE button and you will be asked to perform a short task. Press CONTINUE to proceed. If you have any questions at this point please ask the experimenter now.'*

The program allows the images to be presented one at a time. The left mouse button allows the participant to move backwards and view a previous image the right button moves forwards and moves on to the next image. The first 3 images are trials to ascertain that the participant is confident in the use of the program. An internal timer records the time spent viewing each image.

When the participant is satisfied that they have viewed each image, all of them are presented on the screen as small thumbnail images. The participant is then asked to point and click with the mouse on each image that they found sexually interesting.

The latency or time spent viewing each image is recorded in milliseconds as a vector $\mathbf{t_j}$ (j=1..n), where n is the number of images, and a binary vector $\mathbf{u_j}$ is recorded in which 1 indicates an image identified as sexually interesting and 0 otherwise.

Deriving a measurement model for the viewing time procedure is, necessarily, a rather technical exercise but it is vital if the technique is to be rigorously applied and properly evaluated. The measurement model proposed here is an extension of a model first proposed by Van Breukelen and Roskam (1991) and (Roskam 1996) for timed ability tests. It is a Rasch-type model carrying extremely useful cumulative unidimensional scaling properties allowing the means of testing individual misfit as well as providing a statistically derived 'score'. The model will be built initially using the non-offending sample in order to provide a normative basis for sex offender assessment.

To fit the model, data is required from N participants on n images where N exceeds 200 and n (the number of images in a particular subset) is around 20. The stimuli in each subset are made up of images of one type (ie. prepubescent females, adult males). The total number of images in each subset, that are endorsed as sexually interesting, are recorded in u_{ijs} (where u_{ijs} is a binary response of interest for the i^{th} stimuli by the j^{th} person looking at subset s). By simple summation over i this provides a rudimentary score of sexual interest for that subset, r_{js} (where *r* is the total and *s* indicates the s^{th} stimuli subset and *j* indicates the j^{th} participant). In addition the time spent viewing each image in the topic subset is recorded, t_{ijs} (where *t* is the time in seconds, *i* indicates the i^{th} image). For simplicity the subscript s is dropped from the discussion immediately below as we will assume that the model applies to each stimuli subset independently.

The model assumes that each participant manifests a latent arousal (θ_j) for a particular subset which is monotonic with *r*. In addition, the attractiveness of each image is represented as a latent attractiveness parameter (δ_i) which is monotonic with the proportion of participants endorsing that image.

The idea is that the probability of image endorsement is a function of that person's arousal potential and the attractiveness of the image. This gives rise to a simple dichotomous Rasch model:-

$$p_{ij} = \frac{1}{1 + \exp^{(q_j - d_i)}} \tag{1}$$

where θ_j is The latent arousability of person j and d_i is The latent attractiveness of image i.

However, this model does not take account of the time spent viewing the image. The full model states that the probability of endorsing an image is a function of a) the arousal of the participant, b) the attractiveness of the image and c) the time spent viewing the image. This may be formally expressed as:-

$$P_{ij} = \frac{\exp^{(x_j + t_i - s_i)}}{1 + \exp^{(x_j + t_i - si_i)}} \tag{2}$$

where, P_{ij} is the probability of person j endorsing image i given viewing time t_I, ξ_j is the log of the arousal index (θ) for person j, τ_i is the log of the viewing time on image I and σ_i is the log of the 'attractiveness' index (δ) for image i.

Derivation of the Model

In statistical terms the data to be obtained from the ith respondent is of the form (u_i,t_i), where u_i is a binary vector of endorsements and t_i is a vector of viewing times. More explicitly, u_i and t_i are formed of sub-vectors corresponding to different sets of images. Thus, for four image sets, $u_i=(u_{i1},u_{i2},u_{i3},u_{i4})$, where, e.g. $u_{i1}=(1,0,0,1,1)$ would indicate, for the five images in image set 1, that there was a self-reported arousal from images 1, 4 and 5, but not from images 2 and 3. Correspondingly, $t_i=(t_{i1},t_{i2},t_{i3},t_{i4},)$ where, eg. $t_{i1}=(5.7,1.4,2.0,6.9,4.6)$ would indicate the numbers of seconds spent viewing the five images in set 1. Although somewhat cumbersome, it will be necessary to reintroduce the third subscript into the notation: thus, u_{isj} will represent the binary u-response from respondent i to image j of set s, and t_{isj} likewise for the viewing time. The ranges of the subscripts are as follows: i=1,…,N for N subjects; s=1,…,k for k image sets; j=1,…,n for a set containing n images.

A full likelihood function for the data from N respondents may be constructed as:-

$$L = \prod_{i=1}^{N} p(u_i \mid t_i)\, p(t_i) \qquad 3$$

Where $p(u_I \mid t_i)$ is the conditional probability for u_i given t_i, and $p(t_i)$ is the marginal probability for t_i. The first factor will be expressed as:

$$p(u_i \mid t_i) = \prod_{s=1}^{k} p(u_{is} \mid t_{is}) \qquad 4$$

Assumptions implied by this likelihood function are that the u-responses from the *k* image sets are conditionally independent, and that viewing times from the image sets are unconditionally independent. The first assumption is simply that there is no residual dependence between the u_{is} after allowing for viewing times, a much more elaborate model would be required were we not to make this assumption. However, it would appear that such an assumption is central to Abel's paradigm which states that viewing time is a direct *measurement* parameter of arousal. The second assumption will be supported by using distinct individual parameters in the model for the t_{is}, and tests will be performed for the structure assumed.

A Rasch-type model will be adopted for $p(u_{is}|t_{is})$. Let

$$p_{isj} = p\left(u_{isj} = 1 \mid t_{is}\right) = \left[1 + \exp^{\left(-x_{is} - t_s - b_s t_{isj}\right)}\right]^{-1} \qquad 5$$

Where ξ_{is} represents the latent arousal of respondent i on image set s, τ_s represents the latent attractivness of image set s, and $\beta_s > 0$ governs the strength of the effect of viewing time on the self-report outcome for the image set s. This, then, is equivalent to the model expressed in 2 above. According to this model p_{isj} increases with each of ξ_{is}, τ_s and t_{isj}. Also, given t_{isj}, u_{isj} is taken to be conditionally independent of the other elements of t_{is}. A further assumption to be made is that, given the viewing times t_{is}, the components of u_{is} are conditionally independent:

$$p(u_{is} \mid t_{is}) = \prod_{j=1}^{n} p(u_{isj} \mid t_{is}) = \prod_{j=1}^{n} p_{isj}^{\,u_{isj}} \left(1 - p_{isj}\right)^{1 - u_{isj}}$$

$$= \prod_{j=1}^{n} \left[\frac{\exp^{(1-u_{isj})(-\xi_{is} - \tau_s - \beta_s t_{isj})}}{1 + \exp^{(-\xi_{is} - \tau_s - \beta_s t_{isj})}} \right] \qquad 6$$

where p_{isj} is given by (5). Expression (6) gives the first factor appearing in the full likelihood function (3) in terms of parameters $\tau=(\tau_1,...,\tau_k)$, $\beta=(\beta_1,...,\beta_k)$ and $\xi=(\xi_1,...,\xi_N)$, where $\xi_i=(\xi_{i1},...,\xi_{ik})$: t and b are population parameters, common to all individuals, and ξ_i are individual-specific parameters.

Parameter Estimation

The conditional likelihood

$$\prod_{i=1}^{N} p(u_i \mid t_i)$$

can be used on its own to obtain estimates for the parameters ξ, τ and β as follows. Let Σ_x denote summation over all binary vectors u_{is} such that

$$\sum_{j}^{n} u_{isj} = r_{is}$$

then

$$p(r_{is} \mid t_i, \xi_i) = \sum_{x} p(u_{is} \mid t_i, \xi_i)$$

Hence, dividing (6) by (7)

$$p\left(\boldsymbol{u}_{is} \mid r_{is}, \boldsymbol{t}_i, \xi_i\right) = \frac{p\left(\boldsymbol{u}_{is} \mid \boldsymbol{t}_i, \xi_i\right)}{p\left(r_{is} \mid \boldsymbol{t}_i, \xi_i\right)}$$

$$= \frac{\prod_{j=1}^{n} \exp^{(1-u_{is})(-\xi_{is} - \tau_s - \beta_s t_{isj})}}{\sum_{\xi} \prod_{j=1}^{n} \exp^{(1-u_{is})(-\xi_{is} - \tau_s - \beta_s t_{isj})}}$$

$$= \frac{\exp^{r_{is}\tau_s + \beta_s \sum_{j=1}^{n} u_{isj} t_{isj}}}{\sum_{\xi} \exp^{r_{is}\tau_s + \beta_s \sum_{j=1}^{n} u_{isj} t_{isj}}}$$

The ratio (8) depends upon parameters τ and β, but not on ξ_i, so a conditional likelihood function can be formulated as

$$L(\tau, \beta) = \prod_{i=1}^{N} \prod_{s=1}^{k} p\left(\boldsymbol{u}_{is} \mid r_{is}, \boldsymbol{t}_i\right)$$

and maximised to obtain estimates of τ and β. These may then be used to generate estimates for the individual components of ξ by equating r_{is} to its estimated expectation:

$$r_{is} = E\left(r_{is}\right) = \sum_{\xi} p_{isj}$$

The second factor in the full likelihood (3), $p(t_i)$, will be modelled as:

$$p\left(\boldsymbol{t}_i\right) = \prod_{s=1}^{k} p\left(\boldsymbol{t}_{is}\right)$$

so the viewing time responses from different image sets are assumed to be independent. For parametric inference, the times within image set s will be taken to follow a specified joint distribution. A primary candidate for this is the multivariate log-normal with means μ_s, variances σ_{s2}, and common correlation ρ_s, though other joint distributions will also be assessed.

In an alternative approach, the ξ_{is} may be treated more formally as random effects in a more structured model. A joint distribution will then need to be specified for the components of ξ_i, e.g. a multivariate normal with $E(\xi_i)=0$ and

$$\mathbf{var}\left(\xi_i\right) = \sigma^2 \begin{pmatrix} R & S & S & S \\ S & R & S & S \\ S & S & R & S \\ S & S & S & R \end{pmatrix}$$

Where R=var(ξ_{is}) has diagonal entries 1 and off-diagonal entries ρ_1, and S=cov(ξ_{is},ξ_{it}) has entries all equal to ρ_2. This will provide a framework for frequentist likelihood inference or a Bayesian approach. In the latter case, valid probability predictions can be made about the individual parameters ξ_i.