DOCTOR OF PHILOSOPHY

A comparison of functional assessment methods

Toogood, Alexander

Award date:
1996

Awarding institution:
University of Wales, Bangor

A COMPARISON OF FUNCTIONAL ASSESSMENT METHODS

A COMPARISON OF FUNCTIONAL ASSESSMENT METHODS

P.h.d

ALEXANDER TOOGOOD

UNIVERSITY OF WALES

BANGOR

1996

SUMMARY

Six functional assessments were applied to 121 topographies shown by twenty people with severe learning disabilities. Assessment methods were drawn to include semi-structured clinical interview, two rating scales, staff-use of ABC charts, descriptive and experimental analysis.

Assessment methods were compared on rate of prediction, and the rate at which multi-function complexes were assigned. The informant-based and descriptive methods assigned function for approximately two-thirds of topographies, the ABC chart and experimental methods for approximately one-third. Interview assigned multi-function complexes most frequently. Function was assigned for 98% of assessed topographies although no method exceeded 74%. Informant-based methods assigned function more often for outer-directed behaviour and the descriptive method for the remainder. The analogue method failed to assign for people with greater verbal ability. Context impinged on all methods, but in different ways. Staffs' level of training appeared to influence the selection and observation of topographies; sampling was difficult in the natural setting; there were problems interpreting both descriptive and experimental data. Differential rates of prediction adversely affected the overall rate of convergence between methods. Assessment methods averaged agreement on approximately two topographies in every three assessed, mostly on primary predictions. When non-predictions were included the overall rate of convergence averaged just one-in-four, and agreement on

the functions assigned to each topography was equivalent to one-in-three.

Assessment methods varied across topographies and individuals, and three showed a proclivity for ascribing particular functions. The nature of the setting affected all assessment methods, and each presented methodological problems from topography identification to data collection and analysis.

The search for a universal assessment method appears incomplete, and may even be inappropriate. Results support the routine application of complementary assessment combinations suited to prevailing contextual variables. A pre-assessment of person-related variables, environmental context, topography and hypothesised function, may aid the selection of optimal combinations. More work is required to develop non-empirical assessment methods and extend the present taxonomy of function categories. Staff training and a clear clinical definition of challenging behaviour might improve the clinical utility of informant-based and indirect descriptive assessment methods.

A COMPARISON OF FUNCTIONAL ASSESSMENT METHODS

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

In completing this work I owe a great deal to a great many people. Thanks must go in the first place to the Mental Health Foundation for providing the necessary financial, practical and moral support to see the project through. In this respect I am especially grateful to Justin Russell and Elizabeth Gale. I am grateful also to the Clwydian Community Care NHS Trust for allowing me the time necessary to complete the analysis; in this respect I am especially grateful to Ged Fitzpatrick, Caroline Eayers, Andy Bell and the rest of my team for managing so well without me for a while. Special thanks are due, of course, to those who took part in the study for their robust tolerance of our intrusion into their lives over an extended period of time. For them my hope is that this work will somehow result in real practical benefits when at last they move to their new homes. To the staff and managers of the Hospital I also extend my thanks, for without their help and co-operation none of this would have been possible. Special thanks are due to my supervising committee; Professor Mark Williams (chair), Dr John Everatt (statistician), but especially Dr Robert Jones (task-master general) for keeping me on track with my schedule and my thinking. For their continued and highly skilled assistance in carrying out assessment, data collection and analysis, I must thank Kevin Timlin and Hilary Jaques, but especially Kevin for all his hard work and dedication. I am indebted to Dr Eric Emerson, for his insightful comments and support, and to Professor David Felce who along with Eric Emerson offered a great deal of help over

the initial stages via the Mental Health Foundation's Advisory Group. Thanks also to Dr David Reeves for sharing so many revised and up-dated versions of his computer software. To all my friends, colleagues and members of my family who have believed in me and supported my efforts, I say thank you, but none more than my wife Gill and my daughters Paula, Kate and Sophie who without a doubt have carried many additional burdens during this period of our lives together.

I should like to dedicate this work to my parents John William and June Rita Toogood, and to my younger brother Jonathan who despite serious disability managed over the same period to complete a BA in psychology, sociology and philosophy and gain a place to study for a Masters degree.

Chapter One
INTRODUCTION

I. INTRODUCTION

Challenging behaviour affects the lives of many people, often in a very profound way. The phenomenon has to be better understood if those affected are to be assisted to help themselves to a better quality of life. The work described in this thesis was undertaken with this aim in mind.

Chapter one reviews the origin, nature, definition and prevalence of challenging behaviour together with the social and economic grounds for amelioration. The clinical relevance of functional assessment is assessed in this respect through a review of recent developments in service delivery and published research. Functional assessment methods are described and classified, and the comparison literature is reviewed for evidence of procedural strengths and weaknesses and convergence in the ascribing of functions to behaviour.

II. CHALLENGING BEHAVIOUR

Origin and definition of the term

The term 'challenging behaviour' originated with the Association for Persons with Severe Handicaps (TASH) and became popular in the UK following publication of *Facing the Challenge* (Blunden and Allen, 1987) in which a definition previously proffered by Emerson et al. (1987) was adapted to read;

> 'challenging behaviour refers to behaviour of such an intensity, frequency or duration that the physical safety of the person or others is likely to be placed in serious jeopardy, or behaviour which is likely to seriously limit or delay access to and use of ordinary community facilities.'
>
> Blunden and Allen (1987)

Thus, in its original and most widely adopted form challenging behaviour was defined solely in terms of its personal and social consequences. No account was taken of a behaviour's topography, aetiology, or factors influencing its maintenance. Indeed it was never intended that the term should be defined in this way, the intention was simply to promote awareness of what services needed to do in response to the needs of people whose behaviour could be described as challenging (Blunden and Allen, 1987). Unfortunately, the relativist and rather imprecise nature of the construct (Toogood, 1993; Qureshi, 1993a) has resulted in it being applied in a broad and over-inclusive fashion (Jones and Eayrs, 1993; Qureshi, 1993b). According to Qureshi, (1993a) there are at least three types of definition in current use. First, there is the abstract or conceptual definition as discussed above. Second, there are 'rule of thumb' definitions used by carers and staff on a day to day basis. Third, there are operational definitions for use in research, evaluation and service planning. There is, however, no standardised definition for the latter purpose, and researchers wishing to study the phenomenon have normally to devise an operational definition and then take care to identify and describe the populations they study. This adds considerably to the difficulties of comparing across studies. Moreover, many of the relevant research papers avoid referring to the term at all preferring instead to label the behaviour under investigation according to features such as topography (Lerman, Iwata, Smith, and Vollmer, 1994), presumed aetiology (Pace, Ivancic and Jefferson, 1994) or function (Zarcone, Iwata, Smith, Mazaleski and Lerman, 1994).

Common forms of challenging behaviour

The most common forms of challenging behaviour are reported to include physical aggression, self-injury and destructiveness (Durrand and Crimmins, 1991; Felce and Lowe, 1993; Jones, 1991) together with disruptiveness and other forms of socially inappropriate behaviour (Kiernan and Qureshi, 1993; Qureshi and Alborz, 1992). While self-injury and stereotyped responding have been studied widely (e.g. Iwata, Pace, Dorsey, Zarcone, et al. 1994), it appears from the pattern of referrals to specialist support services that outer-directed aggressive behaviours are more likely to be identified as challenging by service workers (Forrest, Cambridge, Emerson, Mansell, et al. 1995: Lowe and Felce, 1995a; Toogood, Bell, Jacques, Lewis, Sinclair and Wright, 1994). However, many people display more than one challenging behaviour (Murphy et al. 1993: Russell and Harris, 1993), Kiernan and Qureshi (1993), for instance, found that 26.3% of those who showed seriously destructive behaviours also showed serious aggression, and that 25.4% also seriously self-injured. These proportions doubled when lesser challenging behaviours were taken into account.

Prevalence of challenging behaviour

Estimates of the prevalence of challenging behaviour vary between studies according to the definition used, the populations studied, the methods applied, and factors such as settings, age range and the behaviours' form. Qureshi and Alborz, (1992), for example, reported the average prevalence for all challenging behaviours across one health region was 13% for community settings and 30% for institutions. Russell and Harris (1993) reported the prevalence of aggressive behaviour

in a different region ranged from 9.7% in day settings to 38.2% in hospitals (mean = 17.6%) and Murphy et al. (1993) that the prevalence of self-injury among a hospitalised population fell between 5% and 40% depending on age. It is generally believed, however, that 15-30% of people with learning disabilities will display some form of challenging behaviour (Kiernan and Qureshi, 1993; Qureshi, 1994; Qureshi and Alborz, 1992). In summarising recent trends in epidemiological research Jones and Eayrs (1993) commented that challenging behaviour appeared to be more common in institutions than in the community and more prevalent in residential than day services. It was more usual for individuals to display more than one challenging behaviour and the majority were seen in people with severe or profound disabilities aged between 15 and 30 years.

Characteristics of individuals who display challenging behaviour

While none of the behaviours identified as challenging are found solely in people who have learning disabilities there is evidence to suggest certain forms of challenging behaviour may be associated with specific conditions. Particular forms of self-injurious behaviour are, for example, often associated with Lesch-Nyhan and Prada-Willi syndrome, stereotyped hand wringing with Rett's syndrome and aggressive behaviours may be present in persons who have certain kinds of epilepsy. In the vast majority of cases, however, the initial causes of a person's learning disability remain unknown as do the precise reasons for the onset of their challenging behaviours. Nevertheless, it does appear that certain individual characteristics are associated with an increased probability of

challenging behaviour. A comprehensive study undertaken by Murphy, Oliver, Corbett, Crayton, Hales, Head and Hall (1993) reported the characteristics of 596 persons who had been identified as engaging in self-injurious behaviour and a sub-group of 54 people whose self-injury was so severe that their management required the use of artificial protective devices. A substantial proportion of the latter group also showed one or more of fourteen other classes of inappropriate behaviour including aggression (40%), destructiveness (36%), vocalisation (66%) and stereotypy (89%). In this study serious self-injury was characterised by the early onset and enduring presence of multiple topographies which were, on average, more common in people with severe or profound learning disabilities, additional physical or sensory impairments, severe communication difficulties and epilepsy. These findings are consistent with other reports for self-injury and challenging behaviours such as aggression (Johnson and Day 1992; Oliver et al. 1987; Russell and Harris, 1993). In the study by Murhpy et al. (1993) self-injury was generally more common among younger persons aged between ten and twenty years, while Russell and Harris (1993) reported that aggressive behaviours were more likely to be labelled challenging when shown by persons aged 15 to 29 years, a finding supported by Qureshi (1993a) for all challenging behaviours.

The personal consequences of challenging behaviour

The personal and social consequences of challenging behaviour are well documented and include injury, loss, exclusion or rejection, physical abuse, excessive medication, physical or mechanical restraint, deprivation and neglect (Emerson, 1990;

Emerson, Felce et al. 1994; Emerson, Toogood, et al. 1987). Challenging behaviour may give rise to dysfunctional stress among families and carers (Pahl and Quine, 1985; Quine and Pahl, 1985; Qureshi, 1993b), and appears among the reasons cited most often in the breakdown of community placements (Intagliata and Willer, 1982).

The economic costs of challenging behaviour

Although the full economic costs of challenging behaviour are difficult to ascertain (Knapp and Mansell, 1994), the cost of residential placements (Toogood et al. 1988) and service infra-structure (Emerson, 1995) may be inflated where challenging behaviours are present. One American estimate put the additional cost of managing self-injurious behaviours alone at 1 billion US dollars per year (Oliver, 1993).

Challenging behaviour and service development

Challenging behaviour may impede the government's policy objective of providing comprehensive community care. Between 1980 and 1993 the capacity of mental handicap hospitals in the UK reduced by some 26,000 places (Emerson and Hatton, 1994). In Wales the number of people receiving support in their own homes increased from 41 in 1982/83 to 2,110 in 1988/89 (DoH, 1989). Over the same period supported housing increased from 166 to 696 places, short-term care places from 34 to 549, and community-based day care placements from 37 to 1,107. According to current plans a further 10,000 people from England and 1,555 from Wales will be resettled into their community by the turn of this century. The balance of research in this area suggests the majority people have benefited from living in the

community (Bruininks and Lakin, 1985; Emerson and Hatton, 1995; Janicki, Wyngaarden-Krauss and Seltzer, 1988; Heal, Haney and Novak-Amando, 1988; Lakin and Bruininks, 1985). Experience suggests, however, that a significant minority have already been excluded by virtue of their challenging behaviour, and that this trend may continue. As up to 30% of the institutionalised population may display some form of challenging behaviour (Qureshi and Alborz, 1992) progress on the replacement of institutional services both here and abroad could depend ultimately upon how well the new pattern of services responds to challenging behaviour (Blunden and Allen, 1987) and whether the proposed alternatives are affordable.

Conclusion

Challenging behaviour is a socially defined construct which lacks a precise definition. There is compelling evidence, however, to suggest that the lives of many people are adversely affected by the occurrence of challenging behaviour, and that it is the personal, social and economic consequences of challenging behaviours which unite them under a single banner. The most common forms of challenging behaviour include aggression, self-injury, destructiveness, disruption and other socially inappropriate behaviours. Challenging behaviours are generally more common in institutions than in communities and are found most often in people aged between 15 and 35 years who have severe or profound learning disabilities and additional physical, sensory, or communication difficulties. Deinstitutionalisation of residential provision for people with learning disabilities has highlighted the need for effective

assessment and treatment of challenging behaviour.

III. RESEARCH AND SERVICE RESPONSES

Although a socially-defined construct (Jones and Eayrs, 1993; Mansell, 1994; Qureshi, 1993a) assessing and designing interventions for challenging behaviour is undertaken principally by specialist clinicians working in the field. The social and economic imperative has been to ameliorate challenging behaviour and its negative consequences. This has often been sought through the application of pharmacologically based treatments, despite their negative side effects and questionable utility (Clarke, 1993; Kiernan and Qureshi, 1993). Behaviourally based approaches have been much less common in the UK than in the USA, for example (Oliver et al. 1987; Remington, 1993). In one UK study Oliver et al. (1987) found that of 596 people identified as engaging self-injurious behaviour only 2% had any form of written psychological treatment programme (see also: Murphy et al. 1993).

Reasons for the poor uptake of behaviourally based treatments are not entirely clear although the absence of a valid and reliable methodology has been cited, along with the failure of behaviour theory to root in the culture of British psychological research (Oliver, 1993). Recent histories of both the research and service communities tend to support this view.

Recent developments in service provision

During the 1970's a growing interest in civil liberties and human rights issues (Whitehead, 1992) combined with a series of scandals in long-stay hospitals (Martin, 1984) to give impetus to a policy of large-scale deinstitutionalisation planned by the government (DoH, 1989). The 1980's saw the Launch of the *Ordinary Life Series* (King's Fund, 1981), and together with the work of various campaign groups (e.g. CMHERA), these initiatives added momentum to a developing ethos in services which was later to embrace normalisation theory (Wolfensberger, 1991; Wolfensberger and Glenn, 1975; Wolfensberger and Thomas, 1983), but reject explicitly behavioural approaches, especially those which utilised aversive stimuli (Kiernan, 1991). During the latter half the decade ideology drove the development of community-based services. In parts of the UK ordinary housing schemes (Mansell et al. 1987) and supported employment projects (Allen, 1994) developed rapidly. In many cases these new services lacked a methodology with which to structure support for community living. Thus, people whose behaviour was challenging were either excluded, or as was often the case, their community placements failed to maintain over time. Specialist services began to emerge (e.g. Emerson et al. 1987; Toogood et al. 1994) and this new specialism needed a framework with which to understand and explain challenging behaviour and structure their work. Most adopted a behavioural approach (Forrest, 1995), which by the beginning of the 1990's had itself undergone considerable change.

Recent developments in applied behaviour analysis research

During the 1970's many published papers examined the use of positive reinforcement techniques for teaching self-help (Azrin et al. 1971; Horner and Keilitz, 1975) and vocational skills (Bellamy et al. 1979; Gold, 1975) to people with learning disabilities. The efficacy of positive reinforcement techniques was also investigated for securing the maintenance of socially appropriate behaviours (Whitman et al. 1970), increasing levels of engagement (Porterfield et al. 1980; Porterfield and Blunden, 1978) and improving carer behaviours (Panyan, Boozer and Morris, 1970; Quilitch, 1975). The suppressive power of punishers was amply demonstrated in successive reports on the reductive effects obtained by procedures such as time-out from reinforcement (Bostow and Bailey, 1969; Foxx and Shapiro, 1978; Solnick et al. 1977), overcorrection (Epstien et al. 1974; Foxx and Azrin, 1973), and the contingent application of electric shock (Corte et al. 1971; Lovaas and Simmons, 1969). Punishment procedures were evaluated using other aversive stimuli such as water mist, (Dorsey, Iwata et al. 1980), lemon juice, (Sajwaj et al. 1974), and aromatic ammonia (Tanner and Zeiler, 1975). INtervention methods were compared for treatment efficacy; validity was demonstrated by outcome, and reliability through successive replication. Little regard was paid at this time to an analysis of a behaviour's function and most procedures were found to be effective on some occasions but not on others.

The social validity of punishment-based procedures, and their acceptance in clinical settings, was relatively short-lived.

Optimism ebbed as treatment results failed to deliver the speedy and consistent solutions people wanted, to maintain over time or generalise. According to separate accounts by Oliver (1993) and Oliver and Head (1993) dissatisfaction with technique-based approaches grew steadily along with the desire for a return to the scientific roots of applied behaviour analysis. Constructional approaches (Goldiamond, 1974) and the movement against the use of aversive procedures (Guess et al. 1987) gained in popularity and strength throughout the 1980's, particularly in America. These events combined to create the conditions under which change was inevitable and the work of Carr (1977) was seminal in reinstating the importance of understanding behaviour in terms of its function rather than its form. The conceptual shift signalled by Carr (1977) was later operationalised by Iwata et al. (1982) who described an experimental methodology for studying the operant functions of self-injury. Replication and refinement of this methodology and its application to other forms of challenging behaviour has accounted for a significant volume of the research conducted in applied behaviour analysis over the last decade (Iwata et al. 1994).

Conclusion

The centrality of functional assessment in applied behaviour analysis was reaffirmed at a time when clinicians working in the field had begun to address the growing need to work with challenging behaviour in a changing service context. Functional assessment represented one approach with potential for addressing an current issue of considerable clinical and social importance.

IV. FUNCTIONAL ASSESSMENT OF CHALLENGING BEHAVIOUR

Current perspectives in behavioural psychology owe much to the work of Skinner (1953; 1957; 1971) and his delineation of the operant model of conditioning. According to operant theory behaviour is maintained by its consequences. A stimulus which follows behaviour, and has the effect of making it more likely to recur is said to have acquired reinforcing properties in relation to that behaviour. A stimulus may be presented (positive reinforcement) or withdrawn (negative reinforcement). So long as the effect is to increase the probability of a behaviour's recurrence, reinforcement has taken place (figure 1.1).

Figure 1.1
Behaviour - consequence relationship

observed effect	consequence of response	
	stimulus presented	stimulus removed
behaviour increases	positive reinforcer	negative reinforcer
behaviour decreases	positive punisher	negative punisher

after Remington (1991)

The term punisher is used, in a technical sense, to describe a stimulus which when presented or withdrawn contingent on the occurrence of behaviour has the effect of decreasing the probability of its recurrence. Thus, behaviour is maintained (or not) by the consequences which follow (B-C), and the consequences are themselves functionally defined in terms of the effect they are observed to have on the probability of the behaviour's occurrence.

Other stimuli may come to acquire a discriminative function by signalling reinforcement availability. As these stimuli precede the behaviour of interest they are termed antecedents (A-B). Antecedent stimuli are also functionally defined - according to the way they relate to the behaviour of interest, and the consequences which follow. Known as the A-B-C of behaviour, this paradigm has underpinned all attempts to understand challenging behaviour within an operant framework. A believable demonstration of the relationship between behaviour, its antecedents and consequences, is an analysis of behavioural function (Baer, Wolf and Risley, 1968).

Functional assessment of challenging behaviours

Skinner's ambition for the science of human behaviour was that greater understanding should lead mankind to be able to explain, predict and ultimately control human behaviour (Skinner, 1953). Thus far, operant theory has offered plausible explanations for both the occurrence and non-occurrence of challenging behaviours shown by people who have learning disabilities. Carr (1977) and Baumeister (1989), for example, each describe three underlying behavioural processes that could be responsible for the maintenance of challenging behaviour;

- extrinsic socially mediated positive reinforcement, the contingent presentation of a stimulus,
- extrinsic socially mediated negative reinforcement the contingent removal of a stimulus,
- intrinsic non-socially mediated (automatic) reinforcement, e.g. sensory or perceptual consequences.

Extrinsic positive reinforcement may include the contingent provision of social attention (Carr and McDowell, 1980) or access to tangibles such as food or activity materials (Durrand and Crimmins, 1988). Socially mediated negative reinforcement may include escape from task-related demands (Iwata, Pace et al. 1994) or non-task related social contact (Carr, 1994). Non-social sources of reinforcement, such as sensory or perceptual stimulation, are termed automatic (Skinner, 1969), since the reinforcement process (positive versus negative) cannot be reliably determined by direct observation alone (Iwata et al. 1990; Vollmer, 1994). A current taxonomy of reinforcement categories is shown in figure 1.2.

Figure 1.2
Classes of reinforcement used in functional assessment

Process	Source	Stimulus	Function category
Extrinsic (social) positive reinforcement	attention or contact from others	presented	social attention
	tangible (food, drink, activity)		tangible
Extrinsic (social) negative reinforcement	attention or contact from others	removed	social avoidance
	task demand		task avoidance
Intrinsic (non-social) reinforcement	sensory / perceptual positive	presented/ removed	automatic +ve / -ve
	sensory / perceptual negative		

The purpose of functional assessment is to determine which of these underlying behavioural processes exerts most influence on the occurrence of a specific response. The identification of behavioural function has been shown to improve the prospect of selecting successful intervention strategies (Carr and Durand, 1985; Durand, 1990; Repp, Felce and Barton, 1988) and there is good evidence from within the behavioural literature to demonstrate that challenging behaviour can be maintained by operant contingencies. Iwata, et al. (1982), for example,

examined the self-injury of nine children under four different environmental conditions. They found rates of self-injury were highest under conditions of high academic demand for two children (negative reinforcement), social disapproval for one (positive reinforcement), and when alone for three (automatic reinforcement). Thus, in two-thirds of cases a clear operant function was identified. A retrospective review of a further 152 cases in which this methodology was applied (Iwata, Pace, Dorsey, et al. 1994) revealed patterns of self-injury during assessment which were consistent with negative (38%), positive (26%) and automatic (21%) reinforcement hypotheses. In only 10% of cases was the behaviour's function unclear. Derby et al. (1992) used similar but briefer experimental methods to assess a variety of challenging behaviours shown by 79 persons with learning disabilities. Their results identified a negative reinforcement processes in 29% of cases, positive reinforcement in 22%, and automatic reinforcement in 15%. Further evidence for the operant control of challenging behaviour may be found in studies which manipulate the hypothesised maintaining variable. A study by Carr, Newsom and Binkoff (1976) examined the self-injurious behaviour of an eight-year old boy by systematically comparing the behaviour's rate under a variety of different environmental conditions. Self-injury was found to occur more frequently when requests were made; and when the environmental context of manding was altered, by interspersing requests with story telling, the rate of self-injury fell dramatically. In a now classic paper by Carr and Durrand (1985) an experimental design was used to identify the potential controlling variables for a variety of

challenging behaviours shown by four children with learning disabilities. In two cases patterns of responding were consistent with a negative reinforcement hypothesis, in a third positive reinforcement was indicated and in the fourth both processes were suggested. The children were taught functionally relevant and irrelevant verbal behaviours as indicated by functional assessment. In all cases the occurrence of challenging behaviours was significantly reduced when the functionally relevant response elicited the consequence identified in assessment. Operant control has been demonstrated for a variety challenging behaviours including aggression (Mace et al. 1986), self-injury (Iwata et al. 1982), and stereotypy (Crawford et al. 1992), and a variety of reinforcement classes have been considered including demand-avoidance (Iwata et al. 1982), social avoidance (Oliver, 1991), social attention (Carr and McDowell, 1988), tangible reinforcement (Durrand and Crimmins, 1988) and automatic reinforcement (Crawford et al. 1992). Some studies have considered and identified variables in a more idiosyncratic context (Day et al. 1988; Iwata et al. 1990). A number of purely descriptive studies have also provided evidence for the operant control of challenging behaviour. Eldeson, Taubman and Lovaas (1983), for example, observed the occurrence of self-injury shown by twenty people with learning disabilities and various staff behaviours such as demands, denials and praise which were directed towards them. Approximately five hours of observations were made for each person in the natural setting. The self-injury of nineteen persons (95%) was observed to escalate following the

presentation of demands, denials or punishments, suggesting these behaviours may have been negatively reinforced by escape. A more recent study by Hall and Oliver (1992) examined the relationship between the occurrence of self-injury and contact from others. These researchers found evidence to suggest that contact from others was governed by long bursts of self-injury, thus demonstrating the presence of an operant effect not only on the self-injurious behaviours of the individual with learning disabilities, but also the behaviour of others. Emerson et al. (1995) performed lag sequential analyses on naturalistic observational data collected on a range of challenging behaviours shown by six people with severe learning disabilities. A total of 34 discrete topographies were examined and twenty-eight (82%) were found to have occurred under conditions that were judged consistent with either a positive or negative reinforcement hypothesis.

Several of the studies reviewed above were apparently unable to identify a function for the challenging behaviours shown by a small proportion of individuals (Derby et al. 1992: Emerson et al. 1995; Iwata et al. 1982; Iwata et al. 1990). This may have been due to the inappropriate application of an assessment methodology, a failure to correctly identify the functional unit from within a complex behaviour stream, or an incomplete consideration of all potentially salient factors. Recent research suggests a range of factors may impinge on the functional assessment of challenging behaviour, including the effects of biological function; setting conditions; multiple-control; non-dyadic relationships; idiosyncratic

discriminative stimuli; relative reinforcer rates, quality and delay; response force or effort; alternate responses; deprivation and satiation effects; temporally distant events; verbal behaviour; activity sequences; behavioural momentum (Carr, 1994; Horner, 1994; Iwata et al. 1994; Mace, 1994). In addition, theories such as neural oscillator disturbance, developmental stage, neurochemical imbalance and psychodynamics (Baumeister, 1989; Carr, 1977), and the presence of linguistic ability (Jones, Lowe and Williams, 1993) allow that the effects of socially mediated reinforcement may be weakened, modified or even annulled. Thus, there are complexities and limitations involved in the practical application of functional assessment methods which may not be immediately apparent. New models are emerging which seek to integrate the biological, contextual and environmental determinants of behaviour (Hall and Oliver, 1992; Murphy, 1994; Oliver, 1993). To date, however, the clinical application of functional assessment is generally restricted to rudimentary investigations of environmentally determined hypotheses for the occurrence of challenging behaviour.

Conclusions

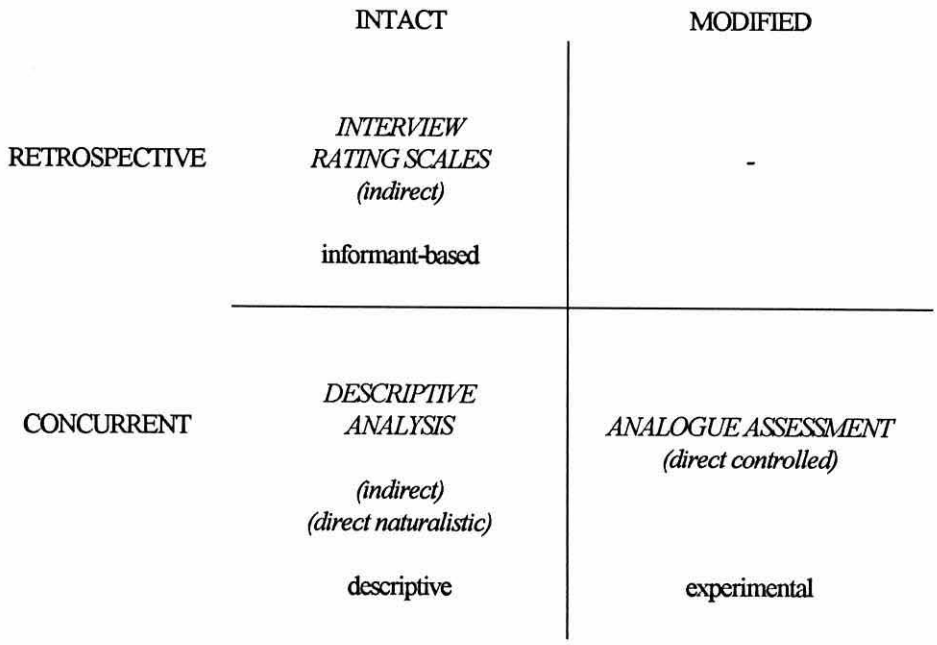
The purpose of functional assessment is to identify which of several underlying behavioural processes may be responsible for the maintenance of a particular response. Research has demonstrated that a prior functional assessment can improve the prospect of successful intervention. The present taxonomy of reinforcement suggests five globally defined classes of reinforcement may be operable. There is good evidence to suggest that challenging behaviours can be maintained by

operant contingencies. Although a range of factors has been identified for consideration in the functional assessment of challenging behaviour, the practical application of functional assessment in clinical settings remains, in general, quite rudimentary.

V. FUNCTIONAL ASSESSMENT METHODS

The functional assessment of challenging behaviour may be approached in a variety of ways. There is a range of assessment methods to choose from and these have been classified in various ways. Durand and Crimmins (1990), for example, classify assessment methods according to whether data are gathered from 'intact' or 'modified' social environments by 'retrospective' or 'concurrent' means. Iwata et al. (1990), on the other hand, suggest assessment methods are either indirect, direct and naturalistic, or direct and controlled (see also: Lerman and Iwata, 1993). Other authors employ a classification system based on whether assessment methods are primarily informant-based, descriptive or experimental (e.g. Emerson et al. 1995). An amalgam of these schemes is used in the present study as represented in figures 1.3 and 1.4. Here semi-structured interview and rating scales are classified as informant-based, indirect, retrospective methods which leave the social environment intact. Descriptive methods are classed direct and naturalistic or indirect and naturalistic depending on the way data gathered, both are concurrent, however, and leave the social environment intact. Analogue assessment is the only experimental method. It is classified direct and controlled and as modifying the social environment.

Figure 1.3
Classification of assessment methods



Based on Durand and Crimmins (1990) and Iwata et al. (1990)

Figure 1.4
The further classification of assessment methods

Classification	Durand and Crimmins (1990)		Iwata et al. (1990)	Method
Informant-based	retrospective	intact	indirect	Interview
	retrospective	intact	indirect	Rating Scales
Descriptive	concurrent	intact	indirect	ABC chart
	concurrent	intact	direct and naturalistic	Descriptive
Experimental	concurrent	modified	direct and controlled	Analogue

* also informant-based where staff collect data

Informant-based, descriptive and experimental methods

Figure 1.5. summarises the strengths and weaknesses of informant-based, descriptive and experimental assessment methods according to a number of recent reviews (Durand and Crimmins, 1990, 1991; Iwata et al. 1990; Iwata, Pace, Dorsey, Zarcone et al. 1994; Lerman and Iwata, 1993; Oliver, 1991; Oliver and Head, 1993; Reed and Head, 1993. Each assessment method is reviewed in more detail taking account of empirical studies that have been made in the field.

Figure 1.5

Strengths and weaknesses identified in prior reviews of informant-based approaches

INSTRUMENT / PROCEDURE	POTENTIAL STRENGTHS	POTENTIAL WEAKNESSES
Semi-structured clinical interview (e.g. O'Neill et al. 1990) ◆ informant-based ◆ retrospective ◆ indirect ◆ intact	◆ popular ◆ quick ◆ easy ◆ economical of resources ◆ can sample large amounts of time ◆ can sample large numbers of events ◆ can overcome problems of low rates of responding ◆ is less likely to overlook important influences ◆ can be setting specific ◆ can take account of setting event, temporally distant antecedents, lean schedules of reinforcement ....all of above, plus	◆ little guidance available ◆ few attempts to standardise ◆ interpretation can be difficult ◆ method not validated ◆ method not empirical ◆ reliability not established ◆ depends on interviewer skills ◆ depends on informant's observational skills ◆ subject to selective recall and other biases
Rating scale (e.g. Durrand and Crimmins, 1988) ◆ informant-based ◆ retrospective ◆ indirect ◆ intact	◆ provides a quantitative measure ◆ some have established measures of reliability and validity	◆ reliability and validity has been questioned ◆ problems reported with understanding the questions and differentiating responses on a seven point scale ◆ questions not applicable to all behaviours ◆ predictive validity not tested

Figure 1.5 (continued)

Strengths and weaknesses identified in prior reviews of direct and indirect descriptive methods

INSTRUMENT / PROCEDURE	POTENTIAL STRENGTHS	POTENTIAL WEAKNESS'
ABC charts (e.g. Evans and Meyer, 1985) ◆ informant-based ◆ concurrent ◆ indirect ◆ intact	◆ more objective than interview ◆ can provide a broadly based quantitative and qualitative data set ◆ can yield estimate of rate, duration etc. ◆ can indicate distribution (e.g. scatterplot) ◆ efficiency and ease of operation ◆ less prone to observer reactivity	◆ little evidence of their ability to predict or agree ◆ selective recording (bias), less busy days, most challenging behaviours, etc. ◆ relies on compliance of care-givers ◆ may be inaccurate or incomplete ◆ interpretation , patterns may be difficult to detect ◆ data may lack detail ◆ temporal relations may not be functional ◆ relationships may be weak or not apparent leading to inconclusive results ◆ contingencies may be masked by irrelevant events
Descriptive analysis by direct observation (e.g. Emerson et al. 1995) ◆ descriptive ◆ concurrent ◆ direct ◆ intact	◆ more objective than interview ◆ can provide fine grain data set ◆ can yield estimate of rate, duration etc. ◆ observes contingencies in the natural environment ◆ does not risk eliciting challenging behaviours ◆ ease of operation with the advent of hand-held computers	◆ results may vary with method of sampling; (interval, momentary, continuous) and category of codes ◆ can be time consuming and tedious, ◆ requires a high level of skill ◆ may be subject to observer reactivity ◆ can be insensitive to lean schedule of intermittent reinforcement ◆ unreliable estimates may lead to invalid conclusions ◆ requires a high number of events in the observational record (behaviour and environmental)

Figure 1.5 (continued)

Strengths and weaknesses identified in prior reviews of experimental methods

INSTRUMENT / PROCEDURE	POTENTIAL STRENGTHS	POTENTIAL WEAKNESS'
Analogue assessment (e.g. Iwata et al. 1982) ◆ experimental ◆ concurrent ◆ direct ◆ modified	◆ empirical (demonstrates hypothesised relations) ◆ predictive validity is established ◆ has objectivity ◆ provides a high degree of control ◆ able to test multiple and combined variables	◆ may cause temporary increase in challenging behaviour ◆ may be too complex for clinical settings ◆ is potentially insensitive to idiosyncratic events ◆ potential risk of introducing new functions ◆ time consuming ◆ needs highly trained staff ◆ is not universally applicable - e.g. dangerous or very low rate behaviours ◆ may be non-productive for up to one-third of individuals ◆ analogue conditions may not replicate natural environment adequately ◆ influence of setting events may be unclear ◆ some difficulty in interpreting data correctly ◆ may only be effective for relatively high rate behaviours

Informant-based methods

Informant-based approaches collect data which are screened by a third party. Common examples include semi-structured clinical interview, (O'Neill et al. 1990) and rating scales (Durand and Crimmins, 1992). These methods are not intrusive, they are administered relatively quickly and present few difficulties for the experienced practitioner. Other advantages include their ability to sample large amounts of data over extended periods of time, to deal with low response rates, and to take setting events into account.

Informant-based interview

Semi-structured interview has an overall framework, but does not prescribe the order or phrasing of questions. Information gained in this way may have a richness and depth not found in wholly structured or un-structured formats (Cline, 1985). However, the method does rely heavily on the skills of the interviewer and the respondent's ability to observe and report events accurately (Durand and Crimmins, 1990; Oliver and Head, 1993; Reed and Head, 1993). Factors which may influence the quality of interview data include the interview setting, rapport building and the perceived relevance of questions (Cline, 1985). There is no standard way of structuring a functional assessment interview or interpreting the data it may generate (Reed and Head, 1993). Useful guidance is provided, however, by O'Neill et al. (1990), LaVigna and Donnellan, (1986), and Sturmey (1991). In the present review no data were found to support the either the reliability or predictive validity of the interview method.

Informant-based rating scales

Unlike interview, rating scales provide a highly standardised method for collecting, analysing and interpreting assessment data. Durand and Crimmins' (1992) Motivation Assessment Scale (MAS), for example, has 16 standardised items which cover four categories of reinforcement. Each item is rated on a Likert-type scale where 0 = never and 6 = always. A mean score is derived for each reinforcement category and the relative ranking of the mean scores is used to determine the (apparent) relative influence of each reinforcement category. Tied or close scores suggest the behaviour may be multiply controlled.

MAS data require less subjective interpretation than interview data. Both the number and nature of the questions are, however, restrictive and may lead to errors of omission (Oliver and Head, 1993) in the ascribing of function. The MAS in particular has been criticised for the narrow relevance of certain items (e.g. the occurrence of aggression when alone), and there are reports of respondents finding some items difficult to rate (Reed and Head, 1993). As presently constructed the MAS fails to distinguish between escape from task-demand and non-demand related social contact. Activity and food related tangible reinforcements are also considered within a single category. Despite this, Durand and Crimmins (1988) reported a high degree of correspondence in the ratings two groups of 35 raters made on the self-injurious behaviours shown by fifty children in classroom settings. Participants in the study were identified on the basis that their behaviour interfered with classroom activity. The cohort of 50 was aged

between 3 and 18 years (mean 14.6 yrs), and included those diagnosed with infantile autism (22), severe learning disabilities (25) and developmental language disorder (3). Mental age, as assessed by the Vineland Social Maturity Scale, was reported to range from 11 to 84 months, and all participants were classed as having moderate to profound intellectual disabilities. Target behaviours were identified according to pre-defined criteria and occurred at an average rate of 15 times per hour. Primary raters were the children's teachers and secondary raters the teachers' assistants. MAS were completed by primary and secondary raters, and again by the primary raters 30 days later. Correlational analyses were used to assess both interrater agreement and stability in the ratings over time. Pearson correlation coefficients on the raw Likert scores ranged from .66 to .92 for interrater agreement and from .89 to .98 on re-test. Correlations for the mean scores ranged from .80 to .95 for inter-rater agreement and from .92 to .98 on re-test. Spearman rank-order correlations were used to examine correspondence in the ranking of function categories and these ranged between .66 and .81 for interrater agreement and .82 to .99 for re-test. What these statistics do not reveal, however, is the proportion of behaviours upon which an exact match occurred in the raw scores, the mean scores, or the relative ranking of function categories. Nor are the effects of the .25 criterion that was used to determine ties and cut-off apparent. Moreover, as a number of subsequent studies failed to obtain similar levels of interrater agreement (Zarcone et al. 1991; Newton and Sturmey, 1991; Sigafoos et al. 1994), the reliability of the device must be regarded as

uncertain. Durrand and Crimmins (1988) evaluated the predictive validity of the MAS by correlating the predictions made for eight children selected from the first study with the results of experimental analyses using analogue methods. Although the correlation was strong, even for the ranking of the potential functions, the reliance on convergence as a test of validity has been criticised (Reed and Head, 1993). A further study by Durand and Carr (1991) did, however, report complete convergence between the MAS and experimental analysis, and provided data on intervention which supported the validity of both assessment findings. Singh et al. (1993) examined the factor structure of the MAS and found it was robust for behaviours that occurred on average 15 or more times per hour. They concluded that the device was clinically useful. Kearney, (1994), evaluated interrater reliability and concluded that it may be influenced by response rate, setting selection and staff training procedures.

Informant-based staff use of ABC recording

The ABC chart is capable of providing information of a more objective kind than either informant-based interview or rating scales. The speed and relative ease of administration (when staff act as observers) makes this a popular approach in clinical practice. A number of texts describe ABC recording by staff (McBrien and Felce, 1992; Meyer and Evans, 1989; Murphy and Oliver, 1987; Presland, 1989), although the method has rarely been evaluated in clinical research. One study by Sasso et al. (1992), trained teachers to collect ABC data as part of a three phase investigation of aggressive and inappropriate verbal responding. The functions ascribed by the ABC method

matched those ascribed by conventional and modified experimental methods. The majority of published studies, however, supply specially trained observers and these findings cannot be extrapolated to cover approaches which use existing care staff, who may be otherwise busy and have received little training in observational methods (Reed and Head, 1993). ABC records rely on subjective judgement in recording and analysis, and can, therefore, be difficult to interpret. There is little published data on the reliability or validity of ABC chart recording by staff (Iwata, 1994; Reed and Head, 1993).

Descriptive analysis of direct observational data

Descriptive analyses aim to identify relationships between behaviour and events as they are observed to occur in the natural (intact) environment. Direct observation usually requires the presence of externally supplied persons, who are specially trained to the task. Observational data may be gathered in variety of ways. Taylor and Romanczyk (1994), for example, report observing the attending behaviours of teachers, rather than the problem behaviours of individual students. These authors correctly predicted how 14 of the 15 students would respond under brief experimental conditions. In another study, Repp and Karsh (1994), made paper-based recording for one subject, and computer-based recording for another. Both methods generated data on treatment integrity as well as effect.

The advent of small hand-held computers has made multiple event recording (Repp and Felce, 1990; Repp, Felce and Karsh, 1991) much more accessible to clinicians. Software is available

which allows events to be sampled at pre-defined intervals (Beasley, Hewson and Mansell, 1989) or continuously in real time (McGill, Hewson and Emerson, 1994). Once encoded, these data can be analysed to provide information about the temporal sequencing of behaviour and events (Bakeman and Gottman, 1986; Sackett, 1979). Hall and Oliver (1992), for example, investigated the relationship between self-injury and social contact by computing the conditional probability of observing social contact in intervals leading up to, during and following bursts of self-injurious behaviour. These authors concluded that contact from others was governed by long bursts of self-injury. When contact was lagged against every occurrence of self-injury, no relationship was apparent. Taking a different approach Emerson et al. (1995) created a series of environmental base states from observational data coded in real time. Each base state was designed to correspond with a specific hypothesised function. A function was ascribed after comparing the conditional probability (of observing each behaviour in each base state) with the unconditional probability (of observing behaviour at any point). Localised temporal relationships were examined using time-based lag sequential analysis (Sackett, 1979). In only two cases was the function ascribed by the first method modified after further analysis. Moreover, convergence occurred on the functions ascribed by experimental methods for 85% of topographies for which both methods made a prediction.

Descriptive methods are favoured for being more objective than informant-based methods (Durand and Crimmins 1990; Iwata et al.

1990) and less intrusive than experimental methods, although reactivity to observer presence can be a problem. The approach has been criticised for being complex and time consuming (Iwata et al. 1990) and clinical application may be limited by the degree of skill and equipment that is required.

Evidence for the predictive validity of descriptive methods is not well developed and many view the non-experimental nature of the approach as a weakness. Lerman and Iwata (1993) identify two problems in this area. First, temporally contiguous events may not be functionally related; a functional relationship may be suggested, therefore, where none exists. Second, functionally significant events which occur infrequently may be masked by the occurrence of more frequent but functionally irrelevant events. Thus, they argue, the results of such analyses may be inconclusive or even erroneous.

The quality of descriptive analyses crucially depends on the protocol developed for coding, and the number of events obtained in the observational record (Bakeman and Gottman, 1986; Moran et al. 1992; Gottman and Roy, 1990). The number of events required for a valid analysis of a multiple coding scheme (Bakeman and Gottman, 1986) may be especially problematic in stable, low-demand environments where there is little activity or social contact. As yet there is no standardised protocol for the collection, analysis or interpretation of observational data. Recent work by Emerson et al. (1995) and Oliver and Hall (1995) appears promising in this respect.

Experimental methods

Experimental methods aim to determine which reinforcers are responsible for maintaining a person's challenging behaviours by manipulating environmental conditions and observing the effect on the behaviour's rate. Challenging behaviours are normally observed across a series of rapidly changing and carefully controlled environmental manipulations which are usually brief and systematically scheduled. Systematic differences in behaviour rate, duration or intensity, which correspond with a specific hypothesis of function suggest the type of reinforcement that is most likely to be operating (Carr and Durand, 1985; Carr, Newsom and Binkoff, 1976; Iwata et al. 1982; Iwata, Pace et al. 1990; Oliver, 1991a; Oliver, 1991b: Parrish et al. 1985). Iwata et al. (1982), for example, described the application of four conditions, each lasting 15 minutes and presented in random order, to test three potential functions of self-injury shown by nine young people. In the first, play materials were available and an adult was present but attended briefly to the child contingent only upon the occurrence of self-injury. An increase in the rate of responding in this condition was interpreted as being consistent with a positive, attention gaining, hypothesis. In the second, the child was instructed to complete a difficult task, instructions were delivered every five seconds following a sequence of graded prompts, and the demands were temporarily withdrawn contingent upon the occurrence of self-injury. An increase in the rate of responding under these conditions was interpreted as being consistent with a negative, task-avoidance, hypothesis. In the third condition, the child

was left alone with no activity materials. Increased responding under low levels of external stimulation was judged consistent with an automatic sensory reinforcement hypothesis. A control condition was devised in which play materials were available and adult attention was delivered every thirty seconds contingent upon the absence of self-injury. A function was ascribed on the basis of differences in the mean rate of responding across analogue conditions being consistent with a specific reinforcement hypothesis. In this study, data differentiated sufficiently to ascribe function in two-thirds of cases. Carr and Durand (1985) investigated the relative effects of social attention and task difficulty on a range of challenging behaviours shown by four children with learning disabilities. Three experimental conditions were used, two comprised the presentation of an easy task and social attention delivered in either 100% or 33% of session intervals, in the third a difficult task was presented along with social attention in 100% of session intervals. A function was assigned for all of the behaviours shown by the four children. Validity was determined by examining patterns of responding after the children had been taught functionally relevant and irrelevant verbal responses as indicated by their assessment. Durand and Carr (1991) used a similar design to develop effective, durable and generalised treatments for challenging behaviours upon which convergence was reported with the functions ascribed by the MAS. Oliver (1991b) developed seven analogue conditions to examine the function of self-injury shown by children with severe learning disabilities. Sessions were sequenced so that only one variable changed at each

cross-over, there were no inter-session intervals. Continuous recording of target and collateral behaviour allowed trends in the behaviour's rate to be plotted both within and across sessions, thus identifying carry-over effects. Brief hypothesis testing conditions were added to further delineate the role of potentially important variables.

Experimental methods are favoured for their ability to provide an empirical demonstration of operant control (Durand and Crimmins, 1990, 1991; Oliver and Head, 1993; Reed and Head, 1993), and although originally developed to examine the function of self-injury, the procedures have been applied to a variety of subject populations, behavioural topographies and programme settings (Iwata, Pace, Dorsey et al. 1994; Iwata, Vollmer and Zarcone, 1990; Mace, Lalli and Lalli, 1991). Reliability has been demonstrated through extensive replications and refinements of the method (see: Iwata, Vollmer and Zarcone, 1990; Iwata et al. 1994), but there is scant evidence on the validity of functions assigned in these studies. Only a very small number of studies have demonstrated the efficacy of treatments based on functional assessment by experimental methods (Carr and Durand, 1985; Durand and Carr, 1991; Repp et al. 1988; Sasso et al. 1992) and studies that examined convergent validity present a very mixed picture with informant-based and descriptive methods showing poor (Burgess, 1987; Crawford et al. 1992; Lerman and Iwata, 1993; Oliver , 1991b) to moderately good (Durand and Crimmins, 1988; Emerson et al. 1995; Lalli et al. 1993; Mace, Lalli and Lalli, 1991; Sasso et al. 1992) levels of agreement.

Analogue methods are criticised in the literature for lacking universal application (e.g. to very dangerous or low rate behaviour), and for having an unacceptably high failure rate. Non-prediction of function occurred for one-third of the cohort in the study carried out by Iwata et al. (1982) and for up to half the subjects in a review of 79 cases by Derby et al. (1992). Although Iwata, Pace, et al. (1994) reported non-prediction occurred for only 4.6% of 152 cases reviewed over an 11 year period, it must be remembered that these assessments were carried out with a highly selective sample. Methodological concerns include risks that analogue assessment may temporarily increase response rates or bring behaviour under the control of new contingencies. Contingencies which operate in analogue conditions may not be identical to those which operate in the natural environment thus diminishing the prospect of transfer to the natural setting. Undifferentiated or artificially low rates of responding may result from a failure to establish important setting events or identify idiosyncratic variables in assessment conditions (Iwata, 1994; Sturmey, 1995). In addition, assessment data may be difficult to interpret where response rates fail to differentiate or the discriminative versus reinforcing properties of a stimulus condition are not clear (Oliver, Crayton, Murphy and Corbett, undated). Many of these problems have been overcome by applying experimental methods to evaluate potential treatment strategies rather than maintaining variables (Repp, Felce and Barton, 1988). Experimental methods have also been criticised, however, for being too complex and time consuming to be of value in a clinical context (Axelrod, 1987; Carr, 1995; Durand

and Crimmins, 1988). Shortened forms of assessment (Cooper et al. 1990; Derby et al. 1994; Harding et al. 1994; Northup et al. 1991), and a technical assistance model (Northup et al. 1994) are relatively recent developments and the application of experimental methods has been described in natural home-based settings (Emerson, Cummings and Barrett, 1989) and as a component of multi-element assessment formats (McBrien and Felce, 1992; O'Neill et al. 1990).

VI. PREVIOUS COMPARISON STUDIES OF ASSESSMENT METHODS

There is limited research on the convergent validity of assessment methods. Crawford et al. (1992) used a rating scale (MAS), direct ABC recording and an experimental method to assess the functions of stereotyped behaviours shown by four people with severe or profound learning disabilities. The MAS and ABC methods each suggested automatic reinforcement while the results of analogue assessment were less conclusive. There were considerable differences, however, in the ratings made by individual members of staff and the two groups (vocational and home-based) who rated behaviour on the MAS. These were reconciled by the authors averaging across the mean domain scores for the raters in each group, and none of the individual ratings were therefore presented. It appears, however, that not all staff rated the sensory function highest although no mention was made of how tied scores and those within 0.25 of highest were interpreted with regard to the possibility of multiple control. Thus, the assigning of automatic reinforcement by the MAS in this study must be regarded tentatively. While direct ABC recording yielded the clearest

separation in functions for all four subjects, the authors correctly indicated that a sampling bias may have occurred by virtue of the generalised low rates of interaction that were apparent within the natural setting. Finally, undifferentiated and paradoxical patterns of responding in the analogue data created difficulties of interpretation. This problem might have been resolved had the analogue conditions been presented on more than two occasions. A good level of agreement was reported for the MAS and experimental analogue assessment in three separate studies (Carr and Durand, 1991; Durand and Carr, 1991; Durand and Crimmins, 1988). It should be noted, however, that the experimental conditions used in these studies differed from those described by Iwata et al. (1982). Sasso et al. (1992) reported good agreement between informant-based ABC recording and two experimental analyses of aggressive and inappropriate vocal behaviours shown by two children diagnosed with autism. This was a two-phase study in which experimental analyses were first conducted by the authors, then ABC data were collected by highly trained teachers who received additional training from the authors, before the teachers themselves conducted a further experimental analysis in the classroom. Following assessment the effects of treatment were evaluated with multiple baseline designs in phase two. Results from each assessment indicated an escape function for all behaviours. Treatment evaluation revealed significant reductions in the levels of challenging behaviour and concurrent increases in on-task behaviour for both children, and these were maintained at follow-up. There is, however, one particular aspect of methodology which may have

influenced agreement in this study which is that all ABC data were collected in 15 minute sessions selected specifically to mirror the analogue conditions. Thus, it could be argued that there had been a passive manipulation of environmental conditions and that it was only the methods of data collection which differed between assessment methods. However, Lalli et al. (1993) report finding complete agreement on the functions ascribed by experimental and descriptive methods in two cases out of three, and partial agreement in a third, while Mace, Lalli and Lalli (1991) found that the experimental method they used agreed one of two functions that had been suggested by the descriptive method. Emerson et al. (1995) found agreement between descriptive and experimental methods was good when comparison was restricted to the topographies upon which both methods ascribed function (14 out of 23 investigated). This was a particularly good study since a wide range of topographies was considered across a relatively large subject sample (n=6). However, the rate of non-prediction by each method was not inconsequential, and agreement between them was poor when non-predictions were taken into account. Lerman and Iwata (1993) compared the functions ascribed by descriptive and experimental assessment methods and concluded that in five out of six cases the descriptive and analogue methods yielded data that would not have led to the same conclusion being drawn about a behaviour's function. Moreover, while the descriptive methods were useful in identifying social and non-social sources of reinforcement, they did not reliably discriminate between positive and negative reinforcements. While this study was extensive in its scope it appeared to disregard completely

the setting and social context within which variables may acquire reinforcing properties and it is unsurprising, therefore, that disagreement was apparent in the results. Oliver, (1991b) examined the functions ascribed by interview, analogue and direct descriptive assessment methods, and found little evidence of convergence in their predictions (see also: Burgess, 1987). This was an interesting study because each assessment method was applied independently by a different clinician who was blind to the others' results. Burgess (1987) compared the functions ascribed to twenty self-injurious topographies shown by five subjects. These data indicated that the interview and direct observation methods agreed completely on the functions assigned to six topographies out of twenty (30%). At least one topography was agreed for each subject, and there were four additional partial agreements. Complete agreement was found for three out of 16 topographies (18.75%) upon which the descriptive and experimental methods were compared. There were two additional partial agreements, but no agreements were apparent for two of five subjects. Finally, although not a comparison study a recent paper by Repp and Karsh (1994) used interview and direct observation to determine the function of tantrum behaviours shown by two children with learning disabilities. While interview suggested an escape function the descriptive methods indicated the presence of a positive reinforcement process involving social attention, even though the target behaviours were observed more frequently under demand conditions. Interventions were implemented based on the social attention hypothesis and these significantly reduced the rate of tantrum behaviours. This study illustrates

how the interpretation of assessment data can lead to different functions being assigned for the same behaviour.

Summary

On the basis of this review it appears informant-based interview methods would be unlikely to converge significantly with descriptive or experimental assessment methods. Experimental and descriptive methods may be expected to converge moderately while the MAS ought to obtain a good level of agreement with both experimental and descriptive assessment methods. Thus, overall agreement between informant-based, descriptive and experimental methods should fall somewhere between moderate and good (figure 1.6).

Figure 1.6
Summary of findings by previous studies

Study	Methods compared	Agreement rating
Crawford et al. (1992)	*informant + experimental*	poor
Burgess (1987) / Oliver (1991b)	*informant + experimental*	poor
Durand & Crimmins (1988)	*informant + experimental*	good
Sasso et al. (1992)	*informant + experimental*	good
Repp and Karsh (1992)	*informant + descriptive*	poor
Burgess (1987 / Oliver (1991b)	*informant + descriptive*	poor
Crawford et al. (1992)	*informant + descriptive*	good
Burgess (1987) / Oliver (1991b)	*experimental + descriptive*	poor
Lerman and Iwata (1993)	*experimental + descriptive*	poor
Mace, Lalli and Lalli (1991)	*experimental + descriptive*	moderate
Emerson et al. (1995)	*experimental + descriptive*	moderate
Lalli et al. (1993)	*experimental + descriptive*	moderate

It should be noted, however, that none of the studies reviewed standardised the selection of subjects, the assessment methodologies chosen for comparison, the classes of topography assessed, or the criteria used to determine agreement. Conclusions regarding convergence must inevitably be regarded as tentative.

VII. A FURTHER COMPARISON OF ASSESSMENT METHODS

In reviewing the comparison studies above a number of limitations were apparent. First, most studies compared only two or three assessment methods. Second, assessment was often (but not always) restricted to a specific response type, e.g. self-injury or stereotypy. Third, the number of the subjects was limited to between two and eight. Fourth, the criteria used to determine agreement was not always specified.

A number of researchers have identified the need to further develop, refine and compare a range of functional assessment methods (Crawford et al. 1992; Durand, 1990; Iwata et al. 1990; Iwata et al. 1993; Sturmey, 1995). The current dearth of comparison studies strongly suggests more work is needed in this area. Multiple assessment formats are increasingly being recommended (Durand and Crimmins, 1990; Felce and McBrien, 1994; Iwata et al. 1990; O'Neill et al. 1990). While Durand and Crimmins (1990) suggest hypotheses are better if two or more methods agree, they also observe that;

> '... although there is a high level of agreement that functional assessment should occur, there is not a corresponding level of agreement as to *how* this should occur' (Chapter 3, page 31).

The present study

The present study was designed to extend previous research by;

1. selecting a larger subject sample of participants,
2. assessing a larger number of topographies,
3. assessing a diverse range of topographies,
4. comparing a wider range of assessment methods,
5. applying the assessments in a clinical context.

A larger subject sample

Twenty people were selected to take part in the present study. The sample was considered large enough to reflect natural variation among the people with learning disabilities whose behaviour is described as challenging and small enough to manage the logistics of conducting multiple functional assessments within a clinical context and the rigours of applied research. The sample was 2.5 times larger than the next largest study covered in the review.

Comparing across a larger number of topographies

The present study identified and assessed 121 discrete topographies. This was approximately five times greater than the number of topographies examined in next largest study reviewed.

A broader range of behavioural topographies

Assessments were applied to a topographically diverse range of challenging behaviours, a total of six classes were established for comparison.

A larger array of assessment methods

Five assessment methods were selected for comparison and, as two versions of the rating scale were applied, a total of six assessments were made of each identified topography. Assessments were selected to include informant-based, experimental and descriptive methods. While not exhaustive, the selection of assessment methods was representative of the range of methods available.

A clinical context

All the assessments were carried out in a clinical context to ensure the ecological validity of the study. Thus, it is hoped that the present study will go some way toward satisfying calls that have been made for more work to be done on investigating the convergent validity of assessment methods, and that the scale of the project, together with the diversity of topographies assessed, means the results will have value in both a clinical and research context.

Chapter Two

METHOD

Survey

and

Functional Assessment Methods

I. SUBJECTS

Initial survey

It was suggested in chapter one that in the absence of a clear clinical definition the term challenging behaviour was open to interpretation and risked being over-inclusive. To minimise ambiguity a population in which the phenomenon is to be studied must first be identified and described as fully and accurately as possible. To identify and describe a population accurately requires the use of an instrument which is both reliable and valid.

The *Hester Adrian Research Centre Behaviour Problems Survey* (HARC BPS) was developed to identify and describe challenging behaviour as it occurs among people with learning disabilities. The instrument has been applied widely and has been shown to obtain acceptable levels of reliability and validity (e.g. Qureshi, 1993a).

The instrument

The *HARC BPS* provides demographic information about the population surveyed together with individual data on social functioning, relationships and challenging behaviour. The instrument is sectional, with the part reserved for challenging behaviours being divided into four sub-categories;

- physical attack,
- self-injurious behaviour,
- destructive behaviour,
- disruptive and other socially inappropriate behaviour.

Respondents are asked to provide a range of identifying and descriptive information, and to rate the severity of challenging behaviours according to whether they present;

- a serious management problem,
- a serious management problem, but controlled in the present setting
- a lesser management problem.

Relevant parts of the *HARC BPS* were used in the present study to i) describe the population as a whole, ii) aid the selection of a cohort for detailed functional assessment, and iii) further describe the characteristics of those selected.

Method

Ninety-two subjects were identified from a cohort to be resettled from hospital. Each person's key-worker was interviewed, together with the appropriate villa manager. At least one respondent knew the person well, having fulfilled their respective role for a minimum of twelve months. Pro-forma were used to record all interview responses. These were later analysed into a series of personal profiles. The personal profiles were arranged to describe the population as a whole, and to allow across subject comparisons to be drawn.

Sample selected for assessment

Fifty-one people were identified as showing at least one 'serious' or 'serious but controlled' behaviour problem. Those whose behaviour occurred at very low rates (n=3) were excluded. Thus, 48 persons were discussed with service managers and staff until twenty were agreed as presenting the most serious

challenging behaviours. For the purposes of this exercise the relative seriousness of the challenging behaviours under review was defined according to a composite of;

- the estimated rate, intensity or duration of the behaviours shown,
- consequences for the individual - such as injury, restraint, diminished opportunities for socialisation and participation in daily activities,
- consequences for others - such as injury, disruption, annoyance, elevated stress or other adverse emotional responses, and,
- the extent to which the behaviours were perceived as posing a major management problem for staff.

Consent

A detailed research proposal was submitted to the appropriate ethics committee. Approval was granted, subject to individual consents being obtained. None of the participants was able to consent to being assessed, or to allowing their assessment results to be compared. Next-of-kin were therefore approached. Nineteen individual consents were obtained. Permission was refused for one. Consent was therefore obtained for a substitute candidate yielding 20 subjects in all who proceeded to the second phase of the study.

Profile - background information

Details of the sample are given in table 2.1. In summary, the average age of the cohort was 38 years (range 24-63). Thirteen were male. All were described as having severe learning disabilities. The average length of stay was 17 years (range = 10-22). Three were diagnosed autistic, one had Down's Syndrome, seven had epileptic seizures. Three had been

TABLE 2.1 SUBJECT CHARACTERISTICS AS DEFINED THROUGH INITIAL SURVEY DATA

Subject	1`	2	3	4	5	6	7	8	9	10	11	12	13	14	`15	16	17	18	19	20
Age	30	32	35	54	32	38	28	48	41	33	63	29	53	36	48	55	33	40	24	28
Sex	M	M	M	F	F	F	F	M	F	F	F	M	M	M	M	M	M	M	M	M
Degree of learning disability	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	M	S	S
Length of stay (years)	21	18	22	18	16	17	13	18	16	20	15	13	14	22	14	23	17	10	20	13
Specific syndrome or condition	N	N	N	N	N	N	N	A	A	A	N	N	N	DS	N	N	N	N	N	N
Presence of Epilepsy (controlled?)	Y	N	N	N	Y	Y	N	Y	?	N	N	N	N	N	N	N	?	Y	Y	Y
Diagnosed psychiatric disorder?	N	N	N	S	N	N	N	S	N	N	N	N	D	N	N	N	N	N	N	S?
ADDITIONAL DISABILITY																				
Visual impairment	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	Y
Hearing impairment	N	N	N	N	N	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N
Impeded mobility	N	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	Y	N	N
Incontinent (urine/faeces/both)	N	B	Y	Y	Y	N	Y	N	Y	N	Y	Y	Y	N	N	N	N	Y	N	Y
ASSISTANCE REQUIRED																				
Feeding themselves	A	A	I	I	A	I	I	I	I	A	I	I	I	A	I	I	I	I	I	A
Washing themselves	A	A	A	A	A	A	I	I	A	A	I	A	A	A	A	A	A	I	A	A
Dressing themselves	A	A	A	A	A	A	I	I	I	A	A	A	A	A	A	A	A	I	A	A
DOMESTIC SKILLS																				
Level of supervision needed	-	-	S	-	-	S	-	-	S	-	N	-	-	-	-	S	-	N	-	-
Is client willing to do domestic task	-	-	Y	-	-	N	N	N	N	S	Y	N	Y	-	-	Y	Y	Y	-	N
Can client keep themselves occupied?	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	Y	N

TABLE 2.1 SUBJECT CHARACTERISTICS AS DEFINED THROUGH INITIAL SURVEY DATA

COMMUNICATIVE USE OF SPEECH AND GESTURE	**1`**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**`15**	**16**	**17**	**18**	**19**	**20**
Regular use of varied phrases or sentences				✓				✓			✓					✓	✓	✓		
Uses only a few words, sounds or gestures			✓						✓			✓		✓	✓				✓	✓
Little or no communication	✓	✓			✓	✓	✓			✓			✓							
UNDERSTANDING COMMUNICATION																				
Understands stories of other peoples experiences (i.e. more abstract ideas)									✓									✓		
Understands comments about his personal needs and experiences (Did you like the bus)				✓										✓	✓	✓				✓
Understands simple practical instructions	✓		✓				✓	✓		✓	✓	✓	✓				✓			
Understands a few simple commands only		✓			✓	✓													✓	
Understands little or nothing																				
ARTICULATION																				
Clear enough to be understood by anyone											✓					✓	✓	✓		✓
Can be understood by people who are close to the client but difficult for strangers			✓	✓				✓							✓					
Difficult to understand even those close to the client , and impossible for strangers														✓						
Not enough speech to rate	✓	✓			✓	✓	✓		✓	✓		✓	✓						✓	
HANDLING MONEY																				
Could go shopping and check change																				
Can use money but cannot check change																				

TABLE 2.1 SUBJECT CHARACTERISTICS AS DEFINED THROUGH INITIAL SURVEY DATA

	1`	2	3	4	5	6	7	8	9	10	11	12	13	14	`15	16	17	18	19	20
Realises money has value but does not use														✓						
Has no idea that money has value	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
BEHAVIOUR IN PRESENCE OF OTHERS																				
Almost always behaves appropriately																✓				
Generally acts appropriately			✓	✓														✓		
Sometimes behaves appropriately	✓							✓			✓	✓			✓		✓			
Interacts to satisfy his/her own needs		✓			✓	✓	✓		✓	✓			✓	✓					✓	✓
BEHAVIOUR IN PRESENCE OF STRANGERS																				
Almost always behaves appropriately																✓				
Generally acts appropriately			✓	✓				✓										✓		
Sometimes behaves appropriately	✓										✓				✓		✓			
Interacts to satisfy his/her own needs		✓				✓			✓	✓		✓		✓					✓	✓
No social responses					✓		✓						✓							
PARTICIPATION IN GROUP ACTIVITIES																				
Initiates group activities																				
Participates in group activities willingly			✓						✓		✓									
Participates if encouraged to do so						✓				✓				✓	✓	✓		✓		
Actively disrupts group activities																	✓			
Does not participate in group activities	✓	✓		✓	✓		✓					✓	✓							✓
Varies too much to say								✓											✓	

TABLE 2.1 SUBJECT CHARACTERISTICS AS DEFINED THROUGH INITIAL SURVEY DATA

FRIENDSHIPS	1`	2	3	4	5	6	7	8	9	10	11	12	13	14	`15	16	17	18	19	20
Does the client have any friends of any type	N	N	N	N	N	N	N	N	N	Y	Y	N	N	N	Y	N	N	Y	N	N
Describe relationship with other residents (positive, negative, indifferent, mixed)	I	I	I	I	I	M	N	I	I	I	P	I	I	I	M	I	N	P	I	I
Describe relationship with staff (positive, negative, indifferent, mixed)	M	I	I	P	I	M	M	I	I	P	P	I	I	M	P	P	M	P	P	I

KEY:

Age	in years	
Sex	M = male:	F= Female
Degree of learning disability	S = severe	P = profound
Length of stay (years)	in years	
Specific syndrome or condition	Y = Yes	N = No
Presence of Epilepsy (controlled?)	Y = Yes	N = No
Diagnosed psychiatric disorder?	Y = Yes	N = No
ADDITIONAL DISABILITY	Y = Yes	N = No
ASSISTANCE REQUIRED	A = assistance	I = independent
DOMESTIC SKILLS	S = supervised	N = no supervision - = unable to assess
COMMUNICATIVE		
USE OF SPEECH AND GESTURE	✓ = indicates nearest appropriate description	
UNDERSTANDING COMMUNICATION	✓ = indicates nearest appropriate description	
ARTICULATION	✓ = indicates nearest appropriate description	
HANDLING MONEY	✓ = indicates nearest appropriate description	
BEHAVIOUR IN PRESENCE OF OTHERS	✓ = indicates nearest appropriate description	
BEHAVIOUR IN PRESENCE OF STRANGERS	✓ = indicates nearest appropriate description	
PARTICIPATION IN GROUP ACTIVITIES	✓ = indicates nearest appropriate description	
FRIENDSHIPS		
Friends	Y = Yes	N = No
Other residents	P = positive:	N = negative: I = indifferent: M = mixed.
Staff	P = positive:	N = negative: I = indifferent: M = mixed.

diagnosed with schizophrenia, and one with a depressive illness. Two were reported visually impaired, two had a known auditory impairment, three needed help to ambulate, and eleven were incontinent of urine or faeces. Six subjects needed help with feeding, sixteen with washing and dressing, and 17 to occupy themselves. Fourteen had no speech or used only a few words; ten had insufficient speech to rate clarity of articulation. Thirteen were rated as understanding only a few simple commands or less. None knew about money, or how to use it. Just over half were rated as interacting only to satisfy their own needs, and as avoiding group activities.

Profile - challenging behaviours

Table 2.2 shows how staff rated challenging behaviour in the original survey. Multiple topographies mean the totals exceed the number of subjects in the cohort.

Table 2.2
Behaviour rated by category using the HARC BPS

BEHAVIOUR SUMMARY	SERIOUS	LESSER	CONTROLLED
Physical attacks on others	9	4	1
Self-injurious behaviour	10	6	-
Destructive behaviour	8	8	-
Other behaviour problems	15	9	1
Total	42	27	2

The largest single category of seriously challenging behaviour was 'disruptive and socially unacceptable behaviour'. This was followed by self-injury, aggression, and disruption of the setting. All twenty subjects displayed additional challenging behaviours rated as 'lesser'. These ratings mirrored those of the larger sample, and were consistent with studies reviewed in chapter one.

Conclusions

Overall, the sample was typical of the hospital population whose behaviour is described as challenging. The subjects' level of functioning was consistent severe or profound learning disabilities. A topographically diverse range of behaviours was present. These distributed between categories in the way other studies would have suggested. Including service managers and staff added a degree of external validity to the selection process.

II. SETTING

All participants lived in the same hospital. They did not, however, share the same living space, nor remain in the same places throughout the day. Many attended day-care on a sessional basis. Leisure activities, such as hill-walking, were arranged periodically.

Accommodation

Participants were accommodated in a range of villas throughout the hospital. Each villa housed up to 15 people. Most were single storey buildings which adjoined in pairs to share kitchen facilities. Each had a large living area, half of which functioned as a dining room, half as a lounge. Most had one or two smaller side-rooms. Some provided additional living space in a second large room. Residents slept in their own bedroom, or in dormitories which had been divided into smaller areas by partitions. Facilities for washing and bathing were communal.

Staffing

Each villa had a senior nurse designated as manager. Staffing establishments included a mix of trained and untrained nurses. Rosters provided between two and four staff at any one time throughout the day. A separate night staff provided waking cover between 8 pm and 7 am. A range of 'hotel' services, e.g. laundry, catering, gardening, was provided by other staff. Psychiatry, psychology and other therapists, were available on a referral basis. A key-worker system was in operation which required that one member of staff came to know one resident especially well.

Activities and materials

Villas were furnished with an ordinary range of tables, chairs, and wall-units. Some had ornaments and pictures. There was a TV and radio in each living area. Few recreational materials were available. The daily routine provided little opportunity for meaningful participation in domestic activity. The general picture was of enduring passivity imposed by a lack of materials, opportunity and assistance.

Conclusions

The accommodation and support arrangements were typical of those found among long-stay hospitals. Shared living spaces were large and impersonal. Support was provided by a mix of qualified and unqualified nurses. 'Hotel' services were provided by others. While day-care was arranged, residents spent most of their waking day in the villa; there was little to do, and few recreational materials available.

III. ASSESSMENT METHODS

Five assessment methods were selected for comparison. However, as two versions of the rating scale were administered six sets of assessment results were available for comparison. Each assessment method is listed in figure 2.1, together with the reasons for its inclusion in the study.

Figure 2.1
Reason for selection of assessment methods

Procedure	Reason selected
Semi-structured interview	*The approach is quick, economical of time and popular with clinicians. As an informant-based method, it is indirect, retrospective and leaves the environment intact. Although commonly used reliability and validity are unknown*
Rating scales	*This approach was included as an alternate informant-based method which shares the advantages of interview. Rating scales are indirect, retrospective and leave the environment intact. Although reliability and validity was examined only in relation to self-injurious behaviour the device has been applied to assess other challenging behaviours, and it has received much attention over recent years. A revised version was of the scale was included to cover social-avoidance independently from task-avoidance..*
ABC charts	*This approach was chosen because of its popular application in clinical settings. The ABC method is similar to other informant-based methods, except that it collects data concurrently. It is similar to other descriptive methods, except data are collected indirectly. The reliability and validity of the ABC chart method has not been demonstrated.*
Descriptive analysis	*A method of recording and analysis was selected for its potential utility in clinical practice, and to contrast with other approaches, e.g. while the ABC method utilises concurrent, indirect observation of intact environments, and the experimental method concurrent, direct observation of modified environments, this method uses concurrent direct observation of intact environments. Evidence of reliability and validity is beginning to emerge.*
Analogue assessment	*An experimental method was selected which is widely known and cited. Experimental methods utilise direct, concurrent observation of modified social environments. Claims for the reliability and validity of experimental procedures appear to be widely accepted, and it has been implied that they constitute the method of choice.*

All assessments considered challenging behaviour as it occurred in the person's daily living environment, and all were conducted by the author and/or one of two clinically experienced assistants. At least one clinician was involved in every assessment applied to a subject's challenging behaviours.

Data were gathered over a period of eighteen months and only one subject was assessed at a time. Assessments were completed over a period of 14 days although the order of presentation varied between individuals as a function of subject, staff and observer availability. This imposed a randomness into the scheduling of assessments which was believed advantageous.

Independence was preserved between the three main categories of assessment (informant-based, descriptive and experimental) by conducting each assessment as a discrete exercise. Informant-based methods were applied discretely though in practice the respondent for each interview and corresponding rating scale was generally the person's key-worker, and the two rating scales were administered simultaneously. To minimise possible cross-over in the analysis and interpretation of assessment results no data were analysed until all assessments had been completed. Assessment data were then analysed in batches of method rather than by subject.

Identification of behaviours for assessment

The identification and description of challenging behaviours formed part of the assessment process for each method. Topographies identified by one assessment method were not therefore automatically carried over for assessment by another. Informant-based methods relied solely upon the clinician's ability to facilitate the identification and description of behaviours for assessment. Experimental and descriptive assessments were preceded by a prior period of direct informal observation and the behaviours identified by observers were described and discussed with staff before a code was assigned.

Informant-based, semi-structured interview

One semi-structured interview was scheduled for each subject. Key-workers served as respondents, sometimes with the villa manager present. Interviews were conducted face to face by the author or his assistant in the respondents' normal place of work. In every case the key-worker or villa manger had known the person for a minimum of twelve months.

Interview outline

One interview covered all topographies identified as challenging. A framework was devised for each interview (fig 2.2) based on previous published work (LaVigna and Donnellan, 1986; O'Neill et al. 1990; Sturmey, 1992).

Figure 2.2
framework for semi-structured interview

General information;

- identifying and general background information,
- response definitions - (identified by respondent)
- health and medical status,
- mood and general mental health,
- communication; expressive and receptive,
- reinforcer preferences; likes and dislikes,
- access and availability of activity materials,
- sociability, preference for groups or being alone,

For each response topography defined by the respondent;

- history, onset and recent changes,
- cyclicity over a day, week, month, etc.
- rate, duration, intensity, bout, co-occurrence,
- specific setting conditions, e.g. people, places, crowding, times of day, hunger, fatigue, boredom, etc.
- specific antecedents, e.g. task-demand, no social contact, etc.
- specific consequences, e.g. task-avoidance, social contact, etc.

The clinical relevance of functional assessment was stressed at the start of each interview. The respondent was asked to identify and describe the person's challenging behaviours and

to provide a range of other generalised information. Each topography identified by the respondent was then considered separately. Open and closed questions were used as appropriate. Respondents were encouraged to talk as much as they liked. Interviewers listened and showed appreciation for the answers given. Non-directive utterances, such as 'uh-huh' and 'I understand' were used as much as possible. Clarification was sought through probing, reflecting and summarising, e.g. 'What would I see if ..... ?', or 'So, is what you're saying .....?'. Where necessary illustrations, examples or scenarios were used, either to explain a question, or check the meaning of an answer, for example, 'Suppose you and I were very close by, but talking only between ourselves - what would happen then?' Where conflicting responses were made, questions were re-phrased and asked again.

Analysis and interpretation

Copious notes were taken throughout each interview and a written synopsis was developed shortly afterwards. Evidence for each behaviour was organised into antecedents and consequences (appendix 2.1) and then assessed for congruence with one or more of the following five classes of reinforcement;

social positive - obtain adult attention
social positive - obtain food or activity
social negative - remove adult attention
social negative - remove demand
non-social - automatic

Where more than one function was assigned the evidence was reviewed again across the categories identified and the relative strength of the supporting evidence was assessed. The

ascribed functions were then ordered first to last in descending order of importance.

Accuracy and integrity

Each synopsis was independently reviewed by a second person. Agreements were noted and any disagreements resolved. This resulted in functions being ascribed with increased confidence regarding their congruence with the available evidence.

Informant-based rating scales

Durrand and Crimmins' (1992) Motivation Assessment Scale was selected as a second informant-based method. However, as the original scale did not consider social-avoidance separately from task-avoidance, four new questions were added to cover this potential function. The original items were re-worded, and the new 20 item scale was combined with the original to make a 36 item questionnaire (appendix 2.2). The scales were administered simultaneously but scored separately.

Administration

Questionnaires were administered by the author or his assistant and key-workers served as respondents, sometimes with help from the villa manger. Respondents were first asked to identify and describe the person's challenging behaviours. One questionnaire was then completed for each challenging behaviour identified by the respondent. Respondents were advised to rate items only in relation to the residential setting, and to consider times of the day when the behaviour was most likely to occur (Durrand and Crimmins, 1992).

Analysis and interpretation

Items on each scale were scored separately. A mean score was obtained for each sub-domain and these were arranged in rank-order. A cut-off of 0.5 from the highest mean score was used to determine which functions should be excluded (Durrand and Crimmins, 1992). Whenever multiple functions were ascribed they were ordered first to last according to the size of the mean score. Thus, taking the following hypothetical ratings as an example;

	sensory	escape	att'n	tangible	soc av
Original	1.5	**4.25**	1.75	**4**	-
Revised	1.5	**4.5**	1.25	3.5	2.5

the rank-order would be as follows;

	1	2	3	4	5
Original	escape	tangible	att'n	sensory	-
Revised	escape	tangible	soc av	sensory	att'n

Here the original version would have ranked escape highest and tangible reinforcement second. Because the difference in the two mean scores was 0.25 both functions would have been ascribed in that order. Social attention ranked third; however, as the difference between scores was greater than 0.5 it would have been excluded. The revised version also ranked escape first and tangible reinforcement second. However, because the difference between the two mean scores was greater than 0.5 only the escape function would have been ascribed.

Accuracy and integrity

The ratings, mean scores and relative rankings for each topography were independently checked by a second rater. All were found to be correct.

Informant-based staff use of the ABC chart

A simple pre-printed five column chart was developed (e.g. Meyer and Evans, 1989) for staff to record;

- date and time (of each occurrence)
- behaviour (as specified in advance)
- antecedent (to occurrence of behaviour)
- consequence (for occurrence of behaviour)
- initials (for each entry)

Response definitions were worked out with staff (usually a key-worker) and brief instructions on how to complete the chart were included (refer to appendix 2.3).

Administration

The process of ABC recording was explained to the villa manager and/or key-worker by the author or his assistant. Challenging behaviours were identified through discussion and the villa manger was then asked to arrange for all staff to keep ABC records on the identified behaviours for two consecutive weeks. Completed charts were collected at the end of the observation period. No further assistance or information was given. All managers agreed that their staff would carry out observations and all charts were returned even when blank.

Analysis and interpretation

The antecedent and consequence columns were reviewed for each recorded behaviour in turn. Antecedents and consequences were classified as being consistent with one or more of five possible reinforcement categories covering;

social positive reinforcement - adult attention
social positive reinforcement - food or activity

social negative reinforcement - adult attention
social negative reinforcement - removing demands
non-social reinforcement.

The relative frequency of entries for each category were then compared and a function was assigned where the proportion of entries for a given category appeared clinically significant in relation to the remainder. Individual entries were then visually scanned for correspondence on the functions suggested by the antecedent and the consequences column. Multiple functions were ordered first to last according to the relative number of assessed entries. The number of entries required for a valid analysis varied between records. For example, five entries would be admitted as valid where each suggested the same function for the same behaviour, but not where a different function was suggested.

Accuracy and integrity

The allocation of antecedent and consequences to each reinforcement category was first independently checked by a second rater and then discussed between the raters until agreement reached 100%.

Descriptive analysis - by direct observation

Descriptive data may be collected and analysed by a variety methods. The method first described by Emerson et al. (1995) was selected for the present study on the grounds that it offered a good level of sophistication without being too complex for use in a clinical context.

Response definition and measurement

A universal coding protocol was developed to cover all potentially salient environmental events (base variables) and

these are summarised in figure 2.3. Challenging behaviours (conditional variables) were identified and described from a prior period of informal observation and discussion with staff. Each topography was assigned a code from among a cluster of keys reserved for conditional events and additional keys were used where necessary to code the occurrence of idiosyncratic behaviours.

Figure 2.3
Summary definition of codes

CODE	MEANING	TYPE	SUMMARY DEFINITION
A	presence of food	duration	*food or drink in the same room as subject - e.g. lunch on serving hatch, drinks on table at break time.*
B	giving of food	event	*the act of presenting food or drink to subject, i.e. actually handing it over*
C	presence of activity	duration	*activity materials in the same room as subject, e.g. recreational or sport equipment*
D	giving of activity	event	*the act of presenting activity materials to subject, i.e.. handing it to the subject*
G	staff proximity	duration	*staff presence within one metre of subject, e.g. standing or sitting or passing by.*
H	client proximity	duration	*client presence within one metre of subject (as above).*
M	positive contact	event	*the act of presenting verbal, physical or gestural social praise to the subject*
N	negative contact	event	*the act of presenting a verbal, physical or gestural scolding*
O	neutral contact	duration	*verbal, physical or gestural contact that has no specific purpose, i.e. not praise, correction or instruction*
P	instruct	event	*the act of presenting verbal, physical or gestural commands, or providing assistance to complete a specific task*
Y	engaged	duration	*subject handling task materials in an appropriate and meaningful way, or attending to personal self-care activities such as washing, eating, dressing etc.*
Z	passive	duration	*doing nothing, handling materials inappropriately, e.g. twirling string - mutually exclusive with above.*
S-W	challenging behaviour	duration/ event	*defined by prior observation*

Data were collected in real-time using software developed for the Psion XPII hand-held computer (McGill, Hewson and Emerson, 1994), figure 2.4 shows the coding template used. Duration

keys were used to log the onset, offset and presence of variables with a durational quality. Event keys were used to log the occurrence of events with no durational quality, and a mutually exclusive set was used to record changes between pre-defined states covering engagement and passivity.

Figure 2.4
Coding template used in direct observation

A food present	B food given	C activity present	D activity given	E	F
G staff proximity	H client proximity	I	J	K	L
M positive staff contact	N negative staff contact	O neutral staff contact	P instruct	Q client to client	R
S stereotyped responding	T physical aggression	U SIB	V Damage	W Pica	X negative vocalisation
Y engaged	Z passive				

Each subject was observed for eight hours in the communal area of their normal daily living environment. All observations were conducted by the author or an assistant. Observations were made in 30 minute time-blocks scheduled over 2-3 days to cover morning, afternoon and early evening. At the commencement of each new observation period staff were advised that the purpose of observation was to record events as they occurred naturally. Staff were asked therefore to ignore the

presence of the observer and behave as they would normally. Observers stood discreetly away from the subject, remaining close enough only to view events clearly as they occurred.

Inter-observer agreement

Two observers simultaneously and independently coded 50 minutes (10%) of all observations made in the natural setting. A chance-corrected index of inter-observer agreement was then computed for each observed event using software developed by Reeves (1994a). This programme allows a pair of behavioural records to be compared and gives a Kappa coefficient (Cohen, 1960) for the onset, offset and presence of each code contained in the record. The Kappa coefficient is recommended as an index of inter-observer agreement which discounts chance agreements (Suen and Ary, 1989). Kappa values range from -1.00 to +1.00, a positive Kappa indicating that observers agree more frequently than would be expected by chance. Bakeman and Gottman (1986) suggest Kappa below 0.7 should be regarded with some concern while Fleiss (1981) suggests values above 0.4 may be considered fair and those above 0.6 good. Suen and Ary (1989) suggest 0.6 as an acceptable cut-off for inter-observer agreement.

All records in the present study were submitted for analysis using a two-second time window. This allowed for differences in observer and machine reactivity by counting codes entered into adjacent one second intervals as agreements. Detailed Kappa statistics are presented for each behavioural code in table 2.3. Overall values for the descriptive method averaged 0.83 for onset, 0.81 for offset and 0.90 for presence.

Table 2.3 Inter-observer agreement for descriptive method

			Kappa		
ID	Code	Event - Behaviour	onset	offset	presence
1	O	Neutral contact	1	-	0.9
1	S	Stereotypy	0.8	0.8	0.97
1	W	Pica	0.67	0.67	0.67
1	Z	Passive	1	-	1
2	Y	Engaged	0.79	0.59	0.79
2	Z	Passive	0.62	0.72	0.05
2	N	Negative contact	0.66	0.66	0.66
2	O	Neutral Contact	0.33	0.33	0.5
2	P	Instruct	0.73	0.73	0.75
2	S	Stereotypy	0.4	0	0.65
2	T	Aggression	0.86	0.72	0.75
2	U	Self-injury	1	1	1
2	V	Disruption	0.97	0.85	0.96
3	Y	Engaged	0.78	0.9	0.91
3	Z	Passive	0.91	0.73	0.7
3	N	Neg Contact	1	1	1
3	P	Instruct	0.9	0.61	0.82
3	Q	Other Contact	0	0.5	0.12
3	S	Stereotypy	0.19	0	0.69
3	T	Aggression	0.5	0.59	0.81
3	V	Disruption	0.75	0.73	0.92

			Kappa		
ID	Code	Event - Behaviour	onset	offset	presence
4	Z	Passive	1	1	1
4	M	Pos Contact	1	1	1
4	O	Neutral Contact	0.67	0.67	0.67
4	G	Staff proximity	0.91	0.7.9	0.74
4	A	Presence food	0	0	0
5	Z	Passive	1	1	1
5	O	Neutral contact	1	1	1
5	U	Self-injury	1	1	1
5	V	Disruption	0.5	1	0.76
5	X	Vocalisation	1	0.94	0.99
6	Y	Engaged	1	1	1
6	Z	Passive	1	1	1
6	M	Positive contact	0.8	0.8	0.8
6	Q	Other contact	1	1	1
6	G	Staff proximity	1	0.75	0.96
6	H	Client proximity	1	1	1
6	A	Food present	0	1	0.94
6	B	Food given	1	1	1
6	S	Stereotypy	1	1	1
6	W	Pica	0.91	0.91	0.92
7	Z	Passive	1	1	1

Table 2.3 Inter-observer agreement for descriptive method

ID	Code	Behaviour	onset	offset	presence
7	P	Instruct	1	1	1
7	G	Staff proximity	1	1	1
7	A	Food present	1	1	1
7	S	Stereotypy	1	1	1
8	Y	Engaged	1	1	1
8	Z	Passive	1	1	1
8	O	Neutral contact	0.83	0.83	0.91
8	P	Instruct	1	1	1
8	G	Staff proximity	0.5	1	0.99
8	H	Client proximity	1	1	1
8	B	Food given	1	1	1
8	U	Self-injury	1	1	1
9	G	Staff proximity	1	1	1
9	N	Negative contact	1	1	1
9	O	Neutral contact	0.5	0.5	..5
9	P	Instruct	0.4	0.4	0.6
9	S	Stereotypy	0.75	0.92	0.97
9	Z	Passive	1	1	1
10	Z	Passive	1	1	1
10	M	Positive contact	1	1	1
10	N	Negative contact	0.67	0.67	0.67
10	O	Neutral contact	0.86	0.86	0.87

ID	Code	Behaviour	onset	offset	presence
10	Q	Client contact	1	1	1
10	G	Staff proximity	0.91	0.72	0.99
10	H	Client proximity	0.61	0.92	0.99
10	X	Vocalisation	0.89	0.89	0.86
10	W	Pica	1	1	1
11	Y	Engaged	0.8	0.8	0.99
11	Z	Passive	0.8	0.8	0.99
11	N	Negative contact	1	1	1
11	O	Neutral contact	0.94	0.94	0.94
11	P	Instruct	1	1	1
11	Q	Client contact	1	1	1
11	G	Staff proximity	1	1	1
11	H	Client proximity	0	1	0.99
11	A	Food present	0	1	1
11	B	Food given	1	1	1
12	Y	Engaged	0	1	0.99
12	Z	Passive	1	0	0.79
12	O	Neutral contact	0.64	0.64	0.69
12	G	Staff proximity	0.67	0.67	0.71
12	R	Tantrum	1	0	0.89
13	A	Food present	1	1	1
13	B	Food given	1	1	1

Table 2.3 Inter-observer agreement for descriptive method

ID	Code	Behaviour	onset	offset	presence
13	G	Staff proximity	0.93	0.93	1
13	H	Client proximity	1	1	1
13	N	Negative contact	1	1	1
13	O	Neutral contact	1	1	1
13	P	Instruct	1	1	1
13	S	Stereotypy	0.89	0.89	0.99
13	U	Self-injury	1	1	1
13	X	Vocalisation	1	1	1
13	Y	Engaged	1	1	1
13	Z	Passive	1	1	1
14	A	Food present	0.8	1	1
14	B	Food given	1	1	1
14	G	Staff proximity	0.66	1	0.98
14	H	Client proximity	1	1	1
14	O	Neutral contact	1	1	1
14	X	Vocalisation	0.91	0.91	0.96
15	Z	Passive	1	1	1
15	M	Positive contact	1	1	1
15	O	Neutral contact	0.93	0.93	0.93
15	P	Instruct	1	1	1
15	G	Staff proximity	1	0.86	1
15	H	Client proximity	1	1	1

ID	Code	Behaviour	onset	offset	presence
16	Z	Passive	1	1	1
16	U	Self-injury	1	0	0.93
16	S	Stereotypy	1	1	1
16	O	Neutral contact	0	0	0.25
16	H	Client proximity	0.67	0.67	86
16	R	Verbal	0.6	0.78	0.9
17	Y	Engaged	0.67	1	1
17	Z	Passive	1	1	1
17	M	Positive contact	1	1	1
17	N	Negative contact	0.86	0.86	0.86
17	O	Neutral contact	0.95	0.95	0.95
17	Q	Client contact	0.5	0.5	0.5
17	G	Staff proximity	0.87	1	1
17	H	Client proximity	1	1	1
17	A	Food present	1	1	1
17	B	Food given	1	1	1
17	X	Vocalisation	0.87	1	0.98
18	Z	Passive	1	1	1
18	M	Positive contact	1	1	1
18	N	Negative contact	0.67	0.67	0.67
18	O	Neutral contact	0.86	0.86	0.87
18	Q	Client contact	1	1	1

Table 2.3 Inter-observer agreement for descriptive method

ID	Code	Behaviour	onset	offset	presence
18	G	Staff proximity	0.91	0.72	0.99
18	H	Client proximity	0.61	0.92	0.99
18	X	Vocalisation	0.89	0.89	0.86
18	W	Pica	1	1	1
19	Y	Engaged	0.5	1	1
19	Z	Passive	1	0.67	1
19	M	Positive contact	0.44	0.44	0.55
19	N	Negative contact	0.76	0.76	0.78
19	O	Neutral contact	0.89	0.89	0.88
19	P	Instruct	0.66	0.66	0.66
19	G	Staff proximity	1	0.39	0.96
19	H	Client proximity	1	1	1

ID	Code	Behaviour	onset	offset	presence
19	A	Food present	1	1	1
19	B	Food given	0.8	0.8	0.8
19	S	Stereotypy	1	1	1
19	T	Aggression	0.39	0.39	0.67
19	V	Disruption	1	0.5	0.99
19	W	Pica	0.5	0.1	0.93
20	S	Stereotypy	0	0.5	0
20	U	Self-injury	0.63	0.45	0.56
20	X	Vocalisation	0.56	0.43	0.72
20	Z	Passive	1	1	1
20	R	Vomiting	1	1	1

Analysis and interpretation

Data files were downloaded onto a Dell 466/MX and analysed by the author using bespoke software developed by Reeves (1994b). Discrete codes were manipulated to create five mutually exclusive base states within the observational record, each was designed to correspond with one of five possible reinforcement categories as outlined in figure 2.6.

Figure 2.6
Summary of base states

Function	Base state
Task demand	*all episodes where instructions were delivered at intervals less than, or equal to, the overall median of the inter-event interval*
Social avoidance	*all episodes where staff were proximate and interacting (staff interaction was discriminative for escape)*
Social Attention	*all episodes where staff were proximate, but not interacting, (staff presence was discriminative for contact)*
Tangible (food)	*all episodes where food, drink were present, but not accessible*
Tangible (activity)	*all episodes where activity materials were present, but not accessible*
Automatic reinforcement	*all episodes where no social contact or materials were available*

The probability of observing each topography was then calculated for i) each base state (P_c), and ii) at any point during the observational record (P_u). A function was assigned for each topography where;

- for a specific base condition the 'z' index indicated the difference between P_c and P_u was significant at $p = 0.01$, and
- differences across all base states were plausibly consistent with the function suggested.

Thus, taking the following hypothetical data as an example;

	instruct	contact	proximity	food	activity	no contact or materials
Stereotypy	-33.68	-5.67	-47.9	-34.26	-2.96	62.75

SIB	4.27	-0.21	10.4	1.75	-0.12	-11.45

it appears that the probability of observing stereotypy was significantly greater than chance under conditions where no activity materials were present and no social contact was available. It was also significantly less likely when staff were nearby, were interacting or giving instructions, and when activities were available. Stereotypy was more likely, therefore, under conditions of low stimulation, a pattern considered consistent with an automatic reinforcement hypothesis.

The probability of observing self-injurious behaviour, on the other hand, was significantly greater than chance under conditions where staff were nearby but not interacting, and when demands were made. The behaviour was also significantly less likely when the person was alone with no activity materials, and was seen at levels near to chance in the presence of tangibles and when contact was non-demanding. This pattern would be consistent with both the social attention and task-avoidance hypotheses. Both functions would therefore be assigned and ordered according to the size of the deviation from chance.

Experimental analysis - analogue assessment

Procedures in the present study were based on those described by Iwata et al. (1982). Five conditions were presented on each of four occasions (Figure 2.7). Session length was held constant at 10 minutes, except for s20 when the severity of his self-injury dictated that session length should be reduced to five minutes.

Figure 2.7
Summary description of experimental conditions

Function	Description of analogue condition
Positive reinforcement -social attention	*The clinician sat adjacent to the subject and assumed the appearance of reading a book. Any form of social contact was avoided. A mild disapproving remark was made following every occurrence of target behaviour.*
Positive reinforcement - access to tangibles	*The clinician sat adjacent to the subject, usually at a table. Task materials were available. At the start of each session a favourite food or drink was sampled, covered, and placed on the table. The subject was encouraged to manipulate the task materials. Non-contingent social contact was delivered every 30 seconds. Attempts to access food or drink were redirected, target behaviours were ignored. Food or drink was presented at the close of each session.*
Negative reinforcement - task escape-avoidance	*The clinician sat opposite the subject, usually at a table. Task materials were presented (six wooden blocks and three colour matched boxes with lids). A variety of specific instructions were delivered every 10 seconds, using a graded prompt sequence. Task difficulty increased on successive trails. The task was simplified after non-compliance occurred on three consecutive trials. Demands were withdrawn contingent only on the occurrence of target behaviour; the clinician turned away for the remainder of the inter-trial interval, plus 10 seconds. Task difficulty was reduced at the subsequent trial, and increased thereafter.*
Negative reinforcement - social escape-avoidance	*The clinician remained approximately one metre distant from the subject and spoke continuously, as if in conversation. The clinician moved away and terminated dialogue contingent upon every occurrence of target behaviour. Contact was reinstated after ten consecutive seconds had elapsed without the occurrence of target behaviour.*
Automatic reinforcement	*No activity or social contact was available for the duration of the session. Staff were asked to refrain from interacting with the subject, and to avoid presenting activity materials.*

Sessions were separated by a minimum of ten minutes, presented in counterbalanced order to control for possible order effects, and scheduled equally across mornings and afternoons.

Settings and personnel

Sessions were conducted in a discreet area of the communal day-room or adjacent side-room. All sessions were conducted by the author and/or one of two assistants each with prior experience of using analogue methods. One assistant always acted as observer, and the author provided training and supervision for the operation of analogue conditions run by other persons.

Safeguards

The assessment procedure was explained to each person's next-of-kin and their prior consent obtained. Independent system advocates and staff representatives were invited to be present at any assessment. A physician was asked to determine a person's fitness to participate whenever their health was in question. Termination criteria were established such that anyone present could call for the immediate cessation of any session for any reason whatsoever.

Response definition and measurement

Response identification and definition was accomplished via a prior period of direct informal observation of each subject in the natural setting and discussion with staff. An observer was present throughout all analogue sessions to log the onset, and where appropriate offset, of all pre-defined challenging behaviours using software developed for the Psion XPII (McGill, Hewson and Emerson, 1994).

Inter-observer agreement

Two observers simultaneously and independently coded one session from each analogue condition (25%). A chance-corrected index of inter-observer agreement was calculated for each behaviour observed during reliability sessions. Kappa for the observed behaviours observed are presented in table 2.4. No behaviours were recorded by either observer during reliability sessions arranged for subjects 1, 8, 9, and 11.

Analysis and interpretation

Observational data were transferred to a Dell 466/MX for analysis. Session totals and condition means were calculated,

Table 2.4 Inter-observer agreement for analogue method

			Kappa		
ID	Code	Behaviour	onset	offset	presence
2	u	Self-injury	1	1	1
3	s	Stereotypy	0.72	0.7	1
4	x	Vocalisation	1	1	1
5	u	Self-injury	0.91	0.91	0.9
5	x	Vocalisation	0.9	0.9	0.91
5	v	Disruption	0.93	0.93	0.95
6	t	Aggression	0.61	0.61	0.72
6	w	Pica	0.54	0.56	0.38
6	v	Damage	0.5	0.5	0.62
6	u	Self-injury	0.66	0.66	0.66
6	s	Stereotypy	0.69	0.69	0.84
7	v	Disruption	1	1	1
10	t	Aggression	0.92	0.92	0.88
10	u	Self-injury	0.54	0.39	0.43
10	x	Vocalisation	0.51	0.35	0.51
10	s	Stereotypy	0.81	0.69	0.82
12	t	Aggression	1	0.48	0.88
12	v	Disruption	0.49	0.49	0.66
12	x	Vocalisation	0.38	0.38	0.56
12	r	Tantrum	0.88	0.71	0.88
13	s	Stereotypy	0.8	0.73	1

			Kappa		
ID	Code	Behaviour	onset	offset	presence
13	x	Vocalisation	0.85	0.7	0.84
14	r	Masturbate	1	1	1
14	x	Vocalisation	0.42	0.14	0.99
15	r	Masturbate	1	1	1
16	r	Verbal	0.84	0.75	0.93
17	w	Vocalisation	0.92	0.92	0.92
17	x	Hyperactivity	1	1	1
18	v	Disruption	1	1	1
18	x	Vocalisation	1	1	1
19	s	Stereotypy	71	0.46	0.82
19	t	Aggression	0.69	0.79	0.73
19	v	Disruption	0.85	0.85	0.95
19	x	Vocalisation	0.81	0.78	0.79
20	s	Stereotypy	0.54	0.32	0.59
20	t	Aggression	0.79	0.79	0.82
20	v	Disruption	1	0.65	0.83
20	x	Vocalisation	0.89	0.66	0.97

and differences in the rates of the observed behaviours, and where appropriate their durations and bout lengths were assessed. Functions were assigned on the basis of differences across conditions being i) consistent with one of five potential hypotheses, and ii) present across two or more sessions. In addition, function was only assigned if the mean rate of responding for the indicative condition was 50% above the overall mean calculated across all sessions and conditions. Where two or more of the assessed behaviours satisfied these conditions they were ordered first to last according to the degree of elevation in the respective means.

IV. COMPARISON OF ASSESSMENT DATA

The functions assigned by each assessment method were arranged by subject into a 726 cell matrix comprising assessment method and topography (appendix 2.4). Non-predictions were indicated by the letters 'IND' and multiple functions were rank-ordered according to the criteria described earlier for each assessment method. The function placed first was deemed to be the behaviour's primary function and all other functions were classed secondary and arranged in descending order of importance. Assessment data were compared in a series of analyses and the methods for each are described fully in the relevant sections of chapters three and four.

Chapter Three

RESULTS

Comparison Analyses

on

functional assessment methods

I. INTRODUCTION

Six functional assessments were completed on 121 discrete topographies shown by twenty people with severe learning disabilities. Results from each functional assessment were compared and the findings are presented here in five parts.

In part one assessment methods are compared on their overall rate of prediction, part two then examines the ascribing of multi-function complexes and in part three convergence is assessed between one base and five comparison methods. Analyses in section three cover i) all functions ascribed in any order, ii) functions ascribed as primary, iii) functions ascribed as secondary, and iv) the relative ranking of multi-function complexes, and each is constructed to investigate agreement over i) all topographies shown by all subjects, ii) topographies for which the base method alone ascribed function, and iii) topographies for which the base and each comparison method ascribed function. In part four analysis is restricted to include only the topographies upon which all six assessment methods ascribed function, that is, with all non-predictions excluded. Finally, in part five, each assessment is designated as the base method in turn and the respective comparison methods are assessed for convergence on the functions ascribed to i) all topographies, 11) topographies for which the base alone method ascribed function, iii) topographies upon which the base and comparison methods each ascribed function. A detailed breakdown of functions ascribed by each assessment method may be found at appendix 2.4.

II. RELATIVE RATE OF PREDICTION

There is some evidence in the literature to suggest that assessment methods may ascribe function at different rates (e.g. Emerson et al. 1995). From a clinician's perspective the extent to which an assessment method is capable of ascribing function will be a matter of considerable importance. Thus, in the first analysis assessment methods were compared on their overall rate of prediction.

Method

The number of topographies for which each assessment ascribed function was identified. A prediction was counted each time an assessment method ascribed one or more functions for a specific behaviour. A prediction rate was calculated for each assessment method and expressed as a proportion of the total number of topographies assessed (n=121). Reasons for each method's non-prediction were then identified and compared.

Results

A function was ascribed by one or more assessment methods for 119 topographies (98.4%). The number of topographies for which all six methods ascribed a function was sixteen (13.2%). The number of topographies for which no prediction was made varied between assessment methods and is summarised in table 3.1.

Table 3.1
Prediction rate by assessment method

	Interview	ABC	MAS1	MAS2	Analogue	Desc.
Predictions	89	40	85	87	50	82
Non-predictions	32	81	36	34	71	39
Probability of prediction	0.74	0.33	0.7	0.72	0.41	0.68
Probability of non-prediction	0.26	0.67	0.3	0.28	0.59	0.32

Interview, the rating scales and the direct descriptive method, each ascribed a function for approximately two-thirds of all topographies, while analogue assessment and ABC charts ascribed for approximately one-third.

Reasons for non-prediction

Reasons for non-prediction were considered using three broad categories. First, a challenging behaviour might not have been identified for analysis. Secondly, while a challenging behaviour might have been identified, the data collected on it may not have been sufficient for analysis. Third, a challenging behaviour might have been identified, and had sufficient data collected for analysis, yet the assessment method may still have failed to identify a positive function for the behaviour. Non-predictions from within the present study were therefore identified and analysed accordingly, and the results are tabled in 3.2.

Table 3.2
Classification of non-prediction by assessment methods

	Interview		ABC		MAS1		MAS2		Analogue		Desc.	
Number of non-predictions	32		81		36		34		71		39	
Reason for non-identification	n	%	n	%	n	%	n	%	n	%	n	%
Non-identification of behaviour	29	90.6	32	39.4	34	94.5	34	100	4	5.6	2	5.3
Insufficient data for analysis	1	3.1	37	45.7	2	5.5	-	-	47	66.2	28	71.7
Inconclusive result	2	6.2	12	14.8	-	-	-	-	20	28.2	9	23

Reasons for non-prediction appeared to vary with the type of assessment method reviewed. Interview and the rating scales failed mainly where topographies were not identified by the respondent as challenging (90.6%), rarely were assessment data by these methods not susceptible to analysis and interpretation (range 0-9.3%). Classes of topography not identified by

respondents were, however, considered potentially challenging and included; disruption (n=9), stereotypy (n=6), self-injury (n=4), negative vocalisations (n=6), verbal abuse (n=2), and a multiple topography outburst (n=1). Aggressive behaviours were identified for assessment on every occasion. While the ABC method also failed to ascribe because of non-identification (39.4%), this method failed more often where insufficient were available for analysis (45.7%). The analogue method also failed for this reason with target behaviours being either too low-rate or not observed in 66.2% of cases where no prediction was made. Non-occurrence or low behaviour rate was the principal reason for non-prediction by the descriptive method (71.7%) with statistically non-significant results being found in just 23% of all non-predictions by this method.

Conclusion

Results from the present study suggest that the rate of prediction varied between classes of assessment method. Retrospective informant-based methods made more predictions than the concurrent method, and the direct descriptive method failed to ascribe fewer times than either the indirect descriptive or experimental methods. Moreover, assessment methods failed for different reasons and it is possible that prediction rate may have been influenced in part by aspects of procedure which defined the methods. Methods which relied on informants were, for example, more prone to non-identification of challenging behaviours by respondents, while methods which relied on direct observation were vulnerable to sampling difficulties due either to poor recording or low rates of responding under the prevailing environmental conditions.

These data must, however, be interpreted with caution since in the present study the validity of the functions ascribed was not tested directly. The 'higher predicting' assessment methods may have generated a greater proportion of false positives than the methods which ascribed at lower rates. With this limitation in mind it seems appropriate to conclude only that more interpretable data were obtained when staff acted as retrospective informants as opposed to concurrent observers, and when direct observations were made in the natural environment rather than in the modified environments of the experimental conditions.

III. PREDICTION OF MULTI-FUNCTION COMPLEXES

Challenging behaviour may serve more than one function (Carr, 1995) and functional assessment should detect multiple control where it exists. Predictions of multiple control may, however, result from an assessment method failing to discriminate relationships that are functionally significant from those which are functionally marginal or even irrelevant. Assessment methods which eliminate only one or two classes of reinforcement, or which fail to identify the most salient variables clearly, will be of limited value to clinicians when developing intervention strategies. The proportion of topographies for which each assessment method ascribed multiple control was therefore examined and compared.

Method

Predictions by each assessment were reviewed. Topographies were classified according to the number of functions ascribed and simple proportions were then calculated and compared. A

ratio of single to multiple prediction was derived for each assessment method and it is these which are reported along with the mean, range and median rate of prediction per topography.

Results

Interview, made the greatest number of multiple predictions, both overall (n=65), and as a proportion of the predictions it made (table 3.3).

Table 3.3
Predictions of single and multiple control by assessment method

Number of functions ascribed	INT	ABC	MAS1	MAS2	Analogue	Desc.
Zero	32	81	36	34	71	39
One	24	24	54	46	21	52
Two	44	11	25	31	25	22
Three	14	5	5	9	4	8
Four	7	0	1	1	0	0

The average ratio of single to multiple predictions for the interview method was 1 to 2.7 topographies. Other informant-based methods averaged 1 to 0.57 (MAS1), and 1 to 0.89 (MAS2). Ratios for the descriptive methods were 1 to 0.60 (ABC chart) and 1 to 0.57 (direct observation). The experimental method ascribed a greater proportion of two function combinations than one which resulted in a ratio of 1 to 1.38.

Interview predicted the greatest number of three (n=14) and four (n=7) function combinations. The mean number of predictions by each assessment method is given in tables 3.4 and 3.5, together with the respective median and range. Median values of zero in table 3.4 result from the proportion of topographies for which the ABC chart and analogue methods made no prediction. Except for interview, the mean number of

predictions per topography varied little. This was especially true when non-predictions were excluded (table 3.5). The elevated median for the analogue method in table 3.5 reflects the ratio of two function combinations made by this method.

Table 3.4
Number of predictions per topography - including non-predictions

	INT	ABC	MAS1	MAS2	Analogue	Desc.
Number of topographies	121	121	121	121	121	121
Number of functions ascribed	182	61	124	139	83	120
Mean	1.5	0.5	1.02	1.15	0.69	0.99
Median	2	0	1	1	0	1
Range	0-4	0-3	0-4	0-4	0-3	0-3

Table 3.5
Number of predictions per topography - excluding non-predictions

	INT	ABC	MAS1	MAS2	Analogue	Desc.
Number of topographies	89	40	85	87	50	82
Number of functions ascribed	182	61	124	139	83	120
Mean	2.04	1.53	1.46	1.6	1.66	1.46
Median	2	1	1	1	2	1
Range	0-4	0-3	0-4	0-4	0-3	0-3

Conclusion

Interview ascribed multiple control more often than other assessment methods and was, therefore, either the method most sensitive to multiply controlled behaviour or the approach least able to discriminate functionally relevant relationships. Definitive conclusions cannot be drawn from the present analysis, as the validity of the predictions is not known. The clinical utility of an assessment method which eliminates only one or two classes of reinforcement may, however, be questionable - even when all the predictions are valid. Unless assessment is very context-specific, a clinician would have great difficulty interpreting the relative significance of one

variable over another. It has already been suggested that functional assessment may benefit from being framed more contextually (Horner, 1994; Jones, 1995) and this may be especially important where multiple control is suspected. No systematic differences were apparent between the types of assessment method discussed in chapter one.

IV. CONVERGENCE ON THE ASCRIBING OF FUNCTION

A major aim of the present study was to investigate the level of convergence apparent in the functions assigned by a range of assessment methods. Given multiple methods were to be compared and that variation was apparent between methods on topography selection and overall prediction rate, a base method was required to assess convergence.

Interview was selected as the base method for three reasons. First, interview was identified as being among the most frequently used assessment methods in clinical practice. Second, results of earlier analyses suggested the interview method ascribed the highest number of functions to the greatest number of topographies, predicted multiple control for the largest proportion of topographies overall, and made relatively more predictions per topography than any of the other assessment methods used. Third, selection of the interview method would allow comparisons to be drawn both within and across the classes of assessment method discussed in chapter one. In the following analyses all other assessment methods served as comparison methods.

Convergence with interview on the non-prediction of function

Non-prediction by each assessment method was described in an earlier analysis covering overall prediction rate. The present analysis was designed to examine convergence on the non-prediction of function between interview and the comparison assessment methods.

Method

Each comparison method was assessed against the base method in turn, and the total number of behaviours determined upon which a prediction and a non-prediction was indicated by i) both methods, and ii) the base or comparison method only. These data were then arrayed into a series of 2 X 2 matrices (table 3.6) and a Phi-square statistic was calculated to assesses the strength of the potential association in the categorical data. Stronger relationships are indicated by values as they approach +1.0.

Results

Results suggest the rating scales were more likely to agree a non-prediction with interview (table 3.6). Although the ABC chart method matched 94% of the 32 non-predictions by the interview method, this has to be seen in the context of a high overall rate of non-prediction by the ABC method.

Table 3.6
Agreement on non-prediction by interview

	ABC		MAS1		MAS2		Analogue		Desc.	
INTERVIEW	predict	non predict	predict	non predict	predict	non predict	predict	non predict	predict	non predict
Predict	38	51	82	7	84	5	37	52	54	35
Non-predict	2	30	3	29	3	29	13	19	28	4
Phi correlation	0.12		0.64		0.7		0		0.06	

The experimental and descriptive assessment methods agreed 19 and four non-predictions with the interview method respectively, and ascribed a function for 13 and 28 of those same behaviours. Thus, it would appear that the sub-sets of behaviours upon which functions were assigned and not assigned varied considerably between the classes of assessment discussed in chapter one, and this was reflected in the Phi-square statistics tabled above.

Conclusion

The retrospective informant-based methods agreed substantially on the non-prediction of function. However, as the results of earlier analyses suggested, non-prediction by these methods arose principally from the non-identification of behaviours as challenging. Perhaps the most significant aspect of the present results lies in the way functions were differentially assigned to behaviours by the informant-based, experimental and descriptive assessment methods. These results suggest that the search for a universal assessment method is far from complete and might even be inappropriate.

Convergence with interview - positive predictions

It has been suggested that more confidence may be held in an assessment where two or more methods agree (Durand and Crimmins, 1990). Moreover, a number of studies have used convergent validity to infer the predictive validity of an assessment method in the absence of supporting of outcome data (e.g. Durand and Crimmins, 1988; Emerson et al. 1995).

Convergence is usually determined by assessing the proportion of topographies upon which two or more methods agree. However,

the presence of multiple predictions and differentials in the rate of non-prediction mean results may vary according to i) the criteria used to determine agreement, and ii) how base and comparison methods are decided. To overcome these difficulties a series of analyses is needed. Data from the present study were therefore subject to such a series of analyses, and a range of criteria was used to determine agreement. Interview served as the base method throughout.

Convergence with interview - all predictions in any order

The present analysis considered all functions ascribed by the base and comparison methods irrespective of order. As such, it represented the least conservative test of agreement between pairs of assessment methods. The approach is relevant, however, as it reflects most closely the way decisions might be made in a clinical context.

Method

The functions ascribed by interview were compared with those ascribed by each comparison method in turn. An agreement was scored for each topography where the comparison method ascribed the same category of reinforcement as the interview method, irrespective of how many other functions were ascribed, or how they were ranked. The proportion of topographies upon which agreement occurred was calculated as a proportion of;

i) the number of topographies for which interview ascribed function,

ii) all topographies assessed; including those for which interview failed to ascribe function, and

iii) topographies for which interview and each of the comparison methods ascribed function.

Results

Interview ascribed 182 functions to 89 topographies (mean = 2.04). The overall mean number of predictions per topography, including non-prediction, was 1.77 (based on 214 predictions for 121 topographies). The proportion of topographies upon which each comparison method ascribed one or more functions that matched those of interview is given in table 3.7.

Table 3.7
Probability of convergence with interview - all predictions any order

	ABC	MAS1	MAS2	Analogue	Desc.
Topographies excluding interview fails	89	89	89	89	89
Probability of agreement	0.29	0.75	0.77	0.31	0.40
Topographies including interview fails	121	121	121	121	121
Probability of agreement	0.46	0.79	0.81	0.38	0.33
Topographies for which both predicted	38	82	84	37	54
Probability of agreement	0.73	0.82	0.82	0.76	0.65

The first panel of table 3.6 suggests the rating scales agreed one or more predictions at a markedly higher rate than either the descriptive or experimental methods. The second panel suggests this was true, even when non-predictions by the base method were included. The third panel reveals an improved level of agreement by the experimental and descriptive methods when analysis was restricted to include only those topographies for which the base and comparison methods both ascribed function. It should be noted, however, the higher levels of agreement obtained for the experimental and descriptive methods occurred for a substantially reduced number of topographies. This finding, which is consistent with Emerson et al. (1995), may be significant for two reasons. First, some agreement levels reported in the literature may be artificially inflated.

Secondly, the number of topographies selected for comparison is rarely a variable in clinical practice.

Conclusion

Rating scales were most likely to ascribe and agree the functions assigned by interview. Like interview, these methods were classified earlier as being informant-based, retrospective, indirect, and leaving the environment intact. Experimental and descriptive methods, on the other hand, obtained an improved rate of convergence only when all non-predictions were removed from the analysis. Thus, the lower rate of convergence obtained by these methods may be explained partially by their relative inability to ascribe functions for the same topographies as interview. Once this had been controlled for convergence with interview was much closer to that of the rating scale.

Convergence with interview - by category of function

A later analysis will show how 29% of all functions assigned by the interview method supported a social attention hypothesis. The next highest proportion of functions assigned was task-avoidance (20%), followed by social-avoidance (18%), automatic reinforcement (18%) and tangible reinforcement (15%). Given social attention accounted for a larger proportion of all assigned functions it appeared possible that agreement by the rating scales had loaded onto this category to inflate overall levels of agreement. To examine this possibility, data were re-analysed by category of function.

Method

Functions ascribed by each assessment method were reviewed. Agreement between the base and comparison method was identified and assessed for each topography. Overall agreement was calculated and expressed as a proportion of the number of times interview ascribed from each reinforcement category.

Results

Table 3.8 suggests the rating scales were more likely than the other methods to agree the function assigned by interview across all categories of function. Agreement was highest on the task-avoidance and automatic reinforcement categories and not social attention.

Table 3.8
Probability of agreement between interview and comparison methods by function category

Function	# INT	ABC	MAS1	MAS2	Analogue	Desc.
task escape-avoidance	36	0.28	0.69	0.64	0.36	0.22
social escape-avoidance	33	0.12	0	0.24	0.24	0.18
social attention	53	0.26	0.32	0.38	0.13	0.25
tangible reinforcement	28	0.14	0.54	0.36	0.18	0.11
automatic reinforcement	32	0.09	0.72	0.69	0.13	0.34
Failed	32	0.94	0.91	0.91	0.59	0.13

As the original rating scale (MAS1) did not consider social escape-avoidance independently of task escape-avoidance, the level of agreement on this domain was zero. The revised version (MAS2), however, agreed a quarter of the social escape-avoidance functions ascribed by the interview method.

Conclusion

Except for the category of social-avoidance, the rating scales agreed all categories of reinforcement more often than either

the experimental or descriptive methods. Agreement did not load onto social attention or any other reinforcement class.

Convergence with interview - primary predictions

Sometimes functional assessment will suggest one category of reinforcement, other times multiple predictions will be made. Where multiple predictions are made, it is likely that one reinforcement category will emerge as being relatively more influential. In clinical terms the most influential (primary) variable would most likely be the one selected for manipulation. Secondary predictions may have a modifying effect or be held in temporary abeyance. Given the relative importance of the primary prediction convergence was examined on the ascribing of primary function alone.

Method

The functions ascribed by each assessment were ordered first to last according to the relative strength of the supporting evidence. Base and comparison methods were assessed for agreement on the function each assigned as primary. An agreement was counted whenever the two assessment methods ascribed an identical primary function for the same topography, irrespective of how many additional functions were assigned. The proportion of topographies upon which agreement occurred was calculated for each comparison method based upon;

i) the number of topographies for which interview ascribed function,

ii) all topographies, including those for which interview failed to ascribe function, and

iii) the number of topographies for which each pair of assessments ascribed function.

Results

The first panel of table 3.9 reveals that a greater level of agreement occurred on the primary prediction of function between interview and the rating scales.

Table 3.9
Probability of convergence with interview - first function only

	ABC	MAS1	MAS2	Analogue	Desc.
Topographies excluding interview fails	89	89	89	89	89
Probability of agreement	0.18	0.50	0.52	0.2	0.18
Topographies including interview fails	121	121	121	121	121
Probability of agreement	0.38	0.61	0.62	0.3	0.17
Topographies for which both predicted	39	82	84	38	55
Probability of agreement	0.41	0.55	0.55	0.47	0.29

This remained true when non-prediction by interview was included in the analysis. Panel three, however, suggests the ABC and analogue methods obtained a marked improvement in agreement when analysis was restricted to those topographies for which both the base and comparison methods had ascribed function. Improvement by the direct descriptive method was evident but much less marked.

Conclusion

Rating scales maintained an improved level of agreement over the descriptive and experimental methods when analysis was confined to the first (primary) prediction of function. Experimental and descriptive assessment methods showed an improved level of agreement on primary prediction when analysis was restricted to include only those topographies for which both methods made a prediction. Thus, the pattern of results

was similar to that reported for *all* predictions, although the values representing agreement were lower.

Convergence with interview - primary by category of function

Because the pattern of results was similar to that reported for all predictions, data were re-analysed across reinforcement categories.

Method

Base and comparison methods were reviewed. An agreement was scored each time the base and comparison method ascribed the same function first. Overall agreement was calculated and expressed as a proportion of the number of predictions interview made for each reinforcement category.

Results

Table 3.10 gives the probability of agreement for each comparison method. The two rating scales agreed task-avoidance and automatic reinforcement at almost twice the rate of the other methods, but agreed the tangible and social-avoidance hypotheses at rates markedly different to one another.

Table 3.10
Probability of agreement between each method and interview

Function	# INT	ABC	MAS1	MAS2	Analogue	Desc.
task escape-avoidance	26	0.27	0.5	0.62	0.31	0.12
social escape-avoidance	6	0.17	0	0.5	0	0
social attention	16	0.25	0.19	0.25	0.06	0.19
tangible reinforcement	13	0.23	0.69	0.38	0.15	0
automatic reinforcement	28	0.04	0.71	0.64	0.14	0.36
Failed	32	0.94	0.91	0.91	0.59	0.13

On only one occasion did MAS1 predict tangible reinforcement where MAS2 ascribed social-avoidance. None of the other methods showed an exceptional level of agreement on any one category.

Conclusion

The two rating scales were consistent in agreeing task-avoidance and automatic reinforcement more often than the other methods. One or other of the rating scales agreed the primacy of tangible reinforcement and social-avoidance. All comparison methods agreed primary predictions at a lower rate than *all* predictions. Agreement for the MAS1 was adversely affected by its failure to consider social-avoidance as a discrete function.

Convergence - secondary predictions of function

In clinical terms the identification of secondary functions may serve to modify the design of intervention based on an analysis of primary function. For example, scheduling social-attention as a 'reinforcer' might be contraindicated where social-avoidance had been identified as a secondary function. Convergence on secondary predictions may also indicate of the level of precision obtained in assessment. Secondary predictions were therefore identified, and subject to analysis of agreement.

Method

Base and comparison methods were reviewed. Secondary predictions were identified and compared. An agreement was counted each time the comparison method ascribed the same function for a given topography as the base method,

irrespective of where it ranked, or how many additional functions were predicted. Proportional agreement was calculated based upon i) the number of topographies for which interview made a multiple prediction, and ii) the number of topographies for which each pair of methods made a multiple prediction.

Results

Interview ascribed 92 functions to 64 topographies in the following combination;

Combination	topographies	predictions
Primary + 1 secondary	43	43
Primary + 2 secondary	14	28
Primary + 3 secondary	7	21
Total secondary	64	92

The mean number of predictions per topography was 1.43. The overall probability of agreement between base and comparison methods is summarised in table 3.11.

Table 3.11
Probability of convergence with interview - secondary function only

	ABC	MAS1	MAS2	Analogue	Desc.
Topographies for which interview predicted multiple control	64	64	64	64	64
Probability of agreement	0.03	0.08	0.2	0.11	0.05
Topographies for which both methods predicted multiple control	15	25	24	15	14
Probability of agreement	0.13	0.16	0.54	0.46	0.21

The number of topographies upon which base and comparison methods each suggested multiple control did not exceed twenty-five (40%). Not surprisingly, therefore, agreement on the nature of secondary reinforcement was extremely low. Convergence was infrequent, however, even when analysis was

restricted to the topographies for which the base and comparison methods each predicted multiple control. This finding suggests a lack of precision between methods in the ascribing of secondary function.

Conclusion

Convergence on the secondary prediction of function was poor for all comparison methods. Differences between the classes of assessment method discussed in chapter one were not apparent.

Convergence by category of function - secondary prediction

Given the poor level of convergence across all assessment methods, no data are presented on agreement by category of reinforcement.

Convergence with interview - relative ranking

Where multiple predictions occur, it is often desirable to know exactly what is the relative influence of the reinforcement categories ascribed. The ordering of multiple predictions does, of course, require a high degree of precision in assessment. Poor rates of convergence for secondary predictions, reported above, suggest the required level of precision may not have been achieved. However, to confirm this two sets of analyses were conducted to examine agreement on the relative ranking of assessment predictions. Analyses differed only in the stringency of the criteria used to determine agreement. In the first analysis, agreement was based on the number of times each comparison method agreed with the base method. So, for example, if the base method ascribed 'A-B-C', then 'A' would agree at the order of one; 'A-B' at the order of two; and 'A-B-C' at three. This analysis assumed assessment

methods would be capable of agreeing a proportion of ascribed functions and their relative influence. In the second analysis, agreement was determined by the predictions of the base method. Thus, to match 'A-B-C' a comparison method would also need to ascribe 'A-B-C'. This second analysis assumed assessment methods would be capable of placing the same number and categories of function in the same order.

Method

Assessment results were ordered first to last, according to the strength of supporting evidence. Where two or more predictions ranked equally, they were placed in the following predetermined order; task-avoidance, social-avoidance, social attention, tangible reinforcement, automatic reinforcement. The number of topographies upon which the base method ascribed single and multiple control was identified and the number of agreements determined for each comparison method. In the first analysis an agreement was counted when, for a given topography, the functions ascribed were identical and arranged in the same order. In the second analysis the functions ascribed had to be identical, arranged in the same order and extend to the same number. Single function predictions were classified as agreeing at an order of one prediction and multi-function predictions at an order of two, three or four.

Results

Data presented in table 3.12 derive from the first analysis. The rating scales agreed with interview at the order of one prediction for about half of the 89 predictions made. Other methods averaged agreement for approximately one topography in

five. There were no differences between comparison assessment methods at the order of two, three or four predictions. Agreement on the 64 topographies for which interview ascribed multiple control was virtually non-existent.

Table 3.12
Probability of agreement: number of ranks determined by comparison method

Number of ranks	INT	ABC	MAS1	MAS2	Analogue	Desc.
One	89	0.18	0.50	0.52	0.2	0.18
Two	43	0.02	0.05	0.13	0	0
Three	14	0.07	0	0	0	0
Four	7	0	0	0	0	0

Table 3.13
Probability of agreement: number of ranks determined by base method

Number of ranks	INT	ABC	MAS1	MAS2	Analogue	Desc.
One	25	0.21	0..46	0.29	0.21	0..29
Two	43	0	0.04	0.17	0	0
Three	14	0.07	0	0	0	0
Four	7	0	0	0	0	0
Overall	89	0.28	0.5	0.46	0.21	0.29

In the second analysis (table 3.13) agreement beyond the order of one prediction was, for all practical purposes, non-existent.

Conclusion

Beyond the order of one prediction, convergence on the relative ranking of assessment predictions was virtually non-existent. This finding was anticipated, given the low rate of convergence reported earlier for secondary predictions of function.

Rating scales were no more likely to agree with interview than were the experimental or descriptive methods. Given interview made a high proportion of multiple predictions, more than any

of the comparison methods, even these mediocre levels of convergence may have been artificially inflated. Overall, the evidence tends to suggest that each assessment method predicted single and multiple function for dissimilar sets of topographies, and that the composition of multi-function complexes was different for a substantial proportion of topographies.

Conclusions regarding the relative ranking of predictions must, however, be regarded cautiously. Firstly, the present study provided no data on the validity of the functions ascribed by any assessment method. Little is known about the predictive validity of the interview method generally. The proportion of false positives among the multiple predictions by the interview method cannot, therefore, be determined. This factor may be especially relevant here, since the interview method made substantially more multiple predictions than the comparison methods. If the proportion of false positives was relatively high for the interview method, then the comparison methods could have been 'right' in failing to agree the sequence of functions it suggested. Secondly, the method used to order assessment predictions may not have been reliable. The relative density of a reinforcement schedule does not, for example, necessarily equate with relative reinforcer potency. It does seem unlikely, however, that a clinician would ignore this dimension entirely. The method used may therefore be considered to have face validity. Thirdly, variance in the ratio of single to multiple predictions reported earlier, would have pre-disposed the comparison methods to agree at a low rate beyond the order of one prediction.

Despite these limitations, however, the exceptionally low rate of convergence on the rank ordering of multiple function suggests a boundary in the precision of functional assessment methodologies. Present results suggest functional assessment may be of limited practical value when taken beyond the ascribing of primary function.

V. CONVERGENCE - EXCLUDING ALL NON-PREDICTION

Thus far, analyses have included all topographies shown by all subjects. It is clear, however, that the rate of non-prediction varied considerably between assessment methods. Emerson et al. (1995) reported convergence on 85% of topographies when analysis was restricted to those behaviours for which the two methods made a prediction. In the present study, account was taken of non-prediction by calculating agreement with and without assessment failure. This approach, however, resulted in assessment methods being compared on a variable number of topographies. In the following analyses the number of topographies was equalised by restricting comparison to those behaviours for which all six methods made a prediction. The number of topographies for which all six methods ascribed function was sixteen (13.2%). Predictions covered seven classes of topography shown by nine subjects;

Topography	Number
Aggression	6
Vocalisations	4
SIB	2
Damage	1
Pica	1
Stereotypy	1
Multiple	1
Total	16

Interview predicted a single function for four topographies, and made multiple predictions for twelve, in combinations of two (8), three (3), and four (1) ascribed functions.

Method

Assessment data were subject to analysis following the methods described earlier. Analyses were conducted to include all functions ascribed, primary functions, secondary functions, and the relative ranking of multiple predictions.

Results

Results for the sub-set of sixteen topographies are presented in table 3.14. Comparing these results with those obtained for the full set of 121 topographies it is apparent that agreement based on all functions ascribed in any order, improved markedly for the previously lower predicting analogue and descriptive methods.

Table 3.14
Results of analyses on topographies for which all six methods predicted

	ABC	MAS1	MAS2	Analogue	Desc.
All predictions	16	16	16	16	16
Probability of agreement	0.81	0.87	0.69	0.81	0.62
Primary predictions	16	16	16	16	16
Probability of agreement	0.5	0.5	0.5	0.43	0.31
Secondary predictions	12	12	12	12	12
Probability of agreement	0.08	0	0..08	0	0
Relative ranking of predictions	(variable)				
Probability at order of one (4)	0.5	0.5	0.25	0.25	0.25
Probability at order of two (8)	0.12	0	0	0	0
Probability at order of two (3)	0.33	0	0	0	0
Probability at order of three (1)	0	0	0	0	0

Thus, differences between classes of informant-based, experimental and descriptive method, were less obvious for the sub-set upon which all methods ascribed. This adds further weight to the suggestion that variation occurred in the identification and selection of topographies as well as the functions ascribed.

Results of the present analyses followed the trends identified in previous analyses. Agreement on primary predictions ranged from approximately one-third to one-half of topographies. Agreement on secondary and rank-ordering of multiple predictions was extremely low.

Conclusion

When all non-predictions were removed from the analysis the comparison methods agreed the functions ascribed by interview at a broadly similar rate. Trends in the data were similar to those reported earlier for the full behaviour set, with agreement occurring infrequently on the secondary and relative ranking of predictions. These findings are consistent with the suggestion that lower rates of convergence may be associated with assessment methods' tendencies to predict function for different groups of topographies. However, while in the present analysis agreement was observed to improve, it covered just 13% of all topographies.

VI. AGREEMENT BETWEEN ASSESSMENT METHODS

Thus far, convergence has been assessed with the functions assigned by interview. Analysis now turns to consider the

extent to which each assessment method would be likely to obtain convergence on the functions it ascribed.

Method

Each assessment served as the based method in turn. Agreement was calculated allowing all ascribed functions into the analysis in any order. The number of topographies upon which agreement occurred was determined for each pair of assessment methods and expressed as a proportion of all possible agreements for;

i) the number of topographies in the full set,

ii) the number of topographies for which the base method predicted,

iii) the number of topographies for which both methods predicted, and

iv) the number of functions assigned by each method.

Results

Table 3.15 summaries the overall probability of agreement between each base method and all comparison methods. The relevant number of topographies appears in square brackets. Results for assessment pairs are given in appendix 3.1.

Convergence across all topographies

The mean probability of agreement across all topographies fell within the range of 0.16 to 0.37 (overall mean = 0.27). The interview and rating scale methods obtained more agreement overall than either the experimental or the descriptive methods. Not surprisingly, the two methods which ascribed at a relatively low rate obtained relatively fewer agreements.

Table 3.15
Probability of agreement between assessment methods - all functions ascribed

all predictions	base method					
comparison methods	Interview	ABC	MAS1	MAS2	Analogue	Desc.
Interview	[121]	0.23	0.55	0.57	0.23	0.29
ABC	0.23	[121]	0.17	0.17	0.1	0.12
MAS1	0.55	0.17	[121]	0.62	0.2	0.23
MAS2	0.57	0.17	0.62	[121]	0.21	0.24
Analogue	0.23	0.1	0.2	0.21	[121]	0.16
Descriptive	0.29	0.12	0.23	0.24	0.16	[121]
MEAN	**0.37**	**0.16**	**0.35**	**0.36**	**0.18**	**0.21**

base method only	Interview	ABC	MAS1	MAS2	Analogue	Desc.
Interview	[89]	0.7	0.79	0.79	0.52	0.43
ABC	0.31	[40]	0.21	0.24	0.24	0.17
MAS1	0.75	0.5	[85]	0.86	0.46	0.34
MAS2	0.78	0.52	0.88	[87]	0.52	0.34
Analogue	0.31	0.3	0.28	0.3	[50]	0.35
Descriptive	0.38	0.35	0.33	0.33	0.38	[82]
MEAN	**0.5**	**0.47**	**0.49**	**0.5**	**0.43**	**0.3**

both methods	Interview	ABC	MAS1	MAS2	Analogue	Desc.
Interview	-	0.74	0.88	0.89	0.76	0.65
ABC	0.74	-	0.57	0.58	0.63	0.44
MAS1	0.88	0.57	-	0.88	0.63	0.53
MAS2	0.89	0.58	0.88	-	0.7	0.53
Analogue	0.76	0.63	0.63	0.7	-	0.49
Descriptive	0.65	0.44	0.53	0.53	0.49	-
MEAN	**0.76**	**0.59**	**0.67**	**0.69**	**0.64**	**0.53**

Convergence across topographies selected by the base method

The overall mean level of agreement improved considerably when analysis was restricted to include only the topographies upon which each base method had ascribed function (range 0.30 to 0.50: overall mean = 0.45). There was good agreement among the informant-based methods, and moderate agreement by the experimental and descriptive methods.

Convergence across topographies selected by the base and comparison methods

When analysis was restricted to include only the topographies upon which the base and comparison methods each ascribed function, the overall mean level of agreement increased to 0.65 (range: 0.53 to 0.76). There was more agreement among the informant-based methods, but the differential between the classes of assessment method was diminished.

Convergence on the functions assigned

Previous results had indicated that the interview ascribed multiple control for a greater proportion of topographies, and ascribed more functions per topography than any other method. The effect of multiple prediction was therefore examined by re-calculating agreement as a proportion of the number of functions assigned by each method.

Results

Results are presented in table 3.16 and the number of functions assigned by each method is shown in square brackets. The mean probability of agreement on the functions assigned ranged from 0.22 to 0.40, and with the exception of the descriptive method did not vary significantly from the overall mean of 0.35.

Table 3.16
Probability of agreement between assessment methods based on the number of functions ascribed

	base method					
comparison method	Interview	ABC	MAS1	MAS2	Analogue	Desc.
Interview	[182]	0.57	0.64	0.59	0.44	0.34
ABC	0.19	[61]	0.17	0.16	0.19	0.08
MAS1	0.44	0.34	[124]	0.64	0.35	0.24
MAS2	0.46	0.38	0.72	[139]	0.37	0.23
Analogue	0.2	0.26	0.23	0.22	[83]	0.2
Descriptive	0.23	0.16	0.23	0.2	0.29	[120]
MEAN	**0.3**	**0.34**	**0.4**	**0.36**	**0.33**	**0.22**

Chance agreement

A proportion of agreement could be expected to occur by chance alone. A reliable estimate of agreement by chance is needed to aid interpretation of actual agreement found in the study. A chance agreement statistic was calculated, therefore, for i) all functions assigned, ii) functions assigned as primary, and iii) functions assigned as secondary.

Method

A chance agreement statistic was calculated for every combination of assessment method within each reinforcement category and these are given in appendix 3.2. The values obtained were averaged to yield for each method a mean probability of agreement by chance. The category means were then summed to give an overall probability of agreement by chance for each assessment method. Thus, for example, if for the interview method the simple probability of task-avoidance being ascribed was 0.29, and for the ABC method it was 0.35, then the chance probability of both methods ascribing task-avoidance for the same topography would be given by 0.29 x 0.35 = 0.10. Repeating the calculation for interview and each of the other assessment methods would yield a series of chance probabilities, from which could be derived an overall mean for the task-avoidance category;

method:			interview		chance
abc	0.35	X	0.29	=	0.1
mas1	0.27	X	0.29	=	0.08
mas2	0.37	X	0.29	=	0.11
analogue	0.5	X	0.29	=	0.15
descriptive	0.12	X	0.29	=	0.03
mean	0.32	X	0.29	=	0.09

Following the same procedure for each reinforcement category would yield a series of means for the interview method;

Task avoidance	= 0.09
Social avoidance	= 0.01
Social attention	= 0.03
Tangible reinforcement	= 0.03
Automatic reinforcement	= 0.07

the sum of which would be equal to the overall mean probability of agreement by chance (interview mean = 0.23).

Results

Table 3.17 summarises the results obtained for each assessment method.

Table 3.17
Mean probability of agreement by chance alone

	INT	ABC	MAS1	MAS2	ANA	DESC	MEAN
All ascribed functions	0.2	0.21	0.21	0.22	0.2	0.2	0.21
Primary functions	0.23	0.21	0.23	0.24	0.24	0.19	0.22
Secondary functions	0.46	0.39	0.31	0.36	0.29	0.26	0.35

Data relating to all ascribed functions suggest that agreement could have occurred by chance alone for approximately one topography in every five assessed, and that a similar level could be expected for primary predictions of function also. Chance agreement was, however, much higher for secondary predictions of function, at a little over one topography in every three assessed.

Actual agreement

Overall mean levels of agreement were presented earlier, and these are re-tabulated here for ease of reference in table 3.18 below.

Table 3.18
Mean probability of agreement for each assessment method

Assessment method	all topographies assessed	base method only	base + comparison methods	number of functions ascribed
Interview	0.37	0.5	0.76	0.3
ABC charts	0.16	0.47	0.59	0.34
MAS1	0.35	0.49	0.67	0.4
MAS2	0.36	0.5	0.69	0.36
Analogue	0.18	0.43	0.64	0.33
Desc	0.21	0.3	0.53	0.22
Overall	**0.27**	**0.45**	**0.65**	**0.32**

These data describe the mean level of agreement obtained by each assessment method across the five other methods, and are shown as a proportion of all topographies assessed, the topographies upon which the base method alone ascribed function, the topographies upon which the base and each comparison method ascribed function, and the number of functions each method ascribed.

After excluding all non-predictions from the analysis agreement between assessment methods occurred on average for two topographies in every three assessed; a clinically significant rate three times greater than the level reported for chance alone. When analysis was based on the functions ascribed by each base method the overall mean rate was just over twice the chance level of agreement at 0.45. However, when analysis included non-prediction the overall mean rate of convergence drew close to chance at 0.27, or a little over one topography in every four assessed. Moreover, when the overall rate of convergence was assessed based on the number of functions that were ascribed agreement was limited to one prediction in every three; a level equivalent to chance.

Conclusion

Acceptable levels of agreement occurred only when analysis was restricted to include topographies upon which the base and comparison method each ascribed function. This suggests failure to ascribe was a major factor influencing agreement. This finding has important implications for the selection of the most appropriate assessment methods.

VII. CHAPTER SUMMARY OF RESULTS

A function was assigned by one assessment method or more for a 98% of all assessed topographies. None of the methods ascribed for more than 73% of topographies, however, and all six methods ascribed for only 13% of topographies. The experimental and ABC chart methods ascribed for the smallest number of topographies and the interview ascribed for the highest. The interview method also ascribed multiple control more times than any other method, and assigned more functions to topographies where multiple-control was indicated.

The rating scales agreed proportionately more of the functions assigned by interview than either the experimental or the descriptive methods. Agreement on secondary function, and the ordering of multiple prediction was, however, universally low.

When assessment methods were compared directly agreement was found to vary according to the criteria applied in analysis. Overall, present results suggest that agreement should not be expected for more than one topography in every four assessed. However, where two or more methods ascribe function for the same behaviours agreement may be expected to average two

topographies in every three assessed, but if multiple function complexes are predicted agreement may be limited to just one function in three.

Chapter Four

RESULTS

The further comparison

of

functional assessment methods

I. INTRODUCTION

Chapter three assessed convergence on the functions ascribed by six assessment methods to 121 topographies shown by twenty individuals with severe learning disabilities. Agreement was moderately good for primary prediction function, when analysis was restricted to include only those topographies upon which the base and comparison methods both ascribed function. Agreement on secondary and multiple predictions was not strong, and was generally depressed by differential rates of non-prediction. No single assessment method ascribed function for more than 73% of topographies, and yet a function was ascribed by one assessment method or more for 98% of all topographies. Why then did some assessment methods ascribe function where other methods did not? And, why should there have been such divergence in the functions ascribed?

The present chapter addresses three specific questions. First, were assessment methods more likely to assign function from a specific category of reinforcement? Second, did assessment methods assign function differentially according to the class of topography assessed? Third, were assessment methods more likely to assign function for behaviour shown by particular individuals?

Accordingly, assessment methods were examined for tendencies to ascribe functions from a particular reinforcement category, or for a specific class of topography. The way in which assessment methods assigned functions to topographies was re-examined on an person by person basis.

II. ASSESSMENT METHOD AND FUNCTIONS ASCRIBED

The aim of the present analysis was to establish whether assessment methods were more likely to ascribe function from one or more categories of reinforcement, and if so whether this could have contributed to the levels of disagreement reported earlier.

Method

Each assessment was designated as the base method in turn. The rate at which the base assessment method ascribed from each reinforcement category was determined and compared with the overall mean rate of the comparison methods. Analyses were conducted for;

- all functions ascribed,
- functions ascribed as primary,
- functions ascribed as secondary,
- all functions excluding non-predictions.

The ordering of multiple functions was preserved for all assessment methods and the probability of each function being ascribed compared.

Results

All predictions

A different assessment method was identified as being most and least likely to ascribe from each reinforcement category (table 4.1). MAS1 ascribed tangible reinforcement at almost twice the mean rate of the comparison methods (0.15), and did not ascribe social-avoidance at all. The ABC method ascribed social attention at approximately 2.3 times the mean rate of the comparison methods (0.19), and was least likely to ascribe automatic or tangible reinforcement. Analogue assessment

ascribed task-avoidance at approximately 1.5 times the mean rate of all other methods (0.25), and was the method least likely to ascribe social attention (see appendix 4.1).

Table 4.1
Probability of ascribing function - all predictions

function	INT	ABC	MAS1	MAS2	Analogue	Desc.
number ascribed	182	61	124	139	83	120
task avoidance	0.2	0.31	0.29	0.28	0.38	0.15
social avoidance	0.18	0.1	0	0.13	0.21	0.22
social attention	0.29	0.44	0.15	0.19	0.14	0.2
tangible reinforcement	0.15	0.1	0.28	0.17	0.14	0.17
automatic reinforcement	0.18	0.05	0.27	0.23	0.11	0.27

The descriptive method was more likely to ascribe social-avoidance and automatic reinforcement, and less likely to ascribe task-avoidance.

Primary predictions

Table 4.2 gives the probability of each assessment method ascribing primary function by category of reinforcement.

Table 4.2
Probability of ascribing function - primary prediction

function	INT	ABC	MAS1	MAS2	Analogue	Desc.
number ascribed	89	40	85	87	50	82
task avoidance	0.29	0.35	0.27	0.37	0.5	0.12
social avoidance	0.07	0.1	0	0.07	0.13	0.2
social attention	0.18	0.38	0.08	0.11	0.12	0.22
tangible reinforcement	0.15	0.15	0.29	0.16	0.14	0.12
automatic reinforcement	0.31	0.03	0.35	0.29	0.19	0.34

ABC charts ascribed social attention for one-third of topographies; approximately five times the rate of MAS1 and 2.7 times the mean of all other methods (0.14). Analogue assessment was the method most likely to ascribe task-avoidance as a primary function. Half the functions ascribed

by this method were task-avoidance. This was four times the rate of the descriptive method, and almost twice the mean of all other methods (0.28). Descriptive methods predicted social-avoidance most often, and MAS1 tangible reinforcement. The descriptive method was eleven times more likely to ascribe automatic reinforcement than the ABC method, and MAS1 ascribed tangible reinforcement at over twice the mean of all other methods (0.14).

Secondary predictions

The probability of each assessment method ascribing secondary function is given by reinforcement category in table 4.3. MAS1 was three times more likely than interview to ascribe task-avoidance. The analogue and interview methods each ascribed social-avoidance at twice the mean rate of all other methods (0.20).

Table 4.3
Probability of ascribing function - secondary prediction

function	INT	ABC	MAS1	MAS2	Analogue	Desc.
number ascribed	64	16	32	41	31	30
task avoidance	0.16	0.31	0.41	0.17	0.23	0.3
social avoidance	0.42	0.13	0	0.32	0.42	0.33
social attention	0.58	0.81	0.38	0.39	0.19	0.2
tangible reinforcement	0.23	0	0.31	0.24	0.19	0.33
automatic reinforcement	0.05	0.13	0.16	0.17	0.06	0.13

The ABC method ascribed social attention at approximately four times the rate of the analogue method, and at 2.3 times the mean rate of other methods (0.35). A different assessment method was most and least likely to ascribe secondary function from each reinforcement category.

Conclusions

Results are further summarised in figure 4.1 which identifies the assessment method most likely to ascribe from each reinforcement category (see appendix 4.2). Analogue assessment was the method most likely to assign task-avoidance.

Figure 4.1
Method most likely to ascribe function

	All functions	Primary function	Secondary function
task avoidance	Analogue	Analogue	MAS1
social avoidance	Descriptive	Descriptive	Analogue: Descriptive
social attention	ABC	ABC	ABC
tangible reinforcement	MAS1	MAS1	Descriptive
automatic reinforcement	MAS1 : Descriptive	MAS1	MAS2

The MAS and descriptive methods were most likely to identify automatic reinforcement, while the descriptive method alone was more likely to predict social-avoidance. The ABC chart method was consistently more likely to ascribe social attention over any other function.

Such variability raises serious questions about the validity of the functions ascribed. It was clear, however, from analyses presented in chapter three, that assessment methods did not ascribe function for identical sub-sets of topographies. Further analysis was required, therefore, to examine the impact of non-prediction on assessment methods' tendencies to ascribe.

Excluding non-predictions

Method

The 16 topographies for which all six assessments ascribed function were identified and subject to analyses described in the preceding section. Assessment methods which ascribed a reinforcement category most frequently in the present analysis

were compared with those reported in the preceding section for the whole set of 121 topographies.

Results

Results from each set of analyses are presented in figure 4.2. Taking all ascribed functions into account, analogue assessment, ABC charts and the rating scales, were the methods most likely to ascribe task-avoidance, social attention and tangible reinforcement respectively.

Figure 4.2
Summary of results - main behaviour set

	MOST LIKELY TO PREDICT		LEAST LIKELY TO PREDICT	
All	whole-set	sub-set	whole-set	sub-set
task-avoidance	Analogue	Analogue	Descriptive	Interview
social-avoidance	Descriptive	MAS1	MAS1	MAS1
social attention	ABC	ABC	Analogue	MAS2
tangible	MAS1	MAS2	ABC	ABC
automatic	Descriptive:MAS1	Interview	ABC	Analogue
Primary				
task-avoidance	Analogue	MAS2	Descriptive	Interview:ABC
social-avoidance	Descriptive	MAS2	MAS1	MAS1
social attention	ABC	ABC	MAS1	MAS2
tangible	MAS1	MAS2	Descriptive	ABC
automatic	MAS1	Interview	ABC	Analogue
Secondary				
task-avoidance	MAS1	MAS1	Interview	Interview
social-avoidance	Analogue:Interview	MAS2	MAS1	MAS1
social attention	ABC	ABC	Analogue	MAS2
tangible	Descriptive	MAS1	ABC	ABC
automatic	MAS2	MAS2	Interview	Analogue:Descritpive

In every analysis the ABC chart was the method most likely to ascribe social attention. Rating scales were more likely to ascribe tangible reinforcement as serving a primary, but not necessarily a secondary, function. The ABC chart was least

likely to ascribe tangible reinforcement in every analysis bar one.

Conclusion

The tendency for analogue, ABC recording and rating scales to assign task-avoidance, social attention and tangible reinforcement, did not alter when analysis was restricted to the 16 topographies for which all assessment ascribed function. Each of these assessment methods, appeared therefore, to show a proclivity for ascribing from a particular category of function, which could not be accounted for by differentials in the topographies selected for prediction.

III. ASSESSMENT METHOD, FUNCTION AND TOPOGRAPHY

Chapter three described how the number of topographies for which a function was ascribed varied between assessment methods within a range of 33% for the ABC charts to 73% for clinical interview. It would be helpful to know, especially in a clinical context, whether the effect of non-prediction was generalised or specific to certain topographies. In the following analyses the functions ascribed by each assessment method were analysed in relation to the classes of topography assessed.

Method

Topographies were arranged to form six discrete classes based on prevalence (table 4.4). A miscellaneous class comprised topographies displayed by < 25% of the cohort. The probability of each assessment method ascribing function was calculated for each class of topography, together with the distribution of the

functions assigned. Data were analysed for all, primary and secondary predictions of function.

Table 4.4
Classes of assessed topography

Class of topography		Number
Aggression directed at others		21
Damage to property		19
Negative vocalisations		18
Self-injurious behaviour		17
Stereotypy		15
Miscellaneous:		31
Pica	5	
Stripping	4	
Verbal abuse	3	
Vomiting	3	
Faecal smearing	3	
Intentional incontinence	3	
Obsessive behaviour	2	
Falling to floor	2	
Non-compliance	1	
Poking rectum	1	
Running around (in bouts)	1	
Spitting	1	
Public masturbation	1	
Hyperactivity	1	
Total	31	121

Results

All predictions

Table 4.5 gives the probability of each assessment method ascribing one or more functions for each class of topography, and describes how the ascribed functions distributed between the available classes of reinforcement. Tables 4.6 and 4.7 cover primary and secondary functions respectively. Highlighted in bold are the methods which ascribed for the greatest number of topographies in each class.

Table 4.5
Function assigned by assessment method and topography class - all functions

		Probability					
Topography class	Method	**Any**	**task-av**	**soc-av**	**soc-attn**	**tangible**	**auto**
Aggression	**Clinical interview**	1	0.52	0.62	0.57	0.38	0.1
	ABC charts	0.48	0.6	0.2	0.8	0.1	0
	MAS1	1	0.43	0	0.24	0.33	0.1
	MAS2	1	0.43	0.48	0.24	0.33	0.1
	Analogue	0.57	0.83	0.5	0.17	0.25	0
	Descriptive	0.52	0.27	0.27	0.64	0.18	0.09
Damage to property	Clinical interview	0.53	0.6	0.4	0.4	0.4	0.1
	ABC charts	0.16	0.22	0.22	0	0.11	0
	MAS1	0.53	0.5	0	0.3	0.3	0.4
	MAS2	0.53	0.4	0.1	0.3	0.3	0.3
	Analogue	0.42	0.25	0.38	0.38	0.13	0.25
	Descriptive	**0.79**	0	0.13	0.33	0.53	0.33
Negative vocalisation	Clinical interview	0.61	0.45	0.27	0.73	0.73	0.27
	ABC charts	0.5	0.44	0.11	0.78	0.22	0.11
	MAS1	0.61	0.82	0	0.18	0.64	0.18
	MAS2	0.61	0.91	0.18	0.27	0.45	0.18
	Analogue	0.5	1	0.33	0.22	0.33	0
	Descriptive	**0.83**	0.53	0.53	0.2	0.27	0.33
Self-injury	Clinical interview	0.71	0.58	0.5	0.67	0.17	0.33
	ABC charts	0.53	0.56	0.22	0.56	0.22	0.11
	MAS1	0.76	0.38	0	0.15	0.69	0.23
	MAS2	0.76	0.62	0.23	0.31	0.38	0.23
	Analogue	0.35	1	0.5	0	0.17	0.17
	Descriptive	**0.82**	0.29	0.43	0.14	0.14	0.43
Stereotypy	Clinical interview	0.6	0	0.22	0.22	0	1
	ABC charts	0.13	0	0	1	0	0.5
	MAS1	0.73	0.09	0	0.18	0	0.82
	MAS2	0.73	0.18	0.09	0.36	0	0.82
	Analogue	0.47	0	0.14	0.57	0.29	0.71
	Descriptive	**1**	0.07	0.13	0.07	0.13	0.87
Miscellaneous	**Clinical interview**	**0.84**	0.27	0.23	0.73	0.23	0.42
	ABC charts	0.23	0.29	0.14	0.71	0	0
	MAS1	0.68	0.24	0	0.14	0.24	0.57
	MAS2	0.74	0.26	0.04	0.3	0.17	0.57
	Analogue	0.26	0.63	0.25	0.13	0.25	0.13
	Descriptive	0.39	0.17	0.42	0.5	0.17	0.17

Table 4.6
Function assigned by assessment method and topography class - primary functions

		Probability					
Topography class	Method	**Any**	**task-av**	**soc-av**	**soc-attn**	**tangible**	**auto**
Aggression	**Clinical interview**	**1**	0.33	0.24	0.19	0.14	0.1
	ABC charts	0.48	0.4	0.2	0.3	0.1	0
	MAS1	**1**	0.38	0	0.19	0.38	0.05
	MAS2	**1**	0.43	0.29	0.1	0.19	0
	Analogue assessment	0.57	0.75	0.08	0	0.17	0
	Descriptive	0.52	0.09	0.27	0.45	0.09	0.09
Damage to property	Clinical interview	0.53	0.4	0.1	0.2	0.2	0.1
	ABC charts	0.16	0.67	0	0	0.33	0
	MAS1	0.53	0.3	0	0.1	0.2	0.4
	MAS2	0.53	0.3	0	0.2	0.3	0.2
	Analogue assessment	0.42	0.25	0.13	0.37	0.13	0.13
	Descriptive	**0.79**	0	0.07	0.27	0.33	0.33
Negative vocalisation	Clinical interview	0.61	0.45	0	0.18	0.27	0.09
	ABC charts	0.5	0.33	0.11	0.33	0.22	0
	MAS1	0.61	0.45	0	0	0.45	0.09
	MAS2	0.61	0.64	0	0	0.27	0.09
	Analogue assessment	0.5	0.78	0	0	0.22	0
	Descriptive	**0.83**	0.4	0.2	0.07	0.13	0.2
Self-injury	Clinical interview	0.71	0.42	0	0.25	0.08	0.25
	ABC charts	0.53	0.33	0	0.44	0.22	0
	MAS1	0.76	0.31	0	0	0.46	0.23
	MAS2	0.76	0.54	0	0.08	0.15	0.23
	Analogue assessment	0.35	0.67	0.17	0	0	0.17
	Descriptive	**0.82**	0.07	0.36	0.14	0	0.43
Stereotypy	Clinical interview	0.6	0	0	0	0	1
	ABC charts	0.13	0	0	0.5	0	0.5
	MAS1	0.6	0	0	0	0	1
	MAS2	0.6	0.11	0	0.11	0	0.78
	Analogue assessment	0.47	0	0.14	0.29	0	0.57
	Descriptive	**1**	0	0.07	0.07	0.07	0.8
Miscellaneous	**Clinical interview**	**0.84**	0.19	0	0.19	0.15	0.46
	ABC charts	0.23	0.29	0.14	0.57	0	0
	MAS1	0.68	0.14	0	0.1	0.19	0.57
	MAS2	0.74	0.22	0	0.17	0.09	0.52
	Analogue assessment	0.26	0.37	0.13	0.13	0.25	0.13
	Descriptive	0.39	0.17	0.25	0.42	0.08	0.08

Table 4.7

Function assigned by assessment method and topography class - secondary functions

		Probability					
Topography class	Method	**Any**	**task-av**	**soc-av**	**soc-attn**	**tangible**	**auto**
Aggression	**Clinical interview**	**0.86**	0.22	0.44	0.44	0.28	0
	ABC charts	0.24	0.4	0	1	0	0
	MAS1	0.48	0.3	0	0.3	0.3	0.3
	MAS2	0.48	0	0.4	0.3	0.3	0.2
	Analogue assessment	0.33	0.14	0.71	0.29	0.29	0
	Descriptive	0.14	0.67	0	0.67	0.33	0
Damage to property	**Clinical interview**	**0.26**	0.4	0.6	0.4	0.4	0
	ABC charts	0	0	0	0	0	0
	Analogue assessment	0.21	0.5	0	0.5	0.25	0
	MAS2	0.21	0.25	0.25	0.25	0	0.25
	Analogue assessment	0.16	0	0.67	0	0	0.33
	Descriptive	**0.26**	0	0.2	0.2	0.6	0
Negative vocalisation	Clinical interview	0.44	0	0.38	0.75	0.63	0.25
	ABC charts	0.28	0.2	0	0.8	0	0.2
	MAS1	0.33	0.67	0	0.33	0.33	0
	MAS2	0.44	0.38	0.25	0.38	0.25	0.13
	Analogue assessment	0.39	0.29	0.43	0.29	0.14	0
	Descriptive	**0.56**	0.3	0.5	0.2	0.2	0.2
Self-injury	**Clinical interview**	**0.59**	0.2	0.6	0.5	0.1	0.1
	ABC charts	0.24	0.5	0.5	0.25	0	0.25
	MAS1	0.29	0.2	0	0.4	0.6	0
	MAS2	0.47	0.13	0.5	0.38	0.38	0
	Analogue assessment	0.24	0.5	0.5	0	0.25	0
	Descriptive	0.29	0.6	0.2	0	0.4	0
Stereotypy	Clinical interview	0.2	0	0.33	0.67	0	0
	ABC charts	0.07	0	0	1	0	0
	MAS1	0.13	0.5	0	1	0	0
	MAS2	0.27	0.25	0.25	0.75	0	0.5
	Analogue assessment	**0.33**	0	0	0.4	0.4	0.2
	Descriptive	0.2	0.33	0.33	0	0.33	0.33
Miscellaneous	**Clinical interview**	**0.65**	0.1	0.3	0.7	0.1	0
	ABC charts	0.03	0	0	1	0	0
	MAS1	0.16	0.4	0	0.2	0.2	0.2
	MAS2	0.23	0.14	0.14	0.43	0.29	0.14
	Analogue assessment	0.1	0.67	0.33	0	0	0
	Descriptive	0.13	0	0.5	0.25	0.25	0.25

Interview assigned functions for all topographies classed aggressive, and for the highest proportion of those classed miscellaneous (table 4.5). Descriptive methods assigned functions for all topographies classed stereotypic, and for the highest proportion of those classed damage to property, negative vocalisation and self-injury.

Interview assigned positive and negative reinforcement in almost equal proportion for all behaviour in the aggressive class. Social-positive and automatic reinforcement were assigned more often for the miscellaneous class. The two rating scales ascribed function in a similar way for both the aggressive and miscellaneous classes of topography. The revised rating scale, however, ascribed social-avoidance for almost half the topographies in the aggressive class, whereas the original did not. The ABC method ascribed social attention most often for all topographies other than those classified as causing damage to property. The analogue method ascribed negative reinforcement most often for all classes of topography other than damage to property and stereotypy. The descriptive method assigned automatic reinforcement to stereotypy in 87% of cases, tangible and automatic reinforcement most frequently for damage to property, social attention and social-avoidance for negative vocalisations, and social-avoidance together with automatic reinforcement for self-injury.

Primary predictions

Interview assigned primary function for all topographies classed aggressive, and for the highest proportion of those classed miscellaneous (table 4.6). Descriptive methods

assigned functions for all topographies in the stereotypy class, and for the highest proportion of those remaining.

Interview and MAS2 assigned primary function in ratios similar one-another. MAS1 differed in the ascribing of tangible reinforcement and social-avoidance. The ABC method rarely identified social-avoidance as serving a primary function, and except for damage to property, ascribed social attention for between 33% and 57% of topographies. The analogue method ascribed task-avoidance for a high proportion of topographies classed aggressive, negative vocalisation and self-injury. The descriptive method assigned social attention most often for behaviour in the aggressive class. Task-avoidance was assigned most often for behaviours classed negative vocalisation and self-injury.

Secondary predictions

Interview was the method most likely to assign a secondary function for all but two classes of topography (table 4.7). Social attention and social-avoidance were ascribed in high proportion across all classes in which the interview method was dominant. At one-third, the analogue method assigned secondary function most often for stereotypy. The functions ascribed divided between automatic and two categories of positive reinforcement. The descriptive method assigned secondary function most often for negative vocalisation. Social-avoidance was ascribed in 50% of cases.

Conclusion

Assessment methods assigned functions differently according to the class of topography assessed. Informant-based methods

ascribed functions more often for topographies classed aggressive and miscellaneous. Descriptive methods assigned functions most often for stereotypy, self-injury, damage to property and negative vocalisation. The relative distribution of functions assigned was also found to vary between assessment methods across classes of topography.

IV. ASSESSMENT METHOD AND SUBJECT

Assessment methods may have systematically varied their rate of prediction and agreement across individuals. The functions ascribed by each assessment method are given by topography in appendix 4.1. Figure 4.3 summarises these data by identifying, on an individual basis, the topographies for which each assessment ascribed function. Two questions arise: First, did assessment methods fail to ascribe function for certain individuals? Second, was agreement on prediction and non-prediction constant across individuals?

Relative rates of prediction

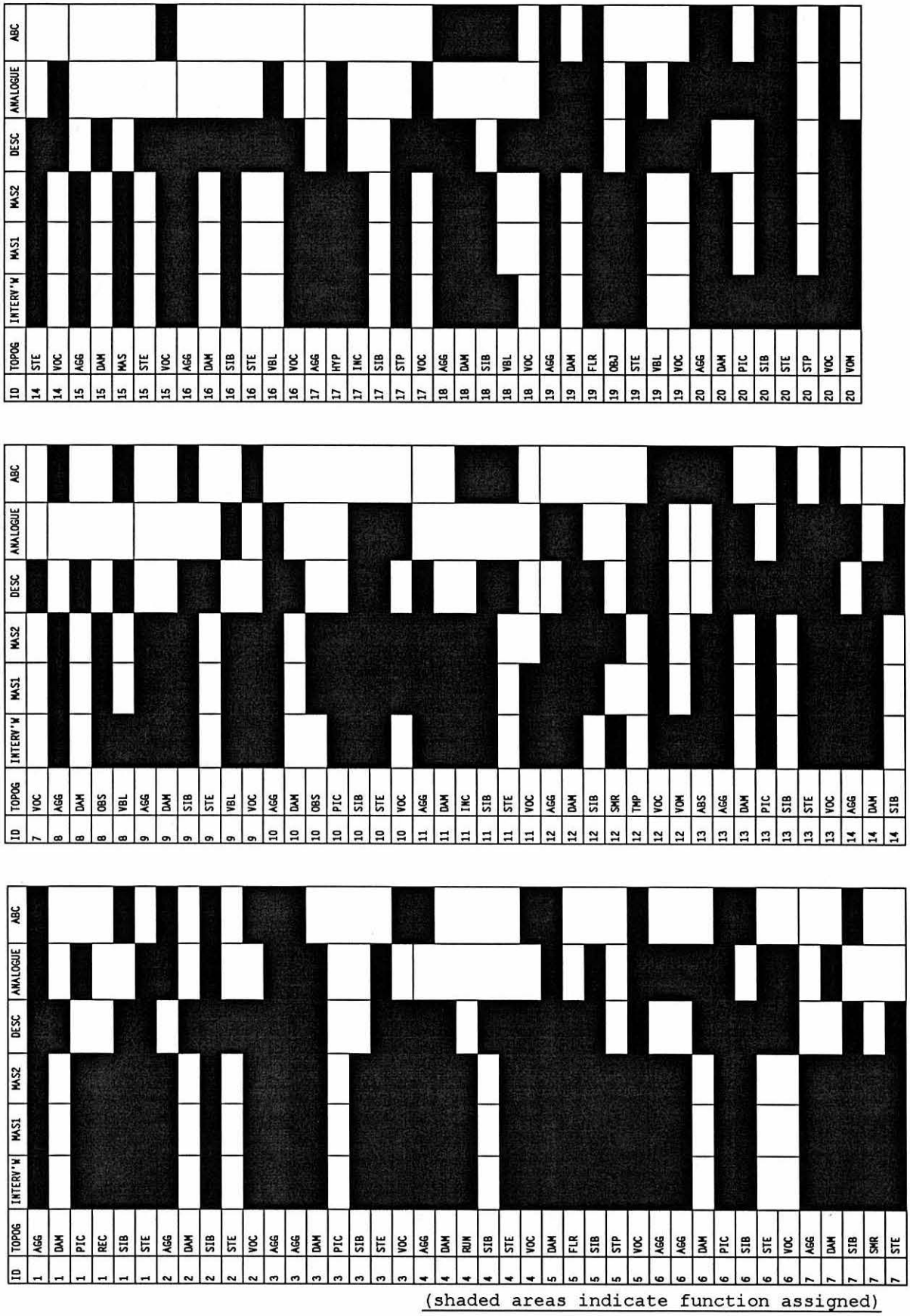
Method

The number of non-predictions was determined for each assessment method and expressed as a proportion of the mean rate of non-prediction by all assessment methods.

Results

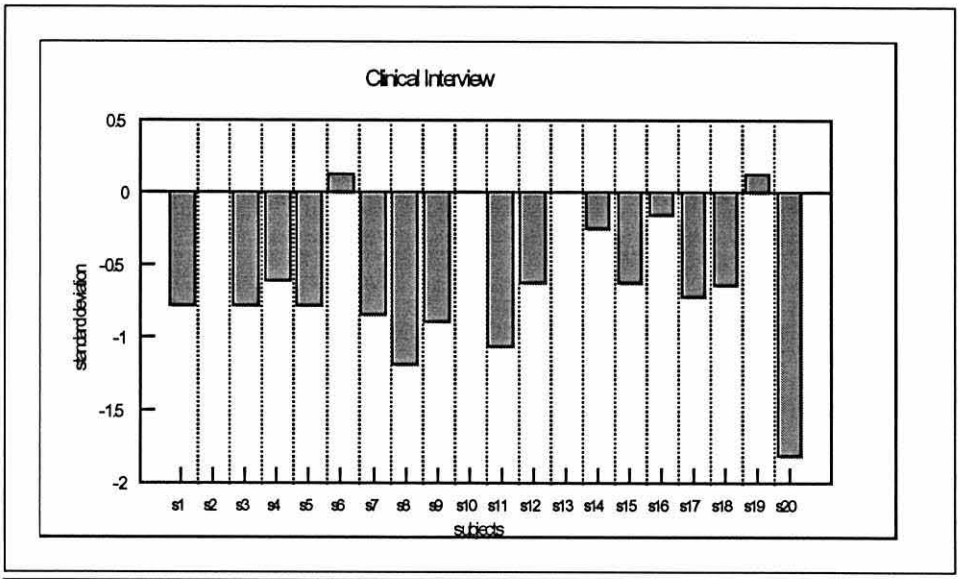
Figure 4.4 plots the standard deviation of non-prediction, for each assessment method, from the overall mean.

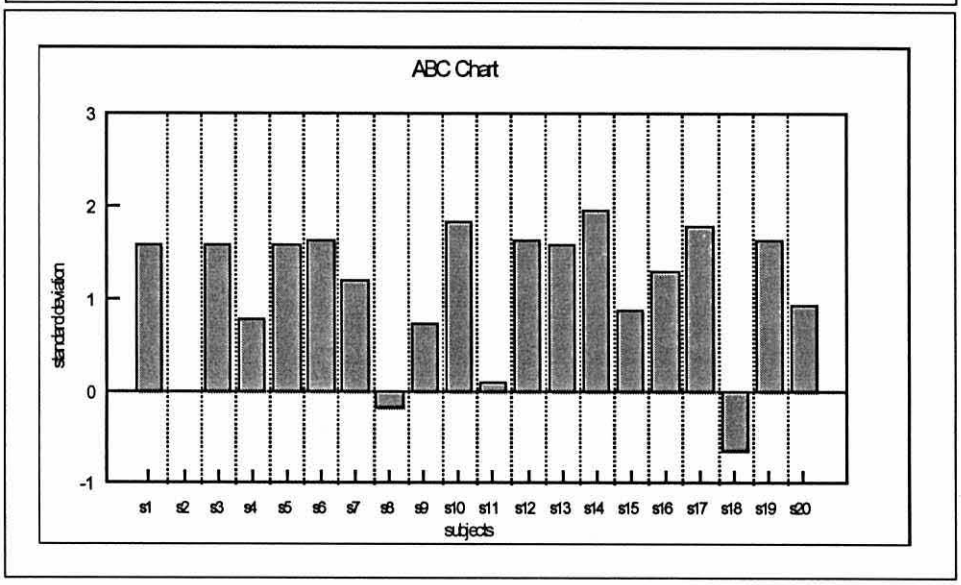
Figure 4.3
Topographies for which function was assigned



(shaded areas indicate function assigned)

Figure 4.4
Distribution of non-prediction by assessment method and individual





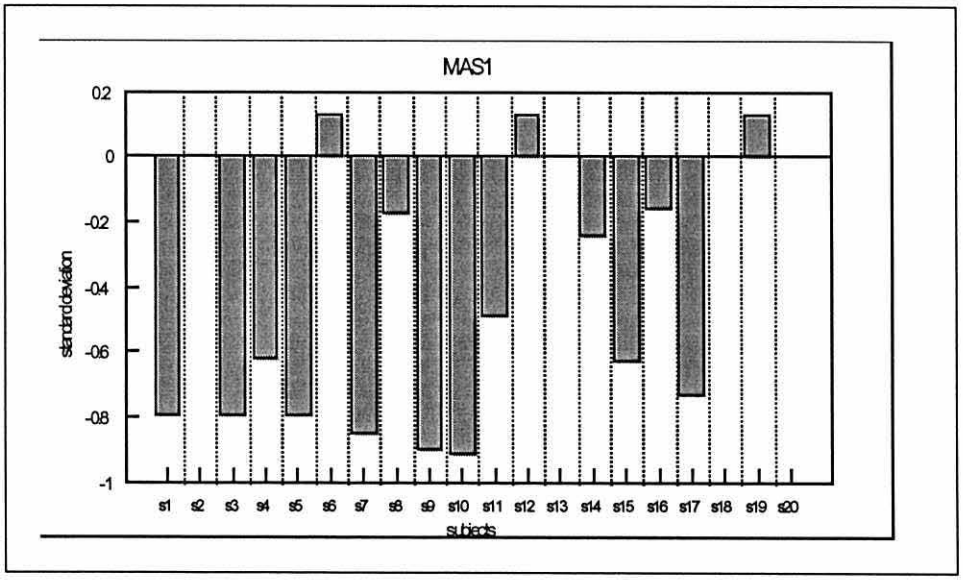
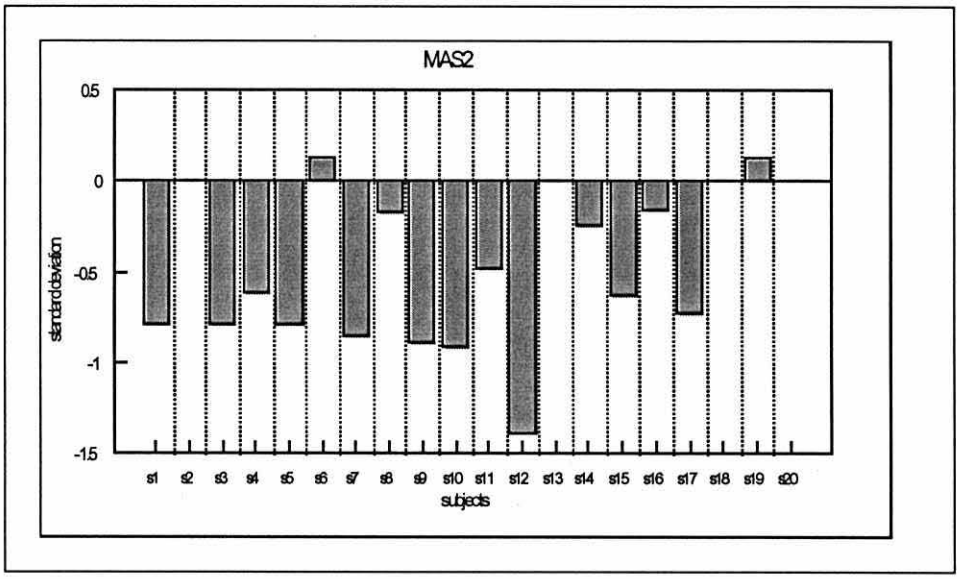
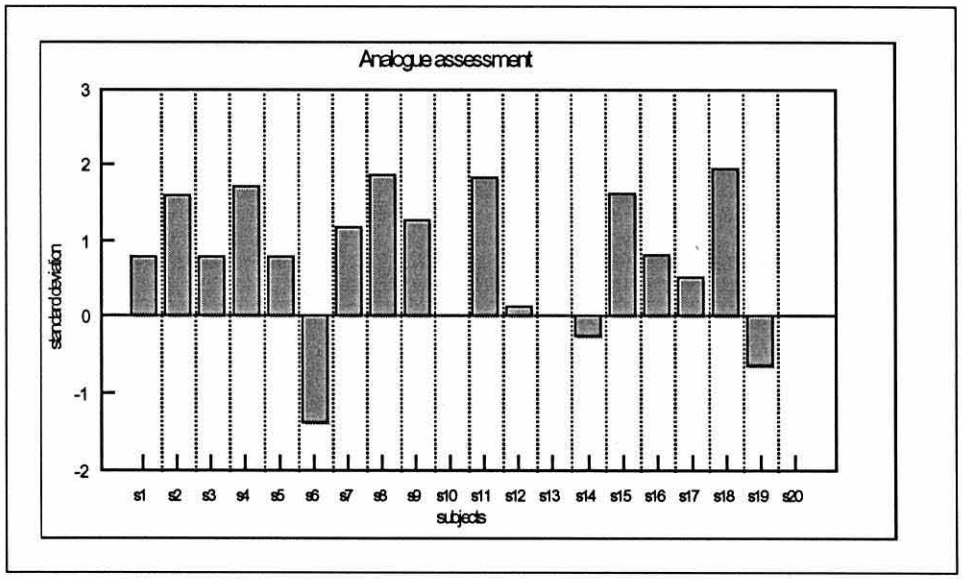


Figure 4.4 continued
Distribution of non-prediction by assessment method and individual





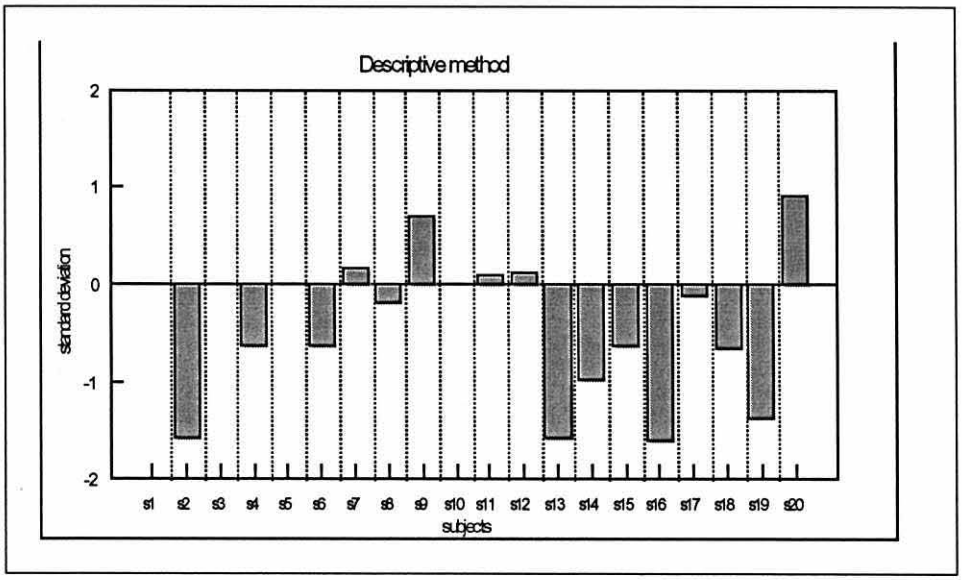


Figure 4.4 shows how interview and the rating scales exceeded the mean rate of non-prediction for s6 and s19 (MAS1 added s12). These assessment methods otherwise ascribed function at a similar rate, and for a similar group of individuals. The ABC method failed to ascribe function above the mean rate for all individuals other than s8 and s18 (n=18), and the analogue method for all but s6, s14 and s19 (n=17). Thus, the low rate of prediction by these methods distributed, relative to other methods, across 85-90% of individuals. The descriptive method failed to assign function above the mean rate for five subjects (s7, s9, s11, s12, s20), a group quite different from the informant-based methods.

Conclusion

Informant-based, experimental and descriptive assessment methods failed to ascribe function, at a rate greater than the mean, for different groups of individuals. While differences in the constellations of topography shown by each individual may account for a proportion of these differences, topographies were sufficiently similar to suggest subject characteristics may have influenced non-prediction.

Convergence on the prediction and non-prediction of function

The last in this series of analyses was designed to examine patterns of agreement across individuals on i) the functions ascribed, and ii) non-prediction.

Method

Each assessment was designated as the base method in turn. Functions ascribed by the comparison methods were examined for convergence with those ascribed by the base method. Agreement

was calculated separately for prediction and non-prediction, and expressed as a proportion of all possible agreement.

Results

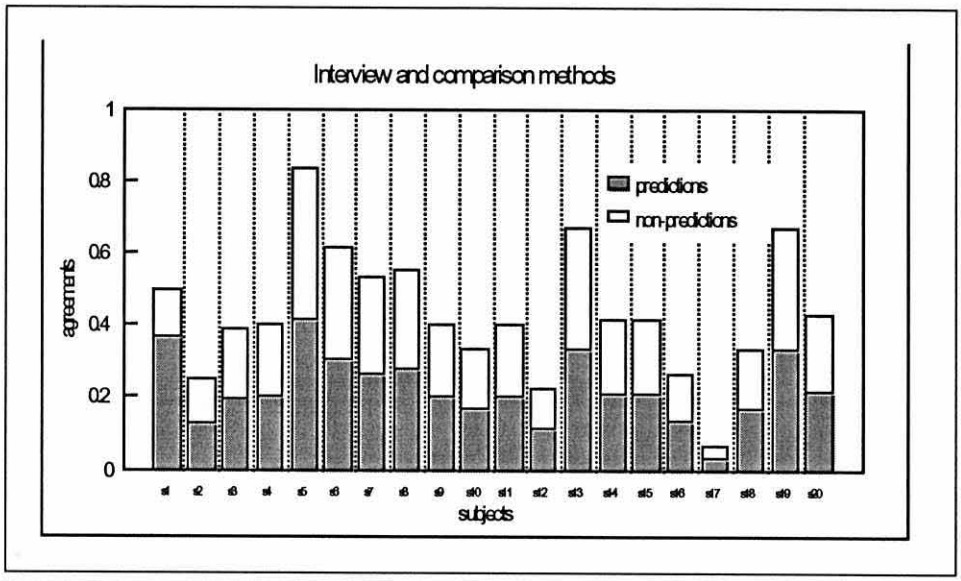
Figure 4.5 shows how agreement on the prediction and non-prediction of function distributed across individuals for each assessment method in turn.

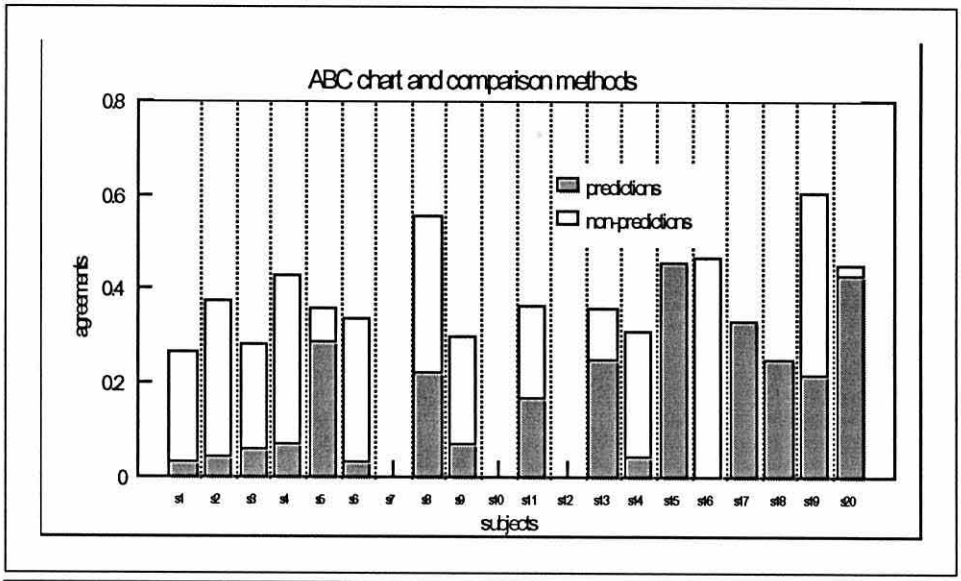
Interview and the rating scales each obtained a level of agreement on the prediction and non-prediction of functions for all twenty subjects. This pattern may have reflected respondent selectivity in identifying topographies for assessment.

The ABC chart method obtained agreement on the prediction of functions for behaviour shown by 16 individuals (80%). Although non-prediction clearly accounted for a greater proportion of agreement, it did not appear to vary systematically across individuals. There was, however, no agreement apparent for three individuals (15%). This kind of pattern might be anticipated where staff observe and record topographies selectively.

Analogue assessment obtained agreement on the non-prediction of function across all twenty subjects. Agreement on ascribed function occurred for behaviour shown by only 11 individuals (55%), suggesting divergence was more likely for certain individuals than others irrespective of topography. This pattern suggests subject characteristics may have influenced prediction rate and/or the overall rate of agreement.

Figure 4.5
Probability of convergence on non-prediction and ascribed function





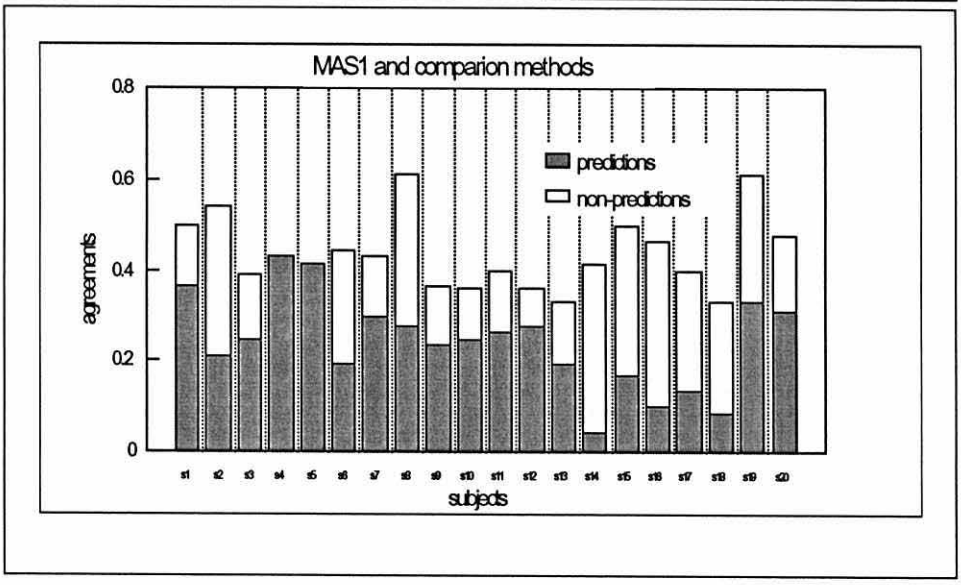
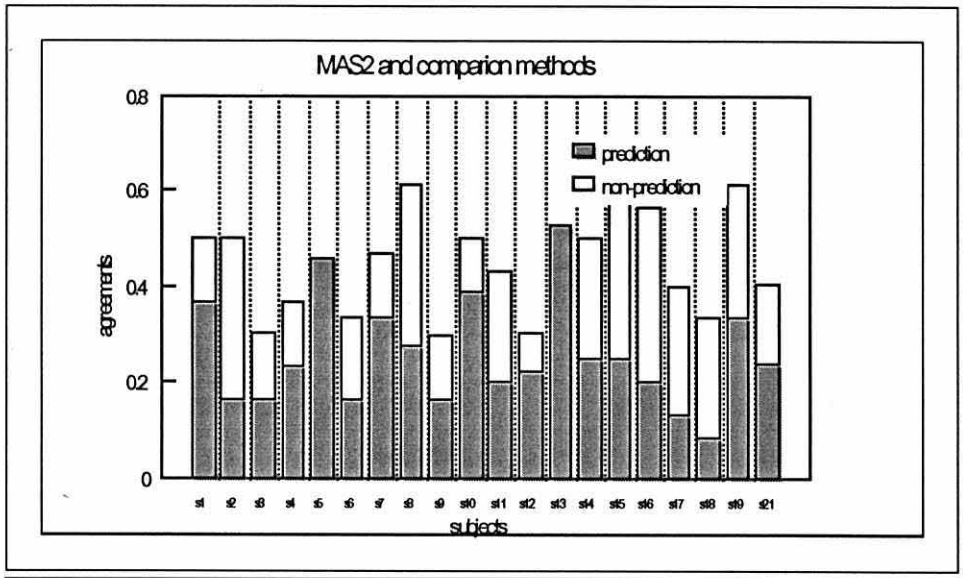
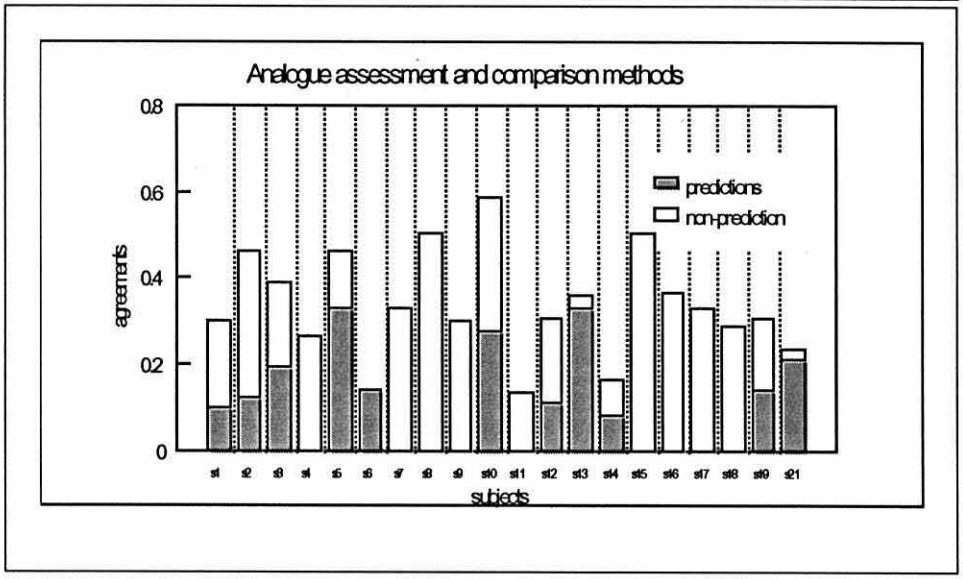
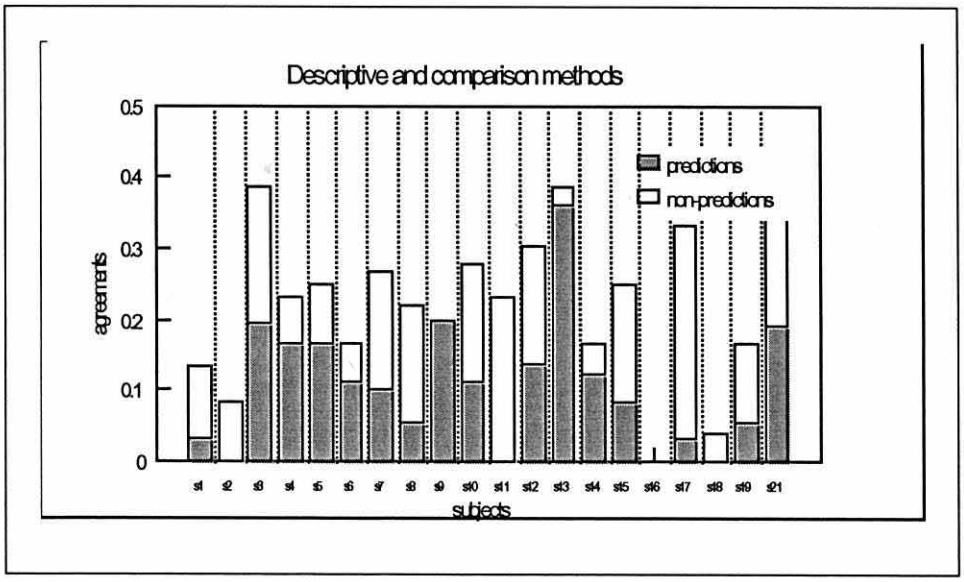


Figure 4.5 (continued)
Probability of convergence on non-prediction and ascribed function







The descriptive method obtained agreement on non-prediction for 19 individuals (95%). Agreement for three individuals was based wholly upon non-prediction and there was no agreement for one individual. Agreement on ascribed function occurred for 16 individuals (80%). The social context of the natural setting may have been more influential with the descriptive method, than factors associated directly with individuals.

Conclusion

Assessment methods revealed different patterns of agreement across individuals both on the prediction and non-prediction of function. Patterns of agreement on non-prediction were consistent with the reasons for assessment failure described in chapter three. On ascribed function, the informant-based methods obtained agreement across all twenty subjects. The ABC chart and descriptive methods each obtained agreement for behaviour shown by 80% of individuals. Agreement for the analogue method alone appeared to vary systematically across individuals, suggesting it may have been more sensitive to person-related variables such as the severity of the learning disability, the presence of additional impairments or the presence and extent of linguistic abilities.

V. CHAPTER SUMMARY OF RESULTS

The pattern of prediction and non-prediction was reviewed for each assessment method. The rating scale (MAS1) was more likely to ascribe tangible reinforcement, the ABC chart showed a marked propensity to ascribe social attention, while the analogue method appeared more likely to ascribe negative reinforcement, especially task-avoidance. It was not clear,

however, whether these assessment methods i) ascribed function erroneously, ii) identified different complexes of function for behaviours that were multiply controlled, or iii) either ascribed more often for behaviours that served the respective functions, or conversely, failed to ascribe function for behaviour which served some other function.

Informant-based methods ascribed function more frequently for topographies classed 'aggressive' or 'miscellaneous'. This may have reflected the selective identification by respondents of topographies for assessment. The direct descriptive method ascribed function more often for all remaining classes of topography, and the classes were consistent with expected sampling effects. The analogue and ABC methods each ascribed function equally across all classes of behaviour. The ABC method failed more often for topographies with poorly defined onsets and offsets (such as stereotypy) while the analogue method failed more often on the relatively low rate behaviours which comprised the miscellaneous class. Some differences were apparent in the functions assigned to classes of topography, with negative reinforcement being assigned more often for aggressive, vocalisation and self-injurious behaviours.

The distribution of prediction and non-prediction of function was assessed across individuals. The pattern of prediction was similar for the informant-based methods. Agreement on both prediction and non-prediction appeared to distribute evenly across individuals. The ABC and descriptive methods also distributed prediction, non-prediction and agreement relatively

evenly, although at markedly different rates. The analogue method alone appeared to ascribe function more selectively.

Overall, these results suggest a complex interaction may have occurred between the three dimensions reviewed to influence the rate of prediction, non-prediction and agreement obtained by each assessment method. It seems likely that the social context of functional assessment may have been influential and this will be discussed more fully in chapter five.

Chapter Five

Discussion

I. INTRODUCTION

Convergence on the functions ascribed by the three classes of assessment method described in chapter one ranged from moderate to good; and was consistent, therefore, with the literature reviewed there. Differential rates of prediction, however, served to reduce the overall rate of convergence to a mean rate of approximately one topography in every four assessed; a level close to chance. Moreover, there was only partial convergence on the functions ascribed to the topographies for which agreement was apparent. Agreement occurred on average for one function in every three ascribed, and was more prevalent among primary predictions. Assessment methods ascribed function differentially across individuals and classes of topography. And, although there was generally more consistency among the informant-based methods, none of the methods was capable of ascribing function universally.

The final chapter of this thesis considers a range of factors that may have contributed to the differential rate of prediction, and the level of divergence in the functions ascribed by each assessment method. Assessment procedures are reviewed, and implications for the clinical application of functional assessment discussed.

II. RELATIVE PREDICTION RATES

The rate of prediction across all assessment methods was high. A function was assigned by one assessment method or more for 98% of all topographies. The average rate of prediction for individual assessment methods was only 60%, however, ranging

from 33% for the ABC method to 73% for the interview method. Convergence reached an acceptable level only when all non-predictions were excluded from analysis. The relative rate of prediction was an important dimension of assessment performance and is worthy, therefore, of further discussion.

Non-prediction by the retrospective informant-based methods

The retrospective informant-based methods ascribed function for a greater proportion of topographies than any of the other assessment methods. The main reason for non-prediction by the retrospective methods was the respondents' failure to identify topographies as presenting a challenge. Respondents generally identified topographies that were discrete and outer-directed, such as aggression. A large proportion of the topographies not identified by respondents were inner-directed or characterised by withdrawal, for example, self-injury and stereotypy. Similar patterns of response identification have been reported elsewhere (e.g. Lowe and Felce, 1995a) and are consistent both with the profile of referrals made to specialist support services (Forrest et al. 1995; Lowe and Felce, 1995b; Toogood et al. 1994), and the topographies identified in epidemiological research as constituting the most challenging behaviours (chapter one).

The fact that respondents failed to identify almost a quarter of topographies as challenging is, however, a matter of some concern. Questions are raised by the present study regarding the observational and reporting skills of front-line staff, their potential habituation to all but the most extreme forms of outer-directed challenging behaviour, and the validity of

applying a socially defined construct to what is essentially a clinical task (Qureshi, 1993a). A number of the studies cited earlier highlight the need for a clear operational definition of challenging behaviour and a comprehensive programme of behavioural training for front-line carers. Other research (Hastings, 1995; Oliver, 1995) suggests that appropriately framed training might improve both the identification and reporting skills of front-line staff and enhance, therefore, the quality of informant-based data collected for the purpose of functional assessment. Staff training and a redefinition of challenging behaviour are indicated by the present results.

Non-prediction by the indirect ABC method

The indirect, concurrent assessment method (ABC chart) ascribed function at less than half the overall rate of the methods classified retrospective and informant-based, or direct and concurrent. The ABC method failed to ascribe function for reasons different to the other methods; that is, many of the observational records provided by staff were either incomplete or non-existent. Even allowing for non-identification by informants, more data were provided for more topographies when staff fulfilled the role of retrospective informant rather than concurrent observer.

This finding is consistent with the procedural reviews cited in chapter one, and may have resulted from a combination of limited training in applied behaviour analysis (Hastings, 1995), and diminished opportunities to observe and record at all times (Reed and Head, 1993). Apparently successful studies cited in the review chapter undertaken by Sasso et al. (1992)

provided additional training for already highly trained teachers, while Crawford et al. (1992) utilised external observers.

Results presented in chapter four suggested fewer ABC records were kept for behaviours which involved minor disruption of the environment and stereotypy. It is possible that, as with the other informant-based approaches, staff failed to identify these topographies as challenging, or as sufficiently challenging to warrant the effort associated with continuous observation and systematic recording. In addition, as stereotypy is often associated with an enduring presence and poorly defined onsets, staff may have found recording antecedents and consequences for some these topographies was simply too difficult. Although ABC recording has been applied to the functional assessment of stereotypy (Crawford et al. 1992), specially trained observers were used.

The relatively high rate of non-prediction by the ABC method in the present study may be attributed in large part, therefore, to problems associated with the way data were gathered, rather than with the method itself. Given ABC recording by staff is a relatively common approach in clinical practice, this interpretation should signal caution. Moreover, together with findings reported for the retrospective methods, the relatively poor performance of the ABC method underscores not only the potential importance of staff training, but of also including a period of direct observation in all functional assessments of challenging behaviour. Should staff training indeed influence the efficacy informant-based assessment methods then a

pre-assessment inventory of staff training or behavioural questionnaire (e.g. Oliver, 1995) might be a useful aid for assessing the potential application these methods in individual cases.

Non-prediction by the experimental and descriptive methods

The same arguments cannot be extended to explain differences in the rate at which the experimental and direct descriptive methods ascribed function. Both methods utilised external observers who were specially trained for the task, and both derived response definitions from a prior period of informal observation. The non-identification of topography was not, however, implicated in the non-prediction of function for either of these assessment methods.

Experimental method

The analogue method failed to ascribe function mainly where low-rate or undifferentiated responding rendered assessment data un-interpretable. Possible reasons for this are discussed more fully later, so only a few general observations will be made at this point. First, the standardised design of the analogue conditions could have rendered the method open to errors of omission (Oliver, 1993; Sturmey, 1995). Second, features of the prevailing social environmental may have exaggerated the effects of introducing novel persons and activity materials into the experimental conditions (Berg and Sasso, 1993). Third, as multiple response topographies were assessed simultaneously, behavioural efficiency may have rendered redundant other members of a response class (Mace, 1994). Fourth, the power of the immediate environmental

contingencies which operated during assessment may have been weakened by the subjects' use of verbal rules (Jones, Lowe and Williams, 1993; Remington, 1991).

Of particular interest at the present time, however, is the relatively low rate at which the analogue method ascribed function overall, and the extent to which this finding might have been anticipated. Extended experimental methods are reported to have failed to ascribe function in between 5% and 33% of individual cases (Iwata et al. 1982; Iwata et al. 1994). In the majority of cases only one topography was assessed for each person, and so the number of topographies reviewed may be assumed to be equal to the number of subjects. Derby et al. (1992) report a success rate of approximately 50% for abbreviated experimental assessment applied in a clinical context. This lower rate could be attributed to the abbreviated nature of assessment, the constraints imposed by clinical application, or the less stringent screening of potential participants. The present study limited the number of sessions to four per condition, applied the conditions in a clinical context, and included a diverse group of individuals as subjects. Thus, a prediction rate similar to Derby et al. (1992) might reasonably be expected. However, the present study also sought to assess multiple topographies in a single assessment, and prediction rate was determined as a proportion of the total number of topographies identified. A function was assigned for at least one presenting topography in 75% of cases, but only 41% of topographies were assigned a function. Present findings do not, therefore, appear to be inconsistent with the published literature, where failure rates may even be

under-represented. Thus, the experimental method may have been justly criticised in the past for ascribing function at a relatively low rate when applied in a wider and less selective clinical context (Sturmey, 1995). While the analogue method clearly has an important role in functional assessment, the relatively low rate of prediction suggests there is also a need for a more diverse range of empirically validated methods.

Direct descriptive method

The descriptive method was classified in chapter one as being direct, naturalistic and concurrent. This method failed mainly where response rates and/or environmental events failed to occur with sufficient frequency during in the period of observation. A key feature of the social environment was its inherent stability; it presented individuals with relatively few demands, very little constructive activity, no materials, and only a meagre amount of social contact. While these conditions may have affected other assessment methods their effects were most obviously apparent for the direct descriptive method. This is evidenced by the way the functions were ascribed across the classes of topography assessed; a function was ascribed for all stereotyped behaviours, but only half the behaviours classed aggressive. Stereotyped behaviours were often observed to occur at a relatively high rate, and independently of alternate activity or social contact. Aggressive behaviours, on the other hand, were observed comparatively infrequently and, by definition, always in the presence of others, which was not itself a frequent event.

Both the analogue and descriptive methods have been deemed unsuitable for assessing relatively low rate behaviour (Reed and Head, 1993). However, referring to behaviour rate *per se* does not offer a complete explanation for the relative rates of non-prediction by these two methods; for the ratio of naturally occurring low and high rate behaviour should have affected both procedures equally, and this was clearly not the case. It will be argued later that the analogue method may have been more susceptible to differences in individual characteristics and potential design problems, while the descriptive method remained vulnerable only to sampling problems associated with the character of the natural environment; even where challenging behaviours were observed to occur.

In the meantime two further differences in the relative prediction rates of the two assessment methods should be noted. First, the direct descriptive method ascribed function for approximately twice the number of topographies of the experimental method. Secondly, the sub-sets of topography upon which a function was ascribed were not similar, even allowing for differences in the overall rate of prediction. Indeed, no two assessment methods assigned function for sub-sets of topography that were exactly alike. This finding is discussed, therefore, in relation to all assessment methods.

Behavioural sub-sets for which functions were assigned

The size and composition of behavioural sub-sets was derived from the number of topographies for which each assessment method ascribed function. Sub-sets of topography were more similar among the retrospective informant-based methods, and

this appeared to derive principally from a high degree of correspondence in the identification by respondents of topographies for assessment. This is not surprising since in most cases the same person served as respondent for all three methods; and in every case the two rating scales were administered simultaneously. In twenty-eight cases no function was assigned by the informant-based methods because the particular topography was not identified by the respondent. In all twenty-eight cases the behaviour was observed to occur in the natural environment; and a function was ascribed by the descriptive method for twenty-four. The descriptive method, however, failed to assign function for thirty-nine topographies. Thus, the sub-sets for these two types of assessment method were very different from one another, and the principal difference between the methods was their data source. The analogue and ABC methods each ascribed at a relatively low rate, and for only a partially similar sub-set of topographies. Interestingly, each ascribed for subjects where the other totally failed.

Taken together these findings suggest that the informant-based, experimental and descriptive methods may have been differentially suited to the assessment of different behaviour shown by different individuals in different social contexts. However, it is at least possible that the higher predicting methods contained a higher ratio of false positives. Indeed, without relating prediction rate to validity, the number of functions ascribed becomes an uncertain measure of efficacy. A weak method which ascribes frequently offers no procedural advantage over a more robust method that ascribes at a lower

rate; indeed, the latter is probably safer. The scale of differential found in the present study strongly suggests, however, that the search for a universal assessment method is far from being complete, and that it may even be inappropriate. Future research might usefully investigate the complementary aspects of different assessment combinations, thereby opening up new avenues for exploring the predictive validity of assessment methods. It may be, for example, that certain methods have poor validity when used in isolation, but are effective when applied as part of a larger complex; the whole being greater than the sum of each of the parts. The need for more than one approach to the functional assessment of challenging behaviour is clearly indicated, as is further work to determine the what are the precise conditions under which any assessment method would be likely to ascribe function. The potential importance of matching assessment selection to a wider range of presenting variables is also highlighted.

III. CONVERGENCE ON THE FUNCTIONS ASCRIBED

Convergence was moderate to good when each assessment was designated as the base method in turn, and agreement was assessed across all comparison methods, with all non-predictions excluded. Agreement was more varied between specific pairs of methods, and none of the pairs reached the near optimum levels that have been reported in a select number of studies (Carr and Durrand, 1985; Durrand and Crimmins, 1988; Sasso et al. 1992). Results from the present study were in fact much more consistent with the majority of previous comparison studies, which have revealed only partial agreement

between the methods compared (Crawford et al. 1992; Emerson et al. 1995; Lalli et al. 1993; Lerman and Iwata, 1993; Mace, Lalli and Lalli, 1991; Oliver, 1991b).

The absence of complete convergence between assessment methods raises potentially serious questions about the validity of the functions ascribed (Oliver, 1991b); although it is possible that each assessment method in the present study was partially correct in ascribing a different multi-function complex for behaviours that were indeed multiply controlled (Carr, 1994). An examination of predictive validity was beyond the scope of the present study, and so no definite conclusions may be drawn on this point. Incomplete convergence does, however, provide a timely reminder that functional assessment serves only to improve the chances of successful intervention; it does not guarantee it (Mace, 1994). Results on convergence strongly support previous calls for the routine application of more than one assessment method (Durrand and Crimmins, 1990; Iwata et al. 1994), and suggest that great care may be needed when selecting from the range of methods available.

Primary, secondary and multiple predictions of function

Convergence occurred principally among the primary predictions of function. This was a welcome finding, since in the majority of cases primary function would most influence intervention. However, the same finding also suggests a boundary of precision existed among the assessment methods used, which were unable to agree secondary predictions or multiple control at an appreciable level. In some cases it may be clinically important to discover the precise nature of secondary function.

For example, the use of social attention may be contraindicated in differential reinforcement of other behaviour (DRO: Jones, Walsh and Sturmey, 1995) where social-avoidance has also been identified. Similarly, the use of time-out procedures (Foxx and Shapiro, 1978) may be less effective for attention maintained behaviour where task or social escape-avoidance is also identified. Thus, the clinical need for an enhancement of assessment methods is clear, and future research might profitably utilise convergence on the ascribing of secondary function, and multiple control, as a bench-mark for evaluating proposed refinements to assessment methodology, including combined approaches.

Informant-based methods

Convergence was greater among the assessment methods classified informant-based and retrospective; although the concurrent ABC chart method also obtained a high proportion of agreement with the interview method when it was designated as the base method. This suggests intra-respondent reliability across this class of assessment methods was generally good, and that an acceptable degree of consistency was achieved in the interpretation of assessment data.

The possible effect of staff beliefs

All data gathered by the informant-based methods were screened by a third-party. It is possible therefore that it was staff beliefs or attributions (Hastings, 1995; Bromley and Emerson, 1995; Lowe and Felce, 1995b) that were reported with a high degree of consistency rather than actual events, whether from memory or observation. Such an explanation would be consistent

with the selectivity respondents showed when identifying topographies for assessment, and with the consistency found in the sub-sets of topography for which the informant-based methods ascribed function. Hastings (1995) found that experienced staff attributed the causes of challenging behaviour in ways that were broadly consistent with current behaviour theory, and that they differentiated between topographies when hypothesising about causes. Oliver (1995) reported statistically significant differences between trained and untrained staff in their responses to a behavioural questionnaire. Staff who served as informants for the present study had a minimum of twelve months experience, and had received an unspecified amount of training. The functions ascribed by the informant-based methods were not the same for all topographies shown by all individuals, nor were all informant-based methods more likely to ascribe from the same reinforcement category. Respondents may therefore have held well developed beliefs about the functions each topography served, just as Hastings' (1995) study suggested they might. If this were the case then it seems highly likely that the same beliefs would have influenced not only topography selection but also the identification of the behavioural processes which potentially underlay each topography. The informant-based assessment methods may have exerted a modifying effect, which is their intended purpose, by helping respondents to consider the evidence for each discrete topography they identified, rather than for the more broadly defined construct of challenging behaviour. The question, which only a study of predictive validity could address, is was this sufficient?

The possible effect of potential errors of omission

Durrand and Crimmins' (1988) rating scale appeared particularly vulnerable to potential errors of omission (Oliver, 1993). On primary predictions just over 20% of disagreements were related in some way to the omission of social-avoidance as a possible function. This finding is returned to at a later time, where the evidence reviewed strongly suggests that this rating scale in particular, and functional assessment in general, might benefit from a wider taxonomy of reinforcement categories (Horner, 1994), and concomitant procedures to examine their relative influence. This notion would, of course, need to be tested empirically, for the level of precision required in clinical application has not yet been fully established, and most clinically-based functional assessment remains rudimentary. Alternative approaches include the elaboration of sequentially refined assessment methods (Carr, 1994), or person-time-context specific analysis (Jones, 1995) which aims to identify what are the very precise stimulus conditions that elicit a specific response.

Potential bias effects

Two of the retrospective informant-based methods in the present study showed no proclivity for ascribing a particular function. ABC charts, on the other hand, showed a marked tendency to ascribe positive reinforcement by social-attention, and one of the rating scales (MAS: Durand and Crimmins, 1992) frequently ascribed tangible reinforcement. These findings are discussed in a later section, and are mentioned here only as a factor likely to impinge on convergence. It will be remembered also that the informant-based methods tended to assign function less

frequently for behaviours that were classified disruptive or stereotyped. Topography selection and possible ascribing bias would, if significant, have the effect of lowering the overall rate of convergence between assessment pairs.

Conclusion

Chapter one delineated the inherent strengths of the informant-based methods, which offer a considerable number practical advantages for the clinician. Present findings, however, suggest these methods should not be relied upon to provide a complete assessment of function. While the level of convergence with the experimental and descriptive methods was good, problems of topography identification and multiple prediction weakened agreement considerably. As with all assessment methods, the informant-based approaches agreed most strongly on primary predictions of function; agreement on secondary predictions and multiple control was poor.

Findings for the informant-based methods underscore the need to include direct observation and an experimental component in a complete functional assessment (e.g. O'Neill et al. 1990). Informant-based methods may, however, form part of a sequentially refined or multiple assessment complex; one tailored quite precisely to the person and the wider social context. Previous work by Hastings (1995), Hastings et al. 1995) and Oliver (1995), strongly suggests that staff training in behaviour analysis might enhance the performance of the informant-based assessment methods. A precursory assessment of respondent knowledge might aid the selection of assessment methods.

Indirect and direct descriptive methods

Carr (1995) recently suggested that ABC chart recording was 'overwhelmingly sufficient' for a clinically-based functional assessment of challenging behaviour, and that more elaborate approaches were appropriate only for the purposes of research. Findings from the present study do not support this view entirely. Moreover, an alternate argument may be made, both in terms of accountability to the client and clinical efficacy, for urging clinicians to aspire toward the rigour and precision that is demanded by the research community. Present results suggest that staff use of ABC recording was not an efficient method of conducting a functional assessment. Some reasons for this have been considered, other factors are to be discussed later. The present area of discussion is restricted therefore to contrasting the performance of the direct and indirect descriptive assessment methods.

Indirect observation

The indirect (ABC) method ascribed function at a relatively low rate, and tended to ascribe social-attention more often than any other function. A range of factors may have acted together to influence the relative rate of prediction and potential bias shown by this method. First, staff beliefs may have affected the identification of topographies in a manner similar to that described earlier for the retrospective informant-based methods (the indirect approach to data collection means the method may also be classed informant-based). In addition, prevailing beliefs about the primary causes of challenging behaviour may have influenced which behaviours and events staff recorded, and could also have coloured the data staff entered onto the logs;

thereby introducing potential selection and sampling bias into the results. Second, competing work demands meant staff would inevitably have been forced to sample events at less busy times of the day. Thus, the ABC records collected by staff would be less likely to offer a complete account than those collected by an external observer (Reed and Head, 1993). Third, staff in the present study received no special training in behaviour theory or observation techniques. The level of previous training among the staff who took part was not known, but must have varied by virtue of job grading. This factor may have resulted in uneven or inconsistent recording. Fourth, given the low rate of prediction by the ABC method, it is possible that the behaviours staff selected were those actually maintained by social-attention; and that the apparent bias was negative rather positive. Social contact was infrequent in the natural setting. A number of persons must therefore have endured a state of perpetual deprivation. While the direct descriptive method revealed no bias in the functions it assigned the method did, like the ABC method, ascribe social attention more often for aggressive behaviours. The direct method, however, showed an improved rate of prediction over the indirect method when ascribing function for stereotyped responding, disruption and self-injury. Thus, the ABC method may have been negatively biased, through omission or failure to ascribe, rather than positively biased in the functions it did ascribe. The absence of validity data makes this interpretation uncertain. The degree of convergence found in the functions assigned by the ABC method after non-prediction was excluded could, however, be interpreted as supporting this

view. It will be argued later that aspects of the presenting topographies may also have contributed to the selectivity staff showed in recording. This point does not, however, detract from the present interpretation.

Direct observation

The direct method assigned function at approximately twice the rate of the indirect method, and showed no proclivity for ascribing social-attention, or any other function. Hence, the main differences between the two assessment methods lay clearly in the rate of prediction, and the presence or absence of potential bias. The direct method utilised externally supplied observers who were trained in behaviour theory, expressed no pre-conceived ideas about maintaining variables, and had no other duties during the observation period. The direct method utilised a more sophisticated approach for data capture and analysis. However, the methods varied principally on the number of topographies for which a sufficient volume of data was generated, and not on the finer points of analytical methodology. This factor may be significant as a substantial proportion of the topographies for which the direct method successfully ascribed function, and for which the indirect method did not, were those which had an enduring presence and a poorly defined onset.

Conclusion

The direct descriptive method was clearly superior in terms of the number of topographies for which a possible function was identified. The method was, however, time-consuming and complex to administer. It is possible that, with a sufficient

training and support, staff could be enabled to collect a greater volume of relevant assessment data, and that Carr's (1995) assertion may yet be realised. Such an initiative could be cost-effective against a background of scarce resources and limited specialist provision.

Experimental method

The experimental method ascribed function at a relatively low rate in the present study. When comparison was restricted, however, the functions assigned by the experimental method converged more often with the informant-based methods than with the direct descriptive method. This was an unexpected finding since the experimental and descriptive methods each used direct observation, which it has been argued thus far, and elsewhere (Berg and Wacker, 1991), is preferable to using third-party informants. The relative rate of convergence with the informant-based methods, however, bodes well for their future development and inclusion as part of a multiple-assessment complex.

Procedural aspects of the analogue methodology appeared to influence prediction rate more than convergence, and the relative effects of procedure are anyway to be discussed later. Factors pertaining to convergence were principally associated with the analysis and interpretation of assessment data, and it is these factors that will be reviewed here.

Analysis and interpretation

Experimental data were interpreted in the way described originally by Iwata et al. (1982). Thus, for example, elevated

responding under the task-demand condition was judged to be consistent with the demand escape-avoidance hypothesis; and that function would be assigned (refer to chapter 2, p. 57 for a full account). Two problems arose when interpreting analogue data that may have served ultimately to diminish convergence. First, in a small number of cases elevation in the task-demand condition created a strong impression that the primary motivation was to gain access to the task materials rather than avoid demands. Contingent withdrawal of the task-demand (and task materials) appeared to increase the rate or intensity of the response. A plausible explanation for this may be derived from the obvious paucity of materials that were available to individuals in the natural setting. However, strict interpretation of elevated responding under demand conditions resulted in an escape-avoidance function being assigned (negative reinforcement), whereas a more impressionistic interpretation would have suggested access to tangibles (positive reinforcement). Competing interpretations of the same assessment data could result in disagreement at a very fundamental level; positive versus negative reinforcement. An incorrectly ascribed reinforcement category would, in a comparison study, diminish convergence - assuming the other methods had themselves ascribed accurately.

A second problem of interpretation, which may also have inhibited convergence, concerned the derivation of meaning and the subsequent classification, of differential rates of responding which were devoid of impressionistic ambiguity. Lovaas (1982), for example, has argued that elevated responding under conditions of increased task-difficulty could be the

function of attempts to moderate over-stimulation rather than task-demand. Thus, under identical stimulus conditions, at least two classifications are possible, and these differ at the level of social versus non-social reinforcement. Similarly, the condition where no social contact and activity materials are present (alone except for the observer) may either establish conditions under which i) these putative reinforcers are sought (positive reinforcement), or ii) where sensory stimulation becomes reinforcing (automatic). Likewise, increased responding under conditions where a person is nearby, but not interacting, (contingent social contact) may be considered consistent with either a socially mediated, positive or negative reinforcement hypothesis (Oliver, Murphy, Crayton, et al. undated). Thus, the experimental method remained vulnerable to errors of commission (Oliver, 1993), by ascribing function according to erroneous interpretations of objective data.

Conclusion

It appears that while experimental analysis may be capable establishing the precise stimulus conditions under a response is more likely, a number of interpretations are possible. By following the classification criteria described by Iwata et al. (1982) an unknown number of errors may have occurred at the interpretation stage. These problems are not, however, insurmountable. One solution would be to seek further confirmation of a given hypothesis by assessing the pattern of responding across all analogue conditions; to ascertain whether they are plausibly consistent with the hypothesis initially suggested. Another would be to allow impressionistic

interpretations to be tested through sequentially refined designs. A third, would be to apply experimental designs to the analysis of intervention strategies rather than the behaviours' maintaining variables (Repp, Felce and Barton, 1988).

IV. FACTORS INFLUENCING PREDICTION RATE AND CONVERGENCE

Possible effects of multiple control

If, as Carr (1995) has recently suggested, challenging behaviours are likely to be multiply controlled, then their function could be expected to vary according to a multiplicity of ever-changing situational contexts. Jones (1995) has already argued that asking 'what overall function a behaviour serves' might not be the right question, suggesting instead that clinicians might more profitably consider 'what function does this behaviour have for this person, in this setting, under these circumstances, at this time?'. This notion has obvious implications for 'snap shot' approaches to assessment, and may go some way toward explaining some of the divergence found in this and previous studies.

Interview made multiple predictions for the greatest number of topographies, and assigned more functions per topography than any other assessment method. Thus, interview was either the method best suited to detecting multi-function complexes, or was the least able to discriminate meaningful functional relationships. None of the other assessment methods assigned multiple function at the same rate as the interview method. A test of assessment validity would be required to examine the

extent of multiple control, and this fell beyond the scope of the present investigation.

At a theoretical level, however, the presence of multi-function complexes might be considered as a factor influencing convergence. An intriguing notion is that differences in assessment methodology could render different assessment methods more or less sensitive to the detection of particular functions served by behaviour that is in reality multiply controlled. Such an idea could mean that none of the functions assigned lacked validity, but that each assessment method simply highlighted a different aspect of operant control. Were this the case, then a multiplicity of assessment methods would be needed. And, as the purpose of assessment is to identify manipulations which are likely to result in clinically significant behavioural changes, each would need to be tested experimentally. The argument may appear circular for if most behaviour is multi-functional then, when extended to its logical conclusion, experimental analysis is the only viable method or functional assessment becomes irrelevant - and, clearly it is not. It may be, however, that contextual assessment (cf. Jones, 1995) is more important than has hitherto been realised; perhaps at least as important as functional assessment.

Conclusion

If the majority of challenging behaviours were multiply controlled, then it does not seem theoretically impossible that assessment methods should detect, or give emphasis to, different aspects of the controlling variables. Without a

rigorous test of validity all consideration is purely speculative. However, future work might usefully centre on the systematic integration of complementary assessment methods, each with an enhanced degree of precision, so that a more broadly-based, but individually tailored assessment is possible. Such an approach might resemble schemes previously proposed by LaVigna and Donnellan (1986) or O'Neill et al. (1990), but would be devoid of the standardised, menu-based approaches that have been recommended until now. There is a risk of technique-based approaches to assessment replacing the technique-based approaches to intervention that were so common during the 1970's, and that the emphasis on analysis will once again be lost.

Possible effects of assessment context

For the purpose of the present discussion assessment context refers to a range of variables that may exert influence over the performance of an assessment method. Areas of particular interest include; informant knowledge of behaviour theory, observational practice and reporting skills; and the nature of the natural setting. Informant knowledge was discussed at length when considering informant-based methods (p. 125), and a pre-assessment of informant knowledge was indicated. To avoid repetition the present discussion will consider only the potential effects of the natural setting on assessment performance.

The setting context

Chapter one described the nature of the natural environments within which assessment was conducted. Environments were

described as barren; being largely devoid of materials, and presenting few opportunities for engagement in meaningful activity or social contact. The further analysis of three randomly selected observational records confirmed this description. The onset of social contact, instruction and staff assistance was observed in fewer than 1% of one-second intervals, and engagement in purposeful activity occupied between 2% and 33% of observed time. If representative, these data suggest participants in the present study must have endured a perpetual state of social and occupational deprivation.

The retrospective informant-based methods may have been insensitive to these conditions; since staff form a major part of each persons' social environment they would be the principal source of the contact that was missing. The descriptive methods sometimes were unable to sample a sufficient number of potentially salient events for analysis. It might reasonably be supposed that under such barren conditions challenging behaviour would be reinforced by the contingent presentation of social contact or activity materials. Their presentation in novel analogue conditions might, however, have quite a different effect. In some cases the mere opportunity for obtaining social contact and materials may have been sufficient to maintain alternate behaviours, thereby resulting in artificially low rates of target behaviour during assessment. In this sense the analogue conditions may not have replicated the natural environment with sufficient integrity, and the brevity of the sessions may have meant that novelty effects were not countered (Sturmey, 1995). This could explain why

proportionately more challenging behaviours were observed in the natural setting, and why the analogue method obtained a relatively low rate of prediction.

How much do we need to know?

Reviewing the contextual basis of functional assessment also raises some, as yet, unanswered questions regarding the level of precision required to safely inform intervention. It could be argued that given such gross deprivation it would be sufficient only to know whether reinforcement was socially mediated, so that material enrichment, for example, could be tailored in the most appropriate way (Vollmer, 1994). On the other hand, it may always be important to know, as far as possible, what are the precise stimulus conditions that elicit a response.

The need for a broader taxonomy of reinforcement categories has been suggested (Horner, 1994). The present system of classifying function according to five major categories of reinforcement means generalisations are often necessary when designing intervention. While it could be argued that incomplete convergence found in the present study arose from the application of an incomplete taxonomy of reinforcement categories, too many of the disagreements occurred at a fundamental level (positive, negative, and non-social) for this to be true. A more likely explanation lies in the potential confounds the social context may have introduced into assessment, or that assessment introduced into the setting.

This question of precision has important implications for clinical decision making, and the selection of assessment methods. Sturmey (1995) has argued in favour of sequentially refined approaches to functional assessment, and that this should be based on a continuum of least to most intrusive and expensive. This essentially pragmatic approach would not be incompatible with variations on a multiple-assessment format which is suggested by the present study, or Jones' (1995) notion of a person-context specific approach.

Conclusion

The character of the natural setting may influence the outcome of assessment in a variety of ways, and should be borne in mind when an assessment combination is being selected. Drawing the findings of the present study together with those of other comparison studies and the reviews cited in chapter one, it would appear that;

- informant-based methods may be best suited to social environments where carers are reasonably well trained in behaviour theory, and that unless carers are well trained and have received prior practice in observation techniques, descriptive methods should be by direct observation,
- but, direct observation will be viable only in dynamic social environments; where demands are made, social contact is frequent, and there are things for people to do, and novelty confounds may be introduced into experimental designs where the natural environment is extremely barren.

A systematic pre-assessment of the social context in which assessment is to be applied may aid the selection of appropriate of assessment methods. Factors to be considered would include staffs' level of training in behaviour theory and

the quality and nature of social interaction and activity levels in the natural environment.

Effects of potential assessment-function relations

Three assessment methods showed a proclivity for ascribing from one of three function categories; the rating scale, ABC recording and analogue assessment. These are discussed in turn.

The rating scale

The rating scale (MAS1: Durand and Crimmins, 1992) ascribed tangible reinforcement more often than any other function, and more frequently than any other assessment method. A study by Thompson and Emerson (1995) revealed a similar pattern of prediction, suggesting that in its present form the device may be susceptible to bias, although in another study Emerson and Bromley (1995) reported self-stimulation and social attention were ascribed more often.

In the present study the revised version (MAS2) did not ascribe tangible reinforcement in the same way as the original (MAS1). This second device comprised a re-worded version of all 16 items, and included four additional items covering social-avoidance. Differences in the ratings may therefore have related in some way to how the questions were phrased or to the availability of an additional domain in MAS2.

MAS2 ascribed social-avoidance six times as a primary function and thirteen times as a secondary function. The MAS1, however, ascribed tangible reinforcement for only one of the six primary, and none of the 13 secondary, functions assigned by

MAS2. Indeed, MAS1 ascribed no function for eight of the 13 secondary social-avoidance functions assigned by MAS2, and distributed the remainder evenly across other reinforcement categories. Thus, the addition of a social-avoidance domain in MAS2 did not appear to relate directly to the way MAS1 assigned tangible reinforcement.

Reed and Head (1993) observed that some respondents found some of the items on the MAS difficult to rate. It was not clear, however, whether this was due to the way the items were phrased or, as Emerson and Bromley (1995) have pointed out, that some items simply do not apply to topographies such as aggression. A detailed content analysis fell beyond the scope of the present study. Present results do indicate, however, that further work may be necessary.

A third point of interest concerns the effects on convergence that the criteria used to assess the cut-off point in the mean domain scores may have had in determining which functions were included. Bromley and Emerson, (1995) found that using the mid-point of the Likert-type scale as a cut-off, produced a different pattern of results than when Durand and Crimmins' (1992) method was used. A more detailed analysis would be required to establish the extent of the potential difference, and this will form the basis of a further study. Previous research does suggest that a different pattern of prediction could have occurred, and that this may have affected the overall rate of convergence.

Conclusion

Although further work is necessary, present results suggest that including social-avoidance covered a potentially important area of omission (Oliver, 1993), and that adding this domain did not subtract directly from the rating of other items. Although Singh et al. (1993) found the factor structure of the MAS was robust, present results indicate the device may be further improved by including a broader array of potential functions (Horner, 1994). This could be achieved by extending the number of global domains, or sub-dividing existing domains. For example, tangible reinforcement could be divided into food and activity related stimuli, and task-demand could consider instruction rate, modality, task complexity, task preference, etc. More work is indicated to see whether the scale could be improved by re-phrasing some of the items and reviewing the way the cut-off point is determined for the mean domain scores.

The ABC method

The ABC method ascribed social-attention most often. This proclivity appeared strong and persisted throughout the analysis of various topographical sub-sets. No comparison was found in the published literature. It was suggested earlier, however, that this apparent bias may have been negative - an artefact of a relatively low rate of prediction. Another proposition was made in which it was suggested that the functions assigned may have reflected staff beliefs about the causes and relative importance of different forms of challenging behaviours. A third possibility, which is to be discussed more fully in a following section, concerns the

possible or perceived relationship between the form and function of behaviour. For present purposes it is sufficient only to recall that staff tended to record antecedent and consequences for behaviours which had a discrete onset and offset. Specifically, most forms of stereotypy were avoided. It is possible, therefore, that a larger proportion of the topographies selected by staff could have been maintained by processes of socially-mediated reinforcement. Moreover, given the levels of deprivation discussed in the previous section, it does not seem unlikely that socially-mediated reinforcement would have included the contingent presentation of social-attention. Thus, the functions ascribed may have been valid, and the apparent bias a mere artefact of selective recording.

Conclusion

Bias by the ABC method may have derived from selective recording, which in turn may have been linked closely with prevailing staff beliefs. The investigation of relationships between staff beliefs and behaviour is a developing area. Further research might profitably consider the relationship between staff beliefs and ABC recording, perhaps as part of a wider investigation into the conditions under which external advice is accepted (Emerson, 1995; Hastings, 1995; Lowe and Felce, 1995b). In the meantime present results suggest that the ascribing of social-attention by the ABC method should, in a clinical context, be regarded with caution.

The analogue method

The analogue method ascribed task-avoidance more frequently than any other function. Iwata et al. (1994) reported that socially-mediated negative reinforcement was ascribed by the analogue method in 38% of cases. In the present study the analogue method assigned task-avoidance to 47% of topographies, and social-avoidance for a further 10.7%. The relatively high ratio of predictions which implicated processes of negative reinforcement could, like the ABC method, be attributed to the relatively low rate at which the analogue ascribed function. The problems of interpretation discussed earlier may also be germane insofar as a proportion of the functions assigned to negative reinforcement category may have been positively reinforced by access to tangibles or social-attention.

Conclusion

The proportion of functions assigned by the analogue method from the categories of negative reinforcement was consistent with the literature reviewed, as was the relative rate of prediction reviewed earlier. A key question, given the relatively low rate of prediction, concerns the proposition that challenging behaviours may be more likely to be maintained by processes involving negative reinforcement. More literature-based reviews, retrospective studies similar to Iwata et al. (1994) and further comparison studies such as the present one, may provide the means for addressing this question. An increased emphasis on predictive validity, in respect of treatment outcome, may also help to delineate the proportion of challenging behaviours that are likely to be

maintained by processes of negative rather than positive reinforcement.

Effects of potential assessment-topography relations

Two of the three assessment types applied in the present study ascribed function more frequently for particular classes of topography. The informant-based methods ascribed function more often for behaviour classified as aggressive and miscellaneous; the direct descriptive method ascribed more often for the remainder. The experimental method ascribed fairly evenly across classes of topography.

Informant-based methods

Ascribing by the informant-based methods was closely linked to the non-identification of topography by respondents. This was largely true for staff use of ABC recording. There is no reason to suppose that the informant-based methods would have failed to ascribe had the missing topographies (or data) been identified. Thus, the primary issue may be considered as one of non-identification. Identification was discussed earlier, and a prior period of direct-observation was suggested as one way of overcoming this potential difficulty. It appears, however, that the informant-based methods did identify and ascribe more readily for discrete outer-directed behaviours; even where these occurred relatively infrequently in comparison with inner-directed or withdrawn behaviours. It was suggested earlier that staff training might improve respondent identification of topography. In the absence of such training clinicians might usefully apply data covering the probability

of prediction based on topography as a guide to assessment selection.

The direct descriptive method

The direct descriptive method ascribed a function for all responses classed stereotypic, and obtained a relatively high rate of prediction for self-injury, disruptive behaviour and negative vocalisation. The descriptive method ascribed function less often for behaviour assigned to the aggressive and miscellaneous categories. Sampling opportunity offered the most plausible explanation for these differences. For example, as stereotypy is often marked by an enduring presence there would be plenty of opportunity in each observation period for sampling. Outer-directed aggressive behaviours occurred less often and were more difficult to capture. Moreover, stereotypy tended to occur more often under conditions where there was no contact or activity available. Aggression tended to occur in the context of social interaction. Thus, there were more opportunities to sample base as well as criterion events.

Conclusion

The informant-based and descriptive assessment methods assigned function at a different rates for different classes of topography. Features associated with topography, such as type, expected rate of occurrence, outer-directedness, and discreteness of onset, may provide a useful indication of potential assessment efficacy, especially when assessment context is taken into account. In chapter four the probability was estimated of each assessment method assigning a function by class of topography, and the relative distribution of

predictions across the five reinforcement categories was compared. This might provide a useful starting point for the further development of a empirically-based guide to assessment selection which is based partially on topographical characteristics.

Potential assessment-topography-function relations

There is a certain intuitive appeal in the suggestion that particular classes of topography may be more likely to serve a particular function. There is a logic, for instance, in linking stereotypy with automatic reinforcement. And, as aggressive behaviours tend to be avoided by most people, in linking these responses with an escape-avoidance function. Emerson and Bromley (1995) investigated potential relationships between the form and function of challenging behaviour. They concluded that, although loose relationships were apparent, form did not provide a good indication of function. This conclusion was based, however, entirely on the functions ascribed the MAS (Durand and Crimmins, 1992), which has limited validity and questionable reliability.

Emerson and Bromley (1995) reported that the functions assigned most often by the MAS were; self-stimulation for self-injury, destructiveness and other challenging behaviours; and social-attention for aggressive behaviours. Results from the present study (chapter four) were in partial accord with these findings. The MAS1 ascribed automatic reinforcement most often for self-injury, disruption and other challenging behaviours; and task-avoidance, together with tangible reinforcement, for aggressive behaviours. Indeed, all except the descriptive

method assigned negative reinforcement predominantly for aggressive responding, and all identified automatic reinforcement more often for stereotyped responding. Thus, a prima-facie case may be made for suggesting a probabilistic link between the form and function of challenging behaviour. Clearly more work is needed. But, knowledge of such relations, should they be found, could aid clinicians to select assessment methods, formulate hypotheses, and interpret assessment results with greater confidence.

Conclusion

Results from the present study were in partial accordance with results of a previously published study. Although the authors of that study concluded that form was not a good predictor of function, data from the present study suggest there could be a link, which if shown to be reliable might aid clinicians to select assessment methods, formulate hypotheses, and interpret assessment data with greater confidence. Further evaluative studies would need to be carried out, and a retrospective review of published studies may help to establish the strength of potential, if limited, relationships between form and function. The contingency tables presented in chapter four (p. 100) might also be developed to examine the likelihood of different assessment types assigning particular functions to particular classes of topography.

Effects of potential person-related variables

The higher predicting informant-based and direct descriptive methods ascribed function for behaviour shown by all twenty participants. The lower predicting methods failed entirely for

certain individuals, although these were not the same persons. The ABC method failed to ascribe for three people, but as was discussed earlier this appeared to result principally from staffs' selective recording practices.

The analogue method, which was standardised across subjects, failed to ascribe a function for all behaviours shown by five persons; mainly because of low rate or undifferentiated high-rate responding. Thus, the analogue method appeared to suit the assessment of behaviour shown by certain individuals but not others, and for reasons which were associated in some way with the person.

Possible effects of assessment design and implementation

Chapter one reviewed three systems for classifying assessment methods in which the following discriminible features were identified;

- basic orientation (informant-based versus experimental versus descriptive),
- approaches to data collection (retrospective versus concurrent & direct versus indirect), and
- impact on the social environment (modified versus intact).

While a number of procedural reviews are available, the final part of the present chapter will consider the relative strengths and weaknesses of assessment methods as they were applied in the present study. The principal aim is to add to the existing reviews by considering the assessment methods used in the related areas of data collection, analysis and interpretation, and to assess their potential effects on prediction rate and convergent validity.

Data collection

Informant-based methods

As the clinical interview was semi-structured neither the order of inquiry nor the phrasing of questions was prescribed at the point of data collection. This was a strength of the method since the interviewer could probe areas of interest and clarify with the respondent any ambiguities that may have arisen. The quality of interview data does, however, crucially depend on the skills of the interviewer and the respondent (Reed and Head, 1993). Rating scales, on the other hand, imposed limits on the order and phrasing of questions. While this may be advantageous for the inexperienced practitioner, respondents were restricted to making the best fit possible between what they knew or believed, and what the range of scenarios offered them. Some respondents had difficulty rating some items, although none stood out as being especially problematic. Respondents also commented that it was sometimes difficult to decide which was the most appropriate frequency-rating. Differences in the way the two versions of the rating scale ascribed tangible reinforcement and social-avoidance suggests data collection by rating scales may have been hampered by;

i) the overly restrictive range of scenarios available,
ii) the way scenarios were phrased, and
iii) an inadequate range of motivational hypotheses.

Further analysis and more detailed research would be required to establish the validity of these indications.

The concurrent informant-based method failed to ascribe function mainly where it failed to generate a sufficient quantity of usable data. Staff appeared to be highly selective

about recording, and the number of entries on ABC charts ranged from 0-77. Although there is no direct evidence to be found in the results, a strong impression was formed over the period of the study that staff operated an unwritten rule about how much data was to be collected. The rule may have been formed on the basis of what staff believed was the minimum amount necessary to satisfy the clinician or fulfil their job description. The perceived relevance of functional assessment may have contributed to the variability found in the quality and rate of recording. At a more practical level there were clearly problems associated with asking staff to record very high rate behaviour, such as stereotypy, which may also have had an unclear onset, and staffs' capacities for recording events were no doubt hampered by other work demands. Had staff been trained as observers, or had the observers been externally supplied, then the ABC method may have generated more usable data. Previous reviews have cited mediated observation as a weakness of the ABC method (Reed and Head, 1993), none has specifically identified the non-selection of behaviour as a factor affecting prediction rate or convergent validity.

Experimental method

Direct observation of the experimental analogue conditions presented no special difficulties. A simple extension of the coding protocol could, however, have enhanced data capture to allow an accurate determination of the temporal sequencing of responses and events, and to establish the integrity of assessment conditions (Sturmey, 1995). There is a strong case for assessing the integrity of assessment conditions. Apart from checking potential violations of procedure, there is a

need to establish beyond doubt that the planned contingencies actually operate during assessment. For instance, the analogue condition which is designed to test the social-attention hypothesis aims to deliver social-attention contingent upon the occurrence of a response. The contingency, however, relies upon the emission of a response in the presence of the person running the session. If no response is emitted the effect is to double the number of sessions in which there a person is present but not interacting (alone), and the social-attention hypothesis remains un-tested.

A question not addressed directly by the present study concerns the relationship between the density of reinforcement in the natural environment and the optimal duration of sessions. The effects of extinction, deprivation and satiation may vary according to the richness of the prevailing schedule in the natural setting, and as Sturmey (1995) points out there may also be implications for the number and sequencing of sessions. The present study limited the presentation of each condition to four ten-minute sessions, presented in a quasi-random order. The possibility remains that the performance of the analogue method might have been improved had the number of sessions, their duration and their ordering, been determined on an individual basis.

Descriptive method

The direct descriptive method utilised a specially tailored coding protocol which allowed potentially salient events to be captured in time as they occurred in the natural setting. The

main weakness of this method lay in sampling given the barren nature of the environment.

Analysis and interpretation

Informant-based methods

The interview and ABC methods each suffered from the absence of an objective and clearly defined method for completing a content analysis; although the level of convergence between these methods suggests a reasonable degree of stability was achieved in this case. Both methods were vulnerable, however, to subjective (mis)interpretation in two key stages of the assessment process; data gathering and data analysis. A clear advantage of the rating scales lay in the relative ease with which informant-based data was interpreted. While the rating scales data source was narrow, the scoring system minimised subjective interpretation. Emerson and Bromley (1995), however, showed how using different criteria produced a different pattern of results, and this ambiguity needs to be removed if convergent validity is to be improved.

Experimental method

A number of difficulties were apparent when interpreting data generated by the experimental method. The rigorous criteria used to the select analogue data meant the potential influence of biological setting conditions, novelty and sequencing effects may not have been fully considered. Uniformly high responding was not interpreted as supporting an automatic reinforcement hypothesis because alternative explanations could not be ruled out. Moreover, strict adherence to the accepted interpretation of elevated responding could have resulted in an

unknown number of erroneous classifications being made. A further difficulty of interpretation was exemplified by the (alone) condition where no contact or activity materials were available. The conventional interpretation of elevated responding would suggest automatic reinforcement, but observer presence could for some people have been discriminative for the onset of behaviour actually maintained by social-attention. The pervasive atmosphere of deprivation found in the natural environment, and observer presence throughout all assessment conditions, could have served to reinforce a discriminative rather than reinforcing function for this condition.

Descriptive method

The objective interpretation of descriptive data was maximised by referencing statistical output to predetermined criteria. However, functional relationships could only be implied from simple temporal relationships, and the nature of the environment made this an uncertain affair. Sampling problems were encountered because the prevailing environment was stable. Moreover, observer effect was not examined. One of the main difficulties with this particular approach, however, to was the complex nature of analysis and the amount of time required for data processing.

V. CHAPTER SUMMARY

Convergent validity ranged from moderate to good. The relative rate of prediction and non-prediction was, however, more problematic and adversely affected the overall rate of convergence. No assessment method was universal in its scope for ascribing function. Informant-based methods ascribed

function at a higher rate for outer-directed behaviours, and the descriptive methods for behaviours with an enduring presence. The experimental and informant-based descriptive methods each ascribed at a relatively even rate across all classes of behaviour. Informant-based methods were adversely affected by poor response identification, the descriptive methods by sampling difficulties, and the experimental method by low-rate or undifferentiated patterns of responding.

Convergence was greater among the primary functions assigned by the retrospective informant-based methods. These methods assigned function for a similar sub-set of topographies. The descriptive method obtained a lower rate of convergence and assigned function for a substantially different sub-set of topographies. None of the assessment methods agreed secondary or multi-function predictions at an appreciable level; and disagreement on primary function was often at a fundamental level. The level of precision required in functional assessment has yet to be determined, especially where material enrichment is indicated by conditions of extreme environmental deprivation.

The social context of assessment appeared to affect assessment methods in different ways. The concurrent and retrospective informant-based methods seemed to be influenced by observer/respondent beliefs and their level of training in behaviour theory. The descriptive and experimental methods appeared to be affected equally but differently by the stability of the highly impoverished natural setting.

Three assessment methods showed a proclivity for ascribing from three reinforcement categories; two were consistent with other reports, and although potential bias could be explained in part by a low prediction rate, this is an area worthy of further investigation.

There was correspondence between assessment methods on the functions ascribed to different classes of topography. Negative reinforcement was more likely to be ascribed for behaviour classed aggressive, and automatic reinforcement for stereotyped responding. Other functions were distributed evenly across other classes of topography. Nonetheless, a prima-facie case was made for the further investigation of potential relationships between the form and function of challenging behaviour.

Finally, aspects of assessment design may have interacted to influence both prediction rate and convergence. Following the system of classification described in chapter one it was apparent that;

- indirect, informant-based, methods ascribed function at a higher rate when they were applied retrospectively as opposed to concurrently,
- direct naturalistic methods ascribed more functions than direct modified methods,
- descriptive methods ascribed at a higher rate when observation was direct rather than indirect.

Cross-validation of ascribed function was stronger among the informant-based retrospective methods. Prediction rate is, however, an incomplete measure of efficacy, and convergent validity does not imply predictive validity.

Summary of implications

Berg and Wacker (1991), are unequivocal in stating that questionnaires and surveys should not be relied upon for a functional assessment of challenging behaviour; if they have a place at all, it is restricted to narrowing the potential field of inquiry. Present results support this position, although it may be framed too strongly. Certain approaches to assessment appeared to be inherently more suited to certain behaviour types or features of the presenting social environment. What may be required is wider range of multi-assessment complexes. Moreover, the relative rate of prediction by the analogue method challenges the view that an experimental analysis would normally be the method of choice, although without referencing prediction rate to predictive validity this is position remains uncertain. Imperfect convergence between methods raises questions about the validity of the functions ascribed. It may be, however, that alternate approaches to functional assessment simply highlight different aspects of behaviour that is multiply controlled, and the function of which varies according to a multiplicity of interacting contextual variables. Assessments methods may be more or less suited to certain individuals, features of behaviour and other situational factors. It may be, of course, that some methods are better suited to particular stages of functional assessment. If so, an examination of the effects of situational variables at each stage of a multi-assessment complex might be worthy of consideration. Assessment methods should be viewed as complementary rather than alternative, and tailored to the person, setting context, topography and hypothesised function.

The validity of functional assessment may be improved by testing sequentially refined hypotheses function, which are based on a broader taxonomy of reinforcement categories, and includes a pre-selection assessment phase to determine;

- person-related variables - such as degree of disability, sensory impairment, and level of verbal functioning. Further work is required to determine which of these may be important,
- topographical variables - such as the expected rate, duration and intensity, discreteness of onset, and a notional or expected probability of each hypothesised function being confirmed by different assessment methods,
- contextual variables - of the natural environment - such as the expected rate of activity, demand, and social contact,
- contextual variables - potential informants - such as their level of training in behaviour theory, their observational and reporting skills; perhaps by using an inventory or questionnaire,
- assessment variables - selecting the optimal assessment complex and matching this with the amount of time, the skills and equipment available.

An alternative would be to adopt a more standardised approach to functional assessment such as that described by O'Neill (1990) and McBrien and Felce (1992). Such an approach might begin with a combination of interview, rating scale and a period of direct observation. This phase would be principally concerned with the generation or elimination of competing hypotheses. Convergence would be sought in the functions assigned. Experimental methods may then be used to test very specific hypotheses about maintaining variables or potential intervention strategies (Repp, Felce and Barton, 1988). What is clear from the present study is that a complete functional assessment cannot be made by one method alone.

VI. CONCLUSIONS

Challenging behaviour touches the lives of many. Chapter one reviewed the construction, definition, and prevalence of challenging behaviour shown by people with learning disabilities. The personal and social consequences of challenging behaviour were also reviewed, and presented together with a recent chronology of developments in the separate but related worlds of applied research and clinical practice. It was suggested there that the potential contribution of applied behaviour analysis has hitherto not been fully realised. Practitioners need a technology with which to operationalise the ideological objective of providing a comprehensive range of high quality community-based services for people with learning disabilities who also display challenging behaviour; if ideology is ever to become a reality. Experience has begun to temper the resistance to a behavioural approach which was borne of a previous era, the door is once again ajar. Functional assessment may offer a substantial component of the technology services require to assist the significant minority of people with learning disabilities whose behaviour presents a serious challenge. However, the absence of a valid and reliable technology has been cited as a reason for the poor uptake of functional analytic approaches to date. Clinicians need to be convinced of the efficacy and accessibility of functional assessment. Present results suggest further work might profitably address the selection, reliability and validity of multiple assessment methods. The development of a pre-assessment inventory might assist this process. Pre-assessment could include a prior analysis of

variables associated with topography, subject and environment. The latter would include the interactional context of the social environment as well as the level of training and education of potential informants. Such a task might be approached in a number of ways; by literature reviews, reviews of programmed research and the further replication of studies such as the present one. The quality of the assessment process may be enhanced by ensuring that an appropriate level of training is provided for staff who work in settings where challenging behaviours are likely to occur.

The limitations of functional assessment must, however, be recognised. Challenging behaviour is a complex phenomenon that occurs in complex and ever changing social environments. Specific topographies may be multiply controlled. Rule governance and biological determinants may mean challenging behaviour is not entirely contingency shaped. The level of precision required to indicate effective treatment in clinical practice has yet to be determined. Current research suggests, however, that functional assessment must stretch beyond the rudimentary analysis of linear relationships between antecedent, behaviour and consequence. Contrary to Carr's (1995) suggestion, the ABC chart does not appear to be overwhelmingly all that is required in clinical settings; at least until the majority of staff working in service settings obtain an improved level of training in behaviour theory, observation, recording and reporting techniques. Whatever the level of sophistication required, it seems clear that a valid functional assessment cannot be conducted without reference to the situational context in which it is to be applied. Jones

(1995) has already suggested that global functional assessment may not be conceptually valid. Horner (1994) has called for the further development and refinement of the current taxonomy of functions, and Carr (1994) suggests sequentially refined assessment may be advantageous. Oliver (1993) and Murphy (1994) have each begun to advance elaborate models that seek to integrate biological and environmental determinants; these need to be applied, tested and developed more fully in an applied service context, together with ways of addressing the role of setting events, multiple control, non-dyadic relations, relative reinforcer rates, reinforcer quality, reinforcer delay, response force or effort, deprivation, satiation, temporally distant events, and behavioural momentum (Carr, 1994; Horner, 1994; Mace, 1994).

Results of the present study support the view that more sophistication will be required - not less. The quest for a universal methodology may be inappropriate, and people who display challenging behaviour may be better served by the field if it adopts principles which inform a conceptually systematic and individually tailored investigation of function, which is itself constructed on an individualised and context specific basis (Baer Wolf and Risely, 1968). The slavish adherence to a single assessment procedure must not be allowed to replace the slavish adherence to intervention strategies that became common practice during the 1970's. Assessment methods clearly have different strengths and weaknesses. All have a potential for error, and some appear better suited to certain individuals, topographies, hypotheses of function and social contexts, than others (Horner, 1994).

Validity is another issue. To date, experimental approaches have received more attention from researchers than any other assessment method. More work is clearly needed to develop alternate assessment methods to the same level as the experimental approaches, and beyond. In the meantime results form the present study suggest that the routine application of a multiple assessment format is desirable (Durrand, 1991; Durrand and Crimmins, 1990), and that this should perhaps include an informant-based, descriptive and experimental component (e.g. O'Neill, 1990; McBrien and Felce, 1994). However, a systematic investigation of the combined validity of different assessment combinations would extend the present knowledge base considerably.

REFERENCES

REFERENCES

Allen. D. (1994) Towards Meaningful Day Time Activity. In E. Emerson, P. McGill, and J. Mansell. (Eds) Severe Learning Disabilities and Challenging Behaviours: Designing High Quality Services. London. Chapman and Hall.

Axelrod. S. (1987) Functional and Structural Analysis of Behavior: Approaches Leading to Reduced Use of Punishment Procedures? Research in Developmental Disabilities. 8, 165-178.

Azrin. N.H., Bugle. C. and O'Brien. F. (1971) Behavioral Engineering: Two Apparatuses for Toilet Training Retarded Children. Journal of Applied Behavior Analysis. 4, 249-253.

Baer. D. M., Wolf. M. M. and Risley. T. R. (1968) Some Current Dimensions of Applied Behaviour Analysis. Journal of Applied Behaviour Analysis. 1, 91-97.

Bakeman. R. and Gottman. J. M. (1986) Observing Interaction: An Introduction to Sequential Analysis. Cambridge, University Press.

Baumeister. A. A. (1989) Etiologies of Self-injurious and Destructive Behavior. Proceedings of the Consensus Conference on the Treatment of Severe Behavior Problems and Developmental Disabilities. National Institutes of Health. Bethesda, Maryland.

Beasley. F., Hewson. S. and Mansell. J. (1989) MTS: A Handbook for Observers. Canterbury. Centre for the Applied Psychology of Social Care. University of Kent.

Bellamy. G. T., Horner. R. H. and Inman. D. (1979) Vocational Habilitation of Severely Retarded Adults: A Direct Service Technology. Baltimore. University Park Press.

Berg. W. K. and Sasso. G. M. (1993) Transferring Implementation of Functional Assessment procedures from the Clinic to Natural Settings. In J. Reichle and D. P. Wacker. (Eds) Communicative Alternatives to Challenging Behaviour: Integrating Functional Assessment and Intervention Strategies. Baltimore. Brookes.

Berg. W. K. and Wacker. D. P. (1991) The Assessment and Evaluation of Reinforcers for Individuals with Severe Mental Handicap. In B. Remington (Eds) The Challenge of Severe Mental Handicap: A Behaviour Analytic Approach. Chichester. Wiley.

Bijou. S. W., Peterson. R. F. and Ault. M. H. (1968) A Method to Integrate Descriptive and Experimental Field Studies at the Level of Data and Empirical Concepts. Journal of Applied Behavior Analysis. 1, 175-191.

Blunden. R. and Allen. D. (1987) Facing the Challenge: An Ordinary Life for People with Learning Difficulties and Challenging Behaviours. London. King's Fund Centre.

Bostow. D. E. and Bailey. J. B. (1969) Modification of Severe Disruptive and Aggressive Behavior Using Brief Timeout and Reinforcement Procedures. Journal of Applied Behavior Analysis. 2, 31-37.

Bromley. J. and Emerson. E. (1995) Beliefs and Emotional Reactions of Care Staff Working with People with Challenging Behaviour. Journal of Intellectual Disability. 39, 341-352.

Bruininks. R. H. and Lakin K. C. (1985) Living and learning in the Least Restrictive Environment. Baltimore. Brookes.

Burgess. A. (1987) Functional Analysis of Self-Injurious Behaviour Using Direct Observation - A Comparison with Interview and Analogue Methods. Unpublished MSc. Research Dissertation. University of Surrey.

Burgio. L. D., Page. T. J. and Capriotti. R. M. (1985) Clinical Behavioural Pharmacology: Methods for Evaluating Medications and Contingency Management. Journal of Applied Behaviour Analysis. 18, 45-59.

Carr. E. G. (1977) The Motivation of Self-injurious Behavior: A Review of Some Hypotheses. Psychological Bulletin. 84, 800-816.
Carr. E. G. (1994) Emerging Themes in the Functional Analysis of Problem Behavior. Journal of Applied Behaviour Analysis. 27, 393-399.

Carr. E. G. (1995) Positive Approaches to Challenging Behaviour: Practical Methods and Update. Paper presented at; The Way Ahead: British Institute of Learning Disabilities International Conference. Oxford.

Carr. E. G. and McDowell. J. J. (1980) Social Control of Self-Injurious Behavior of Organic Etiology. Behavior Therapy. 11, 402-409.

Carr. E. G., Newsom. C. D. and Binkoff. J. A. (1976) Stimulus Control of Self-Destructive Behavior in a Psychotic Child. Journal of Abnormal Child Psychology. 4, 139-153.

Carr. E. G and Durrand. V. M. (1985) Reducing Behaviour Problems Through Functional Communication Training. Journal of Applied Behavior Analysis. 18, 111-126.

Carr. E. G., Robinson. S. and Palumbo. L. W. (1990) The Wrong Issue: Aversive Versus Non-aversive Treatment. The Right Issue: Functional Versus Non-functional Treatment. In A. C. Repp. and N. N. Singh. (Eds) Perspectives on the Use of Nonaversive and Aversive Interventions for Persons with Developmental Disabilities. Sycamore Illanois, Sycamore Publishing Co.

Clarke. D. (1993) Psychopharmacological Approaches to Challenging Behaviour. In I. Fleming and B. Stenfert Kroese. (Eds) People with Learning Disability and Severe Challenging Behaviour: New developments in Services and Therapy. Manchester. Manchester University Press.

Cline. T. (1985) Clinical Judgement in Context: A Review of Situational Factors in Person Perception During Clinical Interviews. Journal of Child Psychology and Psychiatry. 26, 369-380.

Cohen. J. (1960) A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement. 20, 37-46.

Cooper. L. J. Wacker. D. P., Sasso. G., Reimers. T. B. and Donn. L. A. (1990) Using Parents as Therapists to Evaluate Appropriate Behavior. Journal of Applied Behavior Analysis. 23, 285-296.

Corte. H. E., Wolf., M. M. and Locke. B. J. (1971) A Comparison of Procedures for Eliminating Self-injurious Behavior of Retarded Adolescents. Journal of Applied Behavior Analysis. 4, 210-213.

Crawford. J., Brockel. B., Schauss. S. and Miltenberger. R. G. (1992) A Comparison of Methods for the Functional Assessment of Stereotypic Behavior. Journal for the Association of People with Severe Handicaps. 17, 77-86.

Department of Health, Social Security, Welsh Office and Scottish Office (1989) Caring for People: Community Care in the Next Decade and Beyond. (Cm 849). London. HMSO.

Derby. K. M., Wacker. D. P., Sasso. G., Steege. M., Northup. J., Cigrand. K. and Asmus. J. (1992) Brief Functional Assessment Techniques to Evaluate Aberrant Behavior in an Outpatient Setting: A Summary of 79 Cases. Journal of Applied Behavior Analysis. 25, 713-721.

Derby. K. M., Wacker. D. P., Peck. S., Sasso. G., DeRaad. A., Berg. W., Asmus. J. and Ulrich. S. (1994) Functional Analysis of Separate Topographies of Aberrant Behavior. Journal of Applied Behavior Analysis. 27, 267-278.

Dorsey. M. F., Iwata. B. A., Ong. P. and McSween. T. E. (1980) Treatment of Self-injurious Behavior Using Water Mist: Initial Response Suppression and Generalization. Journal of Applied Behavior Analysis. 13, 343-353.

Durand. V. M. (1990) Severe Behaviour Problems: A Functional Communication Training Approach. New York. The Guilford Press.

Durand. V. M. and Carr. E. G. (1991) Functional Communication Training to Reduce Challenging Behavior: Maintenance and Application in New Settings. Journal of Applied Behavior Analysis. 24, 251-264.

Durand. V. M. and Crimmins. D. (1988) Identifying the Variables Maintaining Self-Injurious Behavior. Journal of Autism and Developmental Disorders. 18, 99-117.

Durand. V. M. and Crimmins. D. (1990) Chapter 3. In V. M. Durand. (Auth) Severe Behaviour Problems: A Functional Communication Training Approach. New York. The Guilford Press.

Durand. V. M. and Crimmins. D. (1991) Teaching Functionally Equivalent Responses as an Intervention for Challenging Behaviour. In B. Remington (Eds) The Challenge of Severe Mental Handicap: A Behaviour Analytic Approach. Chichester. Wiley.

Durand. V. M. and Crimmins. D. (1992) The Motivation Assessment Scale (MAS) Administration Guide. Topeka, KS. Monaco and Associates Inc.

Elderson. S. M., Taubman. M. T. and Lovaas. O. I. (1983) Some Social Contexts of Self-Destructive Behavior. Journal of Abnormal Child Psychology. 11, 199-312.

Emerson. E. (1990) Self Injurious Behaviour - Some of the Challenges it Presents. Mental Handicap. 18, 92-98.

Emerson. E. (1995) Service Developments: Peripatetic Support Teams and Community Housing - where are we up to?. Paper presented at: The Way Ahead: British Institute of Learning Disabilities International Conference. Oxford.

Emerson. E., Barrett. S., Bell. C., Cummings. R., McCool. C. Toogood. A. and Mansell. J. (1987) Developing Services for People with Severe Learning Difficulties and Challenging Behaviours. Canterbury. Institute of Social and Applied Psychology. University of Kent.

Emerson. E. and Bromley. J. (1995) The Form and Function of Challenging Behaviours. Journal of Intellectual Disability Research. 39, 388-398.

Emerson. E., Cummings. R. and Barrett. S. (1989) Using Analogue Assessment. Canterbury. Centre for Applied Psychology of Social Care. University of Kent.

Emerson. E. and Hatton. C. (1994) Moving Out: The Impact of Relocation from Hospital to Community on the Quality of Life of People with Learning Disabilities. London. H.M.S.O.

Emerson. E., Felce. D., McGill. P. and Mansell. J. (1994) Introduction. In E. Emerson, P. McGill, and J. Mansell. (Eds) Severe Learning Disabilities and Challenging Behaviours: Designing High Quality Services. London. Chapman and Hall.

Emerson. E., Thomson. S., Reeves. D., Henderson. D. and Robertson. J. (1995) Descriptive Analysis of Multiple Response Topographies of Challenging Behavior Across Two Settings. Research in Developmental Disabilities. 16, 301-329.

Emerson. E., Toogood. A., Mansell. J., Barrett. S., Bell. C., Cummings. R. and McCool. C. (1987) Challenging Behaviour and Community Services 1. Mental Handicap. 15, 166-169.

Epstein. L. H., Doke. L. A., Sajwaj. T. E., Sorrell. S. and Rimmer. B. (1974) Generality and Side Effects of Overcorrection. Journal of Applied Behavior Analysis. 7, 385-390.

Felce. D. (1993a) Ordinary Housing: A Necessary Context for Meeting Service Philosophy and Providing an Effective Therapeutic Environment. In R.S.P. Jones and C.B. Eayrs (Eds) Challenging Behaviour and Intellectual Disability: A Psychological Perspective. Clevedon, Avon. B.I.L.D. Publications.

Felce. D. and Lowe. K. (1993) Supporting People with Severe Learning Disabilities and Challenging Behaviours in Ordinary Housing. In C. Kiernan (Ed) Research to Practice? Implications of Research on the Challenging Behaviour of People with Learning Disability. Clevedon, Avon. B.I.L.D. Publications.

Fleiss. J. L. (1981) Statistical Methods for Rates and Proportions. New York. Wiley. Cited in R. Bakeman and J. M. Gottman (1986) Observing Interaction: An Introduction to Sequential Analysis. Cambridge. Cambridge University Press.

Forrest. J., Cambridge. P., Emerson. E., Mansell. J., Asbury. M. and Beecham. J. (1995) Community Support Teams for People with Learning Disabilities and Challenging Behaviour: Final Report to the Joseph Rowntree Foundation and the Mental Health Foundation. The Tizard Centre, University of Kent at Canterbury.

Foxx. R. M. and Azrin. N.H. (1973) The Elimination of Autistic Self-stimulatory Behavior by Overcorrection. Journal of Applied Behavior Analysis. 6, 1-14.

Foxx. R. M. and Shapiro. S. T. (1978) The Timeout Ribbon: A Nonexclusionary Timeout Procedure. Journal of Applied Behavior Analysis. 11, 125-136.

Gold. M. (1975) Try Another Way. Champaign Research Press.

Goldiamond. I (1974) Toward a Constructional Approach to Social Problems. Behaviorism. 1, 1-84.

Gottman. J. M. and Roy. A. K. (1990) Temporal Form: Detecting Sequences in Social Interaction. New York. Cambridge University Press.

Guess. D., Helmstetter. E., Turnbull. H. R. and Knowlton. S. (1987) Use of Aversive Procedures with Persons who are Disabled: An Historical Review and Critical Analysis. Seattle, Washington. The Association for Persons with Severe Handicaps.

Hall. S. and Oliver. C. (1992) Differential Effects of Severe Self-injurious Behaviour on the Behaviour of Others. Behavioural Psychotherapy. 20, 355-365.

Harding. J., Wacker. D. P., Cooper. L. J., Millard. T. and Jensen-Kovalan. P. (1994) Brief Hierarchical Assessment of Potential Treatment Components with Children in an Outpatient Clinic. Journal of Applied Behavior Analysis. 27, 291--300.

Harris. P. (1993) The Nature and Extent of Aggressive Behaviour Amongst People with Learning Disabilities (Mental Handicap) in a Single Health District. Journal of Intellectual Disabilities Research. 37. 221-242.

Hastings. R. P. (1995) Understanding Factors that Influence Staff Responses to Challenging Behaviours. Mental Handicap Research. 8, 296-320.

Hastings. R. P., Remington. B. and Hopper. G. M. (1995) Experienced and Inexperienced Health Care Workers' Beliefs About Challenging Behaviours. Journal of Intellectual Disability. 39, 474-483.

Heal. L. W., Haney. J. I. and Novak-Amado. A. R. (1999) Integration of Developmentally Disabled Individuals into the Community. Baltimore. Brookes.

Horner. R. H. (1994) Functional Assessment: Contributions and Future Directions. Journal of Applied Behavior Analysis. 27, 401-404.

Horner. R. H. and Keilitz. I. (1975) Training Mentally Retarded Adolescents to Brush their Teeth. Journal of Applied Behavior Analysis. 8, 301-309.

Intagliata. J., and Willer. B. (1982) Reinstitutionalization of Mentally Retarded Persons Successfully Placed into Family-Care and Group Homes. American Journal of Mental Deficiency. 87, 34-39.

Iwata. B. A., Dorsey. M. F., Slifer. K. J., Bauman. K. E. and Richman. G. S. (1982) Toward a Functional Analysis of Self-injury. Analysis and Intervention in Developmental Disabilities. 2, 3-20.

Iwata. B. A. (1994) Functional Analysis Methodology: Some Closing Comments. Journal of Applied Behavior Analysis. 27, 413-418.

Iwata. B. A., Pace. G. M., Dorsey. M. F., Zarcone. J. R., Vollmer. T. R., Smith. R. G., Rodgers. T. A., Lerman. D. C., Shore. B. A., Mazaleski. J. L., Goh. H., Edwards Cowdery. G., Kalsher. M. J., McCosh., K. C. and Willis. K. D. (1994) The Functions of Self-injurious Behavior: An Experimental-Epidemiological Analysis. Journal of Applied Behavior Analysis. 27, 215-240.

Iwata. B. A., Vollmer. T. R. and Zarcone. J. R. (1990) The Experimental (Functional) Analysis of Behavior Disorders: Methodology, Applications and Limitations. In A. C. Repp. and N. N. Singh. (Eds) Perspectives on the Use of Nonaversive and Aversive Interventions for Persons with Developmental Disabilities. Sycamore Illanois, Sycamore Publishing Co.

Janicki. M. P, Wyngaarden-Krauss. M. and Seltzer. M. M. (1988) Community Residences for persons with Developmental Disabilities Here to Stay. Baltimore. Brookes.

Johnson. W. L. and Day. R. M. (1992) The Incidence and Prevalence of Self-Injurious Behavior. In J. K. Luiselli., J. L. Matson. and N. N. Singh (Eds) Self-injurious Behaviour: Analysis, Assessment and Treatment. New York. Springer-Verlag.

Jones. R. S. P. (1991) Reducing Inappropriate Behaviour Using Nonaversive Procedures: Evaluating Differential Reinforcement Schedules. In B. Remington. (Ed) The Challenge of Severe Mental Handicap: A Behaviour Analytic Approach. Chichester. Wiley and Sons.

Jones. R. S. P. (1995) Functional Analysis: Still Cautious After All These Years. Paper Presented at: The St John of God Positive Programming Conference. Dublin.

Jones. R. S. P. and Eayrs. C. B. (1993) Challenging Behaviour and Intellectual Disabilities: An Overview. In R. S. P. Jones and C. B. Eayrs (Eds) Challenging Behaviour and Intellectual Disability: A Psychological Perspective. Clevedon, Avon. B.I.L.D. Publications.

Jones. R. S. P., Lowe. F. and Williams. H. (1993) Verbal Self-Regulation. In I. Fleming and B. Stenfert Kroese. (Eds) People with Learning Disability and Severe Challenging Behaviour: New Developments in Services and Therapy. Manchester. Manchester University Press.

Jones. R. S. P., Walsh. P. G. and Sturmey. P. (1995) Stereotyped Movement Disorders. Chichester. Wiley.

Kearney. C. A. (1994) Interrater Reliability of the Motivation Assessment Scale: Another Closer Look. Journal of the Association for Persons with Severe Handicaps. 19, 139-142.

Kiernan. C. (1991) Professional Ethics: Behavioural Analysis and Normalization. In B. Remington. (Ed) The Challenge of Severe Mental Handicap: A Behaviour Analytic Approach. Chichester. Wiley and Sons.

Kiernan. C. and Qureshi. H. (1993) Challenging Behaviour. In C. Kiernan (Ed) Research to Practice? Implications of Research on the Challenging Behaviour of People with Learning Disability. Clevedon. Avon. B.I.L.D. Publications.

King's Fund Centre (1981) An Ordinary Life. (Project Paper 24). London. King's Fund Centre.

King's Fund Centre (1984) An Ordinary Working Life: Vocational Services for People with Mental Handicap. (Project Paper 50) London. King's Fund Centre.

Knapp. M. and Mansell. J (1994) Assessing Costs and Benefits. In E. Emerson, P. McGill, and J. Mansell. (Eds) Severe Learning Disabilities and Challenging Behaviours: Designing High Quality Services. London. Chapman and Hall.

Lakin. K. C. and Bruininks. R. H. (1985) Strategies for Achieving Community Integration of Developmentally Disabled Citizens. Baltimore. Brookes.

Lalli. J. S., Browder. D. M., Mace. F. C. and Brown. D. K. (1993) Teacher Use of Descriptive Analysis Data to Implement Interventions to Decrease Students' Problem Behavior. Journal of Applied Behavior Analysis. 26, 227-238.

LaVigna. G. W. and Donnellan. A. M. (1986) Alternatives to Punishment: Solving Behavior Problems with Non-aversive Strategies. New York. Irvington.

Lerman. D. C. and Iwata. B. A. (1993) Descriptive and Experimental Analyses of Variables Maintaining Self-injurious Behaviour. Journal of Applied Behavior Analysis. 26, 293-319.

Lerman. D. C., Iwata. B. A., Smith. R. G. and Vollmer. T. R. (1994) Restraint Fading and the Development of Alternative Behaviour in the Treatment of Self-Restraint and Self-Injury. Journal of Intellectual Disability Research. 38, 135-148.

Lovass. O. I. (1982) Comments on Self-Desctructive Behaviors. Analysis and Intervention in Developmental Disabilities. 2, 115-124.

Lovaas. O. I. and Simmons. J. Q. (1969) Manipulation of Self-Destruction in Three Retarded Children. Journal of Applied Behavior Analysis. 2, 143-157.

Lowe. K., and Felce. D. (1995a) Challenging Behaviour in the Community. In Challenging Behaviour: What we Know. Mental Health Foundation.

Lowe. K., and Felce. D. (1995b) How do Carers Assess the Severity of Challenging Behaviour? A Total Population Study. Journal of Intellectual Disability Research. 39, 117-127.

Martin. J. P. (1984) Hospitals in Trouble. Oxford. Blackwell.

Mace. C. F. (1994) The Significance and Future of Functional Analysis Methodologies. Journal of Applied Behavior Analysis. 27, 385-392.

Mace. C. F., Lalli. J. S. and Lalli. E. P. (1991) Functional Analysis and Treatment of Aberrant Behavior. Research in Developmental Disabilities. 12, 155-180.

Mansell. J. (1994) Challenging Behaviour: The Prospect for Change. British Journal of Learning Disabilities. 22, 2-5.

Mansell. J., Felce. D., Jenkins. J., De-Kock. U. and Toogood. S. (1987) Developing Houses for People with Mental Handicaps. Tonbridge Wells. Costello.

Matson. J. L. and Gardner. W. I. (1991) Behavioral Learning Theory and Current Applications to Severe Behavior Problems in Persons with Mental Retardation. Clinical Psychology Review. 11, 175-183.

McBrien. J. and Felce. D. (1992) Working with People Who Have Severe Learning Difficulty and Challenging Behaviour; A Practical Handbook on the Behavioural Approach. Kidderminster. BIMH Publications.

McGill. P., Hewson. S. and Emerson. E. (1994) CTS: A Software Package for the Collection of Observational Data on Psion Organisers. Canterbury. Tizard Centre. University of Kent.

Meyer. L. H. and. Evans. I. M. (1989) Nonaversive Intervention for Behavior Problems: A Manual for Home and Community. Baltimore. Brookes.

Moran. G., Dumas. J. E. and Symons. D. K. (1992) Approaches to Sequential Analysis and the Description of Contingency in Behavioral Interaction. Behavioral Assessment. 14, 65-92.

Murphy. G. H. (1994) Understanding Challenging Behaviour. In E. Emerson, P. McGill, and J. Mansell. (Eds) Severe Learning Disabilities and Challenging Behaviours: Designing High Quality Services. London. Chapman and Hall.

Murphy. G. and Oliver. C. (1987) Decreasing Undesirable Behaviours. In W. Yule and J. Carr. (Eds) Behaviour Modification for People with Mental Handicaps. Beckenham. Croom Helm.

Murphy. G. H., Oliver. C., Corbett. J., Crayton. L, Hales. J., Head. D and Hall. S. (1993) Epidemiology of Self-Injury, Characteristics of People with Severe Self-Injury and Initial Treatment Outcome. In C. Kiernan (Ed) Research to Practice? Implications of Research on the Challenging Behaviour of People with Learning Disability. Clevedon, Avon. B.I.L.D. Publications.

Newton. J. T. and Sturmey. P. (1991) The Motivation Assessment Scale: Interrater Reliability and Internal Consistency in a British Sample. Journal of Mental Deficiency Research. 35, 472-474.

Northup. J. Wacker. D. P., Sasso. G., Steege. M. Cigrand. K. Cook. J. and DeRaad. A. (1991) A Brief Functional Analysis of Aggressive and Alternative Behavior in an Out-clinic Setting. Journal of Applied Behavior Analysis. 24, 509-522.

Northup. J., Wacker. D. P., Berg. W. K., Kelly. L., Sasso. G. and DeRaad. A. (1994) The Treatment of Severe Behaviour Problems in School Settings Using a Technical Assistance Model. Journal of Applied Behavior Analysis. 27, 33-47.

O'Brien. F. and Azrin. N. H. (1972) Developing Proper Mealtime Behaviors of the Institutionalized Retarded. Journal of Applied Behavior Analysis. 5, 389-399.

Oliver. C. (1991a) The Application of Analogue Methodology to the Functional Analysis of Challenging Behaviour. In B. Remington. (Ed) The Challenge of Severe Mental Handicap: A Behaviour Analytic Approach. Chichester. Wiley and Sons.

Oliver. C. (1991b) Self-Injurious Behaviour in People with Mental Handicap. Prevalence, Individual Characteristics and Functional Analysis. Unpublished PhD thesis. University of London.

Oliver. C. (1993) Self Injurious Behaviour: From Response to Strategy. In C. Kiernan (Ed) Research to Practice? Implications of Research on the Challenging Behaviour of People with Learning Disability. Clevedon, Avon. B.I.L.D. Publications.

Oliver. C. (1995) How the Causes of Self-Injurious Behaviour are Perceived by Those in Close Contact with People Who Show Self-Injury. Paper presented at; The Way Ahead: British Institute of Learning Disabilities International Conference. Oxford.

Oliver. C., Crayton. L., Murphy. G. and Corbett. J. (undated) Formal Functional Analysis of Self Injurious Behaviour: The Further Development of Analogue Conditions. Undated Manuscript.

Oliver. C. and Head. H. (1993) Self-Injurious Behaviour: Functional Analysis and Interventions. In R. S. P. Jones and C. B. Eayrs (Eds) Challenging Behaviour and Intellectual Disability: A Psychological Perspective. Clevedon, Avon. B.I.L.D. Publications.

Oliver. C., Murphy. G. and Corbett J. A. (1987) Self-Injurious Behaviour in People with Mental Handicap: A Total Population Study. Journal of Mental Deficiency Research. 31, 147-162.

O'Neill. R. E., Horner. R. H., Albin. R. W., Storey. K. and Sprague. J. R. (1990) Functional Analysis of Problem Behavior: A Practical Assessment Guide. Sycamore. Sycamore Publishing Co.

Pace. G. M., Ivancic. M. T. and Jefferson. G. (1994) Stimulus Fading as Treatment for Obscenity in a Brain-Injured Adult. Journal of Applied Behavior Analysis. 27, 301-305.

Pahl. J. and Quine. L. (1985) Families with Mentally Handicapped Children: A Study of Stress and Service Response. Health Services Research Unit. University of Kent.

Panyan. M., Boozer. H. and Morris. N. (1970) Feedback to Attendants as a Reinforcer for Applying Operant Techniques. Journal of Applied Behavior Analysis. 3. 1-4.

Parrish. J. M., Iwata. B. A., Dorsey. M. F., Bunck. T. J. and Slifer. K. J. (1985) Behavior Analysis: Program Development and Transfer of Control in the Treatment of Self-Injury. Journal of Behavior Therapy and Experimental Psychiatry. 16, 159-168.

Porterfield. J. and Blunden. R. (1978) Establishing an Activity Period and Individual Skill Training within a Day Setting for Profoundly Mentally Handicapped Adults. Journal of Practical Approaches to Developmental Handicap. 2, 10-15.

Porterfield. J., Blunden. R. and Blewitt. E. (1980) Improving Environments for Profoundly Handicapped Adults Using Prompts and Social Attention to Maintain High Group Engagement. Behavior Modification. 4, 225-241.

Porterfield. J. and Gathercole. C. (1985) The Employment of people with Mental Handicap: Progress Towards an Ordinary Working Life. London. King's Fund Centre.

Presland. J. L. (1989) Overcoming Difficult Behaviour: A Guide and Source Book for Helping People with Severe Mental Handicaps. Kidderminster. BIMH Publications.

Quilitch. H. R. (1975) A Comparison of Three Staff-Management Procedures. Journal of Applied Behavior Analysis. 8, 59-66.

Quine. L. and Pahl. J. (1985) Examining the Causes of Stress in Families with Severe Mentally Handicapped Children. British Journal of Social Work. 51, 510-517.

Qureshi. H. (1993a) Prevalence of Challenging Behaviour in Adults. In I. Fleming and B. Stenfert Kroese. (Eds) People with Learning Disability and Severe Challenging Behaviour: New Developments in Services and Therapy. Manchester. Manchester University Press.

Qureshi. H. (1993b) Impact on Families: Young Adults with Learning Disability who Show Challenging Behaviour. In C. Kiernan (Ed) Research to Practice? Implications of Research on the Challenging Behaviour of People with Learning Disability. Clevedon. Avon. B.I.L.D. Publications.

Qureshi. H. (1994) The Size of the Problem. In E. Emerson, P. McGill, and J. Mansell. (Eds) Severe Learning Disabilities and Challenging Behaviours: Designing High Quality Services. London. Chapman and Hall.

Qureshi. H. and Alborz. A. (1992) Epidemiology of Challenging Behaviour. Mental Handicap Research. 5, 130-145.

Reed. J. and Head. D. (1993) The Application of Functional Analysis in the Treatment of Challenging Behaviour. In I. Fleming and B. Stenfert Kroese. (Eds) People with Learning Disability and Severe Challenging Behaviour: New developments in Services and Therapy. Manchester. Manchester University Press.

Reeves. D. (1994a) Assessing Reliability in Time Sequenced Observational Data. Manchester. Hester Adrian Research Centre. University of Manchester.

Reeves. D. (1994b) HARCLAG: A Program for the Sequential Analysis of Observational Data. Manchester. Hester Adrian Research Centre. University of Manchester.

Remington. B. (1991) Behaviour Analysis and Severe Mental Handicap: The Dialogue Between Research and Application. In B. Remington. (Ed) The Challenge of Severe Mental Handicap: A Behaviour Analytic Approach. Chichester. Wiley and Sons.

Remington. B. (1993) Challenging Behaviour in People with Severe Learning Disabilities Behaviour Modification or Behaviour Analysis? In C. Kiernan (Ed) Research to Practice? Implications of Research on the Challenging Behaviour of People with Learning Disability. Clevedon. Avon. B.I.L.D. Publications.

Repp. A. C. and Felce. D. (1990) A Microcomputer System Used for Evaluative and Experimental Behavioural Research in Mental Handicap. Mental Handicap Research. 3, 21-32.

Repp. A. C., Felce. D. and Barton. L. E. (1988) Basing the Treatment of Stereotypic and Self-Injurious Behaviors on Hypotheses of their Causes. Journal of Applied Behavior Analysis. 21, 281-289.

Repp. A. C., and Karsh. K. G. (1994) Hypothesis-Based Interventions for Tantrum Behaviors of Persons with Developmental Disabilities in School Settings. Journal of Applied Behavior Analysis. 27, 21-31.

Repp. A. C., Felce. D. and Karsh. K. G. (1991) The Use of a Portable Microcomputer in the Functional Analysis of Maladaptive Behaviour. In B. Remington. (Ed) The Challenge of Severe Mental Handicap: A Behaviour Analytic Approach. Chichester. Wiley and Sons.

Repp. A. C. and Singh. N. N. (1990) Perspectives on the Use of Nonaversive and Aversive Interventions for Persons with Developmental Disabilities. Sycamore, Illanois. Sycamore Publishing Co.

Russell. O. and Harris. P. (1993) Assessing the Prevalence of Aggressive Behaviour and the Effectiveness of Interventions. In C. Kiernan (Ed) Research to Practice? Implications of Research on the Challenging Behaviour of People with Learning Disability. Clevedon. Avon. B.I.L.D. Publications.

Sackett. G. P. (1979) The Lag Sequential Analysis of Contingency and Cyclicity in Behavioral Interaction Research. In J. D. Osofsky (Ed) Handbook of Infant Development. New York. Wiley.

Sasso. G. M., Reimers. T. M., Cooper. L. J., Wacker. D. Berg. W., Steege. M., Kelly. L. and Allaire. A. (1992) Use of Descriptive and Experimental Analyses to Identify the Functional Properties of Aberrant Behavior in School Settings. Journal of Applied Behavior Analysis. 25, 809-821.

Sajwaj. T., Libet. J. and Agras. S. (1974) Lemon-Juice Therapy: The Control of Life-Threatening Rumination in a Six-Month-Old Infant. Journal of Applied Behavior Analysis. 7, 557-563.

Sigafoos. J., Kerr. M. and Roberts. D. (1994) Interrater Reliability of the Motivation Assessment Scale: Failure to Replicate with Aggressive Behavior. Research in Developmental Disabilities. 15, 333-342.

Singh. N. N., Donatelli. L. S., Best. A., Williams. D. E. Barrera. F. J. Lenz. M. W. Landrum. T. J., Ellis. C. R. and Moe. T. L. (1993) Factor Structure of the Motivation Assessment Scale. Journal of Intellectual Disability Research. 37, 65-74.

Skinner. B. F. (1953) Science and Human Behaviour. New York. McMillan.

Skinner. B. F. (1957) Verbal Behavior. New York. Appleton-Century-Crofts.

Skinner. B. F. (1971) Beyond Freedom and Dignity. London. Penguin Books.

Solnick. J. V., Rincover. R. A. and Peterson. C. R. (1977) Some Determinants of the Reinforcing and Punishing Effects of Timeout. Journal of Applied Behavior Analysis. 10, 415-424.

Sturmey. P. (1991) Assessing Challenging Behaviour Using a Semi-structured Behavioural Interview: A Case Transcript. Mental Handicap. 19, 56-60.

Sturmey. P. (1995) Analog Baselines: A Critical Review of the Methodology. Research in Developmental Disabilities. 16, 269-284.

Suen. H. K. and Ary. D. (1989) Analyzing Quantitative Behavioral Observation Data. New Jersey. Lawrence Earlbaum Associates.

Tanner. B. A. and Zeiler. M. (1975) Punishment of Self-Injurious Behavior Using Aromatic Ammonia as the Aversive Stimulus. Journal of Applied Behavior Analysis. 8, 53-57.

Taylor. J. C. and Romanczyk. R. G. (1994) Generating Hypotheses about the Function of Student Problem Behavior by Observing Teacher Behavior. Journal of Applied Behavior Analysis. 27, 25-256.

Thompson. S. and Emerson. E. (1995) Inter-Informant Agreement on the Motivation Assessment Scale: Another Failure to Replicate. Mental Handicap Research. 8, 203-208.

Toogood. A., Emerson. E., Hughes. H., Barrett. S., Cummings. R. and McCool. C. (1988) Challenging Behaviour and Community Services 3: Planning Individualised Services. Mental Handicap. 16. 70-74.

Toogood. S. (1993) Normalisation, Community Services and Challenging Behaviour. In R. S. P. Jones and C. B. Eayrs (Eds) Challenging Behaviour and Intellectual Disability: A Psychological Perspective. Clevedon. Avon. B.I.L.D. Publications.

Toogood. S., Bell. A., Jaques. H., Lewis. S., Sinclair. C. and Wright. L. (1994) Meeting the Challenge in Clwyd: The Intensive Support Team. British Journal of Learning Disabilities. 22, 18-24.

Touchette. P. E., MacDonald. R. F. and Langer. S. O. (1985) A Scatter Plot for Identifying Stimulus Control of Problem Behavior. Journal of Applied Behavior Analysis. 18, 343-351.

Vollmer. T. R. (1994) The Concept of Automatic Reinforcement: Implications for Behavioral Research in Developmental Disabilities. Research in Developmental Disabilities. 15, 187-204.

Whitehead. S. (1992) The Social Origins of Normalisation. In H. Brown and H. Smith (Eds) Normalisation A Reader for the Nineties. London. Routledge.

Whitman. T. L., Mercurio. J. R. and Caponigri. V. (1970) Development of Social Responses in Two Severely Retarded Children. Journal of Applied Behavior Analysis. 3, 133-138.

Wolfensberger. W. (1991) A Brief Introduction to Social Role Valorisation as a High Order Concept for Structuring Human Services. Syracuse University. Syracuse.

Wolfensberger. W. and Glenn. L.(1975) Program Analysis of Service Systems. Toronto. National Institute on Mental Retardation.

Wolfensberger. W. and Thomas. S. (1983) Program Analysis of Service Systems Implementation of Normalisation Goals. Toronto. National Institute on Mental Retardation.

Zarcone. J. R., Iwata. B. A., Smith. R. G., Mazaleski. J. L. and Lerman. D. C. (1994) Reemergence and Extinction of Self-Injurious Escape Behavior During Stimulus (Instructional) Fading. Journal of Applied Behavior Analysis. 27, 307-316.

Zarcone. J. R., Rodgers. T. A. and Iwata. B. A. (1991) Reliability Analysis of the Motivation Assessment Scale: A Failure to Replicate. Research in Developmental Disabilities. 12, 349-260.

APPENDICES

Appendix 2.1

Interview - specimen synopsis

Synopsis of interview

Challenging behaviours

Aggression	SIB	Damage	Pica
Stereotypy	Vocalisation	Vomit	Stripping

Aggression - hitting out at staff with his hands in a slapping action, will also kick and bite. These behaviours occur in any setting and have formed part of Sxx's behavioural repertoire for many years. There is no change in their frequency which is maintained at a low rate by staff keeping away. The behaviours could be expected to occur many times a day.

Antecedents proximity of staff, requests to do something or go somewhere
Consequences staff avoid contact, instruction followed through sometime
Potential function social avoidance, task avoidance

SIB - slapping his face very hard with his hand or with an object, e.g. a plastic cup. These behaviours occur in bouts, often associated with high intensity stereotyped rocking and jumping up and down from a crouched position. Multiple hits occur and can be high rate - many times a minute. These behaviours are long standing and have not changed over recent times. they will occur mainly in the Villa and are independent of persons present.

Antecedents no staff or activity
Consequences no contact or intervention
Potential function consistent with automatic and social avoidance hypotheses,

Damage - banging doors, displacing objects, picking window sills, scratching glass, throwing objects. These behaviours occur daily, they are long standing and have not changed over recent times. They will occur whoever is on duty. The behaviours occur mostly on the Vila. Sxx will get up suddenly and run through the Villa banging doors etc. He will throw object placed nearby.

Antecedents no activity or contact, being asked to do something or go somewhere, others enter the room he is occupying.
Consequences behaviour is ignored, sometimes demands are followed through if important
Potential function consistent with automatic and social or task avoidance hypotheses

Pica - placing inedible items into his mouth. This behaviour can occur throughout the day, although there are few items in the room Sxx spends most of his time. he has picked wood from the window sills to place in his mouth. The behaviour is long standing, has not changed and does not depend on staff or resident presence or interaction.

Antecedents no activity or contact
Consequences debris removed
Potential function consistent with automatic, social attention hypotheses

Vocalisation - screaming, shouting, speaking loudly, talking to himself with increasing volume. The behaviours occur daily and have been present for many years. They co-occur with stereotyped rocking and when Sxx is crouched on the sofa or plastic mattress with his jumper pulled over his head. The behaviours do not depend on staff or resident presence. Sometimes vocalisations refer directly to demands for coffee.

Antecedent no contact or activity, or persons entering the room
Consequences. vary - ignored, allowed coffee, social/verbal response from staff
Potential function tangible, social attention, automatic reinforcement hypotheses

Stereotypy - rocking to and fro on his knees, bouncing up and down in a crouched position on his knees, twisting his body (while kneeling) from the trunk upwards, turning his head to the left, stretching out his right arm upwards and behind his back, often while crouched, placing his jumper over his head and rocking hi whole body side to side while crouched. These behaviours occur throughout a substantial part of every day. They are long standing and have changed little over time. The behaviours occur mostly in the room Sxx spends most of his day. Music plays all day - some music appears to intensify these responses - at other times responses will intensify apparently independent of the music. Occasionally bouts will become very intense and will co-occur with SIB, vocalisation and other challenging behaviour.

Antecedent no contact or activity
Consequence mostly not responded to by staff
Potential function consistent with automatic and social avoidance hypotheses.

Vomiting - occurs sometimes after meals. Vomiting is believed to be self-induced and is produced in measured quantity. Sxx is very slightly built and occasionally misses meals - he is asked or told meals are available but not bound to partake.

Antecedent eating
Consequence changed into clean clothes
Potential function consistent with automatic and social attention hypotheses

Stripping - removing clothing outer garments, usually shirts or jumper. Can occur on more than daily basis, usually when alone in the room where he spends most of his day. Sxx will spill coffee and food on his clothing which leaves them damp and uncomfortable. he is also incontinent of urine throughout the day.

Antecedent no contact or activity
Consequence re-dressed or changed at some point
Potential functions tangible, social attention

Additional note.

Sxx is socially isolated - occasionally he will choose to spend time in the day room. He spends most of his day in a side room with music playing. Staff deliberately do not interact for fear of setting off challenging behaviours. To this extent social avoidance is implicated as a potential function for all behaviours.

SUMMARY OF FUNCTIONS ASSIGNED BY CLINICAL INTERVIEW

BEHAVIOUR	POTENTIAL FUNCTION
Aggression	Social avoidance, task avoidance
SIB	Automatic, social avoidance
Damage	Automatic, social avoidance, task avoidance
Pica	Automatic, social attention
Vocalisation	Tangible, social attention, automatic
Stereotypy	Automatic, social avoidance
Vomiting	Automatic, social attention
Stripping	Automatic, tangible

Appendix 2.2

Integrated rating scale including four items covering the social avoidance function

Original, re-worded and new scale items

1 If (s)he was left alone for a long period of time, say an hour, is it expected that the behaviour would be repeated continuously throughout?

2 Does the behavior occur following a request to perform a difficult task?

3 Does the behavior seem to occur in response to your talking to other persons in the room?

4 Does the behavior ever occur to get a toy, food or activity that this person has been told that he or she cannot have?

5 Would the behaviour occur when anyone (or a particular person) enters the room?

6 If (s)he was left alone for a long period of time, say an hour, is it expected that the behaviour would be repeated continuously throughout?

7 Does this behaviour occur when (s)he is asked to do something she finds difficult?

8 Does this behaviour occur when you are talking to other people in the room?

9 As a result of this behaviour, does (s)he receive an object, food or activity that she has been previously told she cannot have?

10 Would the behavior occur repeatedly, in the same way, for very long periods of time , if no one was around? (for example, rocking back and forth for over an hour.)

11 Does the behavior occur when any request is made of this person?

12 Does the behavior occur whenever you stop attending to this person?

13 Does the behavior occur when you take away a favorite toy, food, or activity?

14 Does the behaviour follow a particular pattern, repeated over a period of time, such as rocking backwards and forwards?

15 Does this behaviour occur whenever (s)he is asked to do something?

16 Does this behaviour occur when people stop attending to him/her.

17 Does the behaviour occur when a favourite object, food or activity is taken away from him/her?

18 Does the behaviour occur when other people come within about 1 metre of the person?

19 Does it appear to you that this person enjoys performing the behavior? (It feels, tastes, looks, smells and/or sounds pleasing.)

20 Does this person seem to do the behavior to upset or annoy you when you are trying to get him or her to do what you ask?

21 Does this person seem to do the behaviour to upset or annoy you when you are not paying attention to him or her? (For example, if you are sitting in a separate room interacting with another person.)

22 Does the behavior stop occurring shortly after you give this person the toy, food or activity he or she requested?

23 Does (s)he seem to enjoy this behaviour?

24 Does (s)he produce this behaviour in order to upset or annoy other people when they are trying to get him/her to do what they asked?

25 Does (s)he do this behaviour when you are not paying attention to him/her, for example if you were in another room with someone else.

26 Does this behaviour stop when (s)he is given an object, some food or an activity that (s)he wanted?

27 Does the behaviour occur when the person is with other (or a particular) service user(s)?

28 When the behavior is occurring, does this person seem calm and unaware of anything going on around him or her?

29 Does the behavior stop occurring shortly after (one to five minutes) you stop working or making demands on this person?

30 Does this person seem to do the behavior to get you to spend some time with him or her?

31 Does this behavior seem to occur when this person has been told that he or she can't do something he or she had wanted to do?

32 Would the behaviour stop within a short time of other people moving away from the person, (say 1 metre distant) or leaving the room?

33 When this behaviour occurs does (s)he show little awareness of things which are happening around him/her.

34 Does the behaviour stop, even for a short time, when people stop asking him/her to do things?

35 Does (s)he do this behaviour so that people will spend time with him/her?

36 Does the behaviour happen when (s)he is not allowed to do something (s)he had wanted to do?

Question order =

1	2	3	4	5	6	7	8	9
S	E	A	T	*P*	S	E	A	T

10	11	12	13	14	15	16	17	18	19	20
S	E	A	T	S	E	A	T	*P*	S	E

21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	A	T	S	E	A	T	*P*
S	E	A	T	*P*	S	E	A	T														

KEY:
Bold = original version
Standard = English revision

S= sensory feedback
E= escape
A= social attention
T= tangibles
P= person avoidance

Additional items for social avoidance.

* Would the behaviour occur when anyone (or a particular person) enters the room?

* Does the behaviour occur when other people come within about 1 metre of the person?

* Does the behaviour occur when the person is with other (or a particular) service user(s)?

* Would the behaviour stop within a short time of other people moving away from the person, (say 1 metre distant) or leaving the room?

Appendix 2.3

Specimen ABC chart

ABC Chart

Name ..... Sheet Number ..... Record any of the inappropriate behaviours below

Date, time, place	**Antecedents** (what happened before the behaviour)	**Behaviour** (what behaviour did you see?)	**Consequences** (what happened directly after the behaviour?)	**Ints**

Please continue overleaf

ABC Chart

Dates covered by ABC chart: ..... to .....
Times covered by chart: All waking hours

Behaviours to record:

Antecedents to look for:

Consequences to look for:

Filling in the ABC chart

The Behaviour Column

Write out a brief description of the behaviour stating exactly what the person did, e.g. "she slapped me on the arm", "he put his hand through the window and broke the glass" etc. Please try to avoid "fuzzy" descriptions such as "she threw a wobbly" or "he went bananas".

The Antecedent column

Clearly state what events immediately preceded the behaviour. In particular the following should be looked for:

Attention:	Was the person receiving attention from anyone? Was the attention positive, neutral or critical? Has the person been without attention for a length of time? How long? Was another person entering or leaving the room or standing /sitting nearby but not attending to the person?
Activities	Was the person engaged in an activity at the time ? What was the activity? Was there no activity available? Had an activity just ended?
Demands	Were demands being made of the person or a request to do something or go somewhere? Please describe it.
Food/Drink	Was the person waiting for or asking for food or drink? Was food or drink in sight but not available? Had food or drink just been taken away or finished?
Environment	Was the environment very noisy, hot, cold, crowded?

The Consequences Column

Write down in this column exactly what people did immediately after the behaviour occurred and what happened to the client. Note in particular the following:

Attention	Was the person given attention after the behaviour? Was the attention positive, neutral, critical, restraint? Was attention withdrawn?
Activities	Was the activity taken away after the behaviour? Was activity provided after the behaviour?

Continued Overleaf

Consequences column continued

Demands	Did any demand or request cease after the behaviour occurred? Was any demand or request made after the behaviour occurred?
Food/drink	Was food or drink taken away after the behaviour occurred? Was food or drink provided after the behaviour occurred?
Environment	Was the person moved to another room which was quieter, cooler, warmer? Was anything else changed?

This is not an exhaustive list, the objective is to write down all the relevant details.

Example of part of a completed ABC chart

Name: *John Smith* Sheet No *2*

Date & Time	Antecedents	Behaviour	Consequences	Ints
3.6.93 Lounge 5.30 pm	*At the tea table eating his meal. Asked him to eat more slowly.*	*Threatened to hit me and also swore*	*Told him not to swear*	*Kate*
3.6.93 lounge 7.45 pm	*Sitting, Watching T.V.*	*Got up with no warning and hit Jim across the face and laughed.*	*Told him off, asked him why he hit Jim. Sat by him to prevent another attack*	*Steve*

Appendix 2.4

Functions assigned to each topography by each assessment method arranged by subject

NUMBER OF TOPOGRAPHIES IDENTIFIED FOR ASSESSMENT
AND
PERCENT OF PERSONS SHOWING THE TOPOGRAPHY

Topography class	number	% subjects
Aggression directed at others	21*	95
Damage to property	19	95
Negative vocalisations	18	90
Self-injury	17	85
Stereotypy	15	75
Pica	5	25
Stripping	4	20
Verbal abuse	3	15
Vomiting	3	15
Faecal smearing	3	15
Intentional incontinence	3	15
Obsessive behaviour	2	10
Falling to floor	2	10
Non-compliance	1	5
Poking rectum	1	5
Running around (in bouts)	1	5
Spitting	1	5
Public masturbation	1	5
Hyperactivity	1	5

* two persons showed aggressive behaviours that were sufficiently different to warrant being classified independently

All functions assigned by assessment methods:
arranged by subject and topographical class.

Key:

ind	=	no function ascribed
1	=	task escape or avoidance
2	=	social escape or avoidance
3	=	social attention
4	=	tangible reinforcement
5	=	automatic reinforcement

All assigned functions are shown; multi-function complexes are classed primary and secondary and ordered according to the strength of supporting evidence.

Appendix 2.4 Functions assigned by assessment methods

ID	BEH	INT	ABC	MAS1	MAS2	ANA	DESC
1	AGG	1 2	2 1 3	1 4	1	1 3	2
1	SIB	5 3	3 1 2	5 4	5 3	IND	2
1	PIC	5 3	IND	5	5 3	4	IND
1	REC	5 3	IND	5	5	IND	IND
1	STE	5 3	IND	5 3	5 3 1	3 5	2 1
1	DAM	IND	IND	IND	IND	IND	4
2	AGG	1 2	3	4	4	1	IND
2	SIB	3 5	1 2	5	1 2	5 1	5
2	VOC	4	3	4	4 1	IND	5
2	DAM	IND	IND	IND	IND	IND	3
2	STE	IND	IND	IND	IND	IND	4 5 2
3	AGG	5 3	1 3	3	2	1	3
3	AGG	5 3	1 3	3 5 1	2	1	3
3	STE	5	IND	5	3 5 2	5 3	5
3	VOC	5	1 3	5	5 2 1 3	IND	2
3	SIB	5 3	IND	5	5 3	IND	IND
3	DAM	4	IND	5	3 5	3	3
3	PIC	IND	IND	IND	IND	IND	IND
4	STE	5	IND	5 1 3	1 5 3	IND	5

ID	BEH	INT	ABC	MAS1	MAS2	ANA	DESC
4	AGG	3 4	1 3	1	1	IND	4 1 3
4	SIB	IND	IND	IND	IND	IND	2 4
4	DAM	1 4	IND	1 3	1	IND	4 2
4	RUN	3 1	IND	1 3	1	IND	IND
4	VOC	1 3	3 1	1	1	IND	4 5 1
5	SIB	3 2 1	IND	1 4	1	1	3 1 4
5	VOC	1 5 3 4	1 3	1 4	1	1 3	4 2 1
5	FLR	1 2 3	IND	1 4	1	IND	1 5
5	DAM	3	1	5 1	5 1	3	2 4
5	STP	1 2 3	IND	4 1	1	IND	IND
6	STE	IND	IND	IND	IND	5 3	3
6	SIB	4 1 2 3	4	4	1	IND	2
6	DAM	IND	IND	IND	IND	5	4
6	PIC	5	3	5	5 3	5	5
6	VOC	IND	IND	IND	IND	4 1	5
6	AGG	2 1 4	IND	5	2	1 2 4	IND
6	AGG	2 1 4	IND	1	2	1 2 4	IND
7	STE	5 3	IND	5	5	IND	5
7	SIB	1 2	3 1	1	1	IND	5

ID	BEH	INT	ABC	MAS1	MAS2	ANA	DESC
7	DAM	1 2 3 4	IND	5 1 3	3	3	IND
7	VOC	IND	IND	IND	IND	IND	5
7	SMR	5 3	IND	5	5	IND	IND
7	AGG	1 2 3 4	IND	1 5	1 4	IND	IND
8	VBL	4 2	3	IND	IND	IND	3
8	AGG	4 2	4 3 1	42 5	4 2 5	IND	IND
8	OBS	5 1 2	IND	5	5	IND	IND
8	DAM	IND	IND	IND	IND	IND	5 4
9	SIB	1 2	3	4 1	4	IND	5
9	DAM	4	IND	4	4	IND	IND
9	VOC	4 2	3	4 1	4 1	IND	IND
9	STE	IND	IND	IND	IND	IND	5
9	VBL	4 2	IND	4 1	1 4	3	IND
9	AGG	4 2	IND	4	1 4	IND	IND
10	STE	5	IND	5	5	5	5
10	AGG	1 3	IND	4 1	1 4	1	2
10	SIB	1	IND	4	1 4 2	1	5
10	DAM	IND	IND	IND	IND	IND	5
10	PIC	5 3	IND	5	5 4 2	IND	IND
10	VOC	IND	IND	1	1 4	1	IND

ID	BEH	INT	ABC	MAS1	MAS2	ANA	DESC
10	OBS	IND	IND	5	5	IND	IND
11	STE	IND	3	IND	IND	IND	5
11	AGG	3	IND	4 3	4 3	IND	5
11	SIB	3	3	4 3	3 4	IND	2
11	VOC	3	IND	IND	IND	IND	IND
11	INC	3	3	3	3	IND	IND
11	DAM	3	IND	3 4	4 3	IND	IND
12	TMP	IND	IND	IND	IND	2 1	1 2
12	SIB	IND	IND	1	1 4 2	IND	1
12	DAM	1 2	IND	1	1	4 2	4
12	VOM	1 2 3	2 3	IND	3	IND	IND
12	VOC	3 4 2	2 3	4 1 3	1 4 5	1 4	1 2
12	AGG	4 2 1	IND	1	1	4 2	IND
12	SMR	5	IND	IND	5	IND	IND
13	STE	5	IND	5	5 3	5	5
13	AGG	3	3	1 3	3 2 5	1 2	3
13	SIB	IND	1	IND	IND	1 2	3 2
13	DAM	IND	IND	IND	IND	1 2	3
13	PIC	5 3	IND	5	3	IND	3
13	VOC	1 4 3	1	1 4	1	1 2	1 2

ID	BEH	INT	ABC	MAS1	MAS2	ANA	DESC
13	ABS	4 3	3	5	5	IND	IND
14	STE	5	IND	5	5	IND	5 4
14	SIB	IND	IND	IND	IND	1 2	5 1
14	DAM	2 3 1	IND	5	5	IND	4 3
14	VOC	IND	IND	IND	IND	1 3	2 3
14	AGG	1 2	IND	4 1	1	1	IND
15	AGG	2 3	IND	3	2 3	IND	IND
15	STE	IND	IND	IND	IND	IND	5
15	DAM	IND	IND	IND	IND	IND	5
15	VOC	1 4 3	4	1	4 3	IND	1 3
15	MAS	5 3	IND	5	5	IND	IND
16	VBL	IND	IND	IND	IND	4 1	2
16	STE	IND	IND	IND	IND	IND	5
16	AGG	1 2 3 4	IND	1 4 3	1 3	IND	2 1
16	SIB	1 2 3 4	IND	1 4 3	5 1 3	IND	2 1
16	VOC	1 2 3 4	IND	4 1 3 5	1 3	IND	2 1
16	DAM	IND	IND	IND	IND	IND	5
17	AGG	2 3	IND	4	4 2	IND	IND
17	INC	3	IND	4	4 1	IND	IND
17	STP	3	IND	5	5 3	IND	2

ID	BEH	INT	ABC	MAS1	MAS2	ANA	DESC
17	HYP	3 5	IND	5	5	1 2	3
17	VOC	IND	IND	IND	IND	1 2	3 5
17	SIB	IND	IND	IND	IND	IND	IND
18	AGG	1 3	1	4	2	IND	3
18	DAM	1	1	4	4	IND	5
18	VBL	1 3	1	IND	IND	IND	3 4
18	VOC	IND	IND	IND	IND	IND	1 4 2
18	SIB	1 3	1	4	4	IND	IND
19	AGG	3	3	3	3	2 3	1 3 4
19	DAM	IND	IND	IND	IND	2 5	3 4
19	VOC	IND	IND	IND	IND	1	1 4 2
19	FLR	1 4	1	1	1	1	2
19	OBJ	4 3	IND	4	4	IND	IND
19	STE	5	IND	5	5	3 4	5
19	VBL	IND	IND	IND	IND	IND	4 2 3
20	VOM	53	IND	3	3 5	IND	3
20	STE	5 2	5 3	5	5	2 4	5
20	AGG	2 1	2	1 4	1 2	4 1 2	3
20	SIB	5 2	4 3 5	4	1 2	2 1 4	5
20	DAM	5 2 1	4	1	1 2	1	IND

ID	BEH	INT	ABC	MAS1	MAS2	ANA	DESC
20	VOC	4 3 5	4 3 5	4 1	1 2	4 2 1	1
20	STP	5 4	IND	IND	IND	IND	IND
20	PIC	5 3	IND	IND	IND	1	IND

Appendix 3.1

Probability of agreement between each pair of assessment methods

PROBABILITY OF AGREEMENT BETWEEN
EACH PAIR OF ASSESSMENT METHODS

Base method - Interview	ABC	MAS1	MAS2	Analogue	Desc.
Number of topographies (all)	121	121	121	121	121
Probability of convergence	0.23	0.55	0.57	0.23	0.29
No. of topographies base method	89	89	89	89	89
Probability of convergence	0.31	0.75	0.78	0.31	0.38
No. of topographies both methods	38	82	84	37	54
Probability of convergence	0.74	0.88	0.89	0.76	0.65

Base method - ABC	INT	MAS1	MAS2	Analogue	Desc.
No of topographies (all)	121	121	121	121	121
Probability of convergence	0.23	0.17	0.17	0.1	0.12
No. of topographies base method	40	40	40	40	40
Probability of convergence	0.7	0.5	0.52	0.3	0.35
No. of topographies both methods	38	35	36	19	32
Probability of convergence	0.74	0.57	0.58	0.63	0.44

Base method - MAS1	INT	ABC	MAS2	Analogue	Desc.
Number of topographies (all)	121	121	121	121	121
Probability of convergence	0.55	0.17	0.62	0.2	0.23
No. of topographies base method	85	85	85	85	85
Probability of convergence	0.79	0.21	0.88	0.28	0.33
No. of topographies both methods	82	35	85	37	53
Probability of convergence	0.88	0.57	0.88	0.63	0.53

Base method - MAS2	INT	ABC	MAS1	Analogue	Desc.
Number. of topographies (all)	121	121	121	121	121
Probability of convergence	0.57	0.17	0.62	0.21	0.24
No. of topographies base method	87	87	87	87	87
Probability of convergence	0.79	0.24	0.86	0.3	0.33
No. of topographies both methods	84	36	85	37	53
Probability of convergence	0.89	0.58	0.88	0.7	0.53

Base method - Analogue Ass'ment	INT	ABC	MAS1	MAS2	Desc.
Number of topographies (all)	121	121	121	121	121
Probability of convergence	0.23	0.1	0.2	0.21	0.16

No. of topographies base method	50	50	50	50	50
Probability of convergence	0.52	0.24	0.46	0.52	0.38

No. of topographies both methods	37	19	37	37	39
Probability of convergence	0.76	0.63	0.63	0.7	0.49

Base method - Descriptive Analysis	INT	ABC	MAS1	MAS2	Analogue
Number of topographies (all)	121	121	121	121	121
Probability of convergence	0.29	0.12	0.23	0.24	0.16

No. of topographies base method	82	82	82	82	82
Probability of convergence	0.43	0.17	0.34	0.34	0.35

No. of topographies both methods	54	32	53	53	39
Probability of convergence	0.65	0.44	0.53	0.53	0.49

Appendix 3.2

Probabilities of agreement by chance

ALL ASCRIBED FUNCTIONS

interview				abc chart				mas1				mas2				analogue assessment				descriptive analysis			
	task av	int	chance		task av	abc	chance		task av	mas1	chance		task av	mas2	chance		task av	ana	chance		task av	desc	chance
abc	0.31	0.2	0.06	int	0.2	0.31	0.06	int	0.2	0.29	0.06	int	0.2	0.28	0.06	int	0.2	0.38	0.08	int	0.2	0.15	0.03
mas1	0.29	0.2	0.06	mas1	0.29	0.31	0.09	abc	0.31	0.29	0.09	abc	0.31	0.29	0.09	abc	0.31	0.38	0.12	abc	0.31	0.15	0.05
mas2	0.28	0.2	0.06	mas2	0.28	0.31	0.09	mas2	0.28	0.29	0.08	mas1	0.29	0.29	0.08	mas1	0.29	0.38	0.11	mas1	0.29	0.15	0.04
ana	0.38	0.2	0.08	ana	0.38	0.31	0.12	ana	0.38	0.29	0.11	ana	0.38	0.29	0.11	mas2	0.28	0.38	0.11	mas2	0.28	0.15	0.04
desc	0.15	0.2	0.03	desc	0.15	0.31	0.05	desc	0.15	0.29	0.04	desc	0.15	0.29	0.04	desc	0.15	0.38	0.06	ana	0.38	0.15	0.06
mean	0.28	0.2	0.06	mean	0.26	0.31	0.08	mean	0.26	0.29	0.08	mean	0.27	0.29	0.08	mean	0.25	0.38	0.09	mean	0.29	0.15	0.04
	soc av	int	chance		soc av	abc	chance		soc av	mas1	chance		soc av	mas2	chance		soc av	ana	chance		soc av	desc	chance
abc	0.1	0.18	0.02	int	0.18	0.1	0.02	int	0.18	0	0	int	0.18	0.13	0.02	int	0.18	0.21	0.04	int	0.18	0.22	0.04
mas1	0	0.18	0	mas1	0	0.1	0	abc	0.1	0	0	abc	0.1	0.13	0.01	abc	0.1	0.21	0.02	abc	0.1	0.22	0.02
mas2	0.13	0.18	0.02	mas2	0.13	0.1	0.01	mas2	0.13	0	0	mas1	0	0.13	0	mas1	0	0.21	0	mas1	0	0.22	0
ana	0.21	0.18	0.04	ana	0.21	0.1	0.02	ana	0.21	0	0	ana	0.21	0.13	0.03	mas2	0.13	0.21	0.03	mas2	0.13	0.22	0.03
desc	0.22	0.18	0.04	desc	0.22	0.1	0.02	desc	0.22	0	0	desc	0.22	0.13	0.03	desc	0.22	0.21	0.05	ana	0.21	0.22	0.05
mean	0.13	0.18	0.02	mean	0.15	0.1	0.01	mean	0.17	0	0	mean	0.14	0.13	0.02	mean	0.13	0.21	0.03	mean	0.12	0.22	0.03
	attn	int	chance		attn	abc	chance		attn	mas1	chance		attn	mas2	chance		attn	ana	chance		attn	desc	chance
abc	0.44	0.29	0.13	int	0.29	0.44	0.13	int	0.29	0.15	0.04	int	0.29	0.19	0.06	int	0.29	0.14	0.04	int	0.29	0.2	0.06
mas1	0.15	0.29	0.04	mas1	0.15	0.44	0.07	abc	0.44	0.15	0.07	abc	0.44	0.19	0.08	abc	0.44	0.14	0.06	abc	0.44	0.2	0.09
mas2	0.19	0.29	0.06	mas2	0.19	0.44	0.08	mas2	0.19	0.15	0.03	mas1	0.15	0.19	0.03	mas1	0.15	0.14	0.02	mas1	0.15	0.2	0.03
ana	0.14	0.29	0.04	ana	0.14	0.44	0.06	ana	0.14	0.15	0.02	ana	0.14	0.19	0.03	mas2	0.19	0.14	0.03	mas2	0.19	0.2	0.04
desc	0.2	0.29	0.06	desc	0.2	0.44	0.09	desc	0.2	0.15	0.03	desc	0.2	0.19	0.04	desc	0.2	0.14	0.03	ana	0.14	0.2	0.03
mean	0.22	0.29	0.06	mean	0.19	0.44	0.09	mean	0.25	0.15	0.04	mean	0.24	0.19	0.05	mean	0.25	0.14	0.04	mean	0.24	0.2	0.05
	tang	int	chance		tang	abc	chance		tang	mas1	chance		tang	mas2	chance		tang	ana	chance		tang	desc	chance
abc	0.1	0.15	0.02	int	0.15	0.1	0.02	int	0.15	0.28	0.04	int	0.15	0.17	0.03	int	0.15	0.14	0.02	int	0.15	0.17	0.03
mas1	0.28	0.15	0.04	mas1	0.28	0.1	0.03	abc	0.1	0.28	0.03	abc	0.1	0.17	0.02	abc	0.1	0.14	0.01	abc	0.1	0.17	0.02
mas2	0.17	0.15	0.03	mas2	0.17	0.1	0.02	mas2	0.17	0.28	0.05	mas1	0.28	0.17	0.05	mas1	0.28	0.14	0.04	mas1	0.28	0.17	0.05
ana	0.14	0.15	0.02	ana	0.14	0.1	0.01	ana	0.14	0.28	0.04	ana	0.14	0.17	0.02	mas2	0.17	0.14	0.02	mas2	0.17	0.17	0.03
desc	0.17	0.15	0.03	desc	0.17	0.1	0.02	desc	0.17	0.28	0.05	desc	0.17	0.17	0.03	desc	0.17	0.14	0.02	ana	0.14	0.17	0.02
mean	0.17	0.15	0.03	mean	0.18	0.1	0.02	mean	0.15	0.28	0.04	mean	0.17	0.17	0.03	mean	0.17	0.14	0.02	mean	0.17	0.17	0.03
	auto	int	chance		auto	abc	chance		auto	mas1	chance		auto	mas2	chance		auto	ana	chance		auto	desc	chance
abc	0.1	0.18	0.02	int	0.18	0.05	0.01	int	0.18	0.27	0.05	int	0.18	0.23	0.04	int	0.18	0.11	0.02	int	0.18	0.27	0.05
mas1	0.28	0.18	0.05	mas1	0.27	0.05	0.01	abc	0.05	0.27	0.01	abc	0.05	0.23	0.01	abc	0.05	0.11	0.01	abc	0.05	0.27	0.01
mas2	0.17	0.18	0.03	mas2	0.23	0.05	0.01	mas2	0.23	0.27	0.06	mas1	0.27	0.23	0.06	mas1	0.27	0.11	0.03	mas1	0.27	0.27	0.07
ana	0.14	0.18	0.03	ana	0.11	0.05	0.01	ana	0.11	0.27	0.03	ana	0.11	0.23	0.03	mas2	0.23	0.11	0.03	mas2	0.23	0.27	0.06
desc	0.17	0.18	0.03	desc	0.27	0.05	0.01	desc	0.27	0.27	0.07	desc	0.27	0.23	0.06	desc	0.27	0.11	0.03	ana	0.11	0.27	0.03
mean	0.17	0.18	0.03	mean	0.21	0.05	0.01	mean	0.17	0.27	0.05	mean	0.18	0.23	0.04	mean	0.2	0.11	0.02	mean	0.17	0.27	0.05

PRIMARY FUNCTIONS

interview			
	task av	int	chance
abc	0.35	0.29	0.1
mas1	0.27	0.29	0.08
mas2	0.37	0.29	0.11
ana	0.5	0.29	0.15
desc	0.12	0.29	0.03
mean	0.32	0.29	0.09

abc chart			
	task av	abc	chance
int	0.29	0.35	0.1
mas1	0.27	0.35	0.09
mas2	0.37	0.35	0.13
ana	0.5	0.35	0.18
desc	0.12	0.35	0.04
mean	0.31	0.35	0.11

mas1			
	task av	mas1	chance
int	0.29	0.27	0.08
abc	0.35	0.27	0.09
mas2	0.37	0.27	0.1
ana	0.5	0.27	0.14
desc	0.12	0.27	0.03
mean	0.33	0.27	0.09

mas2			
	task av	mas2	chance
int	0.29	0.37	0.11
abc	0.35	0.37	0.13
mas1	0.27	0.37	0.1
ana	0.5	0.37	0.19
desc	0.12	0.37	0.04
mean	0.31	0.37	0.11

analogue assessment			
	task av	ana	chance
int	0.29	0.5	0.15
abc	0.35	0.5	0.18
mas1	0.27	0.5	0.14
mas2	0.37	0.5	0.19
desc	0.12	0.5	0.06
mean	0.28	0.5	0.14

descriptive analysis			
	task av	desc	chance
int	0.29	0.12	0.03
abc	0.35	0.12	0.04
mas1	0.27	0.12	0.03
mas2	0.37	0.12	0.04
ana	0.5	0.12	0.06
mean	0.36	0.12	0.04

	soc av	int	chance
abc	0.1	0.07	0.01
mas1	0	0.07	0
mas2	0.07	0.07	0
ana	0.13	0.07	0.01
desc	0.2	0.07	0.01
mean	0.1	0.07	0.01

	soc av	abc	chance
int	0.07	0.1	0.01
mas1	0	0.1	0
mas2	0.07	0.1	0.01
ana	0.13	0.1	0.01
desc	0.2	0.1	0.02
mean	0.09	0.1	0.01

	soc av	mas1	chance
int	0.07	0	0
abc	0.1	0	0
mas2	0.07	0	0
ana	0.13	0	0
desc	0.2	0	0
mean	0.11	0	0

	soc av	mas2	chance
int	0.07	0.07	0
abc	0.1	0.07	0.01
mas1	0	0.07	0
ana	0.13	0.07	0.01
desc	0.2	0.07	0.01
mean	0.1	0.07	0.01

	soc av	ana	chance
int	0.07	0.13	0.01
abc	0.1	0.13	0.01
mas1	0	0.13	0
mas2	0.07	0.13	0.01
desc	0.2	0.13	0.03
mean	0.09	0.13	0.01

	soc av	desc	chance
int	0.07	0.2	0.01
abc	0.1	0.2	0.02
mas1	0	0.2	0
mas2	0.07	0.2	0.01
ana	0.13	0.2	0.03
mean	0.07	0.2	0.01

	attn	int	chance
abc	0.38	0.18	0.07
mas1	0.08	0.18	0.01
mas2	0.11	0.18	0.02
ana	0.12	0.18	0.02
desc	0.22	0.18	0.04
mean	0.18	0.18	0.03

	attn	abc	chance
abc	0.18	0.38	0.07
mas1	0.08	0.38	0.03
mas2	0.11	0.38	0.04
ana	0.12	0.38	0.05
desc	0.22	0.38	0.08
mean	0.14	0.38	0.05

	attn	mas1	chance
abc	0.18	0.08	0.01
abc	0.38	0.08	0.03
mas2	0.11	0.08	0.01
ana	0.12	0.08	0.01
desc	0.22	0.08	0.02
mean	0.2	0.08	0.02

	attn	mas2	chance
abc	0.18	0.11	0.02
abc	0.38	0.11	0.04
mas1	0.08	0.11	0.01
ana	0.12	0.11	0.01
desc	0.22	0.11	0.02
mean	0.2	0.11	0.02

	attn	ana	chance
abc	0.18	0.12	0.02
abc	0.38	0.12	0.05
mas1	0.08	0.12	0.01
mas2	0.11	0.12	0.01
desc	0.22	0.12	0.03
mean	0.19	0.12	0.02

	attn	ana	chance
abc	0.18	0.22	0.04
abc	0.38	0.22	0.08
mas1	0.08	0.22	0.02
mas2	0.11	0.22	0.02
desc	0.12	0.22	0.03
mean	0.17	0.22	0.04

	tangible	int	chance
abc	0.15	0.15	0.02
mas1	0.29	0.15	0.04
mas2	0.16	0.15	0.02
ana	0.14	0.15	0.02
desc	0.12	0.15	0.02
mean	0.17	0.15	0.03

	tangible	abc	chance
int	0.15	0.15	0.02
mas1	0.29	0.15	0.04
mas2	0.16	0.15	0.02
ana	0.14	0.15	0.02
desc	0.12	0.15	0.02
mean	0.17	0.15	0.03

	tangible	mas1	chance
int	0.15	0.29	0.04
abc	0.15	0.29	0.04
mas2	0.16	0.29	0.05
ana	0.14	0.29	0.04
desc	0.12	0.29	0.03
mean	0.14	0.29	0.04

	tangible	mas2	chance
int	0.15	0.16	0.02
abc	0.15	0.16	0.02
mas1	0.29	0.16	0.05
ana	0.14	0.16	0.02
desc	0.12	0.16	0.02
mean	0.17	0.16	0.03

	tangible	ana	chance
int	0.15	0.14	0.02
abc	0.15	0.14	0.02
mas1	0.29	0.14	0.04
mas2	0.16	0.14	0.02
desc	0.12	0.14	0.02
mean	0.17	0.14	0.02

	tangible	desc	chance
int	0.15	0.12	0.02
abc	0.15	0.12	0.02
mas1	0.29	0.12	0.03
mas2	0.16	0.12	0.02
ana	0.14	0.12	0.02
mean	0.18	0.12	0.02

	auto	int	chance
abc	0.03	0.31	0.01
mas1	0.35	0.31	0.11
mas2	0.29	0.31	0.09
ana	0.19	0.31	0.06
desc	0.34	0.31	0.11
mean	0.24	0.31	0.07

	auto	abc	chance
int	0.31	0.03	0.01
mas1	0.35	0.03	0.01
mas2	0.29	0.03	0.01
ana	0.19	0.03	0.01
desc	0.34	0.03	0.01
mean	0.3	0.03	0.01

	auto	mas1	chance
int	0.31	0.35	0.11
abc	0.03	0.35	0.01
mas2	0.29	0.35	0.1
ana	0.19	0.35	0.07
desc	0.34	0.35	0.12
mean	0.23	0.35	0.08

	auto	mas2	chance
int	0.31	0.29	0.09
abc	0.03	0.29	0.01
mas1	0.35	0.29	0.1
ana	0.19	0.29	0.06
desc	0.34	0.29	0.1
mean	0.24	0.29	0.07

	auto	ana	chance
int	0.31	0.19	0.06
abc	0.03	0.19	0.01
mas1	0.35	0.19	0.07
mas2	0.29	0.19	0.06
desc	0.34	0.19	0.06
mean	0.26	0.19	0.05

	auto	desc	chance
int	0.31	0.34	0.11
abc	0.03	0.34	0.01
mas1	0.35	0.34	0.12
mas2	0.29	0.34	0.1
ana	0.19	0.34	0.06
mean	0.23	0.34	0.08

SECONDARY FUNCTIONS

interview			
	task av	int	chance
abc	0.31	0.16	0.05
mas1	0.41	0.16	0.07
mas2	0.17	0.16	0.03
ana	0.23	0.16	0.04
desc	0.03	0.16	0
mean	0.23	0.16	0.04

abc chart			
	task av	abc	chance
int	0.16	0.31	0.05
mas1	0.41	0.31	0.13
mas2	0.17	0.31	0.05
ana	0.23	0.31	0.07
desc	0.03	0.31	0.01
mean	0.2	0.31	0.06

mas1			
	task av	mas1	chance
int	0.16	0.41	0.07
abc	0.31	0.41	0.13
mas2	0.17	0.41	0.07
ana	0.23	0.41	0.09
desc	0.03	0.41	0.01
mean	0.18	0.41	0.07

mas2			
	task av	mas2	chance
int	0.16	0.17	0.03
abc	0.31	0.17	0.05
mas1	0.41	0.17	0.07
ana	0.23	0.17	0.04
desc	0.03	0.17	0.01
mean	0.23	0.17	0.04

analogue assessment			
	task av	ana	chance
int	0.16	0.23	0.04
abc	0.31	0.23	0.07
mas1	0.41	0.23	0.09
mas2	0.17	0.23	0.04
desc	0.03	0.23	0.01
mean	0.22	0.23	0.05

descriptive analysis			
	task av	desc	chance
int	0.16	0.03	0
abc	0.31	0.03	0.01
mas1	0.41	0.03	0.01
mas2	0.17	0.03	0.01
ana	0.23	0.03	0.01
mean	0.26	0.03	0.01

	soc av	int	chance
abc	0.13	0.42	0.05
mas1	0	0.42	0
mas2	0.32	0.42	0.13
ana	0.42	0.42	0.18
desc	0.33	0.42	0.14
mean	0.24	0.42	0.1

	soc av	abc	chance
int	0.42	0.13	0.05
mas1	0	0.13	0
mas2	0.32	0.13	0.04
ana	0.42	0.13	0.05
desc	0.33	0.13	0.04
mean	0.3	0.13	0.04

	soc av	mas1	chance
int	0.42	0	0
abc	0.13	0	0
mas2	0.32	0	0
ana	0.42	0	0
desc	0.33	0	0
mean	0.32	0	0

	soc av	mas2	chance
int	0.42	0.32	0.13
abc	0.13	0.32	0.04
mas1	0	0.32	0
ana	0.42	0.32	0.13
desc	0.33	0.32	0.11
mean	0.26	0.32	0.08

	soc av	ana	chance
int	0.42	0.42	0.18
abc	0.13	0.42	0.05
mas1	0	0.42	0
mas2	0.32	0.42	0.13
desc	0.33	0.42	0.14
mean	0.24	0.42	0.1

	soc av	desc	chance
inte	0.42	0.33	0.14
abc	0.13	0.33	0.04
mas1	0	0.33	0
mas2	0.32	0.33	0.11
ana	0.42	0.33	0.14
mean	0.26	0.33	0.09

	attn	int	chance
abc	0.81	0.58	0.47
mas1	0.38	0.58	0.22
mas2	0.39	0.58	0.23
ana	0.19	0.58	0.11
desc	0.2	0.58	0.12
mean	0.39	0.58	0.23

	attn	abc	chance
int	0.58	0.81	0.47
mas1	0.38	0.81	0.31
mas2	0.39	0.81	0.32
ana	0.19	0.81	0.15
desc	0.2	0.81	0.16
mean	0.35	0.81	0.28

	attn	mas1	chance
int	0.58	0.38	0.22
abc	0.81	0.38	0.31
mas2	0.39	0.38	0.15
ana	0.19	0.38	0.07
desc	0.2	0.38	0.08
mean	0.43	0.38	0.16

	attn	mas2	chance
int	0.58	0.39	0.23
abc	0.81	0.39	0.32
mas1	0.38	0.39	0.15
ana	0.19	0.39	0.07
desc	0.2	0.39	0.08
mean	0.43	0.39	0.17

	attn	ana	chance
int	0.58	0.19	0.11
abc	0.81	0.19	0.15
mas1	0.38	0.19	0.07
mas2	0.39	0.19	0.07
desc	0.2	0.19	0.04
mean	0.47	0.19	0.09

	attn	desc	chance
int	0.58	0.2	0.12
abc	0.81	0.2	0.16
mas1	0.38	0.2	0.08
mas2	0.39	0.2	0.08
ana	0.19	0.2	0.04
mean	0.47	0.2	0.09

	tangible	int	chance
abc	0	0.23	0
mas1	0.31	0.23	0.07
mas2	0.24	0.23	0.06
ana	0.19	0.23	0.04
desc	0.33	0.23	0.08
mean	0.21	0.23	0.05

	tangible	abc	chance
int	0.23	0	0
mas1	0.31	0	0
mas2	0.24	0	0
ana	0.19	0	0
desc	0.33	0	0
mean	0.26	0	0

	tangible	mas1	chance
int	0.23	0.31	0.07
abc	0	0.31	0
mas2	0.24	0.31	0.07
ana	0.19	0.31	0.06
desc	0.33	0.31	0.1
mean	0.2	0.31	0.06

	tangible	mas2	chance
int	0.23	0.24	0.06
abc	0	0.24	0
mas1	0.31	0.24	0.07
ana	0.19	0.24	0.05
desc	0.33	0.24	0.08
mean	0.21	0.24	0.05

	tangible	ana	chance
int	0.23	0.19	0.04
abc	0	0.19	0
mas1	0.31	0.19	0.06
mas2	0.24	0.19	0.05
desc	0.33	0.19	0.06
mean	0.22	0.19	0.04

	tangible	ana	chance
int	0.23	0.33	0.08
abc	0	0.33	0
mas1	0.31	0.33	0.1
mas2	0.24	0.33	0.08
ana	0.19	0.33	0.06
mean	0.19	0.33	0.06

	auto	int	chance
abc	0.13	0.05	0.01
mas1	0.16	0.05	0.01
mas2	0.17	0.05	0.01
ana	0.06	0.05	0
desc	0.13	0.05	0.01
mean	0.13	0.05	0.01

	auto	abc	chance
int	0.05	0.13	0.01
mas1	0.16	0.13	0.02
mas2	0.17	0.13	0.02
ana	0.06	0.13	0.01
desc	0.13	0.13	0.02
mean	0.11	0.13	0.01

	auto	mas1	chance
int	0.05	0.16	0.01
abc	0.13	0.16	0.02
mas2	0.17	0.16	0.03
ana	0.06	0.16	0.01
desc	0.13	0.16	0.02
mean	0.11	0.16	0.02

	auto	mas2	chance
int	0.05	0.17	0.01
abc	0.13	0.17	0.02
mas1	0.16	0.17	0.03
ana	0.06	0.17	0.01
desc	0.13	0.17	0.02
mean	0.11	0.17	0.02

	auto	ana	chance
int	0.05	0.06	0
abc	0.13	0.06	0.01
mas1	0.16	0.06	0.01
mas2	0.17	0.06	0.01
desc	0.13	0.06	0.01
mean	0.13	0.06	0.01

	auto	desc	chance
inte	0.05	0.13	0.01
abc	0.13	0.13	0.02
mas1	0.16	0.13	0.02
mas2	0.17	0.13	0.02
ana	0.06	0.13	0.01
mean	0.11	0.13	0.01

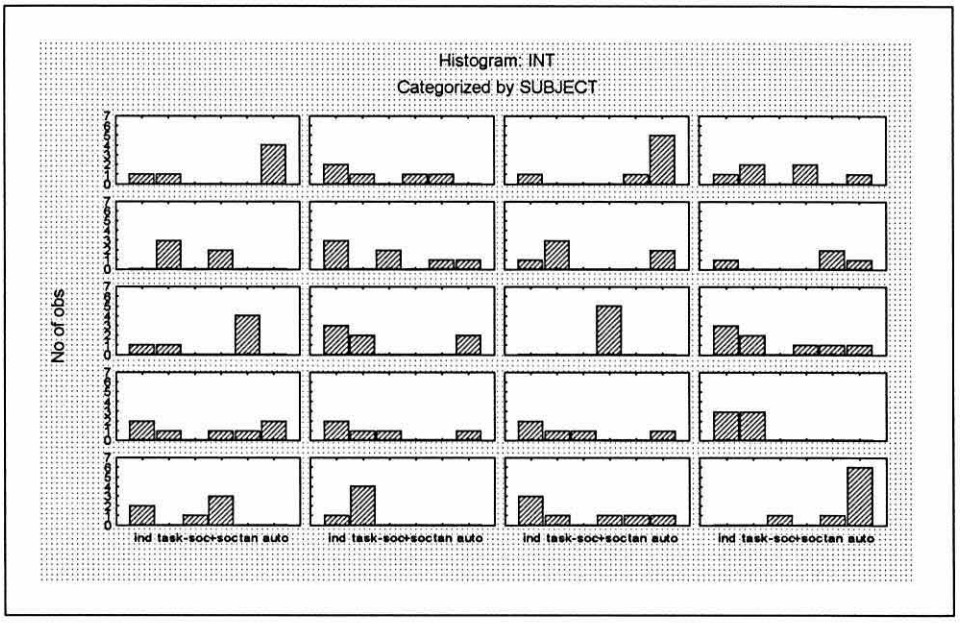
Appendix 4.1

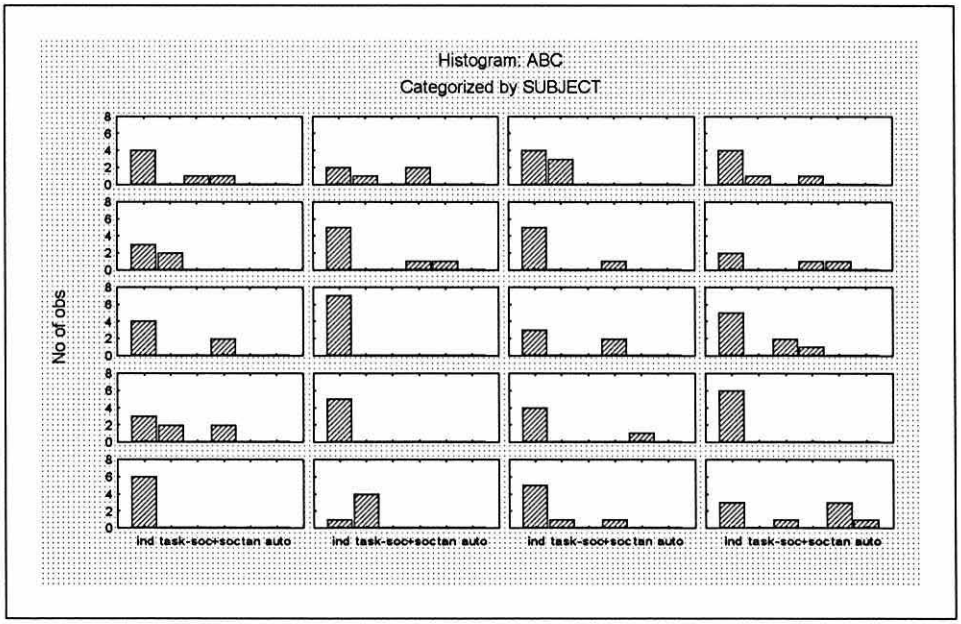
Distribution of functions ascribed

categorised by subject

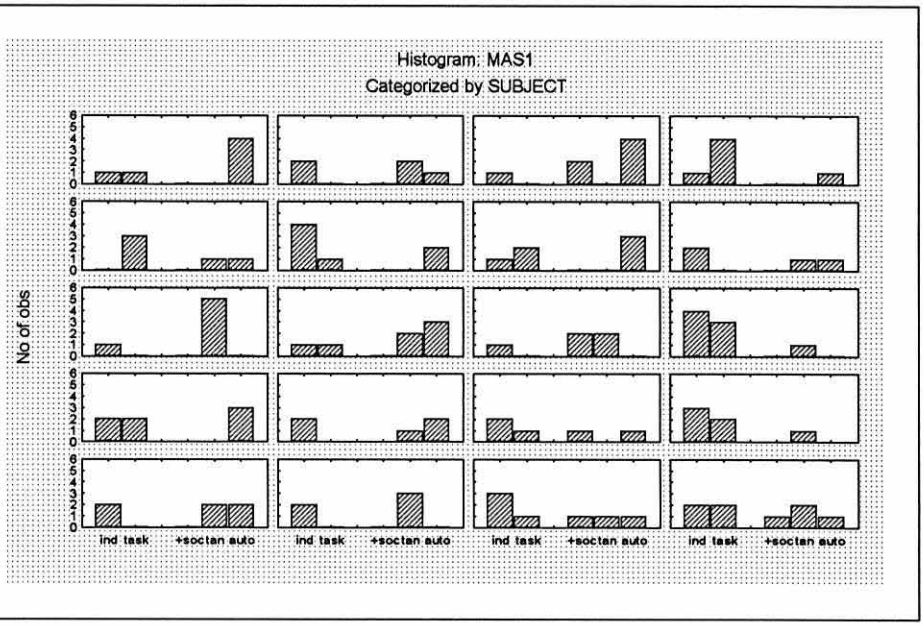
DISTRIBUTION OF FUNCTIONS ASCRIBED CATEGORISED BY SUBJECT

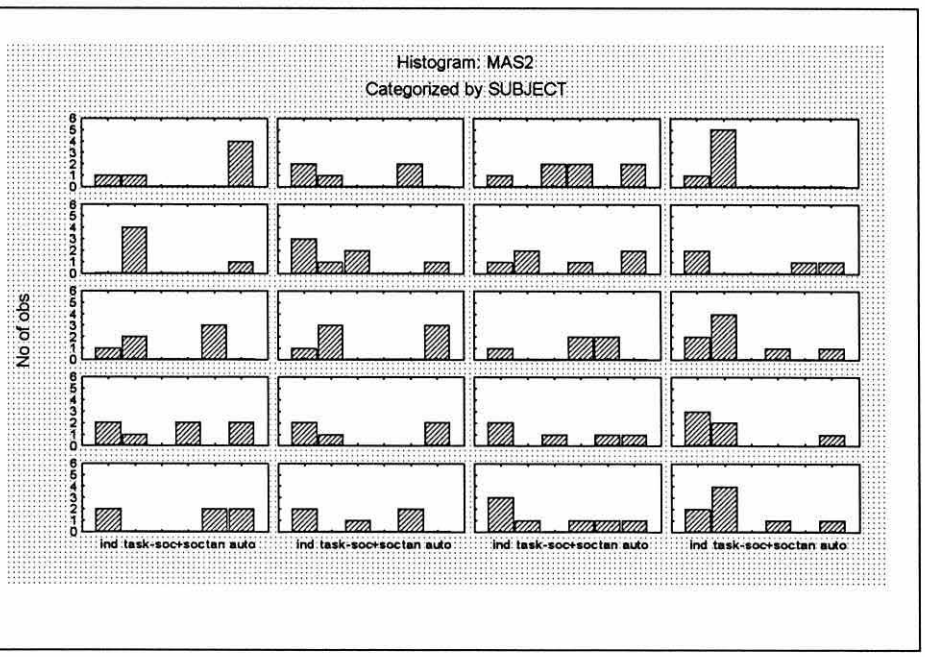
The number of functions each assessment method assigned is shown for each subject across all topographies assessed.

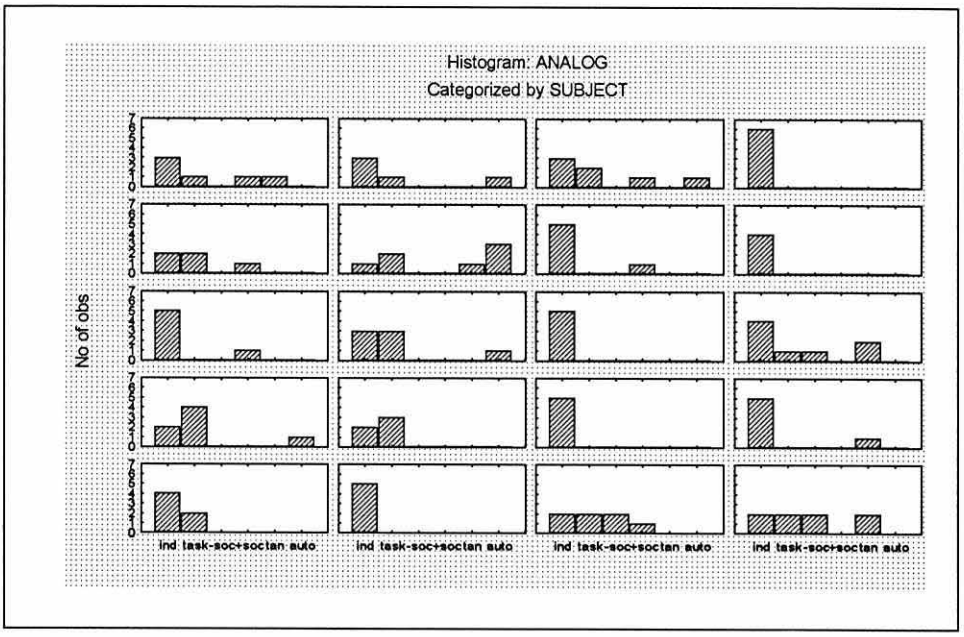


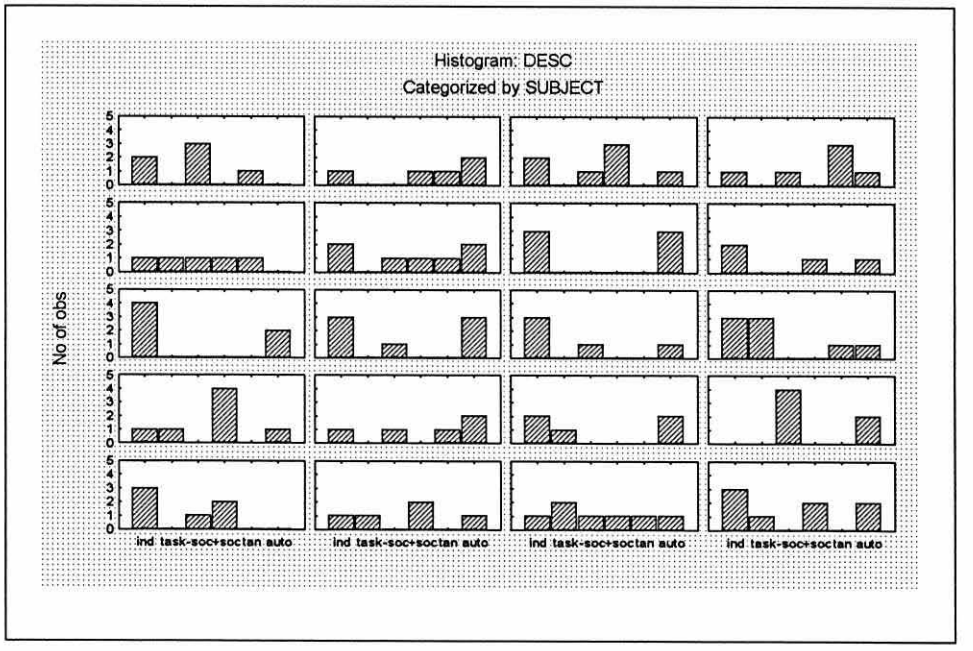


Key: ind = failed to assign a function
task = demand escape or avoidance
-soc = social escape or avoidance
+soc = social attention
tan = tangible reinforcement
auto = perceptual or sensory reinforcement


Histogram: MAS1
Categorized by SUBJECT
No of obs
ind task
+soctan auto


Histogram: MAS2
Categorized by SUBJECT
No of obs
ind task-soc+soctan auto


Histogram: ANALOG
Categorized by SUBJECT
No of obs
ind task-soc+soctan auto
ind task-soc+soctan auto
ind task-soc+soctan auto
ind task-soc+soctan auto


Histogram: DESC
Categorized by SUBJECT
No of obs
ind task-soc+soctan auto
ind task-soc+soctan auto
ind task-soc+soctan auto
ind task-soc+soctan auto

Appendix 4.2

Functions ascribed most and least frequently

by assessment methods

ASSESSMENT METHODS MOST AND LEAST LIKELY
TO ASCRIBE FROM EACH FUNCTION CATEGORY AND THE CATEGORY MEAN

Function category	MOST LIKELY		LEAST LIKELY		MEAN
task avoidance	Analogue	0.38	Descriptive	0.15	0.25
soc ial avoidance	Descriptive	0.22	MAS1	0	0.13
social attention	ABC	0.44	Analogue	0.14	0.19
tangible reinforcement	MAS1	0.28	ABC	0.1	0.15
automatic reinforcement	Descriptive: MAS1	0.27	ABC	0.05	0.14

Function category	MOST LIKELY		LEAST LIKELY		MEAN
task avoidance	Analogue	0.5	Descriptive	0.12	0.28
soc ial avoidance	Descriptive	0..2	MAS1	0	0.7
social attention	ABC	0.38	MAS1	0.08	0.14
tangible reinforcement	MAS1	0.29	Descriptive	0.12	0.14
automatic reinforcement	MAS1	0.35	ABC	0.03	0.23

Function category	MOST LIKELY		LEAST LIKELY		MEAN
task avoidance	MAS1	0.41	Interview	0.16	0.23
soc ial avoidance	Analogue: Interview	0.42	MAS1	0	0.2
social attention	ABC	0.81	Analogue	0.19	0.35
tangible reinforcement	Descriptive	0.33	ABC	0	0.19
automatic reinforcement	MAS2	0.17	Interview	0.05	0.11

FUNCTION CATEGORIES ASCRIBED MOST TO LEAST OFTEN
BY EACH ASSESSMENT METHOD

	INT	ABC	MAS1	MAS2	ANA	DESC
Predictions for all functions in rank order	soc attn	soc attn	task av	task av	task av	auto
	task av	task av	tan	auto	soc av	soc av
	soc av auto	soc av tan	auto	soc att	soc att tan	soc att
			soc att	tan		tan
	tan	auto	soc av	soc av	auto	task av

	INT	ABC	MAS1	MAS2	ANA	DESC
Predictions for primary functions in rank order	auto	soc attn	auto	task av	task av	auto
	task av	task av	tan	auto	auto	soc att
	soc att	tan	task av	tan	tan	soc av
	tan	soc av	soc att	soc att	soc av	task av
	soc av	auto	soc av	soc av	soc att	tan

	INT	ABC	MAS1	MAS2	ANA	DESC
Predictions for secondary functions in rank order	soc att	soc att	task av	soc att	soc av	soc av tan
	soc av	task av soc av	soc att	soc av	task av	
	tan		tan	task av tan	soc att tan	task av
	task av	auto	auto			soc att
	auto	tan	soc av	auto	auto	auto