Journal of Autism and Developmental Disorders, Vol. 34, No. 1, February 2004 (© 2004)

On Mosaics and Melting Pots: Conceptual Considerations of Comparison and Matching Strategies

Jacob A. Burack,[1,2,5] Grace Iarocci,[3] Tara D. Flanagan,[1] and Dermot M. Bowler[4]

Conceptual and pragmatic issues relevant to the study of persons with autism are addressed within the context of comparison groups and matching strategies. We argue that no choice of comparison group or matching strategy is perfect, but rather needs to be determined by specific research objectives and theoretical questions. Thus, strategies can differ between studies in which the goal is to delineate developmental profiles and those in which the focus is the study of a specific aspect of functioning. We promote the notion of a "mosaic," rather than a "melting pot," approach to science in which researchers communicate conservative and precise interpretations of empirical findings.

KEY WORDS: Autism research; comparison groups; matching strategies; methodology; mosaic.

INTRODUCTION

The notion that a meaningful comparison group is informative for understanding characteristics of a specific person or group of persons is neither new nor restricted to science. Rather, it can be found throughout Western history and in everyday parlance. One ancient example occurs in the book of Genesis in which Noah is described as a righteous person within his own generation. Although a dubious compliment for one of the few survivors of a generation that was virtually wiped out because of its decadence, the verse provides a context for evaluating the degree and limitations of Noah's virtue. In contemporary popular culture, the development of children is commonly discussed in relation to that of their peers as parents routinely describe their child's height and birth in terms of national percentiles. The relative rank might be a welcome sign of typical development, a positive indicator of an area of exceptional development, or a cause for concern. From Noah to contemporary culture, some characteristic of the person of primary interest is better understood because of the comparison to some group of other persons.

The choice of the comparison group or groups is essential to understanding the primary character as it provides the context and constraints for interpretation of the individuals' characteristics or level of performance. For example, Noah's contemporaries are an appropriate comparison group because those were the people among whom he lived, but their apparently pervasive wickedness diminishes any evaluation about the extent to which Noah's level of righteousness was exemplary by higher standards. In the example of childhood rankings, the comparison group is typically made up of the national population and provides a relative metric of the children's level of development for the specific measurement of interest. However, comparison groups that are more similar or matched on a specific dimension to the child might provide for a more precise evaluation of development in relation to specific questions or considerations. For example, a child's weight or height rank that is

[1] Department of Educational Psychology, McGill University, Montréal, Quebec, Canada H3A 1Y2.
[2] Canadian Center for Cognitive Research on Neurodevelopmental Disorders, Hôpital Rivière des Prairies, Québec, Canada H1E 1A4.
[3] Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6.
[4] Department of Psychology, City University, London, United Kingdom EC1V 0HB.
[5] Correspondence should be addressed to Jake Burack, Department of Educational Psychology, McGill University, 3700 McTavish Street, Montreal, Québec H3A 1Y2, Canada; e-mail: jake.burack@mcgill.ca

0162-3257/04/0200-0065/0 © 2004 Plenum Publishing Corporation

relevant to his or her gender, ethnicity, socioeconomic status, or type and level of disability is more informative than one based on a comparison with the national population. Thus, the researchers' story can be even more precise with comparison groups that are chosen because they "match" the target person or group on one or more characteristics. Yet, the choices of comparison group and matching measure are only beneficial to the extent that they inform about the questions that are asked and the story that is told.

As our contribution to this special issue, we highlight both the role of matching as a theoretical consideration in the interpretation of research and the necessity of increased precision in delineating comparison groups in order to provide more accurate and fine-tuned "stories" about persons with autism. We focus on two issues: one, the populations from which comparison groups should be chosen, and two, the criteria on which they should be matched to the target group. For both of these issues, the bottom line is that no strategy is a perfect one, and that the choices need to be based on the researchers' objectives and theoretical questions. Researchers need to vary their comparison and matching strategies in relation to the story they wish to tell and, in turn, the outcome of that story needs to be interpreted within the context of the chosen strategies. Accordingly, empirical work on autism might best be considered within a "mosaic" framework that allows for smaller but more precise stories rather than a "melting pot" one with grander but less exact information.

THE QUESTIONS THAT PEOPLE ASK

The primary considerations in determining the optimal choice of comparison group and matching criteria are the questions of researchers. In the behavioral study of persons with autism, these questions are typically focused on identifying characteristics that are both unique to the population and essential to the disorder. The emphasis is typically on sources of impairments, but examples of particular strengths or wellness can especially be informative. Although comparison questions can be framed within a variety of frameworks, they are generally discussed in relation to the somewhat interrelated questions of normalcy and uniqueness (Burack, Iarocci, Bowler, & Mottron, 2002; Wagner, Ganiban, & Cicchetti, 1990; Zelazo, Burack, Benedetto, & Frye, 1996). The normalcy question refers to the extent to which the one or more specific aspects of functioning among persons with autism is similar to or different from that of typically developing persons. The uniqueness question extends the normalcy one as it is focused on assessing the extent to which specific characteristics that are atypical among persons with autism are found exclusively in that group and are not common among other clinical populations. Of course, with regard to both these questions, the rates and pathways of development are not identical for the members of any group or population, so comparisons are based on some averaged index for each of the groups.

Each of the questions provides general frameworks that can be further fine-tuned, but all are associated with separate implications for choices in selecting comparison and matching procedures (Burack, 1997; Wagner *et al.,* 1990). The normalcy question is conceptually more straightforward as it solely involves typically developing persons as the comparison participants. Any group differences might involve the identification of strengths or weaknesses of the target group, but the reference point is typical development. The uniqueness question is more complex because the inherent question can never be exhaustively answered unless all possible populations and groups are tested as comparison groups, and the possible reasons for including a specific comparison group are endless. For example, one comparison group might be useful because they show similarities in behavior to the persons with autism, another because they show patterns of behavior that appear to be in direct contrast, and still others because they display some combination of similarities or differences.

THE COMPLEX CASE OF MATCHING IN AUTISM

Matching and comparison strategies for studies with persons with autism are complicated by at least three characteristics that are pronounced in this group, even if they are not necessarily unique to it. The most defining challenge in the study of autism is the pattern of pronounced strengths and weaknesses, commonly described as peaks and valleys of abilities, in the cognitive, linguistic, and social profiles of persons (Burack *et al.,* 2002). These specific discrepancies are intrinsically linked to the diagnostic criteria and core foundations of the disorder, and are, therefore, not found among other populations, each of which is likely to display its own unique set of profiles (Burack, Pastò, Porporino, Iarocci, Mottron, & Bowler, 2001a; Zelazo, Burack, Benedetto, & Frye, 1996). Therefore, careful matching of persons with autism to those from another group of persons in one aspect or domain of functioning is likely associated with discrepancies in other areas.

The second challenging characteristic is that the range of intellectual functioning is vast and spans levels from severe mental retardation to those that are considerably above average. Although reports of the specific distribution vary across studies, the consensus is that a considerable number of persons fall within the range of mental retardation. This necessitates the assessments of the effects of mental retardation on the manifestation of autism in the low IQ group and the more general consideration of similarities and differences among persons with autism of different IQ levels. As in the cases of other groups who function in the range of mental retardation, low IQ complicates comparison strategies with typically developing children who obviously cannot be matched on both chronological age (CA) and developmental level or mental age (MA) (Burack, Evans, Klaiman, & Iarocci, 2001b; Zigler & Hodapp, 1986). This leads to several questions regarding the equitability of a single mental age that was attained with considerably different rates of developmental progress, the possible role of IQ or related factors on performance, and the role of experience in relation to the CA–MA disparity.

A third consideration is that the life experiences of a person with autism are quite different from those of someone else, regardless of the extent of the discrepancy between MA and CA. As autism is defined by a triad of impairments that are relevant to interactions with people and other aspects of the environment (APA, 2000; Klin & Volkmar, 1997; Wing & Gould, 1979), the social world of the developing person with autism is typically one that differs considerably from those of other people. These social deficits, as well as concomitant ones in cognitive areas of functioning, generally necessitate interventions that, even when administered in the home or in a regular school, are unique to persons with autism. The implications for performance on both matching and experimental measures are likely to be mixed with enhanced performance on aspects of functioning that are emphasized in the intervention programs and deleterious effects on those aspects that may be more relevant to typical developmental experiences.

THE CHALLENGING LANDSCAPE OF PEAKS AND VALLEYS

Autism, like all behavioral disorders, is clinically defined by a specific pattern of characteristics. As with many so-called developmental disorders, these characteristics are associated with profiles of strengths and weaknesses across social and cognitive areas of functioning (Mottron & Burack, 2001; Siegel, Minshew, & Goldstein, 1996). Although persons with autism show considerable within-group variability, their peaks and valleys are both unique to the disorder and especially pronounced in relation to the generally standardized, flat profiles of large groups of typically developing persons. This issue of discrepant areas of functioning provides the central challenge to matching strategies, regardless of whether the intent is to address issues of normalcy in relation to typically developing persons or uniqueness in relation to persons with other types of atypicality.

The primary issue in choosing the matching measure is the extent to which it should be specific to the task at hand. At one extreme, full-scale IQ tests are used as matching measures because they are associated with the benefits of well-standardized tests that provide some metric of functioning that is inclusive of various areas of functioning, valid, reliable, and widely understood (for a discussion of complications regarding the administration of IQ tests as matching measures, please see Mervis & Robinson [1999]). However, regardless of the comparison group, a matching measure of general functioning represents some compromise score across areas of strength and weakness and, therefore, will likely either under- or overestimate the level of functioning in the area of interest. For example, if the area is one of strength for the persons with autism, then the persons with autism matched with another group on a general level of functioning will be likely to perform better because the measure underestimates their level of performance in that area. The converse is true for an area characterized by relatively low functioning.

The alternative to the use of full-scale measures of functioning is to match on measures that are intended to assess specific areas of functioning. This strategy highlights the peaks and valleys of functioning with assessments of functioning that are more fine-tuned for a specific task. For example, a comparison group matched on language processing, an area of general weakness among certain subgroups of persons with autism (Tager-Flusberg & Joseph, 2003), is likely to be considerably lower functioning than one matched on visual-spatial processing, an area of relative strength among persons with autism. One common solution to the problem of the disparity among domains of functioning is to match on a domain of functioning that is relevant to performance on the experimental task. For example, the use of a language matching measure is helpful when the experimental task requires language abilities in order to ensure that any group differences

that are found are due to differences on the ability to complete the task rather than to *a priori* discrepancies in language abilities. This rationale is the basis for the common use of a language measure as a matching tool on theory of mind studies and for measures of nonverbal performance on tasks of perception and attention.

The inherent flaw in the use of more specific measures of matching is that performance typically involves at least some components from various domains, so matching on one domain might leave discrepant levels of functioning between the groups in other relevant ones. This difficulty is further highlighted even for within-domain considerations, as the measures that are best standardized and easiest to administer do not necessarily reflect the complexities of the type of functioning they are intended to measure. For example, the Peabody Picture Vocabulary Test–Revised (PPVT-R) (Dunn & Dunn, 1997) and the British Vocabulary Scale (BPVS) (Dunn, Dunn, Whetton, & Pintillie, 1982), the most commonly used measures for matching in studies of theory of mind and other tasks with language instructions, are tests of one-word receptive language that do not reflect the various linguistic abilities that are necessary to understand and perform the tasks, even when minimal language is required in the response. Thus, differences between groups on the task might be associated with disparities in other areas of language functioning. This is even a likely source of the finding of deficits among persons with autism because performance on the one-word receptive vocabulary tests tends to be considerably higher than on measures of either general or specific aspects of language functioning (for an empirical demonstration, see Mottron, this issue). Accordingly, persons with autism as compared to other groups matched on one-word vocabulary tests would be expected to show lower levels of performance in virtually all areas. The converse may be true in other situations in which the matching measure underestimates the abilities of persons with autism who would, therefore, be likely to display better performance on most areas as compared to the matched groups. However, the former scenario is more common, as persons with autism are typically difficult to test and, therefore, tests on which they seem to do well are preferred.

THE IQ RANGE AND IMPLICATIONS FOR GROUP SELECTION, AND THE CA–MA TRADEOFF

The wide IQ range among persons with autism raises at least two primary issues relevant to the use of comparison groups and matching strategies. One concerns the possible relations between functioning and IQ level and the second concerns the discrepancy between CA and MA that is, by definition, found among persons with mental retardation but obviously not found among typically developing persons with average IQs. The former issue is one that necessitates a comparison among persons with autism of varying levels of cognitive functioning, but matched on developmental level. Among other complications, that, of course, leads to the second issue of the CA–MA discrepancy which is intrinsic to the study of lower functioning persons. In general, this latter issue led to two divergent strategies. One was to ignore this group and essentially focus on the higher functioning persons with autism. However, as many persons with autism function in the range of mental retardation, a second, more inclusive strategy is to compare persons with autism with low IQ to other groups of persons who function in the range of mental retardation and therefore can be matched on both CA and MA.

The strategy of comparison groups of persons with mental retardation necessitates a consideration of the makeup of the comparison group. A group of persons with mental retardation can be either homogeneous or heterogeneous with regard to etiology. As discussed in this special issue and elsewhere, both strategies are used. We advocate the use of homogeneous groups such as those made up exclusively of persons with trisomy 21, Williams syndrome, or fragile X. Needless to say, each of these groupings can be further fine-tuned to a state of deconstruction in which only one person would fit the criteria. However, the idea is to group in a meaningful way with regard to the functioning in the given area of interest. Thus, if persons within an etiological group show some generally defined pattern of functioning, they might prove to be an informative comparison group. The question as to which group to use does not have a correct response. No single group is intrinsically better than another one because they all, by definition, differ significantly from persons with autism, unless the subgroup of persons with autism is a unique one in which all its members belong to a single etiological group. Thus, the rationale for choosing any group needs to be based on the researchers' theories, interests, and questions. The reasons for inclusion for one as opposed to another, in addition to the obvious one of convenience, might just as well be because of certain similarities or differences between the comparison group and persons with autism. Any choice is potentially informative, but clearly the implications of any findings with regard to strengths or weaknesses can only be considered within the context of the specific comparison group.

The findings with regard to strengths or weaknesses can clearly only be considered within the context of the specific comparison group, but these may also be extrapolated to more general considerations within the framework of appropriate theoretical questions. For example, the comparison of patterns of performance on the processing of global–local aspects of nonsocial information and face processing in persons with autism and those with Williams syndrome might highlight a potential disassociation between these two aspects of perception that are thought to be related in typically developing individuals. According to preliminary evidence, face processing is impaired among persons with autism (Klin *et al.,* 1999) but spared among persons with Williams syndrome (Rossen, Klima, Bellugi, Bihrle, & Jones, 1996), although both groups show similarities in terms of intact global processing (for studies of persons with autism, see Mottron, Burack, Iarocci, Belleville, & Enns [2003] and Plaisted, Swettenham, & Rees [1999]; and for a study of persons with Williams syndrome, see Pani, Mervis, & Robinson [1999]). These patterns are unique to these two groups, but the contrast between them can provide insight into the more general issues of the integrity and disassociation of the relation between the two aspects of perceptual functioning.

The limitations of the conclusions between a specific etiological group and the persons with autism as well as the difficulties associated in recruitment of these types of groups are commonly cited as reasons for using a group of persons with mental retardation that are heterogeneous with regard to etiology. The advocates argue for greater generalizability, but this is not the case (Burack, 1990). A heterogeneous group does not reflect a larger group, but rather no group at all (Burack, Evans, Klaiman, & Iarocci, 2001b). The characteristics of heterogeneous groups are simply a function of the particular makeup of the specific group at the time and is, therefore, not likely to be the same as any other group mix of persons with differing etiologies (Burack *et al.,* 2002). A true hodge-podge grouping is particularly problematic with regard to matching strategy because any distinctiveness that may be associated with the specific etiological groups that are represented are likely to be neutralized by the mixture of members. Thus, any area of particular strength or weakness among persons with autism, or any other group, are likely to differ from this "homogenized" profile of the heterogeneous grouping. For these reasons, we do not adhere to the rationale that a heterogenous mix of persons provides a representative comparison group that allows researchers to account for the role of mental retardation (for an earlier incarnation, see Burack & Volkmar [1992]).

THE CONFOUNDING ISSUE OF THE UNIQUE LIFE EXPERIENCES OF PERSONS WITH AUTISM

In addition to the CA–MA discrepancy for many of the persons with autism, the life experience of the entire group is different from that of the typically developing population as well as from persons with other types of atypical developmental history. The symptoms of autism begin in the first years of life, are severe and pervasive, and affect all aspects of social interactions. These characteristics alone would ensure meaningfully discrepant life experiences due to the atypical social and cognitive development that resulted, but these differences are further pronounced due to societal reactions and interventions. Many persons with autism were institutionalized until a decade or two ago, and most continue to be educated in special schools or, at least, special classrooms. Regardless of the particular setting, current programs involve a level of intensive training, of one or more orientations, that is unlikely to be experienced by any other group. As these intensive training and educational programs are considered to be most effective if they begin early, the entire developmental history of a person with autism may be constructed differently than those of his or her peers.

The potential effects of unique developmental and experiential histories on empirical research is highlighted in the work by Zigler and colleagues (Hodapp & Zigler, 1995; Zigler, 1967, 1969; Zigler & Hodapp, 1986) who argue that the performance of persons with mental retardation is largely influenced by personality characteristics and life experiences on everyday life tasks, but even more so in the artificial and unfamiliar experimental settings. The combination of one or more components of a lifelong self-perception of failure and expectancy of failure, educational and residential segregation, and societal stigmatization were associated with deleterious performance strategies and characteristics such as the excessive dependence on cues from the environment, lower levels of intrinsic motivation, and general wariness of others (Zigler, Bennett-Gates, Hodapp, & Henrich, 2002; Zigler & Hodapp, 1986). Some of these issues are also likely to affect deleteriously the performance of persons with autism who share some aspects of life experiences with mental retardation. Conversely, the intensive

intervention programs that are administered to persons with autism may also lead to enhanced performance on some tasks. For example, better than expected performance may be evident on certain measures of social functioning because of intensive intervention involving extensive practice with social scripts rather than as a result of improved understanding of social convention.

CONSIDERATIONS OF ALTERNATIVE MATCHING STRATEGIES

The methodological issues raised here and elsewhere (e.g., Burack *et al.,* 2002; Mervis & Robinson, 1999) with regard to the challenges in choosing comparison groups and matching strategies preclude definitive studies of autism. In that case, what can researchers do other than simply throw up their arms in despair and defeat? We suggest that the response needs to be both pragmatic and conceptual, and that the strategies adopted at each of these levels should essentially be guided by the questions and goals of the researchers. The ultimate goal of this framework is the increased precision of both methodology and interpretation.

Even when the various potential problems regarding comparison groups and matching are considered, the potential strategies are further compromised by practical considerations such as researchers' resources, the likelihood of finding participants who are matched precisely in the way that is intended, and the limitations of available tests with which to match. In this scenario, the questions and intentions of researchers become even more salient in the choice of matching strategies as the choice ultimately determines the extent to which the findings are precise, broad, conservative, and inform about the issues under consideration. The effectiveness of the strategy can be evaluated in relation to the confounding factors that can be eliminated.

Scores of general developmental level based on rigorously constructed and highly valid and reliable full-scale standardized measures of IQ are not particularly useful for matching in a study of a specific component of functioning due to the extreme terrain of peaks and valleys of performance among persons with autism. Rather, this type of index might be most useful for matching in studies that are focused on the profile of strengths and weaknesses at a specific developmental or IQ level. Different scores on these tests provide a general measure that both integrates strengths and weaknesses into a single measure and also allows for matching on specific subcomponents. In this case, the general level of functioning provides a basic metric for evaluating the disparate levels of performance across domains that contribute ironically both to the complex developmental structures and to performance on the measures of development. Thus, matching by developmental measure is useful in negating the notion that the patterns of strengths and weaknesses or level of disparity in functioning across domains occur in relation to one or more specific developmental or IQ levels.

In the case in which the focus is functioning in specific domains and more task-relevant matching measures are needed, the most common strategy is to apply a single test that is easy to administer and can provide an approximate measure of developmental level within a general domain of functioning that is necessary for performance on the experimental task. Accordingly, the PPVT, which is a simple and widely used measure of receptive language, is typically used as the matching measure in studies of theory of mind, which involves rather complex verbal instructions. Similarly, the Raven's Matrices Test and other similar tests of pattern processing are used as matching measures on tests that are thought to entail some aspect of visual-spatial functioning. Paradoxically, the use of these tests, as well as all other standardized tests, as single matching-measures is problematic both because they do not allow for sufficiently precise comparisons with the abilities that are necessary for task performance and because they are not general enough to account for the complexity of the type of processing they are intended to represent or of the more likely scenario of processing across multiple domains. One of the most problematic sequelae is that findings, and therefore the conclusions, can vary considerably in relation to the specific matching measure even with the same comparison population (e.g., Russo, Burack, Flanagan, Zelazo, & Iarocci, in press). These limitations suggest the utility of two contrasting strategies: one involves a broader framework with matching on multiple measures, and the other involves a narrower strategy in which the matching is based on an area of functioning that is closely linked to task performance.

The notion of multiple matching entails the use of different comparison groups from the same population that are each matched on a specific measure (Burack *et al.,* 2002). For example, one group might be matched on a language measure such as the PPVT or mean length utterances, another on a visuo-spatial measure such as the Raven, and yet a third on some aspect of functioning that is similar to the experimental task. This strategy provides a profile of performance regarding

functioning on the specific task in relation to the various matching measures. The pragmatic aspect involves the assessment of the level of uniformity of findings across matching measures with the notion that consistent findings of difference may reflect more pervasive difficulties and that those that are only evident with one or some measure may be tied to more specific aspects of functioning. The more conceptual contribution concerns the relative contributions, or associations, of the various domains reflected in the matching measures to the abilities necessary for task performance. This type of framework can be used to construct extensive representations of the complex developmental relations across domains of functioning. However, the conceptual efficacy is clearly contingent on the extent to which the various measures reflect the domains they are thought to represent. For example, the limitations of the PPVT or Raven may restrict their value in illustrating a developmental profile of task performance in relation to language and visual-spatial functioning, although those are the domains they are thought to reflect. In this context, the value is the identification of contrasting relations among the task and the measures more than among the domains that are implicated conceptually.

In contrast to broadening the scope of matching, an alternative strategy entails increasingly precise matching that involves matching on a measure that is particularly close to the experimental one. The objective is to minimize the likelihood that any group differences in performance on the experimental task can be attributed to *a priori* differences in some aspect of functioning. In this framework, the choice of the matching measure depends on the researchers' questions, interests, and choice of experimental measures. Thus, in a behavioral study of development, the participants might be matched on a developmental precursor or correlate to the ability that is necessary to complete the experimental task. For example, in a study of attention and memory factors involved in solving false-belief tasks, for which the mean verbal mental age for passing among children with autism is approximately 7 years as compared to 4 to $4\frac{1}{2}$ years for typically developing children (Happé, 1995), the groups might be matched on a score associated with passing percentage on false-belief tasks. This would minimize the extent to which any group differences in the specific factors under consideration could be attributed to *a priori* differences in false-belief tasks. Alternatively, the amplitude of an early sensory component (e.g., the N1 or P1) could be used to match subjects for the assessment of the N170 that is associated with face recognition. This would allow the face-specific response, which is thought to emanate from the fusiform gyrus, to be evaluated independently of any sensory differences across individuals. These types of precise matching strategies enhance the confidence in the veracity of group differences and should be considered whenever possible, although they lead to considerably more work for the researchers.

The potential conceptual limitations of this approach are that the identification of an appropriate matching measure is methodologically difficult and that group differences might be obscured as a result of this especially conservative approach. With regard to the methodological difficulties, even similar scores between the persons with autism and the comparison groups on a matching measure that is rather closely associated with the experimental task may not ensure that the chosen strategy is optimal. In particular, the failure to find group differences in performance, whether on matching measures or on experimental paradigms, is often the artifact of the use of tasks that are not optimal for differentiating performance because of their developmental inappropriateness or level of difficulty (Burack *et al.,* 2002). With regard to the implications for the findings associated with this approach, the use of precise measures might be most informative when combined with more general measures that are more likely to elicit group differences, thereby allowing for relatively fine-tuned profiles of relations among domains of functioning.

CONCLUSIONS: A CAUTIONARY STORY OF OPTIMISM

The inherent pragmatic and conceptual problems inherent in all potential matching strategies reflect a particularly problematic component of the scholarship about persons with autism. The notion that no one solution is available to researchers of autism should not be seen as a source of despair but rather as a challenge to rework various aspects of the conceptualizations and representations of the research. Even factors concerning the life experiences and problem behaviors of persons with autism that can never be accurately represented in a comparison group should not be viewed as damning, but rather as another consideration in the quest to better understand persons with autism. Within this framework, we advocate the notion of a "mosaic" rather than a "melting pot" approach to science in which researchers and the journals in which they publish promote more conservative and precise

interpretations of empirical findings. In formal scientific publications and lectures as well as in informal communication and discussions, specific findings about persons with autism should not be translated into broad-brushstroke descriptions of characteristics with little concession to the limitations of the evidence (for a discussion, see Bowler [2001]). Instead, researchers need to couch precisely their findings with regard to the comparison group and matching measures as well as to any of the other factors that may affect the findings.

The mosaic approach is essential for discussing differences between groups. Researchers need to acknowledge that the differences they report might be found only under the specific circumstances of the study and that the extrapolation to some generalized notion of a deficit can only be advanced with substantial convergent evidence. However, this approach is even more essential to conclusions about findings of no differences between persons with autism and other groups that are based on a single or even a few studies. The failure to find differences may be the consequent of any of so many possible factors (Burack *et al.,* 2002) that generalizations can only be forwarded with evidence from multiple sources of convergent findings. In these cases, the emphasis should be focused on similarities in patterns of performance rather than simply on the failure to find statistical differences.

In sum, we argue that problems inherent in the comparisons of persons with autism and persons from other populations, as well as pragmatic considerations, preclude research orthodoxy with only one acceptable solution. Rather, any of several strategies might be used with the constraint that in each case, the group is a meaningfully homogeneous one and the matching measure appropriately reflects the ability for which it is intended to provide a metric. Although they differ with regard to precision and even preference, all may be informative within the context of certain questions, and all need to be carefully interpreted. Within this framework, the focus of the literature on autism needs to be shifted from broad brushstrokes of melting pot approaches to a more precise mosaic in which the findings are discussed within the context of the comparison groups, matching measures, developmental level, experimental task, and other factors.

ACKNOWLEDGMENTS

Jake Burack's work on this paper was supported by a research award from the Social Sciences and Humanities Research Council of Canada; Grace Iarocci's by an award from the Human Early Learning Partnership; Tara Flanagan's by a doctoral research fellowship from the Social Sciences and Humanities Research Council of Canada; and Dermot Bowler's by a research grant from the Wellcome Trust. The authors thank Jim Enns, David Shore, and Laurent Mottron for their conceptual contributions to the ideas presented in this paper.

REFERENCES

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders,* Fourth Edition, Text Revision. Washington, DC: American Psychiatric Association.

Bowler, D. M. (2001). Autism: Specific cognitive deficit or emergent end point of multiple interacting systems? In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 219–235). Mahwah, NJ: Lawrence Erlbaum.

Burack, J. A. (1990). Differentiating mental retardation: The two-group approach and beyond. In R. M. Hodapp, J. A. Burack, & E. Zigler (Eds.), *Issues in the developmental approach to mental retardation* (pp. 27–48). New York: Cambridge University Press.

Burack, J. A. (1997). The study of atypical and typical populations in developmental psychopathology: The quest for a common science. In S. S. Luthar, J. A. Burack, D. Cicchetti, & J. R. Weisz (Eds.), *Developmental psychopathology: Perspectives on adjustment, risk, and disorder* (pp. 139–165). New York: Cambridge University Press.

Burack, J. A., & Volkmar, F. R. (1992). Development of low- and high-functioning autistic children. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 33,* 607–616.

Burack, J. A., Pastò, L., Porporino, M., Iarocci, G., Mottron, L., & Bowler, D. M. (2001a). Applying developmental principles to the study of autism. In E. Schopler, N. Yirmiya, C. Shulman, & L. M. Marcus (Eds.), *The research basis for autism intervention* (pp. 25–41). New York: Kluwer Academic/Plenum Publishers.

Burack, J. A., Evans, D. W., Klaiman, C., & Iarocci, G. (2001b). The mysterious myth of attention deficit and other defect stories: Contemporary issues in the developmental approach to mental retardation. *International Review of Research in Mental Retardation, 24,* 299–320.

Burack, J. A., Iarocci, G., Bowler, D. M., & Mottron, L. (2002). Benefits and pitfalls in the merging of disciplines: The example of developmental psychopathology and the study of persons with autism. *Development and Psychopathology, 14,* 225–237.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test,* 3rd edition. Circle Pines, MN: American Guidance Services.

Dunn, L., Dunn, L., Whetton, C., & Pintillie, D. (1982). *British Picture Vocabulary Scale*. Slough, UK: NFER-Nelson Publishing.

Happé, F. G. E. (1995). The role of age and verbal ability in the Theory of Mind task performance of subjects with autism. *Child Development, 66,* 843–855.

Hodapp, R. M., & Zigler, E. (1995). Past, present, and future issues in the developmental approach to mental retardation and developmental disabilities. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology: Vol. 2. Risk, disorder, and adaptation* (pp. 299–331). New York: Wiley.

Klin, A., & Volkmar, F. R. (1997). The pervasive developmental disorders: Nosology and profiles of development. In S. S. Luthar & J. A. Burack (Eds.), *Developmental psychopathology: Perspectives on adjustment, risk, and disorder* (pp. 208–226). New York: Cambridge University Press.

Klin, A., Sparrow, S. S., de Bildt, A., Cicchetti, D. V., Cohen, D. J., & Volkmar, F. R. (1999). A normed study of face recognition in autism and related disorders. *Journal of Autism and Developmental Disorders, 29,* 499–508.

Mervis, C., & Robinson, B. F. (1999). Methodological issues in cross-syndrome comparisons: Matching procedures, sensitivity, and specificity. *Monographs of the Society for Research in Child Development, 64,* 115–130.

Mottron, L., & Burack, J. A. (2001). Enhanced perceptual functioning in the development of autism. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 131–148). Mahwah NJ: Lawrence Erlbaum.

Mottron, L., Burack, J. A., Iarocci, G., Belleville, S., & Enns, J. T. (2003). Locally oriented perception with intact global processing among adolescents with high-functioning autism: Evidence from multiple paradigms. *Journal of Child Psychology and Psychiatry, 44,* 904–913.

Pani, J. R., Mervis, C. B., & Robinson, B. F. (1999). Global spatial organization by individuals with Williams syndrome. *Psychological Science, 10,* 453–458.

Plaisted, K., Swettenham, J., & Rees, L. (1999). Children with autism show local precedence in a divided attention task and global precedence in a selective attention task. *Journal of Child Psychology and Psychiatry, 40,* 733–742.

Rossen, M., Klima, E. S., Bellugi, U., Bihrle, A., & Jones, W. (1996). Interaction between language and cognition: Evidence from Williams syndrome. In J. H. Beitchman, N. J. Cohen, M. M. Konstantareas, & R. Tannock (Eds.), *Language, learning, and behavior disorders: Developmental, biological, and clinical perspectives* (pp. 367–392). New York: Cambridge University Press.

Russo, N., Burack, J. A., Flanagan, T., Berringer, D., Iarocci, G., & Zelazo, P. D. (in press). Deconstructing the executive function deficit in autism: Implications for cognitive neuroscience. *Brain and Cognition*.

Siegel, D. J., Minshew, N. J., & Goldstein, G. (1996). Wechsler IQ profiles in diagnosis of high-functioning autism. *Journal of Autism and Developmental Disorders, 26,* 389–406.

Tager-Flusberg, H., & Joseph, R. M. (2003). Identifying neurocognitive phenotypes in autism. *Philosphical Transactions of the Royal Society of London. Series B, Biological Sciences, 28,* 303–314.

Wagner, S., Ganiban, J. M., & Cicchetti, D. (1990). Attention, memory and perception in infants with Down syndrome: A review and commentary. In D. Cicchetti & M. Beeghly (Eds.), *Children with Down syndrome: A developmental perspective* (pp. 147–179). New York: Cambridge University Press.

Wing, L., & Gould, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders, 9,* 11–29.

Zelazo, P. D., Burack, J. A., Benedetto, E., & Frye, D. (1996). Theory of mind and rule use in individuals with Down syndrome: A test of uniqueness and specificity claims. *Journal of Child Psychology and Psychiatry, 37,* 479–484.

Zigler, E. (1967). Familial mental retardation: A continuing dilemma. *Science, 155,* 292–298.

Zigler, E. (1969). Developmental versus difference theories of mental retardation and the problem of motivation. *American Journal of Mental Deficiency, 73,* 536–556.

Zigler, E., & Hodapp, R. M. (1986). *Understanding mental retardation*. New York: Cambridge University Press.

Zigler, E., Bennett-Gates, D., Hodapp, R., & Henrich, C. C. (2002). Assessing personality traits of individuals with mental retardation. *American Journal on Mental Retardation, 107,* 181–193.