A New Scale for Measuring Adults' Prosocialness

Article *in* European Journal of Psychological Assessment · January 2005

DOI: 10.1027/1015-5759.21.2.77

4 authors:

Gian Vittorio Caprara
University of Southern Mississippi

Patrizia Steca
Università degli Studi di Milano-Bicocca

Arnaldo Zelli
Italian University of Sport and Movement Sciences "Foro Italico"

Cristina Capanna

A New Scale for Measuring Adults' Prosocialness

G.V. Caprara[1], P. Steca[2], A. Zelli[3], and C. Capanna[1]

[1]University of Rome, "La Sapienza," [2]University of Milan, "Bicocca," [3]University of Motor Sciences, Rome, all Italy

Abstract. In the present study, the authors proposed a novel self-report 16-item scale for assessing individual differences in adult prosocialness and tested its measurement properties by employing an item response theory (IRT) analysis of data collected from a sample of 2,574 Italian adults. Prior work employing classical psychometric methods of analysis had already established the reliability and validity of the instrument. The present study furthered this scrutiny by examining whether the different prosocialness items were equally effective in discriminating people and equally informative; it also examined gender differences in the functioning of the items. The results of IRT analyses strongly supported the measurement effectiveness and sensitivity of the 16 prosocialness items, and findings are discussed for their implications in behavioral assessment research on prosocialness.

Keywords: Prosocialness, assessment, item response theory

Prosocialness is an elusive topic of psychological inquiry. This elusiveness partly derives from complexities in the construct for which, aside from its common sense meaning – the set of voluntary actions one may adopt to help, take care of, assist, or comfort others –, there still exists a lively debate concerning its critical measurement and psychological components (Bar-Tal, 1982; Batson, 1991; Eisenberg & Fabes, 1998; Schroeder, Penner, Dovidio, & Piliavin, 1995). Its elusiveness also, in part, derives from its overall heuristic value. Developmental research has clearly documented that prosocial responding becomes relatively stable during late childhood and early adolescence and that it arises from complex developmental and psychological processes involving attentional and evaluative processes, moral reasoning, social competence, and self-regulatory capacities (Caprara & Pastorelli, 1993; Eisenberg & Fabes, 1998; Krebs & Van Hesteren, 1994). Less knowledge, however, exists with regard to the psychological meaning of prosocialness for personal well-being and adjustment during later stages of life. Lastly, the elusiveness of prosocialness also to some extent derives from the difficulty of reaching a consensus on the best assessment and measurement strategies with the most widespread applicability, especially with respect to the assessment of adults' prosocial responding.

As a result, the study of prosocialness in adulthood still is in its "infancy," and the research contributions that one can compile from the existing literature (Eisenberg et al., 1999, 2002; Koestner, Franz, & Weinberger, 1990) show few commonalities. In the present study, we developed and rigorously evaluated a measure of adults' prosocialness with clear theoretical value and widespread applicability to different research contexts. We first administered a relatively short self-report scale of prosocialness to a large population of Italian adults and then evaluated the new instrument's characteristics, performing analyses that are in accord with the principles and methods of item response theory (IRT), a psychometric approach that in the last two decades has been widely employed in personality/behavioral assessment research (Steinberg & Thissen, 1996; Van der Linden & Hambleton, 1997; Waller & Reise, 1989).

Undoubtedly, prosocial behavior has prognostic value for individuals' personal and social adjustment (Caprara, Barbaranelli, Pastorelli, Bandura, & Zimbardo, 2000; Caprara & Pastorelli, 1993). Research has clearly demonstrated, for instance, that an individual's characteristic prosocial behavior counteracts and protects him/her from experiences of depression and conduct problems, as well as promotes and sustains scholastic achievement throughout adolescence (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Bandura, Caprara, Barbaranelli, Pastorelli & Regalia, 2001; Bandura, Pastorelli, Barbaranelli, & Caprara, 1999; Caprara et al., 2000).

It is in adulthood, however, that prosocial behavior may acquire particular importance, social meaning, and

DOI 10.1027/1015-5759.21.2.77

heuristic value. During adulthood, in fact, one's tendency to act prosocially may be threatened and counteracted by interpersonal experiences (e.g., those encountered in work environments) that draw upon alternative values or goals such as, for instance, competition, personal achievement, and individual power. It also is in adulthood that the need to be assisted, helped, and supported by others might increase in order to overcome the difficulties associated with aging, increasing physical limitations, fewer opportunities for social and interpersonal contact, and loneliness. Importantly, the chances of fulfilling these personal needs may critically rest and depend on an adult's own capacity to provide help, assistance, and support to others, as well as to actively establish or maintain strong social relationships (Eisenberg et al., 2002).

Finally, it is also in adulthood that the relation between gender and prosocial responding might find its fullest meaning (Belansky & Boggiano, 1994; Bussey & Bandura, 1999; Eagly & Crowley, 1986). Men and women tend to display different prosocial behaviors, and these differences are linked to the type of behavior being enacted and the salient social situations or contexts in which one's behavior unfolds (Eisenberg & Fabes, 1998). For instance, men tend to be altruistic and help others especially when the situation calls for quick and decisive actions or when someone is in need of clear help or is in serious danger. On the contrary, women typically enact their altruistic behaviors in the context of familiar and long-term relationships in which a partner, a friend, or a work colleague is in need of nurturance, support, caring, or emotional empathy (Eagly & Crowley, 1986). These and other gender differences in prosocialness probably arise from a host of psychological factors and mechanisms ranging from the social norms associated with gender roles (Eisenberg & Fabes, 1998; Eagly & Crowley, 1986) to the social-cognitive mechanisms regulating gender-linked conduct consistent with gender development and differentiation (Bussey & Bandura, 1999). Regrettably, however, this body of knowledge on gender differences in prosocial responding comes primarily from social psychological experimental studies using adolescents and young adults and scrutinizing a limited spectrum of prosocial behaviors and situations, limiting, therefore, the external validity of the studies (see Eagly & Crowley, 1986, for a careful review; see also Eisenberg & Fabes, 1998).

The Assessment of Prosocial Responding in Adults

In light of the above considerations, it becomes imperative to attain rigor and clarity in the assessment and measurement of individual differences in adults' prosocialness. This task might be particularly daunting given the spectrum of qualitatively different behaviors that prosocial responding may entail (e.g., helping, caring, sharing, perspective-taking, empathy) and its high social desirability. The issue concerning qualitatively different behaviors poses a challenge for traditional methods of measurement in which behaviors are typically sampled with a single instrument and item scores are compiled and aggregated into a composite scale score. This approach assumes that a scale score represents a reliable measure of the construct for all participants, that items belonging to a scale are equally informative about a person's underlying trait, and that the scale can validly discriminate people with differing trait levels.

As the analysis of gender differences presented above has suggested, these assumptions hardly can hold for prosocial responding. Furthermore, as some studies have pointed out (Eisenberg, Carlo, Murphy, & Van Court, 1995), prosocial behavior may partly respond to cultural norms and values stressing solidarity, partnership, and cooperation. Thus, it is critical that assessment of prosocialness relies on analytical strategies allowing the evaluation of discrimination and fidelity of single items compiled in a measurement instrument. In the present investigation, we introduce a novel instrument measuring adult prosocialness. The generating criteria for this instrument relied heavily on the recommendations of existing developmental literature, which have clearly indicated how prosocialness primarily finds expression in actions of helping, sharing, taking care of, and feeling empathic with others (Caprara, Capanna, Steca, & Paciello, in press; Caprara & Pastorelli, 1993).

More importantly, however, we evaluated the properties and characteristics of this new instrument via a core analytical strategy, that of IRT, in which characteristics of items in a test (item parameters) and characteristics of individuals (latent traits) are related to the probability of a positive response (i.e., a trait-consistent endorsement of an item). Over the last decade or so, many researchers have highlighted the advantages that IRT provides over classical test theory (CTT) for constructing adequate personality and cognitive ability instruments and examining their measurement characteristics (Steinberg & Thissen, 1996; Van der Linden & Hambleton, 1997; Waller & Reise, 1989). Traditionally, researchers have focused on dimensional models of behavior and created multi-item instruments to assess individual differences along these dimensions. In particular, an individual's level on the trait dimension is typically determined by averaging or summing the responses to statements or items which are assumed to be reliable and valid indicators of the underlying trait. As noted by Hambleton, Robin, and Xing (2000), this approach is consistent with CTT and is con-

cerned with the estimation and control of error in the assessment process. The theory postulates a relation between subjects' responses on items and the latent trait dimension, and the error is basically operationalized as the differences between the latent trait and the observed performance This error is considered to be randomly distributed across the population of interest, and also within repeated measurements of the same person.

However, classical test procedures have shown some shortcomings. The first, and perhaps most critical problem, is that the theory assumes an equal error of measurement for all examined subjects or, to put it differently, an equal measurement precision across all levels of the trait dimension under study. As noted by Lord (1984), violation of the equal error variance assumption is almost the rule in most assessment contexts. A second problem is that the values of item statistics, as well as reliability and validity statistics, strictly depend on the characteristics of the sample being used; unfortunately, however, samples are rarely truly representative of the population of persons for whom the test is being constructed. Finally, the measurement quality of the dimensional total score, usually evaluated by reliability indices, is affected by the number of items included in the scale, and items on a scale are not equally informative of the underlying trait dimension throughout the trait continuum (i.e., for people with different trait levels).

The IRT analytical methods and procedures overcome these limitations of CTT (Hambleton & Swaminathan, 1985; Hambleton, Swaminathan, & Rogers, 1991; Hulin, Drasgow, & Parsons, 1983). First, IRT analyses can determine a test's measurement precision at any value of the latent trait; that is, they provide an indication of how the trait is precisely measured along its entire continuum. Second, IRT parameters are invariant with respect to the sample characteristics from which they are generated. Third, IRT methods can quantify the information value of both individual items and the overall test, and this information can be evaluated at any level of the latent trait. Finally, IRT methods also permit a direct comparison of different sets of items or entire scales referring to the same construct domain on the basis of their discriminative power.

The Present Study

IRT measurement modeling can bear important implications for the assessment of adult prosocialness. People can differ substantially in their motives to act prosocially, in the behavioral ways they manifest these motives, and in the personal value they assign to the enactment of these behaviors. These differences might emerge at both an individual and a group level (e.g., differences among males and females, across ages).

These considerations are consistent with the IRT notion of uneven information of an item or a test, that is, the fact that measurement precision is not constant across the entire trait range (i.e., is not the same for all people). In the context of the present investigation, for instance, some prosocialness items might be good at differentiating only among people in the high range of prosocialness, whereas other prosocial behavior items might differentiate well only among people at the low or middle range of prosocialness. Other prosocialness items may instead differentiate well among people situated along a relatively broad trait range.

Conventional methods would also be problematic in addressing the issue of whether group (e.g., gender) differences in mean levels of prosocial behavior reflect true differences among the groups or, rather, the fact that items combined in aggregated scores have different measurement properties in different groups. In the present investigation, the properties of a new instrument for measuring adult prosocialness were evaluated according to an IRT analysis of its single items and full scale. Importantly, this instrument was administered to, and data were collected from a very large population of Italian adults, thus, providing robustness to the findings and the study's ecological validity.

In the present study, we were interested in answering four specific questions, which we believe could not be adequately addressed via traditional assessment procedures: (1) Do different prosocialness items behave similarly, that is, are they equally discriminating and difficult? (2) Are these items equally informative across the different levels of the estimated latent trait of prosocialness? (3) What is the information provided by the full scale along the trait continuum? (4) Are there some differences between male and female adults in the items' difficulty and discrimination parameters?

Materials and Methods

Participants

The sample of participants in this study included 2,574 individuals who provided complete data on the variables of interest. The sample was balanced with respect to gender (1.278 males and 1.296 females), and representative of the adult population with respect to age (range = 18 to 92 years old, mean = 44.5 *SD* = 17). Demographically, 58% of the participants were married whereas 33% of the participants were single, and nearly 71% of the participants had earned at least a high-school degree.

Table 1. The sixteen items of the Prosocialness Scale for Adults.

1. I am pleased to help my friends/colleagues in their activities
2. I share the things that I have with my friends
3. I try to help others
4. I am available for volunteer activities to help those who are in need
5. I am emphatic with those who are in need
6. I help immediately those who are in need
7. I do what I can to help others avoid getting into trouble
8. I intensely feel what others feel
9. I am willing to make my knowledge and abilities available to others
10. I try to console those who are sad
11. I easily lend money or other things
12. I easily put myself in the shoes of those who are in discomfort
13. I try to be close to and take care of those who are in need
14. I easily share with friends any good opportunity that comes to me
15. I spend time with those friends who feel lonely
16. I immediately sense my friends' discomfort even when it is not directly communicated to me

The Instrument Measuring Adults' Prosocialness

Table 1 shows the 16 prosocialness items that were included in the adult prosocialness instrument. As the table shows, the items reflect behaviors and feelings that can be traced back to one of four types of actions, namely, sharing, helping, taking care of, and feeling emphatic with others and their needs or requests. While the first three types of actions have typically characterized the measurement of childhood or adolescence prosocial behavior, the inclusion of empathy represents a novel addition to the measurement of prosocialness. Such a decision followed the general hypothesis that, in adulthood, one's empathic motives or predispositions are not merely a correlate of his or her tendency to act prosocially but, rather, an integral part of such a tendency (Caprara et al., in press; Eisenberg & Fabes, 1998).

The 16 prosocialness items were presented to the participants with the following instructions: "The following statements describe a large number of common situations. There are no 'right' or 'wrong' answers; the best answer is the immediate, spontaneous one. Read carefully each phrase and mark the answer that reflects your first reaction." For each prosocialness item, participants indicated on a five-point Likert scale whether the statement was *never/almost never true* (coded as 1), *occasionally true* (coded as 2), *sometimes true* (coded as 3), *often true* (coded as 4), and *almost always/always true* (coded as 5).

IRT and Its Application to the Assessment of Adults' Prosocialness

A core concept of IRT is that the relation between persons' trait levels and their probability of endorsing a given item in a trait-consistent manner is expressed by a response curve with certain characteristics. In particular, this probability is a function of a person parameter (i.e., θ, representing the level of the latent trait) and two item parameters, the item difficulty and the item discrimination, respectively. Illustratively, let us consider the item response curves of two hypothetical dichotomous (e.g., Yes/No) items reported in Figure 1. The levels of the latent trait are plotted on the horizontal axis, whereas the probability of endorsing the correct (i.e., trait-consistent) response for each of the two items is reported on the vertical axis. The item difficulty parameter (β), or "location," represents the latent trait level corresponding to a .50 probability of endorsing the item correctly or in a trait-consistent manner. As Figure 1 shows, the trait level associated with a .50 probability is much lower for Item 1 than for Item 2, thus indicating that Item 1 is "easier" than Item 2.

The item discrimination parameter (α), or "slope," represents instead the item's ability to differentiate between people at contiguous levels of the latent trait. This parameter describes how rapidly the probabilities change with trait levels. As Figure 1 shows, Item 1's change is much "slower" than Item 2's change; thus, Item 1 is less discriminating than Item 2, because its item response probability is relatively less responsive to changes in trait levels.* It is important to note that an item's ability to discriminate between people with similar or contiguous trait levels is highest in the θ trait-level region corresponding to the item difficulty; this indicates that items are not equally informative across the entire trait range. That is, some items might do a good job in discriminating people on the higher end of the trait continuum and be scarcely discriminative at lower ranges of the trait continuum. In other words, an item gives more information when the item difficulty parameter completely matches a person's latent trait (Embretson & Reise, 2000).

The item information expresses the relation linking a trait level to the item difficulty and discrimination parameters and can be computed for all trait values. This allows the construction of the item information curve (IIC) that can be plotted to represent the information as a function of trait level. Across items, the item information curves can be summed to produce an information curve for the full scale, that is, the test information curve

* When the latent trait distribution is normal, the meaning of the item discrimination parameter is similar to the meaning of the factor loading or the item-test correlation of CTT (i.e., an item correlating highly with a scale score is a better indicator of the latent trait in CTT; an item that has a high discrimination value is a better indicator of the latent trait in IRT).

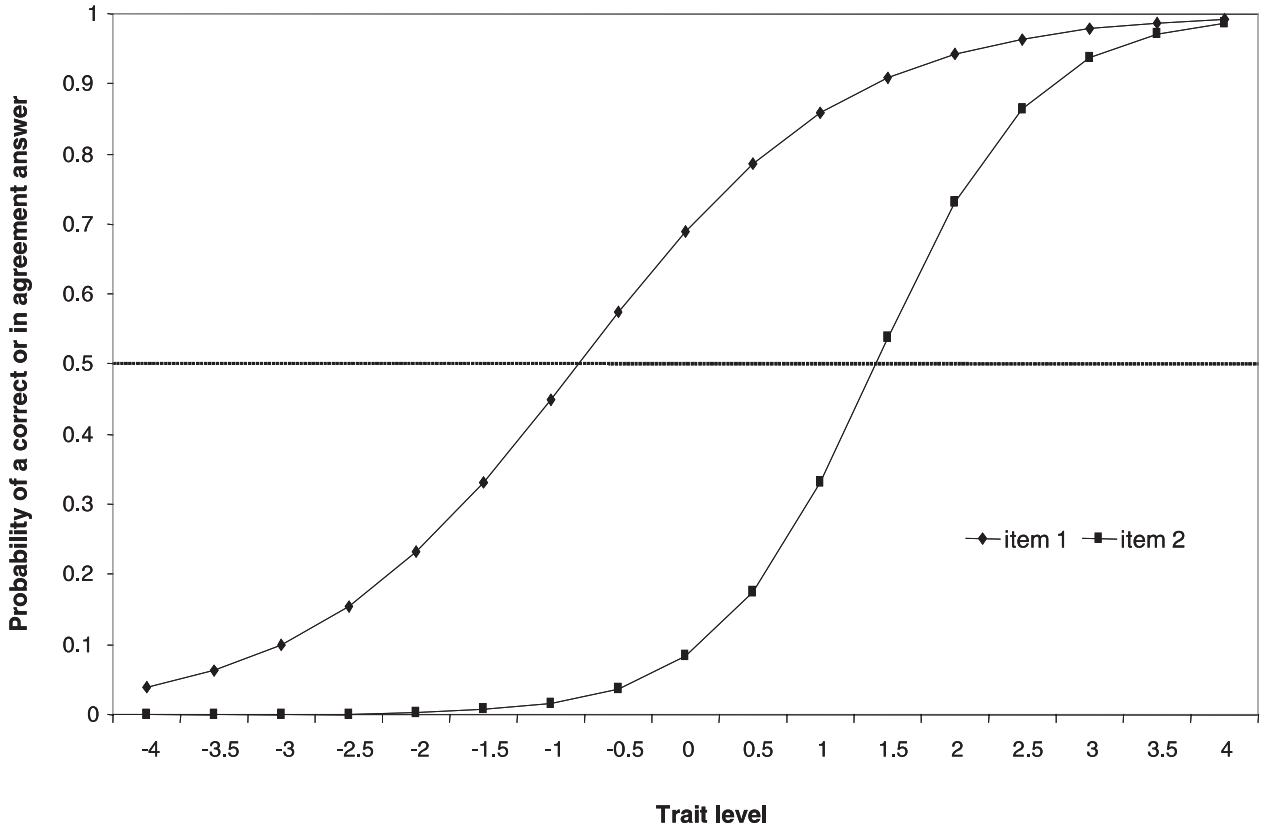

Figure 1. Item characteristic curve of a dichotomous item.

(TIC). The IIC expresses the relative precision of the scale across the different levels of the trait continuum; the inverse square root of the information value at a particular trait level is equal to the standard error of measurement (SEM). Thus, unlike CTT, where measurement precision is generally expressed by a single index (such as Cronbach's α), measurement precision in IRT can differ and be evaluated for people differing in trait levels. Finally, an important characteristic of IRT methods is the possibility of identifying items that are biased for or against a particular group of subjects, through the analysis of the *differential item functioning* (DIF). This DIF index is defined as the different performance on a test item among groups, which is not explainable by the difference across groups in the target trait. Then, a scale item displays DIF if subjects with the same latent trait level have different probabilities of endorsing an item in a particular manner. In IRT analyses, the presence of DIF is evaluated by comparing item parameters estimated for the different groups.

Importantly, measurement instruments and items may have different response scale formats, and IRT methods provide clear guidelines and models for examining the measurement characteristics of items varying in their response scale formats. With polytomous items, the relation between θ and the probability of endorsing an item in a trait-consistent manner is expressed by a series of probability curves, one for each of the response categories (e.g., five curves for a five-point response scale). It is assumed that each response category's probability function provides the probability of endorsing that response for persons at each trait level.

In the present study, the instrument we developed to measure prosocial responding was based on a five-point response scale and, thus, required an IRT method of analysis for polytomous items. For our purposes, we selected the Generalized Partial Credit Model (G-PCM; Muraki, 1992, 1993, 1996), which is a potentially useful item response model that can be adopted when the item responses can be conceptualized as ordered categories (e.g., Likert-type rating scales).

[In the G-PCM a category response curve is estimated for each categorical response that represents the probability of a subject responding in that particular category conditional on trait level. The category response curves can be written as:

$$P_{ix}(\theta) = \frac{\exp \sum_{j=0}^{x} \alpha_i (\theta - \delta_{ij})}{\sum_{r=0}^{M} \left(\exp \sum \alpha_i (\theta - \delta_{ij}) \right)}$$

The δ_{ij} term corresponds to the item-category parameter, and it is often called the item step difficulty associated with a category response of an item; the higher the value of a particular δ_{ij}, the more difficult a particular step is relative to other steps within an item. If the number of response categories is m, only m-1 item-category parameters can be identified. To better understand the model consider the following graphical representation of a five-category item:

1 – – – – – 2 – – – – – 3 – – – – – 4 – – – – – 5
step1 step2 step3 step4

In this item, a subject must complete two steps in order to respond in the middle category and four steps to respond in the highest category. The δ_{ij} parameters represent the relative difficulty of each step. Within an item some steps (or category intersections) may be easier or more difficult than others. The δ_{ij} parameters can be directly interpreted as the points where, on the latent-trait continuum, the category response curves intersect and indicate where the response of one category becomes relatively more likely than the previous category.]

Illustratively, Figure 2 reports the category response curves of a hypothetical item with a five-category response format, as in the case of our instrument measuring adult prosocialness. The figure shows the response curves for each response category. These probability curves can intersect, and the intersection points for contiguous response categories correspond, as in the case of Categories 1 and 2, to the latent trait level at which it becomes more likely to endorse the higher than the lower response category. As Figure 2 shows, this intersection point is located at a low level of the latent trait, indicating the relative easiness of the passage from Category 1 to Category 2. The same reasoning applies to other response categories situated at higher levels of the latent trait (e.g., Categories 4 and 5).

In the present study, item parameters were generated for all 16 prosocial behavior items, and information functions were computed for all items and the full prosocialness scale. Finally, the presence of DIFs between males and females was examined with respect to the difficulty and discrimination parameters. All IRT analyses were conducted using the PARSCALE program (Muraki & Bock, 1993).

Before performing IRT analyses, conventional (i.e., CTT) item and scale statistics were computed. In particular, means, standard deviations, skewness, and kurtosis for prosocial behavior items were first calculated; then, indices of internal consistency (Cronbach α's coefficient and mean corrected item-total correlations) were computed for the entire set of items.

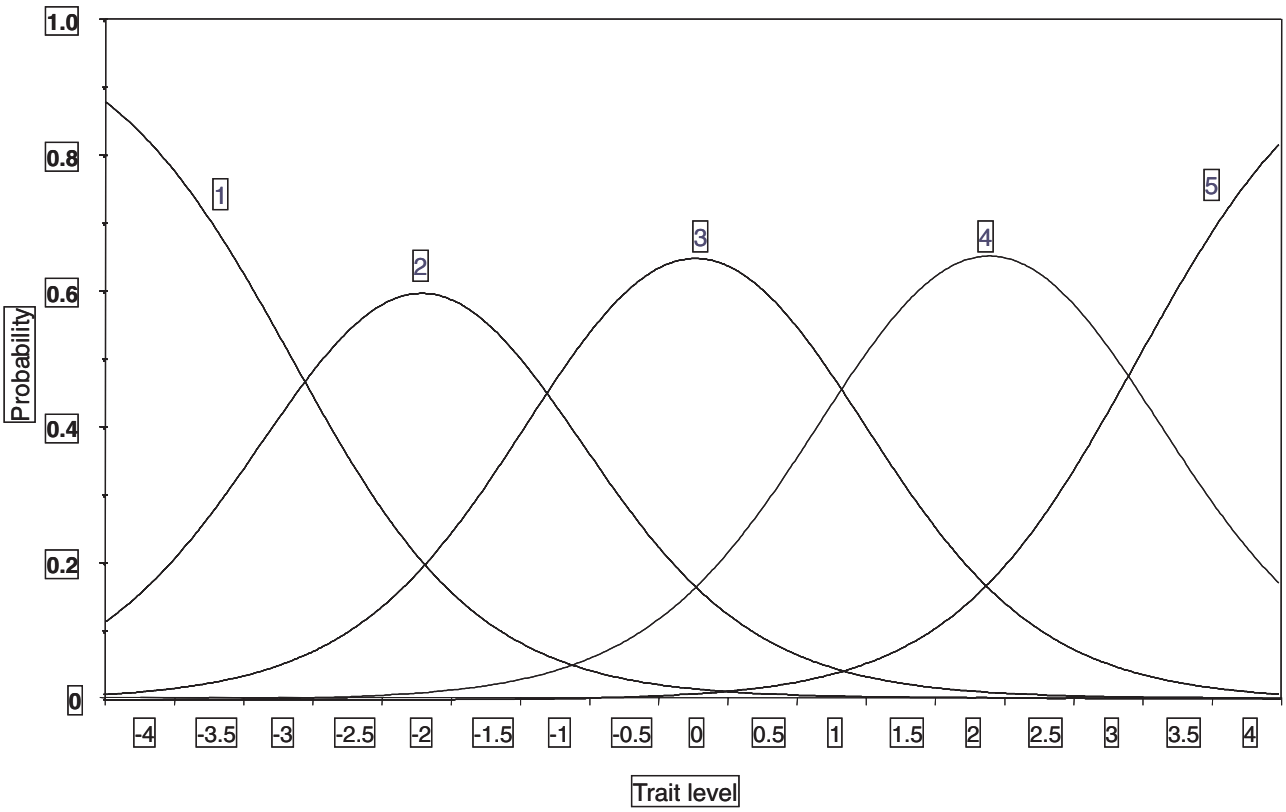

Figure 2. Category response curves of an item witha five-category response format.

Results

CTT Item and Scale Statistics

Table 2 presents a summary of the conventional CTT statistics obtained for the 16 prosocialness items. Across items, the mean responses ranged from 2.96 (Item 11) to 3.79 (Item 9), with an overall mean of 3.52 and standard deviation of .64. Also across items, the average skewness was –.33 and the average kurtosis was –.27. The corrected item-total correlations varied from .47 (Item 2) to .73 (Item 13) and were lower than .50 only in two cases (Items 1 and 2). At the scale level, the Cronbach's α for the entire set of items was .91 and the mean corrected item-total correlation was .59.

This results show the good reliability of the prosocialness scale. However, conventional internal consistency indices provide little information as to the measurement precision of the single items. Furthermore, conventional CTT statistics, which tend to be sample-specific and apply to all subjects, are not particularly sensitive to estimating people's differing latent traits (i.e., traits for people who are situated at different points along the trait continuum) when people's overt responses to items tend to cluster at either of the extreme points of the measurement scale. The IRT approach can, instead, offer a valuable method to examine these measurement issues.

Table 2. Means, standard deviations, skewness, kurtosis, and adjusted item-total correlations for the sixteen prosocialness items ($N = 2574$).

	Mean	*SD*	Skewness	Kurtosis	Item-total correlation
Item 1	3.67	.96	–.50	.08	.48
Item 2	3.70	.95	–.52	.01	.47
Item 3	3.70	.92	–.28	–.46	.70
Item 4	2.97	1.25	.07	–.94	.56
Item 5	3.69	.98	–.45	–.28	.65
Item 6	3.72	.91	–.28	–.38	.60
Item 7	3.28	1.01	–.19	–.37	.65
Item 8	3.47	1.02	–.26	–.49	.63
Item 9	3.79	.93	–.61	.19	.59
Item 10	3.74	.92	–.48	–.05	.67
Item 11	2.96	1.14	–.02	–.71	.52
Item 12	3.65	.96	–.39	–.36	.64
Item 13	3.35	.96	–.19	–.24	.73
Item 14	3.44	.92	–.35	.02	.53
Item 15	3.43	.97	–.32	–.25	.56
Item 16	3.71	.93	–.48	–.03	.52

IRT Estimation of Item Parameters According to the G-PCM

A basic assumption underlying most IRT models is that item covariation arises predominantly from a single dimension (i.e., the unidimensionality assumption). There are different procedures to assess dimensionality; one of the most common and traditional ways is to compare the percentages of variance explained by the first and the second unrotated components in a principal component analysis. In our case, the ratio was about 5:1, attesting to the unidimensionality of the prosocialness scale.

Item Parameters

Next we performed IRT analyses. The G-PCM estimates for the 16 items are shown in Table 3. In particular, the table shows three types of parameters, the items' slope, category, and location parameters. In general, the IRT parameters shown in Table 3 are consistent with the CTT statistics reported in Table 2. As to the items' discrimination, the results showed that 9 out of 16 items were able to detect even slight differences in adults' prosocialness. In particular, Items 3, 10, and 13 yielded the most information and had the highest slope parameters (as well as the highest item-total correlations, see Table 2). Thus, the items referring to people's efforts to help others, console those who are sad, and be close to and take care of those who are in need, are the most effective in detecting slight variations in the trait of prosocialness. The discrimination parameters were also quite acceptable for Items 5, 6, 7, 8, 9, and 12, all of which had a slope parameter greater than 1.

Table 3. IRT parameters estimated under the generalized-partial credit model.

Item	Slope	Category parameters				Location
Item 1	.67	–2.42	–.66	1.88	1.20	–1.06
Item 2	.62	–2.77	–.50	1.43	1.84	–1.22
Item 3	1.90	–1.90	–.69	.66	1.94	–.99
Item 4	.78	–1.11	–.79	.61	1.29	.01
Item 5	1.48	–1.78	–.50	.67	1.61	–.87
Item 6	1.23	–2.11	–.81	.91	2.00	–1.11
Item 7	1.37	–1.91	–.61	.82	1.70	–.33
Item 8	1.25	–1.85	–.62	.68	1.79	–.64
Item 9	1.10	–2.10	–.36	1.00	1.46	–1.07
Item 10	1.68	–1.91	–.47	.77	1.61	–.93
Item 11	.68	–1.95	–.74	1.02	1.67	.12
Item 12	1.34	–2.00	–.50	.57	1.94	–.89
Item 13	2.10	–1.85	–.57	.75	1.67	–.41
Item 14	.81	–2.80	–.64	1.46	1.97	–.64
Item 15	.88	–2.40	–.59	1.02	2.01	–.63
Item 16	.84	–2.40	–.55	1.18	1.78	–1.10

Three of these items (5, 8, and 12) were items referring to adults' emphatic concerns toward others. The parameters for the remaining items (1, 2, 4, 11, 14, 15, and 16) were smaller than 1. As to the items' location parameters (i.e., difficulty level), most of the items (14 out of 16) had a negative value, and the "easiest" items were concerned with sharing (e.g., "I share things that I have with my friends"), helping (e.g., "I am pleased to help my classmates/colleagues in their activities"), and sensibility to others' feelings (e.g., "I immediately sense my friends' discomfort even when it is not directly communicated to me").

Finally, as to the items' category intersections, the parameters were generally spread along the trait continuum but were quite low for all items, especially for Items 2, 6, and 16.

Information Functions for the Items and Full Scale

The G-PCM estimates for the item parameters were then used to produce the information functions for the 16 prosocial behavior items and the overall scale. These functions provide information concerning the measurement precision of the items and the scale and can reveal the point along the trait continuum where measurement precision is highest. Typically, an item's measurement precision – or its ability to detect slight differences in people's trait levels (i.e., its discrimination or slope parameter) – is highest in correspondence to the item's difficulty level. The information functions are generally represented as curves on graphs that have the estimated trait level (θ) on the x-axis and the amount of information provided on the y-axis. The larger the items' discrimination, the more peaked the information function curve will be. In contrast, an item's difficulty determines the position of the curve along the latent trait continuum. In general, easy items provide information in low trait ranges and difficult items provide information in high trait ranges. An inspection of the item information functions revealed that the prosocialness items were, overall, quite informative. There were, however, differences in the degree to which items could detect slight differences in people's prosocialness and discriminate along the entire trait continuum. These results are well illustrated in Figures 3 through 5. As one can see in Figure 3, Items 3, 5, 7, 8, 10, 12, and 13 could discriminate slight differences in prosocialness very well, especially among people with either low or moderate levels of prosocialness. Interestingly, two of these items (3 and 13) also displayed the highest item-total correlations (see Table 2). Overall, these results indicated that items measuring either individuals' empathy (5, 8, and 12) or people's efforts or attempts to help and take care of others (3, 10, and 13) had the highest measurement precision, especially at relatively moderate levels of prosocialness.

Likewise, as Figure 4 shows, the level of information

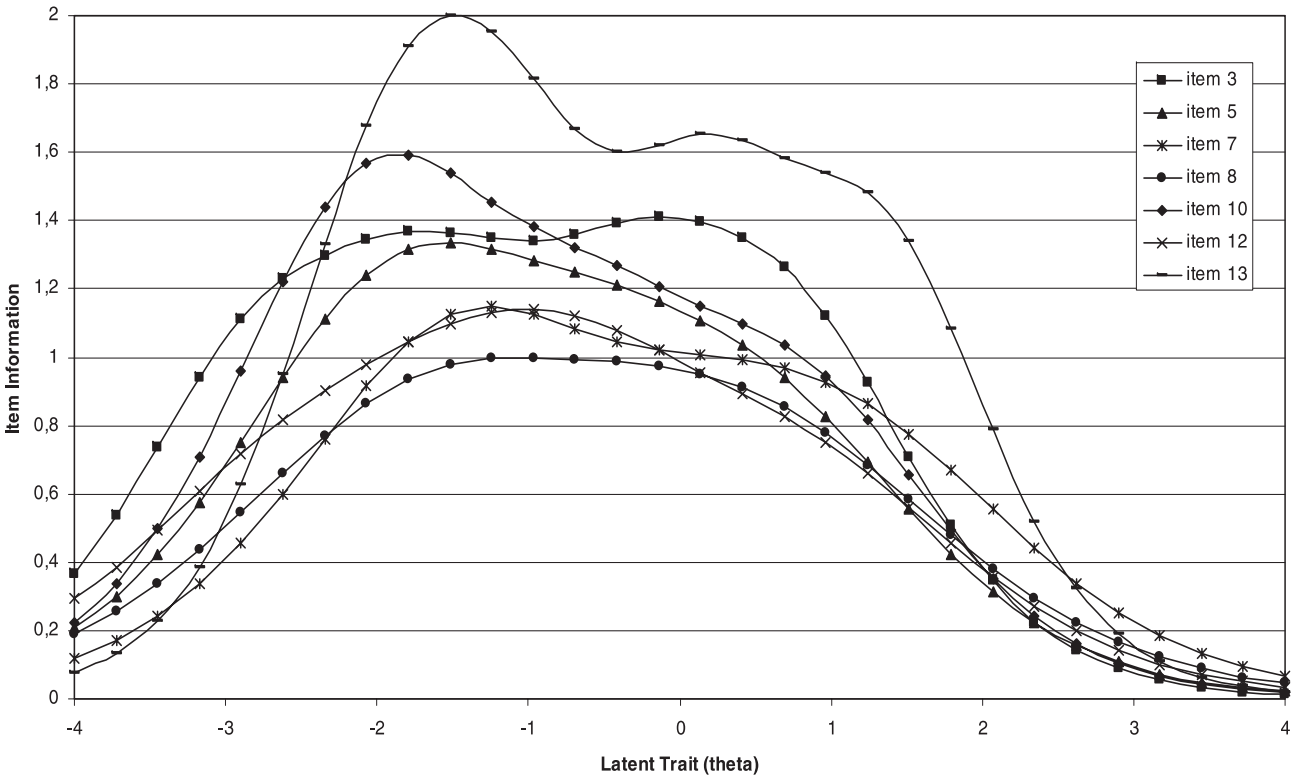

Figure 3. Item information curves for items 3, 5, 7, 8, 10, 12, and 13.

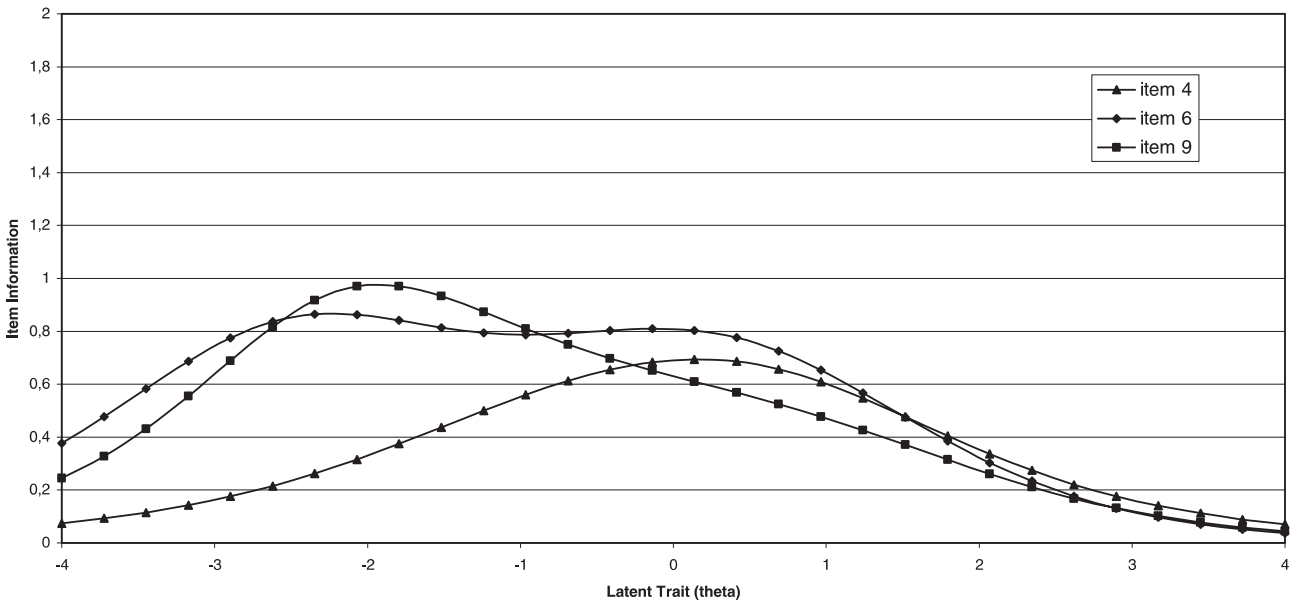

Figure 4. Item information curves for items 4, 6, and 9.

provided by other items (4, 6, and 9) was also quite acceptable and, again, especially at relatively low (6 and 9) or moderate (4) levels of prosocialness. An inspection of these items' content revealed that they measured prosocialness with respect to people's willingness to help and assist those in need or to make themselves available when needed. Finally, as one would have expected from the low item-total correlations and discrimination parameters, Items 1, 2, 11, 14, 15, and 16 provided relatively little information, as Figure 5 shows.

The information and standard error curves for the entire prosocial behavior scale are presented in Figure 6. The mean information value for the entire scale is 8.54, and greater information is concentrated at the low and middle levels of the prosocialness trait continuum. Accordingly, the mean standard error is .43, and standard error is greatest at relatively high trait levels. These results suggest that our prosocialness scale discriminated very well among relatively moderately prosocial individuals and discriminated less well among highly prosocial individuals.

Differential Item Ffunctioning

The last IRT analysis was performed to examine whether the items' functioning varied across male and female adults, that is, whether a differential item functioning existed across gender groups. In particular, we analyzed the *nonuniform* DIF, that is, a DIF concerning the items' location and slope parameters.* A level of $p < .001$ was used as the criterion to evaluate the standardized DIF statistic (SDIF), and the –2 log likelihood statistic was used to evaluate estimation improvement from a no-group-difference model (i.e., one-group model) to a model that hypothesized gender differences in item parameters (i.e., the two-groups model). This statistic is defined as the difference between the –2 log likelihood values (G^2) of the two models (one-group model and two-group models). Under certain conditions, this differ-

Table 4. Item location parameters for males and females and differential item functioning test.

Item	Location parameters and (S.E.)		SDIF	Prob.
	Males	Females		
Item 1	–.917 (.043)	–.679 (.045)	3.779	**.000**
Item 2	–.996 (.048)	–.901 (.051)	1.356	.087
Item 3	–.776 (.022)	–.716 (.023)	1.873	.031
Item 4	.089 (.030)	.199 (.034)	2.413	.008
Item 5	–.571 (.024)	–.789 (.027)	–6.018	**.000**
Item 6	–.929 (.029)	–.789 (.029)	3.420	**.000**
Item 7	–.261 (.025)	–.062 (.024)	5.719	**.000**
Item 8	–.263 (.026)	–.672 (.029)	–10.467	**.000**
Item 9	–.966 (.033)	–.649 (.029)	7.133	**.000**
Item 10	–.621 (.024)	–.852 (.024)	–6.724	**.000**
Item 11	.296 (.038)	.183 (.040)	–2.047	.020
Item 12	–.574 (.026)	–.813 (.028)	–6.254	**.000**
Item 13	–.213 (.020)	–.261 (.020)	–1.702	.044
Item 14	–.524 (.039)	–.341 (.039)	3.329	**.000**
Item 15	–.446 (.036)	–.421 (.035)	.498	.309
Item 16	–.824 (.038)	–.932 (.042)	–.1028	.029

* The category parameters are common to both the examined groups.

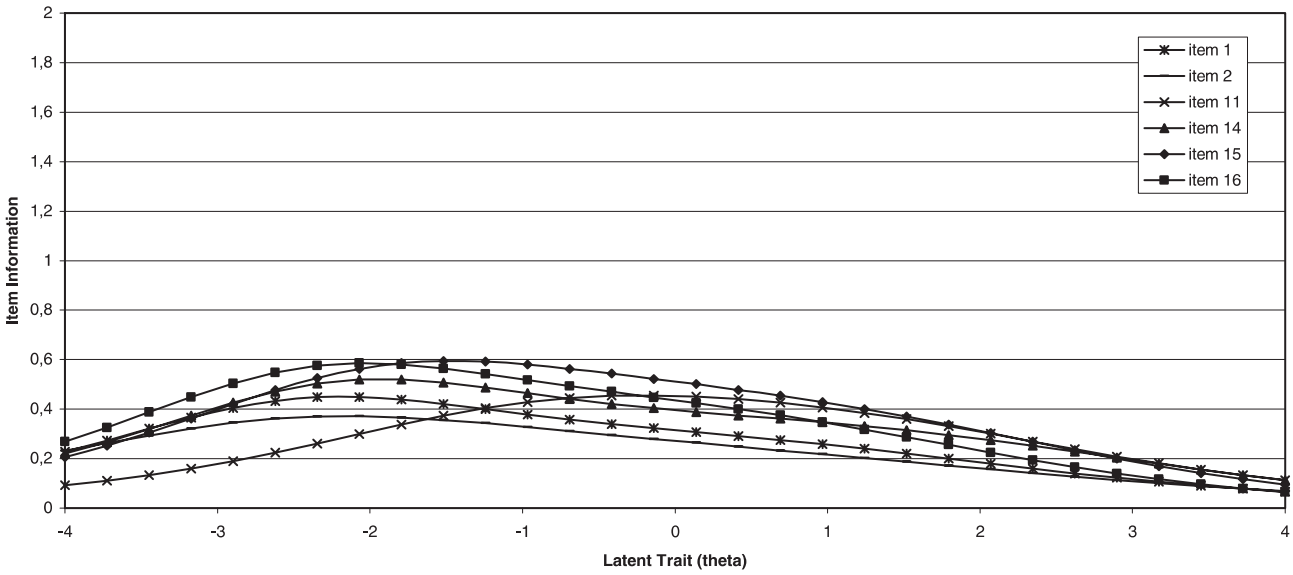

Figure 5. Item information curves for items 1,2,11,14,15, and 16.

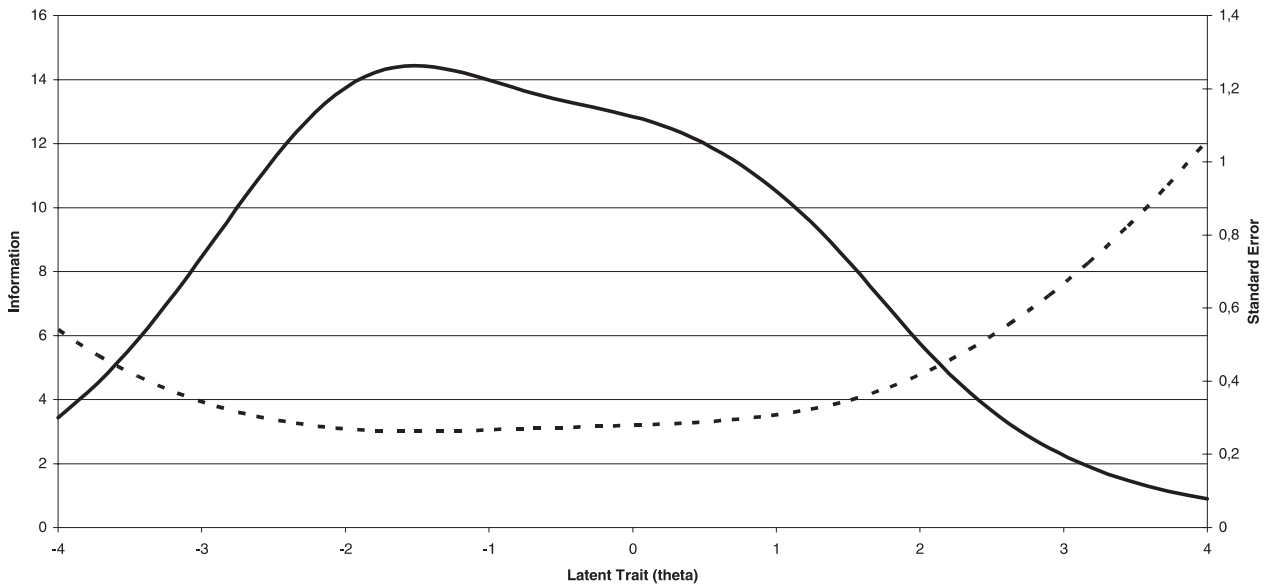

Figure 6. Test information and standard error curves for the prosocial behavior scale.

ence is distributed as chi-square with degrees of freedom equal to the difference in degrees of freedom between the two, the more and the less constrained, models. Large and significant chi-square values would indicate that considering men and women as a unique group, that is, imposing constraints in their item locations and slopes, would worsen the model significantly. The test was performed with 32 degrees of freedom, corresponding to the 32 parameters that are freed in the two-groups model.

As reported in Table 4, the location parameters of Items 1, 6, 7, 9, and 14 resulted significantly higher in women, whereas the location parameters of Items 5, 8, 10 and 12 resulted significantly higher in men. These results were quite in line with the existing literature on gender differences in prosocial behavior (Eagly & Crowley, 1986; Eisenberg & Fabes, 1998). In particular, female adults endorsed items referring to empathy and emotional support (5, 8,10, and 12) more easily (i.e., in larger numbers) than did male adults. In contrast, male adults endorsed items referring to immediate assistance or action (1, 6, 7, and 9) more easily than did female adults.

Regarding the item slope parameters, no difference was found between males and females; this result indicates that the 16 items of the prosocialness scale have an equal capacity to discriminate among male and female

adults with different levels of prosocialness. A significant difference between the two G^2 (295.42; $p < .001$) was found, which indicated that there was a substantial improvement in the model when the item parameters were freely estimated for males and females (i.e., under the hypothesis of gender differences in parameter estimation) relative to when they were constrained to be equal in the two gender groups.

Discussion

In this article, we presented a new scale for evaluating adults' prosocialness and used IRT analyses to evaluate its measurement qualities.

Our IRT measurement scrutiny of the new instrument was justified on both theoretical and methodological grounds. Theoretically, the development of an individual's characteristic tendency to act prosocially arises from a host of socializing experiences and might, during the life course, be associated with very idiosyncratic behaviors. These considerations suggest that the behavioral indicators of prosocialness do not necessarily have the same meaning across individuals, nor that any prosocial behavior is an adequate and reliable indicator of prosocialness along its entire trait range. Methodologically, the assessment of prosocial behavior typically relies on dimensional models that tend to follow CTT. In previous work (Caprara et al., in press), this 16-item prosocialness scale had already demonstrated very adequate psychometric qualities including that of tapping into a single factor or trait dimension of prosocialness, a necessary prerequisite for employing most IRT models. However, these initial psychometric findings followed from methods of CTT and could not address additional important measurement issues, including whether different items were equally informative of adults' underlying trait of prosocialness, and whether uneven information across prosocialness items significantly affects the meaning of an aggregated prosocialness scale. The IRT analyses undertaken in our present investigation primarily addressed these issues.

As to the issue of the prosocial behavior items' informative value, the present findings demonstrated that the majority of the items were moderately discriminative and appropriate to differentiate adults with a middle level of prosocialness. A content inspection of the items showed that the most discriminative items are those that referred to helping and caring behaviors. The negative values of the difficulty parameters suggests that the items were relatively easy to endorse, especially those referring to prosocialness toward friends.

These relations alsoyielded the additional finding that most of the prosocialness items of our scale carried uneven information along the trait range (i.e., the fact that an item was not equally informative about people's prosocialness along the entire trait continuum). Items measuring either empathy or adults' attempt to take care of others were the most informative about people's prosocialness. The information curves of these items were quite peaked, with the highest degree of information associated with Item 13. The information value of these items decreased, but remained quite acceptable, at relatively high levels of the trait continuum. The fidelity in measuring adults' prosocialness was also quite adequate for those items tapping adults' willingness to help and to assist those who are in need. The remaining items, primarily concerning people's willingness to share with others, were less informative about people's underlying prosocialness.

Although this finding requires further inquiry, it is quite likely due to the level of high endorsement that people expressed on these components of prosocial responding (i.e., high mean scores and relatively low standard deviations) and to their highly normative value.

As to the measurement value of the entire prosocialness scale, the implications of the findings above were also clear: As it is, our prosocialness scale is quite effective in measuring systematic individual differences in prosocialness, and this effectiveness holds particularly well for intermediate and low levels of the trait. Thus, the new instrument provides a reliable measure of adults' prosocialness, although its level of measurement fidelity tends to diminish somewhat for people who are either extremely high or low in prosocialness.

Finally, we used IRT analyses to examine potential gender differences in the items' functioning, that is, in the item location and slope parameters. For each item, we basically compared the difficulty and discrimination values obtained in each gender group and sought a statistically significant difference between the two values. Regarding the discrimination parameter results provided additional support to the capacity of all the items of the new instrument to discriminate well among both male and female adults with different levels of prosocialness.

Significant differences were found regarding the difficulty of some items.In this respect, the pattern of findings was particularly clear: Four items concerning empathy and emotional support were more "easily" endorsed by female than by male adults, whereas five items concerning immediate help and assistance for those in need and sharing knowledge and opportunities were more "easily" endorsed by male than by female adults.

While the instrument works equally well for the two genders in discriminating high from low levels of prosocialness, it reveals, for some items, a different functioning in males and females regarding the probability of

endorsing the different category responses. A number of reasons may underlie and explain these differences. The compliance to traditional roles that assign to women tasks of supporting and caring for others and to men tasks of helping and sharing may influence the accessibility of some items' content inducing respondents to easily endorse items referring to specific facets of prosocialness. The same psychosocial process may give rise to self-presentation attitudes that manifest themselves as characteristic styles in endorsing item response categories.

In sum, the present study contributed to the study of adult prosocialness in several respects. Firstly and most importantly, it provided the opportunity to examine the properties of a much needed instrument for measuring adults' prosocialness. The new instrument was conceived to respond to the recommendations of the literature concerning the diverse and complex behavioral dimensions that may characterize an adult's prosocial responding. In particular, the four fundamental aspects of prosocialness, namely, behaviors of helping, sharing, taking care of, and feeling empathic with others were the generating criteria for the proposed instrument. Secondly, it provided strong and novel evidence for the measurement quality of the new instrument. Previous work (Caprara et al., in press) had already demonstrated that the instrument was reliable in measuring individual differences in adults' prosocial responding, that it tapped into a single latent dimension of prosocialness, and that it had high construct validity. Through the novel use of IRT analyses, the present study contributed to this wealth of evidence by showing that most of the instrument's behavioral indicators could measure differences in adults' prosocialness along a very large portion of the trait continuum, and that most of these indicators were also particularly sensitive in discriminating among people with slight differences in latent prosocialness.

Importantly, this evidence was gathered through the study and analysis of data collected on a very large sample of Italian adults, a fact that strengthens the robustness of the findings. Thirdly, the instrument that was scrutinized in the present study may represent an important measurement device with widespread applicability for articulating and furthering the study of adults' prosocial responding, especially when one considers its relatively easy use, brevity, and comprehensibility.

References

Bandura, A., Barbaranelli, C., Caprara, G.V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71*, 364–374.

Bandura, A., Caprara, G.V., Barbaranelli, C., Pastorelli, C., & Regalia, C. (2001). Sociocognitive self-regulatory mechanisms governing transgressive behavior. *Journal of Personality and Social Psychology, 8,* 125–135.

Bandura, A., Pastorelli, C., Barbaranelli, C., & Caprara, G.V. (1999). Self-efficacy pathways to childhood depression. *Journal of Personality and Social Psychology, 2*, 258–269.

Bar-Tal, D. (1982). Sequential development of helping behavior: A cognitive-learning approach. *Developmental Review, 2,* 101–124.

Batson, D. (1991). *The altruism question: Toward a social-psychological answer*. Hillsdale, NJ: Erlbaum.

Belansky, E.S., & Boggiano, A.K. (1994). Predicting helping behaviors: The role of gender and instrumental/expressive self schemata. *Sex Roles, 30*, 647–661.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*, 676–713.

Caprara, G.V., Barbaranelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11*, 302–306.

Caprara, G.V., Capanna, C., Steca, P., & Paciello, M. (in press). Misura e determinanti personali della prosocialità. Un approccio sociale cognitivo [Measurement and personal determinants of prosocialness. A social-cognitive approach]. *Giornale Italiano di Psicologia.*

Caprara, G.V., & Pastorelli, C. (1993). Early emotional instability, prosocial behavior, and aggression: Some methodological aspects. *European Journal of Personality, 7*, 19–36.

Eagly, A.H., & Crowley, M. (1986). Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychological Bulletin, 100,* 283–308.

Eisenberg, N., Carlo G., Murphy B., & Van Court P. (1995). Prosocial development in late adolescence: A longitudinal study. *Child Development, 66,* 1179–1197.

##@LIT = Eisenberg, N., & Fabes, R.A. (1998). Prosocial development. In N. Eisenberg (volume editor), Social, emotional, and personality. In W. Damon (Editor-in-Chief), *Handbook of child psychology,* (Vol. 3, pp. 701–778). New York: Wiley.

Eisenberg, N., Guthrie, I.K., Cumberland, A., Murphy, B.C., Shepard, S.A., Zhou, Q., & Carlo, G. (2002). Prosocial development in early adulthood: A longitudinal study. *Journal of Personality and Social Psychology, 82,* 993–1006.

Eisenberg, N., Guthrie, I.K., Murphy, B.C., Shepard, S.A., Cumberland, A., & Carlo, G. (1999). Consistency and development of prosocial dispositions: A longitudinal study. *Child Development, 70*, 1360–1372.

Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Hambleton, R.K., Robin F., & Xing, D. (2000). Item response models for the analysis of educational and psychological test data. In H.E.A. Tinsley & S.D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 553–585). San Diego, CA: Academic Press.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: Applications to psychological measurement*. Homewood, IL: Dow Jones-Irwin.

Hambleton, R.K., Swaminathan, H., & Rogers H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hulin, C.L., Drasgow, F., & Parsons, C.K. (1983). *Item response*

theory: Applications to psychological measurement. Homewood, IL: Dow Jones-Irwin.

Koestner, R., Franz, C., & Weinberger, J. (1990). The family origins of empathic concern: A 26-year longitudinal study. *Journal of Personality and Social Psychology, 58,* 709–717.

Krebs, D.L., & Van Hesteren, F. (1994). The development of altruism: Toward an integrative model. *Developmental Review, 14,* 103–158.

Lord, F.M. (1984). Standard errors of measurement at different ability levels. *Journal of Educational Measurement, 21*, 239–243.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16,* 159–176.

Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement, 17*, 351–363

Muraki, E. (1996). A generalized partial credit model. In W.J. Van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 153–164). New York: Springer.

Muraki, E., & Bock, R.D. (1993). *PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks*. Chicago: Scientific Software Int.

Schroeder, D.A., Penner, L.A., Dovidio, J.F., & Pilliavin, J.A. (1995). *The psychology of helping and altruism: Problems and puzzle*. New York: McGraw-Hill.

Steinberg, L., & Thissen, D. (1996). Uses of item response theory and the testlet concept in the measurement of psychopathology. *Psychological Methods, 1*, 81–97.

Van der Linden, W.J., & Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer.

Waller, N.G., & Reise, S.P. (1989). Computerized adaptive personality assessment: An illustration with the absorption scale. *Journal of Personality and Social Psychology, 57*, 1051–1058.

Address for correspondence

Gian Vittorio Caprara
University of Rome, "La Sapienza"
Faculty of Psychology
Department of Psychology,
Via dei Marsi 78
I-00185 Rome
Italy
Tel.?? Fax??
E-mail Gianvittorio.Caprara@uniroma1.it