Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v7n1p58

Assessing the relational competence as the core dimension of social worker's perceived self-efficacy through the Rasch model

By Golia and Pedrazza

April 26, 2014

Electronic Journal of Applied Statistical Analysis
Vol. 7, Issue 1, 2014, 58-80
DOI: 10.1285/i20705948v7n1p58

Assessing the relational competence as the core dimension of social worker's perceived self-efficacy through the Rasch model

Silvia Golia*[a] and Monica Pedrazza[b]

[a] *Department of Economics and Management, University of Brescia, Brescia, Italy*
[b] *Department of Philosophy Education and Psychology, University of Verona, Verona, Italy*

April 26, 2014

Perceived self-efficacy is concerned with peoples belief in their ability to succeed in specific situations, and can play a major role in how one approaches goals, tasks, and challenges. There is no all-purpose measure of perceived self-efficacy so it is necessary to derive specific scales for specific contexts. This paper presents the structure of a self-efficacy scale developed for social workers. Using the Rating Scale Model as a measurement model, an objective measurement of the social workers perceived self-efficacy is produced and analyzed. Differential Item Functioning (DIF) analysis was performed in order to determine whether the test questions were fair with respect to some population characteristics.

keywords: Self-efficacy, social work, Rating Scale Model, DIF analysis, social competence, relational competence

1 Introduction and theoretical framework

From the seventies to the nineties the construct of self-efficacy has received increasing attention in the social sciences. Perceived self-efficacy is concerned with peoples belief in their ability to succeed in specific situations (Bandura, 1997, 1982, 1986, 1995): Perceived self-efficacy is defined as peoples judgments of their capabilities to organize and execute courses of action required to attain designated types of performance (Bandura,

*Corresponding author: silvia.golia@unibs.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1986). It can play a major role in how one approaches goals, tasks, and challenges. According to Banduras theory, subjects with high perceived self-efficacy are more likely to view difficult tasks as something to be mastered rather than something to be avoided. Perceived self-efficacy is a belief about what a person can do rather than an evaluation of his/her physical and personality attributes (Bong and Skaalvik, 2003; Zimmerman and Cleary, 2006). This means that self-efficacy is a context-specific rather than a general trait (Bandura, 2012), that is, it does not manifest uniformly across activity domains and situational conditions , but it varies across tasks and contexts. Specifically, Bandura (1997) identifies four sources of self-efficacy beliefs; first, people develop beliefs in their efficacy through mastery experiences; second, efficacy expectations can be formed through social modeling; third, self-efficacy can be the product of social persuasion, that is, individuals who are persuaded to believe in their abilities will persist in the face of obstacles; fourth, self-efficacy judgments rely in part on individual physical and emotional states, that is, high levels of physiological arousal are associated with low efficacy expectations.

In addition according to Banduras cognitive theory (Bandura, 1986, 1997) self efficacy beliefs vary on three dimensions: level of magnitude, strength and generality across tasks and situations. Researchers were interested in what Chen et al. (2001) identifies as a state-like self efficacy construct (SSE). More recently researches have become interested in a trait-like construct developing general self efficacy scales (GSE). According to Eden (1988), there is a positive relationship between GSE and SSE across tasks and situations but, according to Eden and Granat-Flomin (2000), SSE predicted specific domain performance and GSE did not.

Self-efficacy beliefs play a major role in setting goals, they also influence the way people motivate themselves and persevere in the face of difficulties, and peoples resilience to adversity (Pajares, 1997). Perceived self-efficacy shapes peoples emotions: individuals with low self-efficacy feel anxiety and fear, and these emotions in turn influence their cognitive processes (Bandura, 1982).

Stajkovic and Luthans (1998) examined, in a meta-analysis of 114 studies, the relationship between self-efficacy and work-related performance: they demonstrated that self-efficacy is a predictor of performance at different levels of task complexity.

Zimmerman (1995) gave empirical support to the evidence that self-efficacy beliefs exert a strong influence on motivation at work. Self-efficacy, as defined by Bandura (1986), became more and more a very different construct from a generic self-perception of competence. Bandura described the concept as an individuals assessment of their own confidence in their own ability to execute specific and contextually identified skills, in a particular set of circumstances, and thereby to achieve a successful outcome.

1.1 Self efficacy in social work

The Betz and Hacketts self-efficacy measurement package (Betz and Hackett, 1981) assessed educational requirement and job duties regarding social work. That instrument is, however, in our opinion, inadequate for in-depth assessment of self-efficacy about specific and different professional tasks in social work, because it focuses on investigating

prerequisites and rules rather than the perception of self-efficacy at work.

To date, the literature on self-efficacy in social work refers also to two other different scales developed by Holden and colleagues (Holden et al., 1997, 2002). They developed the Hospital Social Work Self-Efficacy Scale (HSWSE), and an extension of it (Holden et al., 1997), to assess social workers confidence in their ability to perform hospital social work. This is a very specific scale and cannot be used to assess social workers self-efficacy in a broad sense, that is relating to competences and actions that the social workers engage in social services outside hospitals. Later, Holden et al. (2002) proposed the Social Work Self-Efficacy Scale (SWSE) in order to assess outcomes in social work education, regarding a broad range of social work tasks. Data collected by Holden et al. (2002) showed a positive change in the perceived self-efficacy of Master of Social Work students, from program entry to the point of graduation. This scale assessed professional practices in general without reference to a particular context or to specific organizational contexts, in contrast to the previous HSWSE, which focused on a particular health-practice setting. This scale is useful for assessing social work students acquired competences, but it is not useful for assessing the perception subjects have relating to their experience as workers employed in a specific social service, so it fails to assess specific and contextualized work competences/abilities. Moreover, Schwarzer and Jerusalem (1995) proposed the so-called Generalized self-efficacy scale. They developed this scale focusing on statements of efficacy regarding general competences and abilities. Holdens and Schwarzer and Jerusalems scales were developed through a top-down process, starting from literature and general knowledge derived from Practice Skills Inventory (OHare and Collins, 1997) on social workers activities and tasks, thereafter they are inadequate for refering to specific contexts and environmental variables. Bandura (2012) notes: strength of self-efficacy is measured across a wide range of performance within an activity domain. In fact a social workers sense of self-efficacy may differ across the different tasks they are required to perform and also across different social services domains and work contexts. Moreover, Bandura (2006) argues for the need to assess capabilities, and not intentions, current ability, or outcome expectancies in the measurement of self-efficacy. I can is a statement of efficacy, I will is a statement of intention. According to Bandura, statements of intention should not be included in a self-efficacy scale.

The above-mentioned scales do not lack in validity or reliability but they lack, in our opinion, in content validity because they refer to the researchers and otherwise given knowledge of a social workers activities. As a result, the developmental process of those scales does not start from the social workers perception of self-efficacy relating to their specific and contextual experience regarding tasks, competences, and activities in an organizational context. As Chen et al. (2001) argue, a GSE scale cannot be a replacement for the SSE. In fact, GSE could be a supplement when the performance under scrutiny can be generalized. For all these reasons a new scale to measure perceived self-efficacy has been proposed by Pedrazza et al. (2013). This scale attempts to assess the necessary competences at work and is based on the identification of the competences and abilities through which a social worker (from now on SW) defines her/his self-efficacy.

The aim of this paper is to present the structure of the self-efficacy scale developed for the SWs and produce an objective measurement of the SWs perceived self-efficacy.

We are interested in a self-efficacy scale matching the specificity of self perceived efficacy to the specificity of performance in the basic public social service. According to Chen et al. (2001), the better the match the greater the predictability. The model used in this study to obtain linear and reproducible measures of SWs perceived self-efficacy, belongs to the family of Rasch models and it is the Rating Scale Model (Andrich, 1978). The context of interest is the so-called basic service, devoted to the familys economic, occupational and health issues. All the items forming the tested questionnaire are related to socio-relational competences classified in two different relational domains: in the first one, the co-workers act and engage in specific activities which can, in complex multi-problematic cases, also involve health and/or law practitioners; the second one is the SW-user relational domain. Up-to-date literature on issues related to relational competences (Mikulincer and Shaver, 2007; Pedrazza and Boccato, 2009, 2012) allows one to look at those two different domains as two dimensions which could lie on the same continuum from the most involving (between SW and user) to a more formal relationship (under co-workers and practitioners belonging to different social and health services).

2 Development of the self-efficacy scale for social workers

Self-efficacy beliefs do not reflect a generic sense of competence or self-esteem, but are instead context-specific. Therefore, self-efficacy should be assessed by using an ad hoc scale, measuring the perception of efficacy of employed SWs referring to their individual behaviors which allows them to exert influence over events that affect their work life. Accordingly, we developed a self-efficacy scale specific for SWs (Pedrazza et al., 2013) through a bottomup qualitative method, that is, we identified a specific SWs institutional context.

In Italy, SWs can be employed in the basic social service or in other specific services (e.g., child welfare, mental health etc.). The basic social service offers help and support to individuals, families, groups, and every social gathering in a specific geographical area (e.g., municipality). The basic social service covers the whole national territory whereas other services (e.g. child welfare) are often combined and each of them may cover large geographic areas referring to different municipalities.

In order to contact the largest number of SWs, it was decided to investigate self-efficacy in basic SWs. In this context the Critical Incident Technique (CIT) (Flanagan, 1954; Mayhew, 1956) was applied. The qualitative bottom up analysis of different critical incidents at work is closer to the original construct of self-efficacy. By an incident is meant any event or situation that is recent, sufficiently complete and representative of ones work. An incident is critical when it makes a significant contribution to work and challenges ones abilities. This technique has been successfully applied to different domains (Butterfield et al., 2005, 2009; Flanagan, 1954; Gremler, 2004; Mayhew, 1956). CIT was applied to two focus groups, each involving eight SWs, selected on the basis of the following criteria: a. they operated in the province of Verona (Italy); b. had at least 10 years of service; c. were highly motivated to give a personal contribution to activities related to permanent education; d. had given written informed consent

to participate in the study. Each session was divided into two stages, both of which adopted semi-structured interviews. At stage 1, participants were asked individually to recall the critical incidents that had been occurring most frequently in their work. They were invited to provide details about the events, problems, and context of these incidents, as well as the behaviors they engaged in to face the critical situation. At stage 2, each participant shared the critical incidents recalled and, through open discussion, participants identified the behaviors they found most effective to manage the incidents. The transcripts of the two focus groups were submitted to a content analysis (with Nvivo 8), which provided 13 key beliefs of self-efficacy, which were used to develop a corresponding number of items (see Table 1).

The SWs who participated in the focus groups, evaluated the correspondence between the 13 items and CIT results, as well as the level of item difficulty. All the items were judged to be consistent with CIT results and easy to understand. It was remarkable that SWs agreed to indicate as critical events and incidents in their daily work different types of dyadic and/or social interaction, related to two different relational domains involving, in the first case, co-workers, in the second one, users. The current worldwide recession has resulted in an increase in requests for social services. The huge number of requests, the often hopeless access of users to the service and the restrictions on frequent turnover applied by public institutions, could have been in part responsible for the number of perceived difficulties in managing relationships intra-service e.g. with users, and inter-services e.g. in interdisciplinary teams. These particular socio-economical circumstances could have brought SWs involved in focus groups to identify as critical events above all relational items and psycho-dynamic related issues, such as holding, acceptance of the user, communication with users and/or coworkers, communication with superiors , communication within interdisciplinary teams, teamwork. The particular type of social service and the particular current social situation render relational issues particularly relevant.

3 Statistical Methods

In this section we revise some theory regarding the family of Rasch models (RM) and Differential Item Functioning (DIF) analysis. The first belongs to the broader class of Item Response Theory models, which aim to measure latent traits. The DIF analysis is a procedure used to determine if test questions are fair with respect to certain population characteristics, that is if test takers with similar levels of a latent trait perform in similar way on individual test questions regardless of their personal characteristics.

3.1 The Rasch model

The family of RM (Rasch, 1960), is a family of measurement models which convert raw scores into linear and reproducible measurements. The models belonging to this family conform to fundamental measurement theory, that is they adhere to the principle of specific objectivity. Specific objectivity requires that the comparison between individuals, based on a latent trait of interest, is independent of which particular test has been

Table 1: List of the items proposed to measure the perceived self-efficacy of the social workers

Item	Indicate the extent to which you agree/disagree with each of the following statements
Acceptance	I am always able to establish a friendly, sympathetic rapport with the user
Powerlessness Management	I am always able to manage the impotence I sometimes feel when dealing with serious situations
Sharing Problems	I always manage to immediately inform / share with my superiors any problems that may arise
Burnout	I am always able to avoid being burdened with the user's problems that I cannot resolve
Holding	I always manage to keep my anxiety levels within certain limits when dealing with serious situations
Limits Recognition	When dealing with complex situations, I am always able to recognize the limits of my competences
Commitment	I am always able to fulfill my commitments to the user
No Judgment	I am always able to refrain from making any type of personal Judgment in my relations with the user
Colleague Support	I am always able to immediately convey to my colleagues my need for support
Inter-professional Support	I am always able to also look for and find support among persons in different professions
Teamwork Support	When dealing with complex cases I am always able to involve people and services from different professions
Redefinition of Targets	When faced with failure, I am always able to redefine objectives and start again from the beginning
Update	I always manage to find enough time to write and update cases reports

used to measure that specific latent trait; symmetrically, the comparison among items measuring the same latent trait is independent of which particular individuals were used for the comparison. For this to happen, the person and item parameters in the measure-

ment model must be additive and separable, and this is a characteristic of the family of RM. These models require unidimensionality, which means that all items forming the questionnaire measure only a single construct, i.e. the latent trait under study, and local independence, which requires that, conditional to the latent trait, the response to a given item is independent of the responses to the other items in the questionnaire. According to the RM, the probability that a subject n answers in a given way, say x, to the item i depends on personal ability, which is the perceived self-efficacy of the SW in the context of this paper, and how difficult the item is to endorse, that is, for this study, the difficulty to cope inter-personal/relational issues in institutional context. In the case of a questionnaire made of polytomously scored items, that is when there are $m + 1$ possible ordered response categories for each item (coded as $x = 0, 1, \ldots, m$), there are two models belonging to the family of RM suitable for handling this type of data, i.e. the Rating Scale Model (RSM) (Andrich, 1978) and the Partial Credit Model (PCM) (Masters, 1982). In the present study, all the items forming the questionnaire used to measure perceived self-efficacy use the same response format. It is, therefore, reasonable to assume that the test constructors, respondents, and test users all perceive the items to share the same rating scale. Moreover, for almost all items, the lowest response categories were chosen by fewer than 10 subjects, therefore the estimates of the parameters related to these response categories are not stable and accurate. As a consequence, the RSM is preferred over the PCM and this is the model used to analyze the data (Linacre, 2000). The subject response probability of the RSM has the following formulation:

$$P(X_{ni} = x) = \frac{exp\left\{x(\beta_n - \delta_i) - \sum_{j=0}^{x} \tau_j\right\}}{\sum_{k=0}^{m} exp\left\{k(\beta_n - \delta_i) - \sum_{j=0}^{k} \tau_j\right\}}, \qquad x = 0, 1, \ldots, m \tag{1}$$

where β_n identifies the ability of person n, δ_i the mean difficulty of item i, and τ_j, called threshold, is the point of equal probability of categories $j - 1$ and j, while $\tau_0 \equiv 0$ for convenience. Thresholds add up to zero, i.e. $\sum_{j=1}^{m} \tau_j = 0$.

If the categories and threshold estimates are ordered, this indicates well-functioning items, thus collapsing to a smaller number of categories will give a poorer fit for the model. If the categories and thresholds are disordered, merging categories may improve item fit and the overall scale, and may reveal the effective number and ordering of categories (Andrich et al., 1997).

In order to validate the scale and evaluate the measure obtained applying the RSM, the person reliability index, the Infit Mean Square (Infit MNSQ) and Outfit Mean Square (Outfit MNSQ) statistics and the principal component analysis (PCA) performed on standardized residuals will be considered (Bond and Fox, 2007).

The person reliability index provides a measure of the replicability of people placement that can be expected if the same sample of people is given another set of items measuring the same latent construct; it is bounded by zero and one, and can also be computed with missing values.

Infit MNSQ and Outfit MNSQ statistics indicate how accurately the data fit the model,

and allow anomalous single responses or abnormal response patterns within individual items to be identified. The infit statistic is an information-weighted fit statistic and it is more sensitive to unexpected behavior affecting responses to items near the subjects measure level. Instead, the outfit statistic is an outlier-sensitive fit statistic, therefore more sensitive to unexpected behavior on items far from the subjects measure level. The expected value for these statistics is one, which corresponds to the perfect fit of data to the model; values less than one indicate that the observations are too predictable whereas values greater than one indicate unpredictability. Reference values for these two statistics are reported in Brentari and Golia (2007) and (Wright and Linacre, 1994).

The PCA of standardized Rasch residuals allows verification of whether the assumption of unidimensionality required by the RSM is satisfied. If all the information in the data is due to the latent construct collected by the RSM, the unexplained part of the data measured by the residuals must be random noise. Therefore, if the first factor, identified by applying the PCA to the standardized Rasch residuals, is at the noise level, then there is no evidence of secondary dimensions. The idea is that the first dimension, called Rasch dimension, has already been removed and one is looking at unwanted secondary dimensions, therefore PCA is applied to the Rasch residuals and not to the original observations. The index used to verify the absence of a second dimension in the Rasch residuals is the first eigenvalue of PCA λ_1 and reference value for the unidimensionality case in Rasch context is 1.4, if the sample size is 500 (Smith, 2002; Brentari and Golia, 2007).

3.2 The Differential Item Functioning analysis

When a test is used to measure a latent trait, such as the perceived self-efficacy of SWs, it is important to ensure that the test itself and the items structure are invariant over population characteristics (such as, for example, age or gender). The DIF analysis examines the relationships among item responses, levels of the latent trait being measured by the questionnaire, and subgroup membership. Formally defined, an item displays DIF if subjects of equal proficiency on the construct to be measured by a test, but belonging to separate subgroups of the population, differ in their expected score on this item (Roussos and Stout, 2004; Penfield and Camilli, 2007). Usually the population is divided into two subgroups, named reference and focal group; the reference group provides the standard performance on the item of interest and in general it is the majority group, whereas the focal group is the group for which a differential performance is to be detected and measured. Two types of DIF can be identified and denoted as uniform and nonuniform. Uniform DIF occurs when the relative advantage of one group over another on a test item is uniform, favoring only one group consistently across the entire scale of ability, whereas nonuniform DIF exists when the conditional dependence of group membership and item performance changes in size or in direction across the entire ability continuum. In this paper we will only consider uniform DIF.

Several methods were proposed to assess DIF in dichotomous and polytomous items (Penfield and Camilli, 2007) and for some of them, a measure of the amount of DIF is available and a classification rule of the level of DIF can be built; a review of the systems

of DIF classification available in literature for polytomously scored items is given in Golia (2012). The usefulness of a classification lies in the fact that, when DIF items are found in the test, it is necessary to make decisions concerning their revision or removal which cannot be based only on the outcome of DIF hypothesis testing statistic, therefore, an interpretable measure of the amount of DIF can help in decisions regarding DIF items.

It is possible to divide the DIF detection methods into two broad classes, distinguishing between parametric and nonparametric methods. Parametric methods require the specification of a parametric functional form for the item response function, that is to identify a specific model for item responding, whereas nonparametric methods detect DIF by comparing item and test scores obtained directly from examinee responses without any model assumption and do not require large samples. The methods used in this study belong to the class of nonparametric DIF detection methods and they are: the Mantel Test (MT) (Mantel, 1963), the Standardized Mean Difference (SMD) statistic (Zwick and Thayer, 1996), the polytomous SIBTEST (Chang et al., 1996) and the Liu-Agresti cumulative common log-odds ratio (L-A LOR) (Liu and Agresti, 1996; Penfield and Algina, 2003). All these methods require the identification of a so-called matching variable, i.e. an explanatory variable able to form comparable groups; this matching variable can be the observed test score (MT, SMD and L-A LOR) or the estimate of a latent variable (polytomous SIBTEST).

The MT (Mantel, 1963) was developed to detect association between matched groups on ordinal variables; it assumes a multivariate hypergeometric distribution for the conditional bivariate distribution of group membership and item response. Under the null hypothesis of no DIF, the MT statistic Mantel-χ^2 is distributed approximately as a chi-square variable with one degree of freedom.

The SMD statistic and the polytomous SIBTEST follow a mean-difference approach, which consists in the evaluation of the conditional between-group difference in the expected value of the item response variable. Given a matching variable with K levels ($k = 1, \ldots, K$), the SMD (Zwick and Thayer, 1996) is defined as: $SMD = \sum_{k=1}^{K} w_{Fk} m_{Fk} - \sum_{k=1}^{K} w_{Fk} m_{Rk}$, where m_{Gk} is the mean item score for the group $G = R, F$ (R stands for reference and F for focal group) in the kth level of the matching variable, and w_{Fk} is the proportion of focal group members who are at kth level of the matching variable. If SMD assumes a negative (positive) value, this means that, conditional on the matching variable, the focal group has a lower (higher) mean item score than the reference group; this implies that the item is more difficult (easier) for the focal group. The SMD divided by the within-group standard deviation of the studied item pooled over the two groups (SD), constitutes a measure of DIF intensity and it is used, in conjunction with the MT, by the Educational Testing Service as a system of DIF classification (Golia, 2012).

The polytomous SIBTEST (Chang et al., 1996) is based on the regression of item performance onto an estimate of ability based on classical test theory of matching variable true score. A subset of items must be presumed approximately unidimensional and DIF-free and they form the base for the estimate of the matching variable. The polytomous SIBTEST statistic is given by $\hat{\beta} = \sum_{k=0}^{n_H} p_k d_k$, where n_H is the maximum possible score of the matching variable, $d_k = \bar{Y}^*_{Rk} - \bar{Y}^*_{Fk}$ and $\bar{Y}^*_{Gk}$ is the mean item score for the group

G ($G = F, R$) with score k adjusted for group mean differences on the latent variable using a correction that involves the regression of true score on observed score, and p_k is a weight equal to the proportion of all examinees with score k. A positive value indicates that the reference group members outperform focal group members of equal true scores, then the item is more difficult for the focal group; a negative value indicates the reverse. A test statistic can be defined by the ratio between $\hat{\beta}$ and its standard error estimate; this ratio, under the null hypothesis of no DIF is distributed approximately as standard normal. The polytomous SIBTEST performs similarly to the methods that use the observed test score as matching variable, but it performs better when the mean abilities of reference and focal group are not equal (Chang et al., 1996). Moreover, the ratio between $\hat{\beta}$ and the difference between the maximum and the minimum possible score on the item, constitutes a measure of the amount of DIF (Golia, 2012).

An alternative approach to testing the presence of DIF is to compute a cumulative common odds ratio. Let us denote the cumulative frequency for response category j at stratum k for the G group N^*_{Gjk}, the frequency at stratum k for the G group N_{Gk} and the frequency of stratum k N_k; the cumulative common odds ratio estimator first proposed by Liu and Agresti (1996) has the following form: $\hat{\psi}_{LA} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{m} A_{jk}D_{jk}/N_k}{\sum_{k=1}^{K}\sum_{j=1}^{m} B_{jk}C_{jk}/N_k}$, where $A_{jk} = N^*_{Rjk}$, $B_{jk} = N_{Rk} - N^*_{Rjk}$, $C_{jk} = N^*_{Fjk}$ and $D_{jk} = N_{Fk} - N^*_{Fjk}$. Given that $\hat{\psi}_{LA}$ reduces to the inverse of the Mantel Haenszel constant odds ratio $\hat{\alpha}_{MH}$ in the case of dichotomous items, $\hat{\psi}_{LA}$ can be seen as a generalization of $\hat{\alpha}_{MH}$. In order to produce an estimator that is consistent with the scale of $\hat{\alpha}_{MH}$, the estimator $\hat{\psi}_{LA}$ is transformed as $\hat{\alpha}_{LA} = 1/\hat{\psi}_{LA}$. An estimate of the L-A LOR is given by $\hat{\lambda}_{LA} = ln(\hat{\alpha}_{LA})$; positive (negative) values of $\hat{\lambda}_{LA}$ indicate DIF in favor of the reference (focal) group, that is the item is easier for the reference (focal) group members. Under the null hypothesis of no DIF, the Liu-Agresti cumulative common log-odds ratio divided by the estimated standard error (LOR-z) is asymptotically normally distributed. Moreover, $\hat{\lambda}_{LA}$ constitutes a measure of the amount of DIF (Golia, 2010).

4 Measuring perceived self-efficacy

As described in Section 2, the use of the CIT allowed the identification of about thirteen critical incidents collected through focus groups of SWs, and the behaviors to solve the issues constitute the items which form the questionnaire proposed to measure the perceived self-efficacy reported in Table 1. For each question in the test, the SWs were asked to indicate their degree of agreement, assigning a score from 1, *complete disagreement*, to 7, *complete agreement*. This questionnaire is part of a survey concerning SWs working in Veneto, one of the regions of Northern Italy. SWs from the regional Social Worker Association were asked to participate in the Departments survey focused on up-to-date identification of professional competences and well-being at work. The survey used the Internet surveying technique CAWI (Computer-assisted web interviewing) and involved 818 SWs (Bressan and Capiluppi, 2011; Minozzo and Ferrari, 2012; Pedrazza et al., 2013). Each participant received by e-mail instructions on how to access the online

questionnaire, securely and anonymously. Subjects could access it during their working time and complete the questionnaire at a time and in conditions of their choosing.

After list-wise deletion, the number of respondents dropped to 769, creating the final sample for this study. A preliminary analysis revealed that the order in the thresholds was respected, so merging categories was not necessary.

The estimates of the parameters involved in the RSM reported in equation (1) were obtained making use of the joint maximum likelihood estimation method (Wright and Masters, 1982), implemented in Winsteps 3.75.0 (Linacre, 2012); moreover, the sum of item difficulty parameters was set equal to 0.0 logits.

The first step of the analysis led to the removal of the items *Update* (Infit MNSQ = 1.5; Outfit MNSQ = 1.53) and *Sharing Problems* (Infit MNSQ = 1.35; Outfit MNSQ = 1.36) because they had poor infit and outfit statistics, indicating that these items could be measuring something other than what was intended. In fact, the content of both items is not specifically related to relational problems but rather to technical and administrative practices related to tools and ways in which information is communicated and organized. The PCA of standardized Rasch residuals obtained from the revised questionnaire highlighted the presence of an unwanted dimension formed by the items *Inter-professional Support*, *Teamwork Support* and *Colleague Support* ($\lambda_1 = 2.3$); this dimension refers to the capability of the SW to look for and find support in other professionals and/or colleagues. Given that inter-professional support, represented by the item *Inter-professional Support*, is the main objective of any type of teamwork in basic social work, we argue that this aspect is included in the item *Teamwork Support*, so we decided to delete the item *Inter-professional Support* in order to weaken the dimension found in the Rasch residuals. The PCA of standardized Rasch residuals drawn from the revised test composed of ten items identified the presence of an unwanted dimension formed by the items *Burnout*, *Holding* and *Powerlessness Management* ($\lambda_1 = 2$); these items refer to the SWs ability to manage negative emotions that arise when dealing with serious situations. The SWs ability to avoid being burdened with the users problems (*Burnout*) is related to the ability to manage impotence and to keep anxiety levels within acceptable limits; as a consequence we decided to eliminate the item *Burnout* in order to weaken this dimension found in the Rasch residuals. The item *Colleague Support* was also removed because of misfit (Infit MNSQ = 1.39; Outfit MNSQ = 1.38).

The resulting final scale consisted of eight items which highlight relational competences as the core competences in the measurement of the perceived self-efficacy of SWs. The person reliability index value (0.82) implies a good level of confidence that the worker placement would be reproducible with a different instrument measuring the same construct and the Cronbachs Alpha equal to 0.84 reveals a good internal consistency for the test. Table 2 displays the estimate of the threshold parameters τ_j while Table 3 reports the estimate of the item difficulty parameters δ_i and the infit and outfit statistics.

We found that the items *Powerlessness Management* and *Holding* are the most difficult to endorse, in other words it is hard for SWs to agree with the statements that they are able to manage the perceived powerlessness (*Powerlessness Management*) and to control the anxiety (*Holding*) in serious situations. This result is convergent with the up-to-date literature related to emotion regulation competences and the abilities required in order

Table 2: Treshold parameters

Threshold	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6
Estimate	-2.48	-1.94	-0.55	0.01	1.61	3.35

Table 3: Item difficulty measures and fit statistics

Item	Measure	Infit MNSQ	Outfit MNSQ
Powerlessness Management	0.81	1.01	1.03
Holding	0.49	1.01	1.00
Redefinition of Targets	0.35	0.77	0.78
No Judgment	0.03	1.03	1.01
Limits Recognition	-0.11	0.83	0.82
Commitment	-0.32	1.01	1.01
Acceptance	-0.45	1.09	1.10
Teamwork Support	-0.81	1.26	1.24

to manage relationships in every social domain. Therefore, a support related to these two aspects could help in increasing the perceived self-efficacy of SWs. On the contrary, the items *Teamwork Support* and *Acceptance* are the easiest to endorse; in other words it is easy for the SW to agree with the statements that they are able to involve various forms of expertize and services to deal with complex situations (*Teamwork Support*) and to be welcoming and sympathetic towards the end users (*Acceptance*). It is important to emphasize that issues on relational competences and techniques related to acceptance of the user on the one hand, and on team-work related issues on the other hand, are relevant parts of SW training. The estimates of personal abilities and item difficulties are expressed in the same unit of measurement, so it is possible to represent them on a single graph called Item Person Map (IPM); Figure 1 shows the IPM for the sample under study. This map shows simultaneously the distribution of person perceived self-efficacy (on the left side of the vertical line) and item difficulties (on the right side of the vertical line); M marks the person and item mean, S is one sample standard deviation away from the mean and T is two sample standard deviations away from the mean. The SWs located at the upper end of the scale have higher levels of perceived self-efficacy; those located at the lower end have lower levels of perceived self-efficacy. Items close to the bottom are frequently endorsed, so most workers consider them easy to endorse, while items higher on the construct are endorsed less frequently, so fewer workers consider

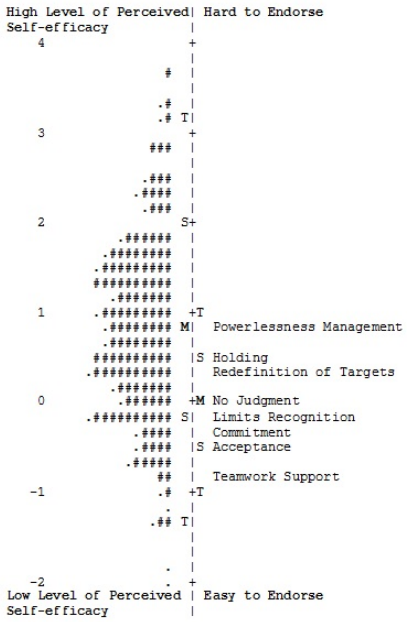

Figure 1: Item person map (each # corresponds to 5 social workers, each . represents one to four social workers)

them easy to endorse.

The average level of estimated perceived self-efficacy (0.86 logit) is higher than the item difficulty mean set to 0 logit; this suggests that the SWs involved in this study have a good perception of their professional self-efficacy. In order to verify whether the test questions are fair with respect of some population characteristics, all the methods described in Section 3.2 were applied, taking into account the final test. The population characteristics considered in this study are age, type of employment contract, type of service and working years. Table 4 reports the frequency distributions of these characteristics.

The computation of Mantel-χ^2, $\hat{\lambda}_{LA}$ and LOR-z were obtained using the DIFAS computer program (Penfield, 2005), the estimation of $\hat{\beta}$ and the associated test statistic were achieved using the DIF-Pack computer program (Stout, 2005) whereas the SMD statistics were computed using the software R 2.14 (R Core Team, 2011). Tables 5 to 8 report the results of these statistics; in almost all the cases the DIF indices highlight the presence of the same DIF items. Positive values of SMD and $\hat{\lambda}_{LA}$ together with a negative value of $\hat{\beta}$ indicate that the item under study is easier for members of the focal group than for members of the reference group of equal level of perceived self-efficacy, whereas positive values of SMD and $\hat{\lambda}_{LA}$ together with a negative value of $\hat{\beta}$ indicate the reverse.

Let us consider the population characteristic *age*; the members of the focal group are SWs over 40 years old. The results reported in Table 5 identify two DIF items, that is

Table 4: List of the items proposed to measure the perceived self-efficacy of the social workers

Social worker characteristics	Number	Percentage	Social worker characteristics	Number	Percentage
Age (years)			*Type of service*		
≤ 40	474	61.64	Families-basic	324	42.13
≥ 41	289	37.58	Other	445	57.87
Missing	6	0.78	Missing	0	0.00
Employment contract			*Working years*		
Permanent	606	78.80	≤ 10	374	48.63
Fixed Term	107	13.91	$11 - 46$	339	44.08
Missing	56	7.28	Missing	56	7.28

Table 5: Differential item functioning by social workers age

Method	Poly SIBTEST		Mantel-χ^2	SMD/SD	Liu-Agresti LOR	
Item	p-value	$\hat{\beta}/m$	p-value		LOR-z	$\hat{\lambda}_{LA}$
Acceptance	0.016	0.018	0.032	-0.138	2.207	0.331
Powerlessness Management	0.094	-0.022	0.119	0.082	-1.526	-0.232
Holding	0.671	-0.006	0.729	0.021	-0.356	-0.053
Limits Recognition	0.841	-0.002	0.535	0.045	-0.633	-0.095
Commitment	0.007	0.033	0.031	-0.132	2.177	0.320
No Judgment	0.554	-0.008	0.502	0.043	-0.673	-0.099
Teamwork Support	0.928	0.001	0.899	-0.011	0.118	0.018
Redefinition of Targets	0.344	-0.012	0.258	0.067	-1.092	-0.167

Acceptance and *Commitment*, and for both these items the younger workers outperform the older, given the same level of perceived self-efficacy. This means that, at the same perceived self-efficacy level, a younger SW finds it easier to establish a friendly and sympathetic rapport with users and to fulfill her/his commitments to them. Recent research involving similar helping profession practitioners such as educators, gave similar results: the older population scored lower than the younger one in the relevance they attributed to empathy (Pedrazza et al., 2008). Moreover, younger SWs are typically more involved in work activities and their commitment is higher than their older colleagues because they feel more involved in the realization of their ideal professional role (Zucchermaglio and Alby, 2006). Nevertheless, following the indications regarding the classification rule

of the level of DIF revised in Golia (2012), the difference in item functioning found in the data can be classified as negligible, implying that it is not high enough to affect the obtained measures.

Table 6: Differential item functioning by social workers type of employment contract

Method	Poly SIBTEST		Mantel-χ^2	SMD/SD	Liu-Agresti LOR	
Item	p-value	$\hat{\beta}/m$	p-value		LOR-z	$\hat{\lambda}_{LA}$
Acceptance	0.000	-0.072	0.000	0.363	-3.676	-0.827
Powerlessness Management	0.163	-0.028	0.363	0.090	-0.917	-0.189
Holding	0.666	-0.009	0.641	0.021	-0.459	-0.096
Limits Recognition	0.543	-0.011	0.676	0.025	-0.406	-0.086
Commitment	0.163	-0.028	0.183	0.103	-1.369	-0.278
No Judgment	0.141	0.030	0.096	-0.123	1.702	0.349
Teamwork Support	0.001	0.065	0.001	-0.275	3.345	0.689
Redefinition of Targets	0.244	0.026	0.038	-0.158	2.053	0.423

The second population characteristic taken into account is the *type of employment contract*; in this case the members of the focal group are SWs with a fixed term contract. All the methods identify two DIF items, that is *Acceptance* and *Teamwork Support*, whereas MT and L-A LOR found another DIF item, i.e. *Redefinition of Targets*, as shown in Table 6. Considering the DIF statistics, it is possible to state that, at the same level of perceived self-efficacy, a SW with a fixed term contract finds it easier to establish a friendly and sympathetic relationship with users and she/he finds it more difficult, when dealing with complex cases, to involve people and services from different professions; these differences between the two groups of SWs are considerable. The analysis of the *Redefinition of Targets* item shows that, at the same level of perceived self-efficacy, SWs with a fixed term contract find it more difficult, when faced with failure, to redefine objectives and start again from the beginning; nevertheless this difference in item functioning is not relevant. The presence of two items with a non-negligible level of DIF in the test could cause a distortion in the estimation of the subject-perceived self-efficacy measure, nevertheless simulation studies (Golia, 2010) have shown that the presence of a set of items, for which the DIF sign and size compensate each other, does not compromise the quality of the estimated abilities. Moreover, the evaluation of their compliance to the construct that was intended to measure by experts ensures the maintaining of these items in the questionnaire to preserve the integrity of the content specifications mandated for the test. In fact, the ability to accept users problems and personality, and to be empathic (*Acceptance*), is a relevant part of the SWs training. We therefore expect SWs with a fixed term contract, who are usually beginners, to be more comfortable in all activities related to the acceptance and understanding of the user (Pedrazza et al., 2008). Moreover, SWs with a permanent contract will have already

experienced difficulties and failures related to *Teamwork Support* and *Redefinition of Targets* so, facing failure, they are better able to involve practitioners from different services and to redefine objectives. Both *Teamwork Support* and *Redefinition of Targets* are relevant issues in public services in the context in which we developed our research. In fact, in the Italian context the SWs are responsible to multidisciplinary teams where clinicians, neuropsychiatrists, and lawyers typically exert a very relevant influence on them. Moreover, *Redefinition of Target* occurs above all in cases where users fail to adhere to the project they agreed with the SWs, so this item renders failure salient for SWs with a fixed term contract. It is difficult for them to accept its meaning and to admit thereby failure.

Table 7: Differential item functioning by social workers type of service

Method	Poly SIBTEST		Mantel-χ^2	SMD/SD	Liu-Agresti LOR	
Item	p-value	$\hat{\beta}/m$	p-value		LOR-z	$\hat{\lambda}_{LA}$
Acceptance	0.839	-0.002	0.887	-0.008	-0.142	-0.021
Powerlessness Management	0.027	-0.029	0.056	0.105	-1.915	-0.272
Holding	0.247	-0.015	0.267	0.043	-1.110	-0.161
Limits Recognition	0.221	-0.014	0.233	0.044	-1.204	-0.177
Commitment	0.251	0.014	0.205	-0.066	1.257	0.181
No Judgment	0.083	0.022	0.066	-0.117	1.861	0.268
Teamwork Support	0.228	0.015	0.063	-0.128	1.844	0.271
Redefinition of Targets	0.447	-0.008	0.521	0.022	-0.646	-0.095

The third population characteristic taken into account is the *type of service carried out*; the members of the focal group are SWs who provide basic services to families. The results reported in Table 7 identify only one DIF item, *Powerlessness Management*, through the polytomous SIBTEST. At the same level of perceived self-efficacy, a SW who provides a basic-families service, finds it easier to manage the impotence she/he sometimes feels when dealing with serious situations. In fact, this type of service focuses on a familys economic, occupational and health issues and, as a consequence, practitioners are typically involved mainly in administrative and organizational activities rather than deeply involved in a relationship with users. Other services such as childwelfare, social services for drug addiction (SERT, it.), family services for abused women, family services for the elderly, are expected to involve SWs deeply, above all from an interpersonal and emotional point of view. Thus we assume that SWs, whose daily practice is mainly related to administrative and organizational issues, are not overwhelmed from the sense of impotence in the face of serious situations, as usually occur in other types of services such as the above-mentioned. Nevertheless, following the indications regarding the classification rule of the level of DIF revised in Golia (2012), the difference in item functioning found in the data can be classified as negligible, so it does not affect the

estimated measures.

Table 8: Differential item functioning by social workers working years

Method	Poly SIBTEST		Mantel-χ^2	SMD/SD	Liu-Agresti LOR	
Item	p-value	$\hat{\beta}/m$	p-value		LOR-z	$\hat{\lambda}_{LA}$
Acceptance	0.000	0.049	0.001	-0.213	3.342	0.508
Powerlessness Management	0.631	-0.007	0.428	0.035	-0.782	-0.115
Holding	0.350	-0.013	0.308	0.049	-1.014	-0.150
Limits Recognition	0.104	-0.019	0.127	0.082	-1.556	-0.235
Commitment	0.019	0.028	0.023	-0.140	2.299	0.338
No Judgment	0.693	0.005	0.823	-0.007	0.221	0.033
Teamwork Support	0.586	-0.007	0.545	0.029	-0.588	-0.090
Redefinition of Targets	0.095	-0.019	0.050	0.084	-1.935	0.296

The last population characteristic taken into account is the *number of working years*; the members of the focal group are SWs who have worked for more than 10 years. All the methods identify two DIF items, *Acceptance* and *Commitment*, as shown in Table 8. Considering the sign of the DIF statistics, it is possible to state that, at the same level of perceived self-efficacy, the senior SWs find it more difficult to establish a friendly and sympathetic rapport with users and this difference between the two groups of SWs is considerable. On the other hand, at the same level of perceived self-efficacy, for the senior SWs it is more difficult to fulfill their commitments to the users but this difference between the two groups of SWs is not relevant. The presence of only one item with a non-negligible level of DIF in the test does not compromise the quality of the estimated abilities and the compliance to the construct that was intended to measure by experts justifies retaining the item in the questionnaire. In fact, recent research involving educators gave the same results: the population with a longer work-experience scored lower than the beginners in the relevance they attributed to empathy (*Acceptance*) (Pedrazza et al., 2008). Beginners are also typically more involved in work activities, and their commitment is higher than their more experienced colleagues because they feel more involved in the realization of their ideal professional role.

5 Discussion and conclusions

The paper presented a self-efficacy scale developed for SWs. The use of the CIT allowed the identification of about thirteen critical incidents collected through focus groups of SWs, and the behaviors to solve these issues constituted the items which formed the initial questionnaire proposed to measure perceived self-efficacy. Making use of the RSM as a measurement model, the original test was revised and the final questionnaire consisted of eight items. Starting from this final test, an objective measurement of social workers

perceived self-efficacy was produced and analyzed. From the resulting final scale, it is possible to recognize that the SWs identify relational competences as core competences ascribable to the set of items they included in the unidimensional self-efficacy scale. There are two possible explanations for this fact. The first relates to the current world-wide recession which led to an increase in requests for social services. The huge number of users requests, the often hopeless access of users to the service and the restrictions on frequent turn-over applied in public welfare-institutions, induced a rising number of perceived difficulties in managing intra-service relationships e.g. with users, and inter-service e.g. in interdisciplinary teams. This could have led our sample to identify as the most critical events relational items and psycho-dynamic related issues such as holding, acceptance of the user, communication with users and/or co-workers, communication with superiors, communication within interdisciplinary teams, and teamwork. The particular type of social service, the context of this survey, and the particular current social situation render relational issues salient. The specific type of SW educational requirements at both basic and advanced educational level could be identified as a second reason connected with a perceived rising importance of relational problems at work. In fact, in the survey context, the main training issues in SW studies are related to sociological/juridical and social work disciplines, so our sample would obtain minimal benefit from specific psychological training. Moreover, Italian social work courses do not point out uniformities in psychological training at either bachelor or master level. In addition this unidimensional scale is interesting because it refers to intra-personal and inter-personal processes in relationship management. According to attachment theory (Dozier et al., 2003; Mikulincer and Shaver, 2007) applied to helping professions, the unidimensionality of the scale runs from the one pole: *Powerlessness Management, Holding* and *Acceptance* which are above all related to intra-personal psycho-dynamics, *No Judgment*, *Recognition of Limits*, and *Commitment*, which relate both to intra- and to inter-personal dynamics; to the opposite pole defined by *Redefinition of Targets* and *Teamwork Support* which are more concerned with social/interpersonal interactions. The up-to-date applications of attachment theory in helping professional context indicates that attachment style (secure, anxious, or avoidant) differ in type and means by which subjects manage the interpersonal distance from the generalized-other (Mikulincer and Shaver, 2007). Therefore we argue that this unidimensional scale could be very useful for identifying self-efficacy perception of the relational competence of social workers at work. Correlational studies simultaneously assessing attachment style and self-perception of efficacy in relational competences could be useful to give SWs an insight into training opportunities to improve well-being at work. A DIF analysis was performed in order to determine whether the questions forming the final scale were fair with respect to some population characteristics. From the analysis, the item Acceptance has shown non-negligible differences in its functioning with respect to the type of employment contract and the number of working years, whereas for the item Teamwork Support a non-negligible level of DIF has been observed with reference to the type of employment contract. Given the same perceived level of self-efficacy, SWs with a permanent contract find it more difficult to establish a friendly and sympathetic rapport with users and easier to involve people and services from different professions, when dealing with complex cases, than SWs with a fixed term

contract. Moreover, at the same perceived level of self-efficacy, a SW who has worked fewer than 10 years, finds it easier to establish a friendly and sympathetic relationship with users than a senior SW. The presence of two items with a non-negligible level of DIF, with respect to the type of employment contract, could cause a distortion in the estimation of the perceived self-efficacy; nevertheless simulation studies (Golia, 2010) have shown that the presence of a set of items, for which the DIF sign and size compensate each other, as in this case, does not compromise the quality of the estimated abilities. Moreover, the analysis of these items by some experts has suggested that the cause of the observed DIF corresponds to a construct that was intended to be measured by the test and then the items are retained in the test to preserve the integrity of the content specifications mandated for the test. The information coming from the DIF analysis can be used to describe, in a better way, the phenomenon under study, that is the perceived self-efficacy.

Aknowledgment

The authors wish to thank F. Bressan and the two anonymous reviewers for their useful comments.

The paper was partially supported by MIUR-PRIN 2007 project 20075XA9CJ 005.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.

Andrich, D., de Jong, J.H.A.L. and Sheridan, B.E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost and R. Langeheine (Eds.) *Application of latent trait and latent class models in the social science.* Munster: Waxmann.

Bandura, A. (1977). Self-Efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122-147.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1995).
Self-efficacy in changing societies. Cambridge, UK: Cambridge University Press.

Bandura, A. (1997). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review*, 84, 191-215.

Bandura, A. (2006). Guide to the Construction of Self-efficacy Scales. In F. Pajares and T. Urdan (Eds.) *Self-Efficacy Beliefs of Adolescents* (pp. 307-337). Greenwich: Information Age Publishing.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, 38, 9-44.

Betz, N.E. and Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men. *Journal of Counseling Psychology*, 28, 399-410.

Bond, T.G. and Fox, C.M. (2007). *Applying the Rasch Model. Fundamental measurement in the human sciences, 2nd edn.* Mahwah, NJ: Lawrence Erlbaum Associates.

Bong, M. and Skaalvik, E.M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15, 1-40.

Brentari, E. and Golia, S. (2007). Unidimensionality in the Rasch model: how to detect and interpret. *Statistica*, 6, 253-261.

Bressan, F. and Capiluppi, C. (2011). Progettazione e andamento dellindagine sugli assistenti sociali nel Veneto. In F. Bressan, M. Pedrazza and E. Neve (Eds.) *Il percorso formativo dellassistente sociale: autovalutazione e benessere professionale* (pp.71-95). Milano:Franco Angeli.

Butterfield, L.D., Borgen, W.A., Amundson, N.E. and Maglio, A.T. (2005). Fifty years of the critical incident technique: 1954-2004 and beyond. *Qualitative Research*, 5, 475-497.

Butterfield, L.D., Borgen, W.A., Maglio, A.T. and Amundson, N.E. (2009). Using the Enhanced Critical Incident Technique in counselling psychology research. *Canadian Journal of Counselling*, 43(4), 265-282.

Chang, H., Mazzeo, J. and Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33, 333-353.

Chen, G., Gully, S.M. and Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods*, 4(1), 62-83.

Dozier, M., Cue, K.L. and Barnett, L. (2003). Clinicians as caregivers: Role of attachment organization in treatment. *Journal of Consulting and Clinical Psychology*, 62(4), 793-800.

Eden, D. (1988). Pygmalion, goal setting, and expectancy: compatible ways to raise productivity. *Academy of Management Review*, 13, 639-652.

Eden, D. and Granat-Flomin, R. (2000). Augmenting means efficay to improve service performance among computer users. *Paper presented at the* 15^{th} *Annual Meeting of the Society for Industrial and Organizational Psychology*, New Orleans, LA:

Flanagan, J.C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327-358.

Golia, S. (2010). The assessment of DIF on Rasch measures with an application to job satisfaction. *Electronic Journal of Applied Statistical Analysis Decision Support System and Service Evaluation*, 1(1), 16-25.

Golia, S. (2012). Differential Item Functioning classification for polytomously scored items. *Electronic Journal of Applied Statistical Analysis*, 5(3), 367-373.

Gremler, D.D. (2004). The critical incident technique in service research. *Journal of Service Research*, 7(1), 65-89.

Holden, G., Cuzzi, L., Spitzer, W., Rutter, S., Chernack, P. and Rosenberg, G. (1997). The hospital social work self-efficacy scale: a partial replication and extension. *Health & Social Work*, 22(4), 256-63.

Holden, G., Meenaghan, T., Anastas, J. and Metrey, G. (2002). Outcomes of social work education: The case for social work self-efficacy. *Journal of Social Work Education*, 38, 115-133.

Linacre, J.M . (2000) Comparing Partial Credit Models" (PCM) and "Rating Scale Models" (RSM). *Rasch Measurement Transactions*, 14(3), 768.

Linacre, J.M. (2012). Winsteps (Version 3.75.0) [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved January 1, 2012. Available from http://www.winsteps.com/

Liu, I.-M. and Agresti, A. (1996). Mantel-Haenszel-type inference for cumulative odds ratios with a stratified ordinal response. *Biometrics*, 52, 1223-1234.

Mayhew, L.B. (1956). The critical incident technique in educational evaluation. *Journal of Educational Research*, 49(4), 591-598.

Mantel, N. (1963). Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, 58, 690-700.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Mikulincer, M. and Shaver, P.R. (2007) *Attachment in adulthood: structure, dynamics, and change*. New York: Guilford Press.

Minozzo, M. and Ferrari, C. (2012). Binomial factor analysis with the MCEM algorithm. *Electronic Journal of Applied Statistical Analysis*, 5(3), 346-352.

OHare, T. and Collins, P. (1997). Development and validation of a scale for measuring social work practice skills. *Research on Social Work Practice*, 7, 228-238.

Pajares, F. (1997). Current Directions in Self-efficacy Research. In M.L. Maehr and P.R. Pintrich (Eds.) *Advances in Motivation and Achievement, vol.10* (pp.149). Greenwich, CT: JAI Press

Pedrazza, M., Boccato, G., Berlanda, S. and Parisi, G. (2008). De la formation a laction : Leducateur extrascolaire, le role ideal et le role percu. In N. Pettersen, J.S. Boudrias and A. Savoie (Eds.) *Comment transformons nous lunivers du travail?* Quebec: Presses de LUniversit du Quebec.

Pedrazza, M. and Boccato, G. (2009). The tension between empathy and assertiveness and its correlation with self efficacy. DiPAV - QUADERNI, 25, 131-142.

Pedrazza, M. and Boccato, G. (2012). Training on attachment style increases secure adults attribution of control. TPM - Testing Psychometrics and Methodology in Applied Psychology, 19, 69-79.

Pedrazza, M., Trifiletti, E., Berlanda, S. and Di Bernardo, G.A. (2013). Self -efficacy in social work: development and initial validation of the self-efficacy scale for social

workers. *Social Sciences*, 2, 191-207.

Penfield, R.D. (2005). DIFAS: Differential item functioning analysis system. *Applied Psychological Measurement*, 29, 150-151.

Penfield, R.D. and Algina, J. (2003). Applying the Liu-Agresti estimator of the cumulative common odds ratio to DIF detection in polytomous items. *Journal of Educational Measurement*, 40, 353-370.

Penfield, R.D. and Camilli, G. (2007). Differential item functioning and item bias. In S. Sinharay and C.R. Rao (Eds.) *Handbook of statistics, vol. 26: Psychometrics* (pp. 125-167). Elsevier.

R Core Team (2011). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* (Reprint, with Foreword and Afterword by B. D. Wright, Chicago: University of Chicago Press, 1980). Copenhagen: Danmarks Paedogogiske Institut.

Roussos, L.A. and Stout, W. (2004). Differential item functioning analysis: Detecting DIF item and testing DIF hypotheses. In D. Kaplan (Ed.) *The Sage handbook of quantitative methodology for the social sciences* (pp.107-116). Thousand Oaks, CA: Sage.

Schwarzer, R. and Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright and M. Johnston (Eds.) *Measure in health psychology: A users portfolio. Causal and control beliefs* (pp.35-37). Windsor, UK: Nfer-Nelson.

Smith, Jr.E.V. (2002). Detecting and evaluating the impact of multidimensionality using item fitstatistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 205-231.

Stajkovic, A.D. and Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, 124, 240-261.

Stout, W. (2005). DIF-pack [Computer Software]. Available from http://psychometrictools.measuredprogress.org/dif1

Wright, B.D. and Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370.

Wright, B.D. and Masters, G.N. (1982). *Rating scale analysis.* Chicago: MESA Press.

Zimmerman, B.J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.) *Self-efficacy in changing societies* (pp. 202-231). New York: Cambridge University Press.

Zimmerman, B.J. and Cleary, T.J. (2006). Adolescents Development of Personal Agency. The Role of Self-efficacy Beliefs and Self-regulatory Skills. In F. Pajares and T. Urdan (Eds.) *Self-Efficacy Beliefs of Adolescents* (pp. 45-69). Greenwich: Information Age Publishing.

Zucchermaglio, C. and Alby, F. (2006). Psicologia culturale delle organizzazioni Roma, Carocci.

Zwick, R. and Thayer, D.T. (1996). Evaluating the magnitude of differential item func-

tioning in polytomous items. *Journal of Educational and Behavioral Statistics*, 21, 187-201.