Europe's Journal of Psychology
ejop.psychopen.eu | 1841-0413

Research Reports

Evaluation of the Psychometric Properties of the Internet Addiction Test (IAT) in a Sample of Cypriot High School Students: The Rasch Measurement Perspective

Panayiotis Panayides*[a], Miranda Jane Walker[a]

[a] Lyceum of Polemidia, Limassol, Cyprus.

Abstract

As Greek Cypriot senior high school teachers, the researchers believe that instruments assessing Internet addiction should be developed and validated for use wherever there are adolescents (the most at-risk population) and Internet access. The purpose of the study was to evaluate the psychometric properties of the Internet Addiction Test (IAT). A sample of 604 randomly selected high school students from five high schools in Limassol, Cyprus participated in the study. The Rasch Rating Scale Model was used for the analyses of the data collected. Results suggested the modification of the IAT in two ways. First, the 5-point rating scale was replaced by a 3-point scale, which was found to be optimal in the pilot study. Second, item 8 was replaced by a self-rating item because it was found to be identical to item 6 both statistically and semantically. The respondents' reliability was satisfactory (0.86) and item reliability very high (0.99). All 20 items were sufficiently spread out and describe distinct levels along the variable and do define a linear continuum of increasing difficulty. All the evidence collected supports the unidimensionality and the high degree of construct validity of the scale. Finally four recommendations for the modification of the scale and future research are proposed.

Keywords: internet addiction, Internet Addiction Test, high school students, Rasch measurement, psychometric properties, unidimensionality

Europe's Journal of Psychology, 2012, Vol. 8(3), 327–351, doi:10.5964/ejop.v8i3.474

Received: 2012-01-02. Accepted: 2012-05-30. Published: 2012-08-29.

*Corresponding author at: Nikou Kavadia 1, K. Polemidia, 4152, Limassol, Cyprus, email: p.panayides@cytanet.com.cy.

Kandell (1998) defined Internet addiction as "a psychological dependence on the Internet, regardless of the activity once logged on" (p.12). Shaw and Black (2008) stated that Internet addiction is "characterized by excessive or poorly controlled preoccupations, urges or behaviours regarding computer use and internet access that lead to impairment or distress" (p.353).

Internet related dependency has been termed *Internet Addiction Disorder* (e.g. Goldberg, 1996), *Internet addiction* (e.g. Chou & Hsiao, 2000; Scherer & Bost, 1997; Young, 1998a), *Internet dependency* (e.g. Lin & Tsai, 2002; Scherer & Bost, 1997), *Internet pathological use* (e.g. Davis, 2001; Morahan-Martin & Schumacher, 2000) and *Problematic Internet Use* (Davis, Flett, & Besser, 2002; Odacı & Kalkan, 2010). Despite the lack of universal agreement in terminology and definition, common indicators concerning this disorder can be found in the literature such as excessive time on the Internet, distress or irritability when the Internet is not available and the feeling of needing to spend more time online (Young & Rodgers, 1998).

Griffiths (2000) observes the scepticism among the academic community regarding the concept of 'Internet Addiction' but points out the acceptance of pathological gambling as an addiction has created a precedent for

other excessive behaviours, such as Internet addiction. In addition Widyanto and Griffiths (2006) state that Internet addiction has frequently been conceptualised as a behavioural addiction, operating on a modified principle of classic addiction models, but further note that the validity and clinical worth of these claims has been questioned. They emphasise the lack of theoretical basis for the construct despite the number of studies which have been undertaken on Internet Addiction. Davis (2001) proposed a model of the etiology of pathological Internet use, the main assumption of which is that it arises from "problematic cognitions coupled with behaviours that intensify or maintain maladaptive responses" (cited in Widyanto & Griffiths, 2006, p.45).

Internet Addiction and Adolescents

Various studies accentuate the importance of examining the impact of problematic Internet use on the most vulnerable to this, adolescents (Ferraro, Caci, D'Amico, & Di Blasi, 2007; Johansson & Götestam, 2004). In general, adolescents are at a critical period of addiction vulnerability, based on their social and also neurobiological factors (Jang, Hwang, & Choi, 2008; Lam-Figueroa et al., 2011; Pallanti, Bernardi, & Quercioli, 2006). With regard to the Internet they are more vulnerable and at risk as they have easy access to the Internet and flexible timetables (Moore, cited in Widyanto & Griffiths, 2006). Furthermore they tend to be less self-regulative (Fu, Chan, Wong, & Yip, 2010), and also have less ability to control their enthusiasm for Internet activities (Yen, Ko, Yen, Chang & Cheng, 2009). More specifically, research indicates Internet use is highest in the 16-24 age groups (Kandell, 1998; Öztürk, Odabasioglu, Eraslan, Genç, & Kalyoncu, 2007). Odacı & Kalkan (2010) suggest this implies a potential risk of Internet dependence among this age group. Internet addiction has been reported to be negatively correlated with academic performance including poor grades, tardiness and procrastination (Chang & Law, 2008; Chou & Hsiao, 2000; Scherer & Bost, 1997; Yen et al., 2009). Furthermore it has been linked to time distortion (Odacı & Kalkan, 2010) and shown to adversely affect sleep habits (Choi et al. 2009; Kesici & Sahin, 2010).

Internet Addiction Scales

Many scales have been developed to identify the level of Internet addiction in users. Goldberg (1996) developed the Internet Addictive Disorder (IAD) scale, with seven diagnostic criteria, mainly adapted from the 1994 edition of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-IV). Young (1998a) suggests pathological gambling is the most akin disorder to the pathological nature of Internet use. She stated that "by using Pathological Gambling as a model, Internet addiction can be defined as an impulse-control disorder which does not involve an intoxicant" (p.238). Young (1998a) introduced a Diagnostic Questionnaire (YDQ) for 'Internet addiction', with eight dichotomous items, adapted from DSM-IV, from the criteria used for pathological gambling. She suggested a cut-off score of five, arguing that this cut off score is consistent with the number of criteria used for pathological gambling and is seen as an adequate number of criteria to differentiate normal from pathological addictive Internet use. Brenner (1997) developed the IRABI (Internet-Related Addictive Behavior Inventory) scale, with 32 true-false items addressing excessive Internet use. In 1998 Young expanded on her YDQ and developed the 20-item Internet Addiction Test (IAT). Respondents are asked to answer the 20 items on a 5-point Likert scale (scored from one to five) indicating the degree to which Internet usage affects their daily routine, social life, productivity, sleeping pattern, and feelings. The higher the score, the greater the problems caused by Internet usage. Young extended the cut-off score of five out of eight criteria of the original 8-item YDQ to the IAT. She suggested a score of 20-39 indicated 'no problems'; 40-69 'frequent problems'; 70-100 'significant

problems' for the user. Morahan-Martin and Schumacher (2000) introduced their Pathological Internet Use (PIU) scale, with 13 items, largely based on the DSM-IV criteria for gambling.

Factor Structure

The overwhelming majority of Internet addiction scales developed have been shown to be multidimensional. The factorial complexity of the measures from these scales varies widely ranging from one (Siomos, Dafouli, Braimiotis, Mouzas, & Angelopoulos, 2008) to as many as seven factors (Caplan 2002).

There are several reasons for such diverse factor structures for Internet addiction. First, the construct has not been consistently defined across the various studies. Jia and Jia (2009) argue that a critical step towards discovering the true factor structure of Internet addiction is achieving a consensus definition. This definition would determine the domain of the construct and the item pool. Second, there are several instruments with varied lengths (from eight to 36 items) in the literature that appear to be measuring the construct. Jia and Jia (2009) also argue that the factor analytic techniques and the decision heuristics used in developing these scales have a direct impact on the structure obtained. Finally confirmatory factor analysis (CFA) was used in some studies to confirm the factor structures. Kline (2000) points out some of the problems associated with this method and emphasises that "The fact that a model is confirmed, … means only that this particular model fits the data. It does not mean that other models might not fit and fit better" (p. 183).

The noticeable inconsistency of various studies related to the factor structure is not always a result of the different scales used. Even in studies where the IAT (Young, 1998b) was used, different factor structures were reported. A 3-factor structure of the IAT was reported by Law and Chang (2007) and Chang and Law (2008). Widyanto, Griffiths, and Brunsden (2011), in comparing the IAT with the IRPS, also extracted 3 factors for the IAT using exploratory factor analysis (EFA). However, in a study a few years earlier, Widyanto and McMurran (2004) reported six factors as did Ferraro et al. (2007) with the Italian version of the IAT.

If all the scales mentioned in this study, including the IAT, are multidimensional then the following question arises: Can the scores on the individual items be summed to give a total score which will be used to identify the severity of the Internet addiction of any respondent?

Some of the studies that used the IAT did report significant inter-factor correlations, perhaps implying (but not stating) the possibility of a unidimensional scale, and this would justify the use of a total score for measuring Internet addiction (e.g. Chang & Law, 2008; Widyanto & McMurran, 2004). Other studies, however, did not report such correlations (e.g. Choi et al., 2009; Law & Chang, 2007)

Various instruments are proposed for studying Internet addiction but it is crucial to establish the validity and reliability of these instruments. "Good measurement is a pre-condition for building up knowledge in the research area of Internet addiction" (Law & Chang, 2007, p.8).

Rasch Measurement

The Rasch model asserts that a person with higher endorsability (i.e. higher position on the Internet addiction continuum) always has a higher probability of endorsing any item than a person with lower endorsability, and a more difficult (to endorse) item has a lower probability of endorsement than a less difficult item, regardless of

person position on the Internet addiction continuum. The original breakthrough by Rasch in 1960 has been developed and extended to address every reasonable observational situation in the social sciences. If the test has a single type of item, with the same number of marks available (as with the Likert scales), then the Rating Scale Model (RSM) applies (Andrich, 1978).

According to the model the probability of a person n responding in category x to item i, is given by:

$$P_{xni} = \frac{\exp \sum_{j=0}^{x} [\beta_n - (\delta_i + \tau_j)]}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} [\beta_n - (\delta_i + \tau_j)]} \qquad x = 0,1,...,m$$

where $\tau_o = 0$ so that

$$\exp \sum_{j=0}^{0} [\beta_n - (\delta_i + \tau_j)] = 1$$

β_n is the person's position on the variable, δ_i is the scale value (difficulty to endorse) estimated for each item i and $\tau_1, \tau_2, \ldots, \tau_m$ are the m response thresholds estimated for the m + 1 rating categories.

Panayides, Robinson, and Tymms (2010) reported a selection of applications of Rasch measurement showing the diversity of situations in the social sciences in which the Rasch approach can be used productively, including construction and evaluation of psychometric scales. For example, Prieto, Roset, and Badia (2001) have used the Rasch dichotomous model to assess the metric properties of the Spanish version of the assessment of Growth hormone deficiency in adults and to confirm its unidimensionality and construct validity. Massof and Fletcher (2001) have used the model to evaluate the validity of and to improve the visual functioning questionnaire which is designed to assess health-related quality of life of patients with visual impairment. Chen, Bezruczko, and Ryan-Henry (2006), have used Rasch analyses to describe mothers' effectiveness in caregiving for their adult children with intellectual disabilities and Myford and Wolfe (2002) examined a procedure for identifying and resolving discrepancies in examiners' ratings.

Unidimensionality

The Rasch model constructs a one-dimensional measurement system from ordinal data regardless of the dimensionality of the data. However, more than one latent dimension will always contribute to empirical data. Multidimensionality will become a real concern when the response patterns indicate the presence of two or more dimensions so disparate that it is no longer clear what latent dimension the Rasch dimension operationalizes.

Factor analysis is widely used in psychometrics to investigate the dimensionality of empirical data. However it "is confused by ordinal variables and highly correlated factors. Rasch analysis excels at constructing linearity out of ordinality and at aiding the identification of the core construct inside a fog of collinearity." (Schumacker & Linacre, 1996, p.470). Linacre (1998) showed that Rasch analysis followed by PCA of standardized residuals was always

more effective at both constructing measures and identifying multidimensionality than direct factor analysis of the original response-level data.

A key issue in the identification of a second dimension is the choice of the critical value of the eigenvalue. Researchers have suggested various critical values. Smith and Miao (1994) and Raiche (2005) suggested 1.4 whereas Smith (2004a) 1.5. Linacre (2005) however, argues convincingly that an eigenvalue less than 2 indicates that the implied dimension in the data has less than the strength of two items, and so, however powerful it may be diagnostically, it has little strength in the data.

Fit Statistics

The Rasch model "analyzes the data as though they are unidimensional, and then the fit statistics report how well the data match the mathematically unidimensional framework that the Rasch analysis has constructed" (Linacre 2011, para. 6). Therefore, the fit statistics report the degree to which the observations meet this vital specification of measurement. Smith (1996) emphasises that items (or persons) that do not fit the model "are not automatically rejected, but are examined to identify in what way, and why, they fall short ... Then the decision is made to accept, reject or modify the data" (p.516).

Linacre and Wright (1994) explain that the outfit statistic is dominated by unexpected outlying, off-target, low information responses and is outlier-sensitive. The infit statistic is an information-weighted sum, introduced to reduce the influence of outliers. It is dominated by unexpected inlying patterns among informative, on-target observations and is inlier-sensitive.

This Study

Research has shown that Internet use is highest among adolescents making this age group the most at risk of Internet dependence. Also, Internet addiction has been reported to be negatively associated with academic performance and grades (Chang & Law, 2008; Chou & Hsiao, 2000; Scherer & Bost, 1997; Yen et al., 2009). The Internet is nonetheless an important teaching and learning resource in education when used properly, and "an indelible feature of modern life" (Young, 1998b, p.1). As Greek Cypriot senior high school teachers, the researchers believe that instruments assessing Internet addiction among this specific student population should be developed.

Furthermore, many of the studies that have investigated Internet addiction reported a multi-factor structure for the construct. There is however, no universal agreement on the number of factors, or if indeed the factors identified were highly correlated possibly resulting in considering the scales used as unidimensional. If the factors are indeed highly correlated so that they could work together to form a single meaningful scale that measures Internet addiction, then the Rasch model would develop an equal-interval measure that would remain invariant (within standard error) for diagnosing the various levels of Internet addiction. According to Koronczai et al. (2011) there are very few psychometric data on the IAT, the most widely used Internet Test. Therefore, the purpose of this study was to evaluate the psychometric properties of the IAT for a sample of Cypriot adolescents through the investigation of the following four research questions:

1. Is the 5-point rating scale psychometrically optimal?

2. Does the IAT provide reliable measures? (the term ‘measures’ is used rather than ‘scores’ to distinguish between linear measures obtained from using the Rasch models and ordinal raw scores obtained from counting observed scores)
3. Do the 20 items define a theoretical linear continuum of increasing difficulty?
4. Do the 20 items define a single construct of Internet addiction?

Young’s IAT can be found online at: http://www.netaddiction.com

Methodology

Participants

The present study involved a total of 604 second and third grade senior high school students (ages 17-18) from five lyceums in Limassol, Cyprus. Four of the lyceums were selected at random (from a total of 10), the fifth being the one where both researchers are members of staff. Following comprehensive explanations of the purpose of the study, permission to administer questionnaires was sought and attained from the relevant head-teachers, all of whom were willing to offer their assistance.

The Instrument

Permission was also sought and attained from Dr. Kimberly Young for the use of her IAT for the purposes of this study. The researchers drew up a questionnaire comprising of 28 items. These were the original 20 items from Young’s IAT; a self diagnostic question; and a further seven questions of a personal nature such as gender, grades and sleep habits. The self diagnostic question asked students to rate the extent to which they thought they were addicted to the Internet on a 5-point scale (1 = none, 2 = a little extent, 3 = a moderate extent, 4 = a fair extent and 5 = a great extent). Widyanto et al. (2011) showed significant correlations of such a question with two Internet Addiction scales and argue that "participants are fairly accurate at evaluating their own level of problems with the Internet" (p. 148).

The questionnaire was translated from English into Greek by the researchers, and subsequently back into English by an independent and experienced English language expert who had not previously seen the original questionnaire. The two English versions were then compared by the researchers who concluded that the meaning of the items had not been altered in the translation.

Oral explanations related to the questionnaire were given to the teachers whose classes had been randomly selected. The researchers also explained the purpose of the study to the students and the voluntary basis for participation.

Selection of the Rasch Rating Scale Model (RSM)

The Rasch RSM was selected for the analysis of the IAT data for the following reasons. First, the Rasch models are the only models that accept the raw scores of the respondents to be a sufficient statistic for the estimation of their underlying position on the variable continuum thus maintaining the score order of students. Since raw scores are the basis for reporting results throughout all the studies on Internet addiction, the Rasch models are consistent with practice. Second, the Rasch models are easier to work with, to understand and to interpret, because they involve fewer parameters. Third, there are fewer parameter estimation problems than with the more general models. The Rasch models give stable item estimates with smaller samples than other models (Thissen & Wainer, 1982). Fourth, the person measures and item calibrations have a unique ordering on a common logit scale (Bond

& Fox, 2001, 2007; Wright & Masters, 1982) making it easy to see relations between them. The item-person map provided by the Rash software is very attractive to users. Fifth, validity and reliability issues can be addressed through the use of the Rasch models (Smith, 2004b).

Most importantly however, the Rasch model is based on a different philosophy from other approaches. This philosophy dictates the structure of the data including the fact that unidimensionality is a must for the measurement process. Other models are driven by a desire to model all of the characteristics observed in the data, regardless of whether they have any contribution to the measurement process. So, the difference is between measurement and modelling. If the aim is to construct a good measure then the items comprising the scale should be constrained to the principles of measurement, thus the Rasch model is highly appropriate.

Selection of the Fit Statistics

The infit mean square and the outfit mean square have been used to estimate the degree of misfit of the items in this study. These two fit statistics were preferred over a large number of fit statistics for their exploratory nature (Douglas, 1990). They can identify a wide range of potential sources of unexpected response patterns and this is an advantage in the sense that a fit statistic that focuses on a specific type of unexpectedness may not have enough power to identify other types, thus missing 'bad' items. Also, the infit and outfit mean squares have been used successfully to assess the fit of the Rasch models for many years (e.g. Curtis, 2004; Smith, 1990; Wright & Masters, 1982), and this encourages their use in the context of the Rasch models. Furthermore, these statistics are computationally simpler and they stand up well in comparison with possibly more precise tests, therefore there is no practical reason to use anything more complicated (Smith, 1990). Finally, they are utilized by most of the available software packages for Rasch calibrations (e.g. Quest, Winsteps, Facets) and are familiar to many researchers.

Critical Values for the Fit Statistics

Wright, Linacre, Gustafson, and Martin-Lof (1994) provide a table of reasonable item mean square fit values and suggest infit and outfit values of 0.6 – 1.4 for scales. Values of 1.4 indicate 40% more variability and values of 0.6 indicate 40% less variability than predicted by the Rasch model. Bond and Fox (2001, 2007) suggest the same values as Wright et al., whereas Curtis (2004) and Glas and Meijer (2003), suggest using simulated data according to an IRT model based on the estimated parameters and then determining the critical values empirically.

However, Lamprianou (2006) argues that misfit is not a dichotomous 'yes'/'no' property but rather a matter of degree and as such it can be considered too large for one study and satisfactory for another depending on the aims of the researchers. Therefore, for the purposes of this study, the researchers decided to consider items with infit or outfit greater than (the widely used cut-off value) 1.4 as ones needing re-examination before deciding to maintain or remove them from the scale, as suggested by Wright et al. (1994) and Bond and Fox (2001, 2007).

Pilot Study

The administration of the questionnaires was divided into two phases. In the first phase, the pilot study, the researchers investigated the appropriateness of the number of categories in the Likert scale used in the original IAT by administering 290 questionnaires to second and third grade senior high school students.

Rasch Diagnostics for the Optimal Number of Categories

Rating scale categories should be well defined, mutually exclusive and exhaustive. In practice the categories of a scale inevitably contain an element of arbitrariness and depend on whether the scale designer has done a good or poor job of the scale definition. The respondents may use the scale effectively (in an informative way) or ineffectively (in an uninformative way) according to their own understanding of the category labels. Wright and Linacre (1992) point out that it is the analyst's task to extract the maximum amount of useful meaning from the responses observed by combining (or even splitting), if necessary categories as suggested by the results of careful analysis. Furthermore Wright and Linacre (1992) advise researchers that in combining two or more categories they must be sure it is reasonable to do so and that both the statistical and substantive validity of the results is improved. Royal, Ellis, Ensslen, and Homan (2010) echo Wright's and Linacre's points and, even though they warn readers that sometimes collapsing categories can alter the meaning of the rating scale, in their study they did so thus improving rating scale optimization,

The researchers followed the Rasch measurement diagnostics suggested by Linacre (2002) and Bond and Fox (2001, 2007) for determining the optimal number of categories. First, categories with low frequencies (Linacre recommends 10 as the minimum number) are described as problematic because they do not provide enough observations for estimating stable threshold values. Second, the average measures (the average of the ability estimates of all persons in the sample who chose a particular category) are expected to increase monotonically in size as the variable increases. This indicates that on average, those with higher scores on the Internet addiction variable endorse the higher categories. Third, the thresholds, or step calibrations (the difficulties estimated for choosing one response category over another) should also increase monotonically across the rating scale. If they do not, they are considered disordered. Fourth, the magnitudes of the distances between adjacent threshold estimates should indicate that each step defines a distinct range on the variable. That is, the estimates should be neither too close together, nor too far apart. Linacre (1999) suggests that thresholds should increase by at least 1.4 logits, to show distinction between categories, but not more than 5 logits, so as to avoid large gaps in the variable. Step disordering and very narrow distances between thresholds "can indicate that a category represents too narrow a segment of the latent variable or corresponds to a concept that is poorly defined in the minds of the respondents" (Linacre, 2002, p. 98). Finally, the fit statistics provide another criterion for assessing the quality of a rating scale. Outfit greater than 2 indicates more misinformation than information, thus the category introduces noise into the measurement process.

Second Phase

In the second phase 314 questionnaires were administered giving a total of 604. Eight classes were selected from the researchers' school (with a population of 574 second and third graders) and five classes from each of the remaining four schools (with corresponding populations varying from 336 to 407) thus giving a proportional sample from the five schools. The total number of second and third graders in the five schools was 2093 and the sample was 28.9% of the population. Approximately 48.1% of the respondents were male and 51.9% female.

Combining the Two Samples

The scoring on the 20 items of the first 290 questionnaires was changed to 1 to 3 (as explained in the results section), thus changing the total scores. The means and standard deviations of the total scores of the questionnaires collected in the second phase were then compared with those from the changed scores of the first.

Furthermore, the correlation between the Rasch item calibrations from the two samples was calculated. Both investigations justified the combination of the two sub-samples into one larger sample thus giving more reliable results, smaller standard errors and more stable item estimates.

Unidimensionality

The dimensionality of the data was investigated through various studies, as suggested by Linacre (1998). First item correlations with the total scores were calculated; second, dimensionality was examined through principal components analysis (PCA) of the standardised residuals; third, fit statistics were calculated.

The meaningfulness of the item ordering was investigated through comparisons with the item ordering as derived through the opinions of four experts (three high school student consultants-career advisors, all with psychology degrees or training and one independent psychologist). The experts had to rate the difficulty of each IAT item on a scale from 0 to 4, where 0 was the easiest item and 4 the hardest. Comparisons were carried out with the use of two correlation coefficients, the product moment correlation coefficient (r) and Spearman's rank correlation coefficient (rho). Rho assesses how well the item order is maintained among the two orderings.

The stability of the item ordering was investigated through comparisons of item calibrations from two groups of students, the higher and lower scorers.

Reliability Indices

The person estimate reliability (R_p) is an indication of the precision of the instrument and shows how well the instrument can distinguish individuals. It can often be replaced by a person separation index (G_p) which ranges from 0 to infinity and indicates the spread of person measures in standard error units. Another useful calculation is that of strata calculated by $[(4G_p + 1)/3]$. Strata are used to determine the number of statistically distinct levels, separated by at least 3 errors of measurement, of person ability that the items have distinguished (Wright & Masters, 1982).

Finally, the item estimate reliability shows how well the items that form the scale are discriminated by the sample of respondents. Wright and Masters (1982) argue that good item separation is a necessary condition for effective measurement.

All Rasch analyses were performed on WINSTEPS (Linacre, 2005).

Results

Pilot Study – Rating Scale Functioning

The data collected from the 290 questionnaires were analysed with emphasis, at this stage on the Rasch diagnostics for the optimal number of categories. Table 1 shows these diagnostics for the original scale with the five categories.

There is a large number of observations in each category (minimum 420, in category 5), the average measure increases monotonically (-1.58, -0.92, -0.35, 0.02 and 0.53 for categories 1, 2, 3, 4 and 5 respectively) and the outfit values are all close to 1 (from 0.85 to 1.20).

Table 1

Summary of Category Structure

Category labels	Observed count (%)	Average Measure	Infit mnsq	Outfit mnsq	Step calibrations
1	2446 (43%)	-1.58	1.06	1.16	None
2	1194 (21%)	-0.92	0.85	0.80	-0.51
3	1036 (18%)	-0.35	0.90	0.89	-0.51
4	643 (11%)	0.02	0.99	1.11	0.30
5	420 (7%)	0.53	1.13	1.20	0.72

However, categories 2 and 3 are disordered. The threshold between categories 1 and 2 is the same as between categories 2 and 3 (-0.51). Also the distance between the first and the last thresholds is only 1.23 logits which perhaps indicates that the scale should have only 3 categories.

One visual and perhaps easier method of inspecting the distinction between thresholds is to examine the probability curves. These curves show the probability of endorsing a given category for every 'person agreeability minus item endorsability' (Ability-Difficulty) estimate. Figure 1 shows the probability curve for the original scale.

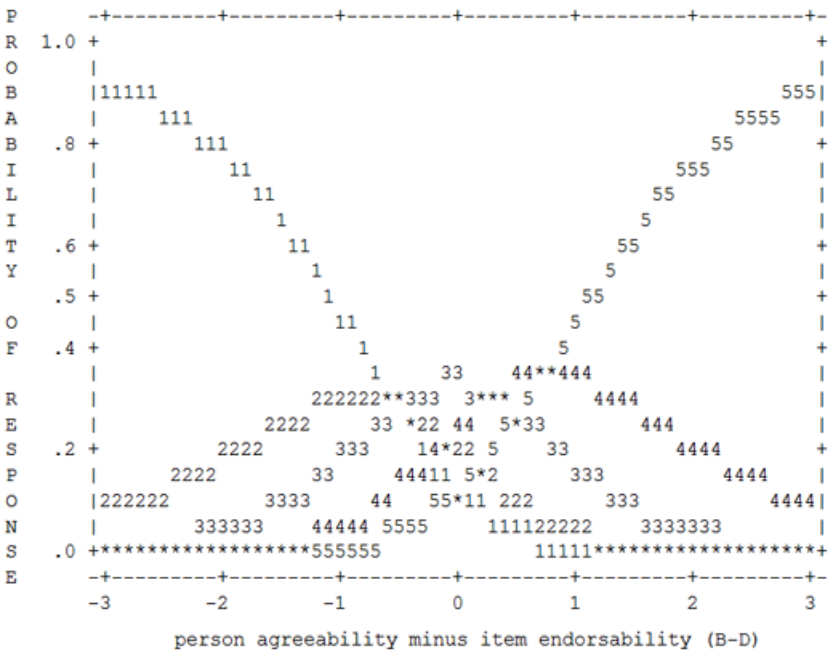

Figure 1. *Category Probability Curves*

Each category should have a distinct peak in the probability curve graph, illustrating that each is indeed the most probable response category for some portion of the measured variable. In this case category 2 never emerges and categories 3 and 4 only peak for a very small range of the variable.

This pattern suggests the need to reconsider both the number of and the corresponding labels of the response options. This led to the collapsing of categories and the use of two different models, first the 12234 model (collapsing categories 2 and 3) and then the 12334 model (collapsing categories 3 and 4).

In the 12234 model, category 3 did not peak at all and the thresholds were again disordered (-1.41, 0.71 and 0.70). Similarly, in the 12334 model, category 2 did not peak and the thresholds were also disordered (-0.69, -0.94 and 1.63).

The above findings led to further collapsing the 4 categories into 3. The final model used was the 12223 model, where categories 2, 3 and 4 of the original model were combined into 1 category. Figure 2 shows the probability curve of this final analysis.

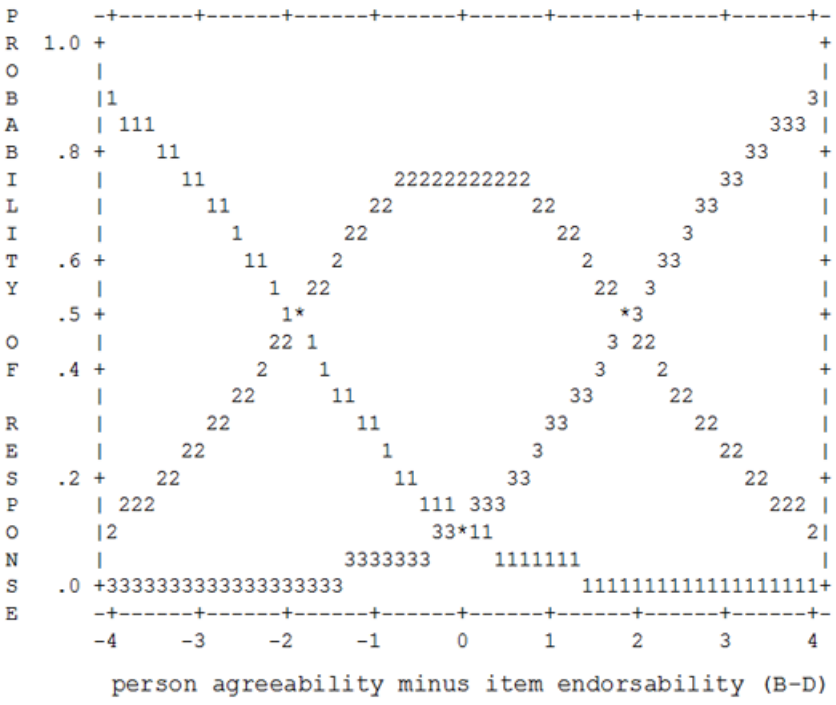

Figure 2. *Category Probabilities, Model 12334*

There is a large number of observations in each category, the observed average measure is monotonically increasing (-2.68, -0.88 and 0.86) and the outfit values for all categories are all very close to 1 (1.04, 0.94 and 1.03). Most importantly however, the thresholds are not disordered. They are now monotonically increasing (-1.91 and 1.91) and there is a distance of 3.82 logits amongst these thresholds and this distance is well inside the optimal range. Furthermore each category peaks in a distinct range illustrating that each is indeed the most probable response category for that distinct range of the measured variable. Finally, Table 2 shows the reliability estimates for each of the models investigated.

There are no differences in the reliability indices among the four models. However, the first three models have disordered categories. These analyses suggest that the original 5-point Likert scale (rarely, occasionally, frequently, often, always) is not optimal for this sample. Instead, a 3-point Likert scale (rarely, frequently, always) should be used. It seems that the distinction between “occasionally”, “frequently” and “often” was not clear in the minds of the respondents and therefore the three categories were combined into one labelled “frequently”. Such combinations of categories can be found in the survey of perceived fears by Stone and Wright (1994). They showed that combining five ordered categories into three increased the test reliability for the sample. In another study, on the

Table 2

Reliability Estimates for Each Model.

	Models			
	12345	**12234**	**12334**	**12223**
Person Reliability	0.87	0.87	0.88	0.87
Separation	2.61	2.61	2.69	2.54
Strata	3.81	3.81	3.92	3.72
Categories	Disordered	Disordered	Disordered	Ordered

evaluation of the diabetes self-care scale, Lee and Fisher (2005) found that a 3-point rating scale was optimal instead of the original 6-point rating scale. Similarly Schulman and Wolfe (2000) found that the seven original categories represented more levels of the self-efficacy variable than the respondents were capable to distinguish and decided that the optimal number was five.

Comparing the Data Collected From the Two Phases

Table 3 shows the results of the statistical tests for differences between the means and standard deviations of the IAT scores from the two phases.

Table 3

Comparisons Between Total Addiction Scores

		Standard Deviations			Means		
	N	**S.D.**	**F**	**p-value**	**Mean**	**t**	**p-value**
Pilot study	290	6.65			32.80		
Phase 2	314	7.58	3.405	0.066	31.98	1.407	0.160

The F-test revealed no differences between the standard deviations ($p = 0.066$) and the t-test no differences between the means ($p = 0.160$).

Furthermore, the two sets of Rasch item calibrations had a correlation of 0.983 ($n = 20$, $p < 0.005$). The non-significant statistical tests and the highly significant correlation between the item calibrations justify the combination of the two samples into one (as in Lee & Fisher 2005).

Investigating the Dimensionality of the Scale

Table 4 shows the item statistics of the Rasch analyses in misfit order.

All item-total correlations are positive and significant ranging from 0.43 to 0.65. At the same time all the items fit the Rasch model very well (except from item 7 which has a marginally higher outfit value of 1.48).

Table 5 shows the results of the PCA of the standardised residuals and Figure 3 the resulting plot of the first factor extracted.

Table 4

Item Statistics in Misfit Order

Items	Item measure	Model S.E.	Infit mnsq	Outfit mnsq	Correlation
7	-0.21	0.08	1.39	1.48	0.43
9	0.38	0.08	1.23	1.29	0.47
4	0.32	0.08	1.19	1.24	0.45
3	1.54	0.10	1.09	1.22	0.45
5	-0.77	0.08	1.02	1.18	0.60
12	-0.63	0.08	1.17	1.16	0.59
10	-0.22	0.08	1.00	1.10	0.52
18	0.86	0.09	1.08	0.95	0.55
17	-0.06	0.08	1.06	1.05	0.55
19	1.60	0.10	1.04	1.00	0.51
6	0.21	0.08	0.98	0.93	0.59
20	1.14	0.09	0.95	0.80	0.58
15	1.02	0.09	0.94	0.91	0.59
13	-0.28	0.08	0.92	0.90	0.63
14	-0.40	0.08	0.91	0.90	0.61
16	-1.39	0.07	0.88	0.88	0.62
8	0.20	0.08	0.87	0.84	0.61
11	-0.63	0.08	0.81	0.81	0.65
1	-1.83	0.07	0.79	0.80	0.59
2	-0.84	0.07	0.79	0.78	0.63
Mean	0.00	0.08	1.01	1.01	
S.D.	0.90	0.01	0.15	0.19	

Table 5

Standardized Residual Variance (in Eigenvalue Units)

	Empirical	(%)		Modeled (%)
Total raw variance in observations	33.8	100.0		100.0
Raw variance explained by measures	13.8	40.8		40.9
Raw variance explained by persons	9.0	26.6		26.6
Raw Variance explained by items	4.8	14.2		14.2
Raw unexplained variance (total)	20.0	59.2	100.0%	59.1
Unexplained variance in 1st factor	1.8	5.5	9.2%	

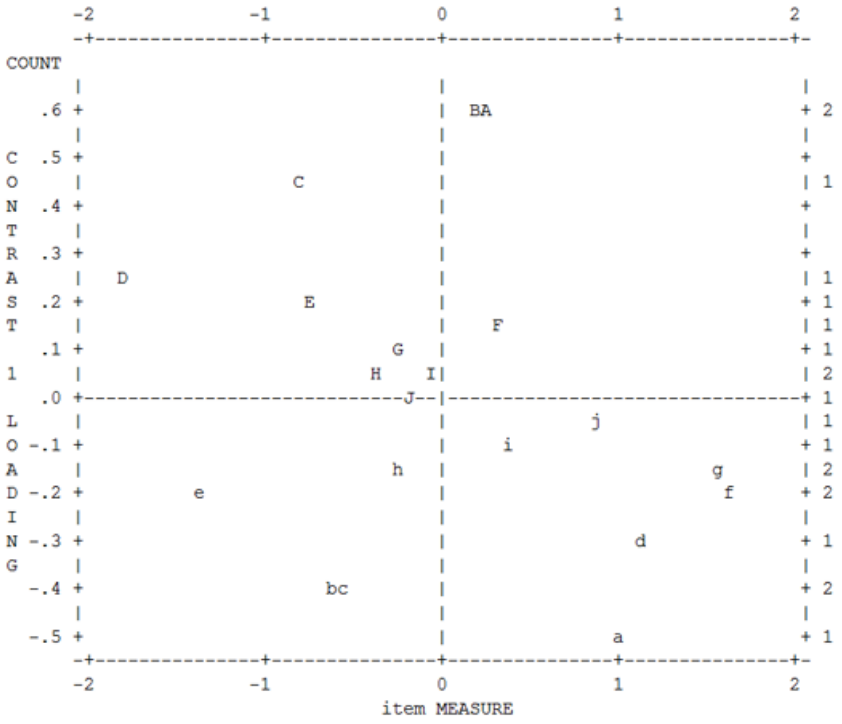

Figure 3. *Standardized Residual-Factor 1 Plot*

To judge the strength of the measurement dimension, the researchers looked at the variance explained by the measure. It was found to be 40.8% of the total variance in the data (eigenvalue 13.8). The first factor has an eigenvalue of 1.8 and the strength of less than two items. Also the variance explained by the first factor is 9.2% of the unexplained variance and only 5.5% of the total variance.

The figure shows the item loadings on the first factor against item measures. The two items with the highest loadings on this factor, items 6 and 8, are very close together. Further investigation was undertaken on these items and Table 6 shows their statistics.

Table 6

Item 6 – Item 8 Statistics

Statistics	Item 6	Item 8
1st factor loading	0.58	0.60
Item measure	0.21	0.20
Standard error	0.08	0.08
Item-total correlation	0.59	0.61

The statistics of the two items are almost identical. Further inspection revealed that the wording of the two items was semantically indistinguishable for the students. The two items were:

Item 6: How often do your grades or school work suffer because of the amount of time you spend on-line?

Item 8: How often does your performance or productivity suffer because of the Internet?

Item 6 was, in the researchers' opinion, clearer. Performance and productivity in the minds of high school students relates to school work.

Item 8 was therefore removed and Rasch analyses were performed on the 19-item IAT. After these analyses the item "To what extent do you think you are addicted to the Internet?" was added to the scale making it into a 20-item scale again. To be consistent, the 5 categories of the Likert scale of this item were changed into three by combining, as with the other items, categories 2, 3 and 4 into one thus changing the scoring from to 1 to 3. Therefore the final IAT consisted of 20 items each with three options. The extra item was added to the scale for three reasons. First, its correlation with the total score of the original 20 items was high ($r = 0.658$, $p < 0.01$). Second, there was no question in the scale requiring the respondents to self-rate the extent of their possible Internet addiction level, and third to make the results of these analyses comparable with results from other studies. Table 7 shows the results of the analyses of the three different versions of the IAT used: the original 20-item, the 19-item and the modified 20-item scale.

Table 7

Results of Analyses of the Three Scales.

	Original 20 items	19 items	Modified 20 items
Person Reliability	0.86	0.85	0.86
Separation	2.47	2.40	2.48
Strata	3.63	3.53	3.64
Variance by Measures	13.8 (40.8%)	13.3 (41.1%)	14 (41.1%)
Unexplained Var.	20 (59.2%)	19 (58.9%)	20 (58.9%)
1st factor			
Eigenvalue	1.8	1.7	1.7
% of unexplained var.	9.2%	8.9%	8.5%
% of Total variance	5.5%	5.2%	5.0%
Item 7:			
Infit	1.39	1.37	1.41
Outfit	1.48	1.47	1.48

The three versions of the scale are almost identical statistically. The researchers decided that the modified IAT was the most favourable because it does not include two items with the same content and a 20-item scale is more preferable than a 19-item one for the purpose of comparisons with other studies. More importantly perhaps the dimensionality investigation is slightly more convincing for the last scale. Even though the eigenvalues of the first factor extracted in all cases are less than 2 (showing strength of less than two items and suggesting no presence of a second dimension) the percentages of variance explained by the first factor are slightly smaller (8.5% of the unexplained and only 5.0% of the total variance). Finally, the ratio of variance explained by the measures to variance explained by the first factor was 8.2:1.

The fit of the items to the model were very good (infit mean value = 1.01 and outfit mean value = 1.01) with only item 7 having infit = 1.41 and outfit = 1.48. Item 7 was "How often do you check your email before something else you need to do?"

Further investigation revealed that the marginal misfit was caused by unexpectedly high responses by four low scorers. Once the responses of those four students were removed from the dataset the infit and outfit values of

item 7 dropped to 1.36 and 1.37 respectively, under the cut-off value of 1.4. The item was therefore not removed because it was only marginally misfitting and its misfit was caused by only four unexpected responses.

Reliability Indices

The person reliability was high at 0.86 (Cronbach's alpha was 0.89) and the separation was 2.48. This separation indicates that the instrument identifies approximately four (3.64) statistically distinct strata of Internet Addiction levels. Furthermore the item reliability was 0.99 indicating that the items are discriminated very well by the sample of respondents and the item separation was 11.07 meaning that the spread of items is about 11 standard errors.

Item Person Map

Figure 4 shows an item-person map slightly different from the WINSTEPS output map.

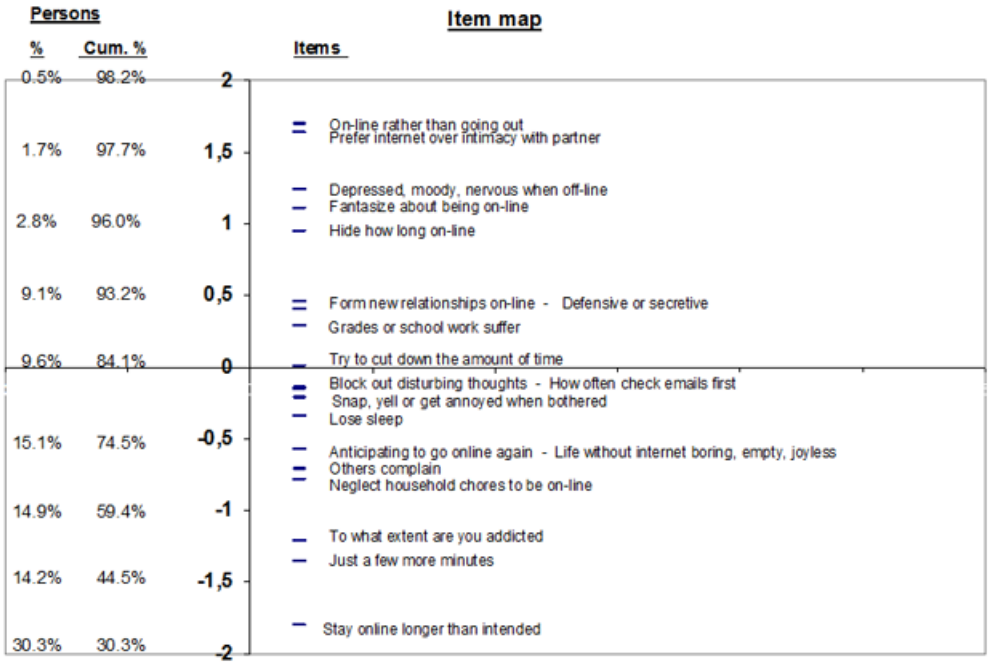

Figure 4. *Item-Person Map*

On the right of the continuum (the logit scale) the item hierarchy is displayed. Item calibrations range from -1.80 to 1.70 logits and they are evenly spread with nine of them above the average item measure (0.0) and 11 below. This spread of items shows a good coverage of the construct under investigation.

The item hierarchy shows that the items relating to preferring the Internet over going out (item 19, measure 1.70) or being intimate with their partner (item 3, measure 1.64) are the most difficult to endorse. The items about staying, (item 1, measure -1.80) or wanting to stay (item 16, measure -1.35) longer than intended, together with their rating of their own level of Internet addiction were the easiest to endorse. The order of the items is in good agreement with the ordering of the experts who rated item 19 and item 3 as the most difficult (mean difficulty 3 and 2.75 respectively) and item 16 as the easiest (mean difficulty 0.5) and item 1 as the third easiest (mean difficulty 1). The correlation coefficient between the experts' ratings and the item difficulties was 0.89 ($p < 0.005$) and rho was 0.92 ($p < 0.005$).

On the right of the continuum the percentages and cumulative percentages of the students with various addiction levels are displayed. Thirty percent of them have a measure below -2. The two percentages are displayed for every 0.5 logits. For example, 15.1% of the students have a measure between -1.0 and -0.5 and 74.5% of them have a measure of -0.5 or lower. The spread of person measures varies from -6.06 to 6.08 (mean -1.31 and S.D. 0.49). One important result is that approximately 84% of the students have a measure below 0 logits and only 16% above. This indicates that the 20-item scale is a little off-target. However, the researchers believe that this is not a disadvantage of the instrument; this result was expected by the researchers because the scale is designed to identify Internet addicts and the percentage of students addicted to the Internet is low (as reported in other studies). The spread of items however is very good on the continuum of the construct which seems to be well defined by the items.

To investigate the stability of the item ordering the respondents were divided into two equal-sized groups, the 302 students with the highest measures and the 302 students with the lowest measures. The item estimates of the two groups were then plotted. Figure 5 is the plot of the two estimates.

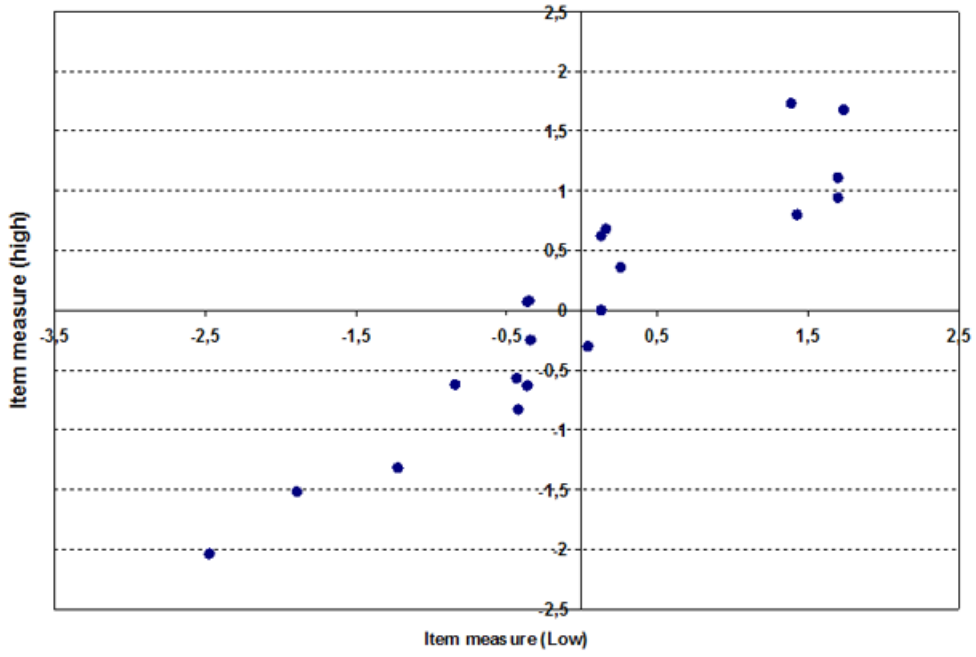

Figure 5. *Plot of Item Estimates*

The correlation between the two item calibrations was 0.94 ($p < 0.005$) and rho was 0.93 ($p < 0.005$), both very high and supportive of the invariant structure of the IAT.

Investigating the Correlation of Person Measures With Other Variables

Table 8 shows the correlation of person measures with other variables.

Table 8

Correlations of Variables with Person Measures

	Correlation	p-value
IAT Total score	0.972	0.000
Monthly family income	0.026	0.557
Average Grade	-0.134	0.001
Hours of sleep – weekdays	-0.236	0.000
Hours of sleep – weekend	-0.121	0.003

There is a very high correlation (0.972) between the person measures and the raw scores. There is no association between person measures and monthly family income and significant negative correlations between person measures and average grade (the academic performance variable), hours of weekday sleep and hours of weekend sleep.

Discussion

The main objective of this study was to evaluate the psychometric properties of the IAT with the use of the Rasch RSM. The translated version of the IAT (into Greek) was administered to a random sample of 604 students from 5 (out of the ten) lyceums in Limassol, Cyprus.

Sechrest, Fay, and Hafeez Zaidi (1972) emphasised the importance of "equivalence in terms of experiences and concepts" (p. 41) when translating questionnaires. Despite the researchers' efforts to achieve this, statistical analyses suggest that items 6 and 8 were impossible for the students to semantically differentiate. Retaining items with identical statistics, and in this case identical meaning too, entails the risk of inflating reliability. Therefore item 8 was removed and the subjective item "To what extent do you think you are addicted to the Internet?" was added. This modified version of the IAT was used for the final analyses.

Research Question 1: Is the 5-Point Rating Scale Psychometrically Optimal?

The 5-point rating scale was not found to be psychometrically optimal. Results from the pilot study showed that the students were unable to distinguish between the Greek equivalents of the original IAT categories "occasionally", "frequently" and "often". Therefore, analyses showed that collapsing the three middle categories into one, labelled "frequently", gave the optimal number of categories which was three.

The researchers cannot tell whether this change from a 5-point to a 3-point rating scale was necessary as a result of possible semantic obstacles encountered through the translation, as suggested by Sechrest et al. (1972), or due to problems with the original construction of the 5-point scale.

Research Question 2: Does the IAT Provide Reliable Measures?

Findings of this study support the high degree of the reliability of the measures produced by the IAT. Reliability indices for the modified 20-item version of the IAT were 0.86, 2.48 and 3.64 for person reliability, person separation and strata respectively. Furthermore item reliability was 0.99 and item separation 11.07. This good item separation is supportive of effective measurement.

Research question 3: Do the 20 Items Define a Theoretical Linear Continuum of Increasing Difficulty?

The 20 items were evenly spread along the linear continuum with a range of difficulties from -1.80 to 1.70 logits. The item hierarchy created by the item calibrations forms a ladder with even steps of easier to endorse items on the bottom and harder to endorse on the top.

This item hierarchy was meaningful and in agreement (highly significant correlations) with the item ordering resulting from the experts' opinions. Also, the stability of the item hierarchy was supported by the highly significant correlations between the item calibrations from two equally sized distinct groups: the higher scorers and the lower scorers.

Finally the highly satisfactory item reliability of 0.99 indicates a good separation of the 20 items along the variable which they define. It is therefore safe to conclude that indeed the item calibrations are sufficiently spread out to define distinct levels along the variable and the 20 items do define a linear continuum of increasing difficulty.

Research Question 4: Do the 20 Items Define a Single Construct of Internet Addiction?

For the dimensionality and the construct validity investigation of the scale the following evidence was collected.

- Item-total correlations were all highly significant (0.43 to 0.65).
- All items fitted the Rasch model well with infit and outfit mean square values below the cut-off score of 1.4.
- PCA of the standardised residuals showed that the variance explained by the measures was 41.1%.
- More importantly however, the first factor extracted after the contribution of the measures to the data had been removed, had an eigenvalue of 1.7 and this shows the strength of less than two items.
- The variance explained by the first factor was 8.5% of the unexplained variance and only 5.0% of the total variance.
- The ratio of variance explained by the measures to variance explained by the first factor was 8.2:1.
- The item hierarchy was in agreement with the order derived through the experts' opinions.
- The correlations of the item calibrations derived from the analyses from two distinct groups of respondents were highly significant supporting the invariant structure of the IAT and the fact that the construct has the same meaning across the two groups.

Finally significant negative correlations of the person measures were found with students' average grade (as a measure of academic performance), as reported by Chang and Law (2008), Lay (1988), Chou and Hsiao (2000), Scherer and Bost (1997) and Yen et al. (2009) and with the number of hours of weekday and weekend sleep, as reported by Choi et al. (2009), Kesici and Sahin, (2010).

All the evidence collected support the unidimensional structure of the IAT and its high degree of construct validity.

Limitations

The sample of 604 high school students is large enough for reliable results but generalization to the whole population of Cyprus is risky since the sample can only be representative of the population from which it was drawn, namely high school students of Limassol.

Furthermore, despite the efforts of the researchers for an accurate translation of the instrument, they cannot rule out the possibility of problems with the "equivalence in terms of experiences and concepts" (Sechrest et al., 1972, p. 41).

Recommendations

Based on the results and limitations of this study the following five recommendations are made:

- To remove item 8 “How often does your performance or productivity suffer because of the Internet?” and to add “To what extent do you think you are addicted to the Internet?”
- To replace the 5-point rating scale with a 3-point one since the latter was found to be psychometrically optimal.
- To further evaluate the modified IAT with a more representative sample of the overall population of Cypriot high school students.
- To evaluate the psychometric properties of the original IAT, in English, using the Rasch model, with emphasis on the number of categories of the rating scale. The researchers maintain that the three middle categories may not be clearly distinguishable in the minds of high school students in other countries too.
- To transform the person logit measures into a more convenient, easier to interpret and with non-negative values scale. They could be transformed to a scale from 1 to 20, the widely used grading scale in the educational system in Cyprus or to a scale from 1 to 100, the most widely used scale internationally.

Concluding Remark

The successful fit of the modified IAT data to the Rasch model, the model of fundamental measurement, provides support that Internet addiction is a rigorously quantitative and unidimensional variable and that the IAT has a high degree of validity.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573. doi:10.1007/BF02293814

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental measurement in the human sciences* (1st ed.). Mahwah: Lawrence Erlbaum Associates.

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

Brenner, V. (1997). Psychology of computer use XLVII. Parameters of Internet use, abuse and addiction: The first 90 days of the Internet usage survey. *Psychological Reports, 80(3),* 879-882. doi:10.2466/pr0.1997.80.3.879

Caplan, S. E. (2002). Problematic Internet use and psychosocial well-being: Development of a theory-based cognitive-behavioral measurement instrument. *Computers in Human Behavior, 18(5),* 553-575. doi:10.1016/S0747-5632(02)00004-3

Chang, M. K., & Law, S. P. M. (2008). Factor structure for Young’s Internet Addiction Test: A confirmatory study. *Computers in Human Behavior, 24,* 2597-2619. doi:10.1016/j.chb.2008.03.001

Chen, S. P., Bezruczko, N., & Ryan-Henry, S. (2006). Rasch analysis of a new construct: Functional caregiving for adult children with intellectual disabilities. *Journal of Applied Measurement, 7(2),* 141-159.

Choi, K., Son, H., Park, M., Han, J., Kim, K., Lee, B., & Gwark, H. (2009). Internet overuse and excessive daytime sleepiness in adolescents. *Psychiatry and Clinical Neurosciences, 63,* 455-462. doi:10.1111/j.1440-1819.2009.01925.x

Chou, C., & Hsiao, M.-C. (2000). Internet addiction, usage, gratification, and pleasure experience: The Taiwan college students’ case. *Computers & Education, 35(1),* 65-80. doi:10.1016/S0360-1315(00)00019-1

Curtis, D. D. (2004). Person misfit in attitude surveys: Influences, impacts and implications. *International Education Journal, 5(2),* 125-144.

Davis, R. A. (2001). A cognitive-behavioral model of pathological Internet use. *Computers in Human Behavior, 17,* 187-195. doi:10.1016/S0747-5632(00)00041-8

Davis, R. A., Flett, G. L., & Besser, A. (2002). Validation of a new scale for measuring problematic internet use: Implications for pre-employment screening. *CyberPsychology & Behavior, 5,* 331-345. doi:10.1089/109493102760275581

Douglas, G. A. (1990). Response patterns and their probabilities. *Rasch Measurement Transactions*, 3(4), 75. Retrieved from http://www.rasch.org/rmt/rmt34a.htm

Ferraro, G., Caci, B., D'Amico, A., & Di Blasi, M. (2007). Internet Addiction Disorder: An Italian Study. *CyberPsychology & Behavior, 10(2),* 170-175. doi:10.1089/cpb.2006.9972

Fu, K.-w., Chan, W. S. C., Wong, P. W. C., & Yip, P. S. F. (2010). Internet addiction: Prevalence, discriminant validity and correlates among adolescents in Hong Kong. *The British Journal of Psychiatry, 196,* 486-492. doi:10.1192/bjp.bp.109.075002

Glas, C. A. W., & Meijer, R. R. (2003). A Bayesian approach to person fit analysis in item response theory models. *Applied Psychological Measurement, 27(3),* 217-233. doi:10.1177/0146621603027003003

Goldberg, I. (1996, July). *Internet addiction disorder* [Electronic mailing list message]. Retrieved from http://users.rider.edu/~suler/psycyber/supportgp.html

Griffiths, M. (2000). Internet Addiction – Time to be taken seriously? *Addiction Research, 8(5),* 413-418. doi:10.3109/16066350009005587

Jang, K. S., Hwang, S. Y., & Choi, J. Y. (2008). Internet addiction and psychiatric symptoms among Korean adolescents. *The Journal of School Health, 78(3),* 165-171. doi:10.1111/j.1746-1561.2007.00279.x

Jia, R., & Jia, H. H. (2009). Factorial validity of problematic Internet use scales. *Computers in Human Behavior, 25,* 1335-1342. doi:10.1016/j.chb.2009.06.004

Johansson, A., & Götestam, K. G. (2004). Internet addiction: Characteristics of a questionnaire and prevalence in Norwegian youth (12-18 years). *Scandinavian Journal of Psychology, 45,* 223-229. doi:10.1111/j.1467-9450.2004.00398.x

Kandell, J. J. (1998). Internet addiction on campus: The vulnerability of college students. *CyberPsychology & Behavior, 1,* 11-17. doi:10.1089/cpb.1998.1.11

Kesici, S., & Sahin, I. (2010). Turkish adaptation study of Internet Addiction Scale. *Cyberpsychology, Behavior, and Social Networking, 13(2),* 185-189. doi:10.1089/cyber.2009.0067

Kline, P. (2000). *An easy guide to factor analysis*. London: Routledge.

Koronczai, B., Urbán, R., Kökönyei, G., Paksi, B., Papp, K., Kun, B., ... Demetrovics, Z. (2011). Confirmation of the three-factor model of problematic Internet use on off-line adolescent and adult samples. *Cyberpsychology, Behavior, and Social Networking, 14(11),* 657-664. doi:10.1089/cyber.2010.0345

Lam-Figueroa, N., Contreras-Pulache, H., Mori-Quispe, E., Nizama-Valladolid, M., Gutiérrez, C., Hinostroza-Camposano, W., ... Hinostroza-Camposano, W. D. (2011). Adicción a Internet: Desarrollo y validación de un instrumento en escolares

adolescentes de Lima, Perú. *Revista Peruana de Medicina Experimental y Salud Pública, 28(3),* 462-469. doi:10.1590/S1726-46342011000300009

Lamprianou, I. (2006). The stability of marker characteristics across tests of the same subject and across subjects. *Journal of Applied Measurement, 7(2),* 192-205.

Law, S. P. M., & Chang, M. K. (2007) *Factor structure for the Internet Addiction Test: A confirmatory approach*. Paper presented at the International DSI & APDSI Conference 2007, Bangkok, Thailand. Retrieved from: http://iceb.nccu.edu.tw/proceedings/APDSI/2007/papers/Final_82.pdf

Lay, C. H. (1988). The relationship of procrastination and optimism to judgments of time to complete an essay and anticipation of setbacks. *Journal of Social Behavior and Personality, 3,* 201-214.

Lee, N. P., & Fisher, W. P., Jr. (2005). Evaluation of the Diabetes Self-Care Scale. *Journal of Applied Measurement, 6(4),* 366-381.

Lin, S. S. J., & Tsai, C.-C. (2002). Sensation seeking and internet dependence of Taiwanese high school adolescents. *Computers in Human Behavior, 18(4),* 411-426. doi:10.1016/S0747-5632(01)00056-5

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement, 2(3),* 266-283.

Linacre, J. M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement, 3(2),* 103-122.

Linacre, J. M. (2002). Understanding Rasch Measurement. Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3(1),* 85-106.

Linacre, J. M. (2005). WINSTEPS Rasch measurement computer program (3.65) [Computer software]. Chicago: Winsteps.com.

Linacre, J. M. (2011). Rasch Measurement and Unidimensionality. *Rasch Measurement Transactions, 24*(4), 1310. Retrieved from http://www.rasch.org/rmt/rmt82a.htm

Linacre, J. M., & Wright, B. D. (1994). Chi-square fit statistics. *Rasch Measurement Transactions, 8*(2), 360. Retrieved from http://www.rasch.org/rmt/rmt82a.htm

Massof, R. W., & Fletcher, D. C. (2001). Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. *Vision Research, 41(3),* 397-413. doi:10.1016/S0042-6989(00)00249-2

Morahan-Martin, J. M., & Schumacher, P. (2000). Incidence and correlates of pathological Internet use among college students. *Computers in Human Behavior, 16,* 13-29. doi:10.1016/S0747-5632(99)00049-7

Myford, C. M., & Wolfe, E. W. (2002). When raters disagree, then what: Examining a third rating discrepancy resolution procedure and its utility for identifying unusual patterns of ratings. *Journal of Applied Measurement, 3(3),* 300-324.

Odacı, H., & Kalkan, M. (2010). Problematic Internet use, loneliness and dating anxiety among young adult university students. *Computers & Education, 55(3),* 1091-1097. doi:10.1016/j.compedu.2010.05.006

Öztürk, O., Odabasioglu, G., Eraslan, D., Genç, Y., & Kalyoncu, Ö. A. (2007). Internet addiction: Clinical aspects and treatment strategies. *Bağımlılık Dergisi, 8,* 36-41.

Pallanti, S., Bernardi, S., & Quercioli, L. (2006). The Shorter PROMIS Questionnaire and the Internet Addiction Scale in the assessment of multiple addictions in a high-school population: Prevalence and related disability. *CNS Spectrums, 11(12),* 966-974.

Panayides, P., Robinson, C., & Tymms, P. (2010). The assessment revolution that has passed England by: Rasch measurement. *British Educational Research Journal, 36(4),* 611-626. doi:10.1080/01411920903018182

Prieto, L., Roset, M., & Badia, X. (2001). Rasch measurement in the assessment of growth hormone deficiency in adult patients. *Journal of Applied Measurement, 2(1),* 48-64.

Raiche, G. (2005). Critical eigenvalue sizes in standardized residual principal components analysis. *Rasch Measurement Transactions, 19*(1), 1012. Retrieved from http://www.rasch.org/rmt/rmt191h.htm

Royal, K. D., Ellis, A., Ensslen, A., & Homan, A. (2010). Rating scale optimization in survey research: An application of the Rasch Rating Scale Model. *Journal of Applied Quantitative Methods, 5(4),* 586-596.

Scherer, K., & Bost, J. (1997). *Internet use patterns: Is there internet dependency on campus?* Paper presented at the 105th Annual Convention of the American Psychological Association, Chicago, IL.

Schulman, J. A., & Wolfe, E. W. (2000). Development of a nutrition self-efficacy scale for prospective physicians. *Journal of Applied Measurement, 1(2),* 107-130.

Schumacker, R. E., & Linacre, J. M. (1996). Factor Analysis and Rasch. *Rasch Measurement Transactions. 9*(4), 470 Retrieved from http://rasch.org/rmt/rmt94k.htm

Sechrest, L., Fay, T. L., & Hafeez Zaidi, S. M. (1972). Problems of translation in cross-cultural research. *Journal of Cross-Cultural Psychology, 3(1),* 41-56. doi:10.1177/002202217200300103

Shaw, M., & Black, D. W. (2008). Internet addiction: Definition, assessment, epidemiology and clinical management. *CNS Drugs, 22(5),* 353-365. doi:10.2165/00023210-200822050-00001

Siomos, K. E., Dafouli, E. D., Braimiotis, D. A., Mouzas, O. D., & Angelopoulos, N. V. (2008). Internet addiction among Greek adolescent students. *CyberPsychology & Behavior, 11(6),* 653-657. doi:10.1089/cpb.2008.0088

Smith, E. V., Jr. (2004a). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal components analysis of residuals. In E. V Smith Jr. & R. M. Smith (Eds), *Introduction to Rasch Measurement* (pp. 575-599). Minnesota: JAM Press.

Smith, E. V., Jr. (2004b). Evidence for the reliability of measures and validity of measure interpretations: A Rasch measurement perspective. In E. V Smith Jr. & R. M. Smith (Eds), *Introduction to Rasch Measurement* (pp. 93-122). Minnesota: JAM Press.

Smith, R. M. (1990). Theory and practice of fit. *Rasch Measurement Transactions, 3*(4), 78. Retrieved from http://www.rasch.org/rmt/rmt34b.htm

Smith, R. M. (1996). Polytomous mean-square fit statistics. *Rasch Measurement Transactions, 10*(3), 516-517. Retrieved from http://www.rasch.org/rmt/rmt103a.htm

Smith, R. M., & Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective measurement: Theory into practice (Vol. 2, pp. 316-327).* Norwood, NJ: Ablex.

Stone, M. H., & Wright, B. D. (1994). Maximizing rating scale information. *Rasch Measurement Transactions, 8*(3), 386. Retrieved from http://www.rasch.org/rmt/rmt83r.htm

Thissen, D., & Wainer, H. (1982). Some standard errors in item response theory. *Psychometrika, 47,* 397-412. doi:10.1007/BF02293705

Widyanto, L., & Griffiths, M. (2006). 'Internet addiction': A critical review. *International Journal of Mental Health and Addiction, 4,* 31-51. doi:10.1007/s11469-006-9009-9

Widyanto, L., Griffiths, M. D., & Brunsden, V. (2011). A psychometric comparison of the Internet Addiction Test, the Internet-Related Problem Scale, and self-diagnosis. *Cyberpsychology, Behavior, and Social Networking, 14(3),* 141-149. doi:10.1089/cyber.2010.0151

Widyanto, L., & McMurran, M. (2004). The psychometric properties of the Internet Addiction Test. *Cyberpsychology, Behavior, and Social Networking, 7(4),* 443-450.

Wright, B. D., & Linacre, J. M. (1992). Combining and splitting categories. *Rasch Measurement Transactions, 6*(3), 233-235. Retrieved from http://www.rasch.org/rmt/rmt63f.htm

Wright, B. D., Linacre, J. M., Gustafson, J.-E., & Martin-Lof, P. (1994). Reasonable mean square fit values. *Rasch measurement transactions, 8*(3), 370. Retrieved from http://www.rasch.org/rmt/rmt83b.htm

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

Yen, C.-F., Ko, C. H., Yen, J.-Y., Chang, Y.-P., & Cheng, C.-P. (2009). Multi-dimensional discriminative factors for Internet addiction among adolescents regarding gender and age. *Psychiatry and Clinical Neurosciences, 63,* 357-364. doi:10.1111/j.1440-1819.2009.01969.x

Young, K. S. (1998a). Internet addiction: The emergence of a new clinical disorder. *CyberPsychology & Behavior, 1(3),* 237-244. doi:10.1089/cpb.1998.1.237

Young, K. S. (1998b). *Caught in the Net: How to recognize the signs of Internet addiction – and a winning strategy*. New York: John Wiley & Sons.

Young, K. S., & Rodgers, R. C. (1998). *Internet addiction: Personality traits associated with its development.* Poster presented at 69th annual meeting of the Eastern Psychological Association. Boston, MA.

About the Authors

Panayiotis Panayides holds a BSc in Statistics with Mathematics (Queen Mary College, University of London), an MSc in Educational Testing (Middlesex University, UK) and a PHD in Educational Measurement (University of Durham, UK). He is currently an assistant headmaster and head of the Mathematics department at the Lyceum of Polemidia, Limassol, Cyprus. His research interests include educational and psychological measurement and research into mathematics education.

Miranda Jane Walker holds a BA in Hispanic Studies and Modern Greek (King's College, University of London) a BA in English Language and Literature (University of Cyprus) and an MA in Education Leadership and Management (Open University, UK). She teaches Spanish, at the Lyceum of Polemidia in Limassol, Cyprus. Her

research interests include high school students' Internet habits and attitudes; teacher and student motivation in the foreign language classroom and educational leadership and management.

Europe's Journal of Psychology
2012, Vol. 8(3), 327–351
doi:10.5964/ejop.v8i3.474