Running Head: SHOULD DECISIONS ABOUT

An assessment of the predictive validity of impact factor scores: Implications for academic employment decisions in social work

Gary Holden, DSW

New York University

Gary Rosenberg, Ph.D

Mount Sinai School of Medicine

Kathleen Barker, Ph.D.

The City University of New York: Medgar Evers College

Patrick Onghena, Ph.D.

Katholieke Universiteit Leuven

Correct citation for final manuscript:

Holden, G., Rosenberg, G., Barker, K., & Onghena, P. (2006). An assessment of the predictive validity of impact factor scores: Implications for academic employment decisions in social work. *Research on Social Work Practice, 16,* 613-624.

Keywords:

scientific communication, decision making, personnel promotion, personnel selection, teacher tenure, educational personnel, occupational tenure, career development, higher education, citation analysis, impact factor scores, validity, faculty, hiring, reappointment, social work education

The first author may be contacted at: Room 407, MC: 6112, New York University: School of Social Work, 1 Washington Square North, New York, NY 10003, gary.holden@nyu.edu

Abstract

Objective: Bibliometrics is a method of examining scholarly communications. Concerns regarding the utility of bibliometrics in general, and the impact factor score (IFS) in particular, have been discussed across disciplines including social work. While there are frequent mentions in the literature of the IFS as an indicator of the impact or quality of scholars' work, little empirical work has been published regarding the validity of such use.

Method: A proportionate, stratified, random sample, of n=323 articles was selected from 17 Web of Science listed social work journals published during the 1992-1994 period.

Results: The relationship between journals' impact factor scores and the actual impact of articles published in those journals (predictive validity) was $r = .41$ (short term) and $r = .42$ (long term).

Conclusion: The practice of using the IFS as a proxy indicator of article impact merits significant concern as well as further empirical investigation.

Introduction

"Let's dump impact factors" exclaimed Abbasi in 2004 (no p.). This negative sentiment regarding a measure that assesses the impact of scientific journals seems to be shared by a number of individuals involved in academic scholarship. Referring to a discussion of academic medicine, Abbasi wrote:

> Emphasis on where research is published—relying on impact factors to reward academic work with funding or promotion—is ripping the soul out of academia. . . . Impact factors have much to answer for, as do deans, sponsors, government agencies, and employment panels who use them as a convenient—but flawed—performance measure. How can a score count for so much when *it is understood by so few and its value is so uncertain*? (2004, no page, emphasis added).

Monastersky (2005a) reported discussions with a number of scholars and editors echoing these themes. Are these authors and others (e.g., Balaram, 2000) correct that we need to be concerned about the uses of the impact factor score (IFS)? There are a number of concerns regarding problems with bibliometrics in general, and the IFS in particular, that have been discussed both within and outside of social work, although these concerns have not always been empirically supported (e.g., Baker, 1990; 1991; Borgman & Furner, 2002; Cnaan, Caputo & Shmuely, 1994; Cole, 2000; Cole & Cole, 1971; Garfield, 1997; Holden, Rosenberg & Barker, 2005a; 2005b; 2005c; Kirk,1984; Kostoff, 1998; 2002; Krueger; 1993; Lindsey, 1976; 1978; 1980; 1982; 1989; MacRoberts & MacRoberts, 1989; 1997; Moed, 2005; Phelan, 1999; Plomp, 1990; Porter, 1977; Rosenberg, Holden & Barker, 2005; Thyer, 2004). With the persistent use of the IFS

across many fields and the continuing correspondent concerns, it seems relevant to empirically examine these issues. Given that social work education is a subset of social work practice and that the IFS has been or may be used in the evaluation of social work faculty, the salience of this topic for the readers of this journal is clear. In addition, *Research on Social Work Practice* has hosted an ongoing evaluative bibliometric discussion which focused on the scholarly impact of social work editorial boards and reviewers (e.g., Pardeck, et al., 1991; Pardeck, 1992; Pardeck, Chung & Murphy, 1995; Pardeck & Meinart, 1999). Finally, given the political context of post-secondary education in the United States – that it is indeed a profession in transition (Finkelstein, Seal & Schuster, 1998), marked by decreasing numbers of full-time and tenure-track opportunities (Barker, 1998; 2002; 2003) and faculty restructuring (Rhoades, 1998) – examination of the use of any indicator of faculty "quality" is deemed critically important.

The IFS

The IFS was developed by Garfield and Sher in the early 1960's (Garfield, 1999). The IFS is based on journals in the Web of Science (WoS) databases. The IFS for journal *X* for 2005 is calculated by dividing the number of citations during 2005 to articles published in journal *X* in 2003 and 2004 by the number of articles in the journal *X* in 2003-04 (ISI, 2004).

The IFS as an evaluative bibliometric indicator

Evaluative bibliometrics, a component of bibliometrics, examines scholarship at the individual or institutional level (e.g., Baker, 1991). The primary concern in this

paper is the use of the IFS in evaluative bibliometrics. Of course, one needs to be cautious regarding the psychometric properties of any indicator. The idea of assessing the psychometric properties of the impact factor is not new and continues to recent times (e.g., Decker, Beutel & Brahler, 2004). Although the IFS received support as an indicator of journal quality (e.g., Christenson & Sigelman, 1985; Saha, Saint, & Christakis, 2003), it has been criticized (e.g., Frank, 2003; Glanzel & Moed, 2002; Linardi, Coelho & Costa, 1996), as have misuses of it (e.g., Adam, 2002; Frank, 2003; Garfield, 1996; Smith, 1998; Staff, 2004). Although Garfield (2003) noted ways editors could legitimately increase the IFS of their journals, instances of less legitimate attempts to manipulate the IF have also been noted (e.g., Sevinc, 2004; Smith, 1997).

In the course of responding to a set of questions about the IFS and its use in evaluation, Garfield (2003) stated:

> The journal impact factor enters the picture when an individual's most recent papers have not yet had time to be cited. So the journal impact factor becomes the surrogate. The basic assumption is that if you get published in high impact journals that probably says something about the general quality of your paper. However, it is no guarantee that it will be cited (p. 365).

While many articles noted the use of the IFS as an indicator of the eventual impact or quality of scholar's work, few provided evidence of the prevalence of its use for such decisions (e.g., Fava & Ottolini, 2000; Moed, 2005; Sellers, Mathiesen, Perry & Smith, 2004). Two educational researchers (Braxton & Del Favero, 2002) seem to advocate for its use. "Journal impact scores are well suited for the scholarship of

discovery as an indicator of quality because of its emphasis on advancements in knowledge and the development of theory" (p 21). Alternatively, others decry its use for evaluative purposes (e.g., Garfield, 1996; Fava & Ottolini, 2000; Moed, 2005; Smith, 1998) while others suggested how it might be used in such situations. For instance, Weigel, et al., (2004), proposed a system where faculty members (in departments of surgery) would have their publications weighted by the IFS of the journal in which they were published.

In terms of actual data, Fassoulaki, Sarantopoulos, Papilas, Patris and Melemeni (2001) surveyed 438 European and North American anesthesiologists regarding their views of the IFS. In terms of academic appointments, Fassoulaki et al., reported that slightly more than 30% answered that a minimum IFS value was used for assessment of candidate's publications and approximately 65% answered that a summated IFS was a criterion for academic appointment. Europeans were significantly more likely to indicate usage of the IFS in these ways. This is an indirect measure of actual use of the IFS. Lower response rates in the US (where the sample consisted of department heads) than in Europe (sample there was department members and journal editors in chief) might have exacerbated observed North American – European differences (and the overall use estimate). Yet, one would likely consider this a relatively informed sample and, even if these percentages were halved, use of the IFS in 15-30% of academic employment decisions in this field, as well as anecdotal evidence of use in other fields, represents a situation worth investigating.

In social work

In their study of the relative impact of social work journals Lindsey and Kirk (1992) found that *Social Work* had the highest IFS of social work journals during the 1981-89 period (mean = .70). Lindsey and Kirk (1992) suggested very different levels of distribution of *Social Work* (e.g., in addition to subscriptions all National Association of Social Workers' members receive the journal as part of their membership, current n of members 150,000+) compared to other journals in the profession may have played a role in this finding (c.f., Howard & Howard, 1992). Furr (1995) examined the 'core influence' (intra group impact of a core set of social work journals on each other) in relation to the IFS for these social work journals (which by definition includes non social work journals). His findings which compared the rankings of these social work journals on both the IFS and core influence suggested that the IFS may not reflect impact within the journal's discipline.

More recently, Sellers, Mathiesen, Perry and Smith (2004) contrasted social work journal rankings on different indices: IFS (for the year 2000) and ratings of the quality and prestige (a combination of familiarity and quality; Cnaan, Caputo & Shmuely, 1994). Their survey focused on 38 journals (n=556, response rate = 26%). Their data revealed a statistically significant Spearman correlation of r_s = .45 (p < .01) between the journals' impact factor score ranking and the rankings of journal quality (n = 32).

Using quality ratings of journals (especially when it is unclear as to what time period respondents are rating) based on low response rate surveys, conducted at specific points in time (compared to impact factor scores computed yearly) would seem

to be even less valid than using the IFS as an indicator of the eventual impact of a scholar's work. Furthermore, using either group level measure (quality ratings of journals or the IFS) as a proxy measure of the eventual impact of an individual article risks an incorrect inference because of the *ecological fallacy* -- inferring something about an individual based exclusively on data from a group to which they belong (e.g., Holden, Rosenberg & Barker, 2005b). Despite this logic, Seipel's (2003) sample of 189 social work educators (from MSW programs) reported that they viewed publishing in first-tier journals as more important for gaining tenure than publishing in second-tier journals, and both as more important than publishing in third-tier journals.

Outside of social work. In terms of work on the IFS as a bibliometric indicator in the assessment of individual scholars outside of social work, Kurmis (2003) provided a recent overview of the IFS and its limitations as have others (e.g., Garfield, 1983; 1996; Gastel, 2001). Based on this review, Kurmis concluded in part that "[e]xtension of journal-impact-factor data to individual articles and authors is inappropriate and should be avoided" (p. 2453). Or, as Zwahlen, Junker and Egger put it, "the impact factor of the journal that publishes a given article is a poor predictor of the number of citations the article is likely to receive" (p. 20; c.f. Seglen, 1992).

Weale, Baily and Lear (2004) have recently shown that the distribution of citations (by 10/03) to articles published in 2002 in immunology and surgery were typically non-normal. These authors suggested that the IFS should perhaps be adjusted to account for such non-normal distributions. It is a basic statistical truism that means are not the best descriptor of the center of a distribution when that distribution is

skewed. Furthermore, Frank (2003) asserted that because of inter- and intra-journal variability, citations to a scholar's articles were a better indicator of the impact of that scholar's work than the IFS of the journal in which that article is published (cf., Furr, 1995; Garfield, 1996; 1999; Gastel, 2001; Seglen, 1997; Whitehouse, 2001).

Seglen (1994) conducted an important conceptual forerunner to the current study. He studied the 'lifetime' publication lists of 16 senior Norwegian biomedical scholars. This resulted in a sample of 907 articles for which an IFS could be obtained from JCR. Across the entire sample of articles, the correlation between the average number of citations to the article per year (across years 2-4 after publication) and the IFS for the journal for the year in which the article was published was .41 (c.f., Hansen & Henriksen, 1997).

Dahwan and Gupta's (2004) examined 902 Indian physics research papers published in 29 physics journals in 1997 (and listed in the WoS) for which the first author had an Indian institutional affiliation. The correlation between the 1997 IFS for the journal and the number of citations the paper received during the 1997 to July 2003 citation window was r = .21.

Summary

The behavior of using IFSs rather than actual rates of citation deserves explication. Probably most of us have heard some version of the refrain 'she usually publishes in good/better/high/top quality journals' (c.f., Lawrence, 2003). How does the speaker make the determination of which journals are better? How evidence based are these determinations? Using the IFS to inform such assessments at least brings some

empirical basis to the decision regarding the relative position of journals. While it may be an improvement over the subjective impressions of individual academic employment decision makers, a key question remains. Is the predictive validity of the IFS sufficient for the serious decisions for which it has been used?

The IFS was not designed to be used in evaluative bibliometric studies and yet it seems to be used for this purpose in some quarters. No evidence was found regarding the frequency of such use in social work. Assuming that some administrators, or tenure and promotion committees, might be interested in such use – as opposed to more directly assessing the impact of a scholar's work via examination of actual citations to her work – it seems relevant to assemble some evidence regarding the utility of a journal's IFS as a proxy measure for the eventual impact of a scholar's article (operationalized as citations) in that journal.

Seglen (1997) has provided perhaps the most comprehensive summary of concerns regarding use of the impact factor in such situations. He stated:

- Journal impact factors are not statistically representative of individual journal articles
- *Journal impact factors correlate poorly with actual citations of individual articles*
- Authors use many criteria other than impact when submitting to journals
- Citations to "non-citable" items are erroneously included in the database
- Self citations are not corrected for
- Review articles are heavily cited and inflate the impact factor of journals
- Long articles collect many citations and give high journal impact factors

- Short publication lag allows many short term journal self citations and gives a high journal impact factor
- Citations in the national language of the journal are preferred by the journal's authors
- Selective journal self citation: articles tend to preferentially cite other articles in the same journal
- Coverage of the database is not complete
- Books are not included in the database as a source for citations
- Database has an English language bias
- Database is dominated by American publications
- Journal set in database may vary from year to year
- Impact factor is a function of the number of references per article in the research field
- Research fields with literature that rapidly becomes obsolete are favoured
- Impact factor depends on dynamics (expansion or contraction) of the research field
- Small research fields tend to lack journals with high impact
- Relations between fields (clinical *v* basic research, for example) strongly determine the journal impact factor
- Citation rate of article determines journal impact, but not vice versa (Seglen, 1997, p. 499, emphasis added)

While many with experience with bibliometrics would agree with Seglen's point above (in italics) – that the IFS of a journal is a poor predictor of the frequency of citations to particular articles in that journal -- what do the data tell us regarding the validity of the IFS as a proxy for the impact of an individual article (and by extension – a scholar's work) in social work?

It is this predictive relationship that is the focus of the current study which can be considered a partial replication and extension of Seglen's (1994) work. It is a partial replication of Seglen in that the current study will provide an estimate of the short-term predictive validity of the IFS within a particular body of literature. It is an extension of the Seglen study in that a) the analysis is being applied to a new and different field; and b) the investigation will utilize a longer follow-up citation window so as to permit an assessment of the longer term predictive validity of the IFS.

The primary goals for this study were to provide estimates of the size of the relationship between the IFSs of journals and the short term impact (total number of citations for the publication year plus four subsequent years) and the long term impact (total number of citations for the publication year plus ten subsequent years) of the articles selected from those journals that were in the sample. A statistical problem arises in this conceptualization because articles are nested within journals. The assumption of independent errors is violated when correlating the IFS with short and long term impact. Therefore the Pearson product-moment correlation will only be used as a descriptive device to make a comparison with the figures published in the literature

possible (see literature review). For the inferential statistics, a multilevel model is used to take account of the nesting structure: articles within journals.

Next, a multi-year IFS will be employed. This modified version of the IFS (the mean of the publication year IFS and the two previous years' IFSs) will be examined as a potential improvement over the single-year IFS as a predictor of article impact. Three other predictive factors will also be explored. One study of a social work journal found that both the number of pages and the number of references per article were positively correlated with impact (Rosenberg, Holden & Barker, 2005). There have been other indications in the literature that there may be a relationship between the number of authors and the amount of impact of (citations to) an article (e.g., Lindsey, 1978a; Meittunen & Nieminen, 2003; Oromaner, 1974).

Finally, given the practical problem of estimating the potential impact of a scholar's work for academic decision makers, we will explore the possibility that a model utilizing multiple predictor variables (one that would be easily available to academic employment decision makers) might offer better prediction of the impact of specific articles. In other words does a combination of the IFS of the journal in which an article is published, the length of the article, the number of references in its reference list and the number of authors provide better prediction of its impact than the IFS alone? The multilevel predictive model does not depend on the assumption of independent errors and therefore will allow inferential conclusions. The exploration of these relationships should help social work educators to make better academic employment decisions.

Method

Sample

The sample for this study was comprised of a random sample of articles published in social work journals covered in ISI's *Journal Citation Reports* (JCR) during the 1992-94 period.

Time period. This 1992-94 period was selected for analysis because it was the most recent possible three year period that would allow a subsequent ten year window for potential citation of each article (for the most recent article from end of 1994).

Sampling frame. A sampling frame of articles and reviews (e.g., book reviews, editorials, meeting abstracts, corrections, letters, and notes were excluded) was developed from ISI's Web of Science (WoS) (ISI, 2004). Our goal was to select high quality journals (operationalized here as journals included in the JCR) that were published and covered in the ISI databases during 1992-94 and 2004 time periods. The most recent version of JCR (2003) at the time of sample selection identified 29 journals in the subject category social work. Thomson ISI includes the *American Journal of Community Psychology* and the *Journal of Community Psychology* in this group although most social work academics would likely consider them to be psychology journals. Therefore we dropped these two journals, leaving a potential sample of 27. Ten of the remaining journals were not contained in the JCR for the 1992-94 period, further reducing the sample of journals to 17. These 17 journals (some/one of which would likely be considered multidisciplinary such as *Family Relations*), and the number of articles for each journal for the 1992-94 period, are displayed in Table 1.

Once the sampling frame of articles and reviews (referred to as articles henceforth) was constructed from the WoS, it was double checked against the table of contents of the journals or the journal itself if no electronic table of contents was available. This was done to insure accuracy of transcription from the Web of Science (WoS) and check for articles missing from the WoS.

The 1,924 articles summarized in Table 1 comprised the final sampling frame for the study. Although a few articles in the sampling frame might be considered research notes or editorial material the decision was made to use the WoS categorization. This choice should make the current study more comparable to other studies in this area.

Sample size. Although our primary interest was describing the size of the correlation between the IFS and the short and long term number of citations, a power analysis was conducted to determine the sample size. The two studies that focused on this issue provided effect size estimates or r = .21 and r = .41 (Dahwan & Gupta, 2004; Seglen, 1997). Although there was no effect size estimate for the long term relationship in the literature, an effect size estimate for the power analysis that was smaller than the smaller effect size in the literature was used. The power analysis used Ouwehand's (2002) Power Calculator [parameters: effect size of r = .20, alpha = .01, 1-tailed, power = .9] and produced a desired sample size of 320.684 articles for the study.

Sampling Process. Tables listing the articles for each journal for each year in the time period were created. Using a sampling fraction of 16.892% (325/1924 – 325 used to compensate for rounding errors) a proportionate, stratified, random sample was selected using Research Randomizer (Urbaniak & Plous, 2005). In other words, the

sampling frame was stratified by journal and year of publication, then 16.892% of the articles from each resulting cell were randomly selected into the final sample. The final sample totaled 323 articles due to rounding. The n's of articles per journal for the sampling frame and the final sample are shown in Table 1.

Measures

The data were obtained from multiple sources. Journal impact factor scores were obtained from the associated year of *Journal Citation Reports* (c.f., ISI, 2004). The number of authors, length of each article (an admittedly crude indicator given the varying formats of journals), number of references it contained and the number of citations to it were obtained from the Web of Science (WoS, http://isi4.isiknowledge.com/portal.cgi/wos).

A modification of the IFS was also employed. In academic employment decisions, one might be most tempted to employ the IFS as a proxy indicator for articles that have been recently published (since no conclusive record of citations would be available). Yet it is clear from an examination of *Journal Citation Reports* (http://scientific.thomson.com/products/jcr/) that the IFS may vary within journals over time. Therefore for each year of publication of each journal we created a second IFS proxy measure which was a three year average measure of the IFS for the publication and the two previous years (c.f., Hansen & Henriksen, 1997). Our goal was to test whether or not this multi-year mean IFS was a better predictor (than the single year IFS) of long term impact. This multi-year approach could be relatively easily employed for academic employment decisions.

Short term impact was operationalized as the total number of citations in WoS journals during the publication year and four subsequent years. *Long term impact* was operationalized as the total number of citations received by the article during the publication year and 10 subsequent years. Both to replicate Seglen, and because in our view they should conceptually be included, self-citations were not excluded from these analyses.

Procedure

The data for the study were obtained from the WoS which is available from Thomson's Institute for Scientific Information (http://www.isinet.com). In October 2005, the WoS database covered approximately 8700 journals. For each article selected into the sample, a General Search in WoS was performed. Use of the General Search feature does mean that there is the possibility of missing citations that have incorrect information regarding the cited article (e.g., incorrect publication years, volumes, pages numbers, etc.). Such mistakes may be discovered by using the WoS Cited Reference search. This was not done here as it would have required inferences beyond the knowledge of the authors of the current study. The authors assumed that any General Search related errors would likely be random across this population. Data from the WoS searches were then entered in to SPSS 13.0 (2004) and SAS 9.1 (2003) for further analyses.

Insert Table 1 about here

Results

Descriptive

Descriptive statistics for the sample of articles (n = 323) are provided in Table 1. In addition to the individual variation at the level of journals in Table 1, it can be seen that across journals, the mean number of authors per article was 1.72. The average article was: 12.06 pages in length; had 30.91 references; was cited 3.07 times in the short term and 7.26 times in the long term. There were a total of 993 citations during the short term period and 2346 citations during the long term period to this sample of articles.

In the bibliometrics literature, the descriptive statistics that are sometimes used include *concentration* (the percentage of papers that receive 50% of the citations), *citedness* (the number of citations an article needs to be in the top 1% of papers), and *uncitedness* (percentage of papers that had not been cited in the study period) were computed as well (ISI, 1999; Katz, 1999). For the current study (the percentages are approximate due to the distributions), *concentration* analyses revealed that 13% of the papers received 51.5% of the short term citations and 12.7% of the papers received 50.1% of the long term citations. *Citedness* analyses showed that a paper needed to have 19 citations to be in the top 1% in the short term while the corresponding number of citations for the long term was 44. The proportion of papers that were not cited (*uncitedness*) in the short term was 28.2% and in the long term it was 16.4%.

Inferential Analyses

A primary goal of this study was to provide estimates of the predictive validity of the IFS. Predictive validity is demonstrated by the size of the association between the

IFS of journals and either the short term or longer term number of citations to articles published in those journals. Correlations of Pearson's $r = .41$ and $r = .42$ were observed for these two sets of relationships, respectively (intraclass correlation coefficients were .1656 and .1732, respectively). Analysis of the predictive validity of the modified IFS revealed no gain as a result of using this measure ($r = .42$ and $r = .41$, respectively). Given these results, the modified IFS was dropped from subsequent analyses.

Next, three additional predictors were examined. The number of pages was not related to short ($r = .003$) or long term impact ($r = .01$). The number of references per article was positively related to short ($r = .17$; $p < .01$) and long term impact ($r = .15$; $p < .01$). The number of authors was positively related to short ($r = .12$; $p < .05$) and long term impact ($r = .12$; $p < .05$).

Finally, it seemed obvious to ask: Is there a combination of variables that an academic employment decision-making group would have access to at the point of their decision, which would predict the short and or long term impact of a specific set of articles? Construction of two multilevel models (using the MIXED procedure in SAS) revealed that the IFS was the strongest predictor of both short term and long term impact (Table 2). The only other significant contributor to both models was the number of references. The proportion of variance accounted for by the fixed factors in a multilevel analysis has been discussed vehemently in the multilevel literature of the past ten years. The discussion arose because the notion of 'proportion of variance accounted for' is not uniquely defined due to the different types of variance and covariance components at the different levels. Table 3 presents the results multilevel

analysis with stepwise addition of predictors based on the estimation formula of Snijders and Bosker (1994, 1999). As can be seen, although the IFS and the number of references represented significant contributors to both the short term and long term models, the proportion of variance explained only ranged from .10 to .12 (depending on the specific variables involved). Although these values can be considered of medium size as compared to validity coefficients of psychological tests (see e.g., Cohen, 1988, p. 80), they clearly indicate that short term and long term impact cannot be reduced to a few simple indicators like IFS without losing most of the information.

Insert Table 2 & 3 about here

One possible anomaly was discovered that could potentially impact on these findings. Two articles (listed in the WoS as reviews) were selected into the sample that one might not consider to be typical reviews (Campling, 1991; 1992). Given that these articles listed a very high number of references, and no citations, we reran the analyses to see if these outliers had any impact. No substantial impact was detected with the exception of the correlation between number of references and short term impact (full sample: r = .17; without Campling: r = .22) the correlation between number of references and long term impact (full sample: r = .15; without Campling: r = .20). In terms of the multilevel model, the IFS and the number of references remained the two significant predictors, but the removal of the two Campling articles actually strengthened the relationship for number of references.

Discussion

Summary of findings

The primary goal of the current study was to examine the predictive validity of the IFSs of social work journals that were covered in the WoS during the time period studied. Moderate to large positive correlations were observed between the IFS for journals in the sample and the subsequent (short and long term) impact of articles published in those journals. A possible alternative to the publication year IFS (a three year average IFS) was tested and then abandoned as it did not offer any advantage. In terms of joint prediction, the IFS and the number of references per article were the only two variables that achieved significance in the multilevel model. A key strength of the current study is that it is based on a stratified random sample that can be considered representative of the population of interest because the randomization was not corrupted by a less than 100% response rate (as is virtually always the case in studies involving human subjects). Furthermore, it covers a broad range of social work journals from three consecutive years and assesses both the short term and long term impact of articles published in those journals.

Caveats

The results of this study should also be considered in the light of its weaknesses. The current study used a sample of particular social work journals, from a particular time period and while the sample may likely be representative of all articles in those journals from that time period, the degree to which the results generalize beyond these parameters is unknown. For instance, would the IFS be more predictive of

article impact in a field with different citation norms, such as psychology? In addition, the bivariate correlations between the IFS and short and long term impact should be considered descriptive only because articles are nested within journals and therefore the assumption of independent errors necessary for inferential purposes is violated. Beyond these issues there is the lack of empirical evidence justifying the undertaking of current study. Some if not many readers will have heard of instances where the IFS was used in the evaluation of a scholar. Such anecdotal evidence is supported by the survey of Fassoulaki, et al., (2001). The use of hearsay evidence for the value of publishing in this or that journal can be vexing for scholars, especially for those whose work is novel and/or challenging to dominant paradigms. Similarly, the lack of more comprehensive evidence regarding the extent of using hearsay to make evaluations is a concern. Regardless of how prevalent these practices, the use of indicators such as the IFS in academic employment decision-making should be informed by evidence regarding the measure's predictive validity, not its assumed validity.

Finally, the current study did not assess the accuracy of categorization of articles in the WoS. While our impression is that there was a relatively high level of accuracy, the actual level was not determined. The possible miscategorization of the two Campling articles as review articles was analyzed and noted above. While our decision to include them in the overall sample might be a problem in one sense, it does allow us to compare our results with other studies that utilized the WoS categories of 'articles' and 'reviews'. Regardless, research on the accuracy of categorization of social work materials in the WoS and estimation as to how systematic vs. random they are would be

useful. There are a number of sources with more general discussions of problems encountered in bibliometric studies (e.g., Holden, Rosenberg & Barker, 2005, Moed, 2005).

Predictive validity of IFS

What do the findings regarding the predictive validity of the IFS mean? Cohen (1988) suggested that one might describe an *r* of .1 as 'small', an *r* of .3 as 'medium' and an *r* of .5 as 'large'. So these estimates from the current study could be described as moderate to large. But how would that translate in the specific application of academic employment decisions? Do Cohen's general guidelines make sense for predictive validity coefficients? A tenure and promotion committee, using the IFS of the journals that a candidate had published in as an indication, would be working with an indicator that on average explained approximately 18% of the variance in the number of citations that the candidates articles would receive over the short and long term.

The current study straddles the domain of scholarship and the domain of politics (Moed, 2005). It examined the relationship of journal and article factors with an outcome of scholarship (citations) located within the increasingly political context of academic employment and funding decisions (e.g., Adam, 2002) and changing employment relations in the United States (Barker, 1995; 1998, 2003; Barker & Christensen, 1998a, 1998b) and a rising labor discontent (Aronowitz, 2000; Rhoades, 1998). The ongoing attack on the institution of tenure (e.g., Hutcheson, 1997) and so far unmeasured corporatization of higher education (Rhoades & Rhodes, 2005; Slaughter & Leslie, 1997) are highly salient features of this political context. This attack can be seen

in both the conversion of full-time tenured lines to non-tenured arrangements (Barker, 2001), the ratcheting up of tenure requirements (Burgan, 1998), attempts to lengthen the tenure track time period at public universities (e.g., Barker, 2003) and private universities ("Tenure Modifications" 2005) -- instances of requirements that shifts higher levels of risk (Beck, 2000) to a new generation of professoriate. This situation is perceptible, if not clear to our junior colleagues.

Our primary focus in this area of scholarship has been on finding ways to improve hiring, reappointment, tenure and promotion decisions in the academy (Holden, Rosenberg & Barker, 2005c). While the use of the IFS in academic employment decisions might be an improvement over the subjective goal of 'publishing in high quality journals' that has been (still is?) used, it seems logical that further improvements can be realized. Whether such improvements will come through more qualified use of the IFS based on research such as the current study or through the use of alternatives to the IFS (e.g., Moed, 2005; Monastersky, 2005b; Revolutionizing peer review, 2005) remains to be seen. Alternatively, Colquhoun (2003) has suggested that:

> Perhaps one way to cope with the problem is to turn it on its head. Candidates can judge institutions by the questions they ask, rather than the other way round. Any selection or promotion committee that asks you for impact factors is probably a second-rate organization. A good place will want to know about the quality of what you have written, not where you published it — and will be aware that the two things are uncorrelated (p. 479).

The current study examined the predictive validity of the IFSs for social work journals. These results have implications for any social work faculty member who is facing an initial hiring, reappointment, tenure or promotion decision. use the impact factor as a proxy indicator of the impact of your scholarship? Our answer is no, unless the limitations of this indicator are clearly understood by the decision makers and no better alternative is available.

References

Abbasi, K. (2004). Let's dump impact factors. *British Medical Journal, 329.* Retrieved October 16, 2004 from: http://bmj.bmjjournals.com/cgi/content/full/329/7471/0-h

Adam, D. (2002). Citation analysis: The counting house. *Nature, 415,* 726-729

Anseel, F., Duyck, W., & DeBaene, W. (2004). Journal impact factors and self-citations: Implications for psychology journals. *American Psychologist, 59, 1,* 49-51.

Aronowitz, S. (2005). *The Knowledge Factory: Dismantling the Corporate University and Creating True Higher Learning*. Boston: Beacon Press.

Baker, D. R. (1990). Citation analysis: A methodological review. *Social Work Research & Abstracts, 26,* 3-10.

Baker, D. R. (1991). On-line bibliometric analysis for researchers and educators. *Journal of Social Work Education, 27,* 41-47.

Balaram, P. (2000). The impact of publication lists. *Current Science, 78,* 1177-1178.

Barker, K. (1995). Contingent work: Research issues and the lens of moral exclusion. In L. Tetrick & J. Barling (Eds.), *Changing employment relations: Behavioral and social perspectives* (pp. 31-60). Washington, DC: American Psychological Association.

Barker, K. (2001). Standard and Nonstandard Employment in Higher Education: Findings from the 1990's. In *Changing Career Paths in Science and Engineering. Report of a CPST Workshop* (pp. 12-30). Washington DC: AAAS, Commission on

Professionals in Science and Technology. (Retrieved April 15, 2006 from: http://www.cpst.org/Changing.pdf

Barker, K. (1998). Toiling for piece-rates and accumulating deficits: Contingent work in higher education. In K. Barker & K. Christensen (Eds.), *Contingent work: American employment relations in transition* (pp. 195-220). Cornell, NY: ILR/Cornell University Press.

Barker, K. (2003). Contingent work in the United States. In *Supplement to the 19th Edition of the Encyclopedia of Social Work*. (pp. 9-19). Washington, DC: NASW.

Barker, K., & Christensen, K. (1998a). Introduction: Controversy and challenges raised by contingent work arrangements. In K. Barker & K. Christensen (Eds.), *Contingent work: American employment relations in transition* (pp. 1-18). Cornell, NY: ILR/Cornell University Press.

Barker, K., & Christensen, K. (1998b). Charting future research. In K. Barker & K. Christensen (Eds.), *Contingent work: American employment relations in transition* (pp.306-320). Cornell, NY: ILR/Cornell University Press.

Beck, U. (2000). *The brave new world of work*: *The political economy of insecurity.* Cambridge, UK: Cambridge University Press.

Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. In B. Cronin (Ed.), *Annual Review of Information Science and Technology, 36* (pp. 3-72). Medford, NJ: Information Today.

Braxton, J. M., & Del Favero, M. (2002). Evaluating scholarship performance: Traditional and emergent assessment templates. *New Directions for Institutional Research, 114,* 19-31.

Burgan, M. (1998, March-April). What we talk about when we talk about tenure. *Academe, 84*(2), 2.

Campling, J. (1992). Social-policy digest 81. *Journal of Social Policy, 21,* 99-118.

Campling, J. (1993). Social-policy digest 85. *Journal of Social Policy, 22,* 99-118.

Christenson, J.A., & Sigelman, L. (1985). Accrediting knowledge: journal stature and citation impact in social science. *Social Science Quarterly 66,* 964-975

Cliff, N. (1987). *Analyzing multivariate data*. New York: Harcourt Brace.

Cnaan, R. A., Caputo, R. K. & Shmuely, Y. (1994). Senior faculty perceptions of social work journals. *Journal of Social Work Education, 30,* 185-199.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cole, J. R. (2000). A short history of the use of citations as a measure of the impact of scientific and scholarly work. In B. Cronin & H. B. Atkins (Eds.), *The Web of Knowledge: A Festchrift in Honor of Eugene Garfield.* (pp. 281-300). Medford, NY: Information Today.

Cole, J. & Cole, S. (1971). Measuring the quality of sociological research: Problems in the use of the *Science Citation Index. The American Sociologist, 6,* 23-29.

Colquhoun, D. (2003). Challenging the tyranny of impact factors. *Nature, 423,* 479.

Dahwan, S. M., & Gupta, B. M. (2004). Comparative evaluation of Indian physics research: Impact factor vs citations frequency. *Current Science, 86, 9,* 1194-1195.

Decker, O., Beutel, M. E., & Brahler, E. (2004). Deep impact – evaluation in the sciences. *Sozial Und Praventivmedizin, 49, 1,* 10-14.

Fava, G. A., & Ottolini, F. (2000) Impact factors versus actual citations. *Psychotherapy and Psychosomatics,* 69, 6, 285-286.

Fassoulaki, A., Sarantopoulos, C., Papilas, K., Patris, K., & Melemeni, A. (2001). Academic anesthesiologists' views on the importance of the impact factor of scientific journals: a North American and European survey. *Canadian Journal of Anesthesia,* 48, 953-957.

Frank, M. (2003). Impact factors: Arbiter of excellence. *Journal of the American Medical Library Association, 91,* 4-6 .

Furr, A. L. (1995). The relative influence of social work journals: Impact factors vs. core influence. *Journal of Social Work Education, 31,* 38-45.

Garfield, E. (1983). How to use citation analysis for faculty evaluations, and when is it relevant? Part 2. *Current Comments, 45,* 5-14.

Garfield, E. (1996). How can impact factors be improved? *British Medical Journal, 313, 7054,* 411-413.

Garfield, E. (1997). Validation of citation analysis. *Journal of the American Society for Information Science, 48,* 962-964.

Garfield, E. (1999). Journal impact factor: A brief review. *Canadian Medical Association Journal, 161,* 979-980.

Garfield, E. (2003). The meaning of the Impact Factor. *International Journal of Clinical and Health Psychology, 3,* 363-369.

Gastel, B. (2001). Assessing the impact of investigators' work: Beyond impact factors. *Canadian Journal of Anesthesia, 48,* 941-945.

Glanzel, W. & Moed, H. F. (2002). Journal impact measures in bibliometric research. *Scientometrics, 53,* 171-193.

Hansen, H. B. & Henriksen, J. H. (1997). How well does journal 'impact' work in the assessment of papers on clinical physiology and nuclear medicine. *Clinical Physiology, 17,* 409-418.

Holden, G., Rosenberg, G., & Barker, K (2005a). Tracing thought through time and space: A selective review of bibliometrics in social work. *Social Work in Health Care, 41(3/4),* 1-34.

Holden, G., Rosenberg, G., & Barker, K. (2005b). Shallow science or meta-cognitive insights: A few thoughts on bibliometrics. *Social Work in Health Care, 41(3/4),* 129-148.

Holden, G., Rosenberg, G., & Barker, K (2005c). Bibliometrics: A potential decision making aid in hiring, reappointment, tenure and promotion decisions. *Social Work in Health Care, 41(3/4),* 67-92.

Howard, M. O. & Howard, D. A. (1992). Citation analysis of 541 articles published in drug and alcohol journals: 1984-1988. *Journal of Studies on Alcohol, 53,* 427-434.

Hutcheson, P. (1997). The corrosion of tenure: A bibliography. *The NEA Higher Education Journal, Fall,* 89-106.

ISI (1999). Citations reveal concentrated influence: Some fields have it, but what does it mean? *Science Watch, 10(1),* 1-2.

ISI (2004). *Journal Citation Reports.* Retrieved 11/12/04 from: http://isi4.isiknowledge.com/portal.cgi?DestApp=WOK_HOME&Func=Frame

Katz, J. S. (1999). Bibliometric indicators and the social sciences. Retrieved 7/27/03 from: http://www.sussex.ac.uk/Users/sylvank/pubs/ESRC.pdf

Kirk, S. A. (1984). Methodological issues in the comparative study of schools of social work. *Journal of Social Service Research, 7, 3,* 59-73.

Kostoff, R. N. (1998). The use and misuse of citation analysis in research evaluation. *Scientometrics, 43, 1,* 27-43.

Kostoff, R. N. (2002). Citation analysis of research performer quality. *Scientometrics, 53, 1,* 49-71.

Kurmis, A. P. (2003). Understanding the limitations of the journal impact factor. *The Journal of Bone and Joint Surgery, 85A,* 2449-2454.

Lawrence, P. H. (2003). The politics of publication. *Nature, 422,* 259-261.

Linardi, P. M., Coelho, P. M. Z., &Costa, H. M. A. (1996). The "impact factor" as a criterion for the quality of scientific production is a relative, not absolute, measure. *Brazilian Journal of Medical and Biological Research, 29,* 555-561.

Lindsey, D. (1976). Distinction, achievement and editorial board membership. *American Psychologist, 31,* 799- 804.

Lindsey, D. (1978). *The scientific publication system in social science.* San Francisco: Jossey-Bass.

Lindsey, D. (1980). Production and citation measures in the sociology of science: The problem of multiple authorship. *Social Studies of Science, 10,* 145-162.

Lindsey, D. (1982). Further evidence for adjusting for multiple authorship. *Scientometrics, 4,* 389-395.

Lindsey, D. (1989). Using citation counts as a measure of quality in science: Measuring what's measurable rather than what's valid. *Scientometrics, 15,* 189-203.

Lindsey, D., & Kirk, S. A. (1992). The role of social work journals in the development of a knowledge base for the profession. *Social Service Review, 66,* 295-310.

MacRoberts, M. H. & MacRoberts, B. R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science, 40,* 342-349.

MacRoberts, M. H. & MacRoberts, B. R. (1997). Citation content analysis of a botany journal. *Journal of the American Society for Information Science, 48,* 274-275.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht, Netherlands: Springer.

Monastersky, R. (2005a, Oct. 14). The number that's devouring science. *Chronicle of Higher Education, 52, 8,* A12.

Monastersky, R. (2005b, Oct. 14). Impact factors run into competition. *Chronicle of Higher Education, 52, 8,* A17.

Ouwehand, A. (2002) Power calculator. Retrieved March 31, 2005 from: http://calculators.stat.ucla.edu/powercalc/

Pardeck, J. T., Arndt, B. J., Light, D. B., Mosley, G. F., Thomas, S. D., Werner, M. A., & Wilson, K. E. (1991). Distinction and achievement levels of editorial board

members of psychology and social work journals. *Psychological Reports, 68,* 523-527.

Pardeck, J. T. (1992). Are social work journal editorial boards competent? Some disquieting data with implications for research on social work practice. *Research on Social Work Practice, 2,* 487-496.

Pardeck, J. T., Chung, W. S., & Murphy, J. W. (1995). An examination of the scholarly productivity of social work journal editorial board members and guest reviewers. *Research on Social Work Practice, 5,* 223-234.

Pardeck, J. T., & Meinart, R. G. (1999). Scholarly achievements of the *Social Work* editorial board and consulting editors: A commentary. *Research on Social Work Practice, 9,* 86-91.

Phelan, T. J. (1999). A compendium of issues for citation analysis. *Scientometrics, 45,* 117-136.

Plomp, R. (1990). The significance of the number of highly cited papers as an indicator of scientific prolificacy. *Scientometrics, 19,* 185-197.

Porter, A. L. (1977). Citation analysis: Queries and caveats. *Social Studies of Science, 7,* 257-67.

Revolutionizing peer review? (2005). Editorial. *Nature Neuroscience, 8,* 397.

Rhoades, G. (1998). *Managed professionals: Unionized faculty and restructuring academic labor.* Albany: State University of New York Press.

Rhoads, R.A., & Rhoades, G. (2005). Graduate employee unionization as symbol of and challenge to the corporatization of U.S. research universities. *Journal of Higher Education, 76,* 243-275.

Rosenberg, G., Holden, G., & Barker, K (2005). What happens to our ideas? A bibliometric analysis of articles in *Social Work in Health Care* in the 1990s. *Social Work in Health Care, 41(3/4),* 35-66.

Saha, S., Saint, S., & Christakis, D. A. (2003). Impact factor: A valid measure of journal quality? *Journal of the Medical Library Association, 91,* 42-46.

SAS (2003). SAS 9.1 for Windows [Computer program]. Cary, NC: SAS Institute Inc.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464.

Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science, 43,* 628-638.

Seglen, P. O. (1994). Causal relationship between article citedness and journal impact. *Journal of the American Society for Information Science, 45, 1,* 1-11.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal, 314.* Retrieved 2/17/03 from: http://bmj.com/cgi/content/full/314/7079/497

Seipel, M. M. O. (2003). Assessing publication for tenure. *Journal of Social Work Education, 39, 1,* 79-88.

Sellers, S. L., Mathiesen, S. G., Perry, R., & Smith, T. (2004). Evaluation of social work journal quality: Citation versus reputation approaches. *Journal of Social Work Education, 40,* 143-160.

Sevinc, A. (2004). Manipulating impact factor: An unethical issue or an Editor's choice. *Swiss Medical Weekly, 134,* 410.

Slaughter, S., & Leslie, L.L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Baltimore: Johns Hopkins University Press.

Smith, R. (1997). Journal accused of manipulating impact factor. *British Medical Journal, 314*, 461.

Smith, R. (1998). Unscientific practice flourishes in science. *British Medical Journal, 316,* 1036-1040.

Snijders, T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods & Research, 22,* 342-363.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London, UK: Sage.

SPSS (2004). SPSS for Windows, 13.0. Chicago, IL: SPSS Inc.

Staff (2004). The undue importance of the impact factor. *European Journal of Cancer, 40,* 919.

Thyer, B. A. (2004). Evaluating social work journal publication systems. *Journal of Social Work Education, 40,* 361-363.

Tsay, M., & Chen, Y. (2004). Journals of general and internal medicine and surgery: An analysis and comparison of citation. *Scientometrics, 64, 1,* 17-30.

Tenure Modifications (2005). Editorial. New York University: Faculty Senators Council, p. 3.

Urbaniak, G. C., & Plous, S. (2005). Research Randomizer. Retrieved 7/31/2005 from: http://www.randomizer.org/

Weale, A. R., Bailey, M., & Lear, P. A. (2004). The level of non-citation of articles within a journal as a measure of quality: a comparison to the impact factor. *BMC Medical Research Methodology*, doi:10.1186/1471-2288-4-14.

Weigel, R. J., Dracon, G., Radhakrishnan, R., Rho. Y., Sevgen, F., & Dafoe, D, C. (2004). Incentive systems for academic productivity in a depart of surgery. *Journal Of The American College Of Surgeons, 199, 2,* 300-307.

Whitehouse, G. H. (2001). Citation rates and impact factors: Should they matter? *The British Journal of Radiology, 74,* 1-3.

Zwahlen, M., Junker, C., & Egger, M. (2004). Commentary II. The journal impact factor in the evaluation of research quality: Villain, scapegoat or innocent bystander? *Sozial Und Praventivmedizin, 49, 1,* 19-22.

Table 1.

Description of the random, stratified, proportionate sample (n = 323).

	Sampling Frame	Final Sample				Number	Page	Number	Short	Long
			1992	1993	1994	Authors	Length	References	Term	Term
	N	n	IFS	IFS	IFS	M	M	M	Impact[1]	Impact[2]
Administration in Social Work	73	12	.100	.164	.121	1.5	15.8	25.0	1.92	5.17
British Journal of Social Work	95	16	.250	.152	.306	1.5	14.9	30.4	2.44	4.88
Child Abuse & Neglect	227	38	1.447	1.021	1.474	2.4	11.1	31.0	8.87	20.87
Child Welfare	121	21	.359	.363	.756	1.8	13.5	23.2	3.14	8.24
Children & Youth Services Review	83	13	.559	.500	.561	1.2	16.5	28.6	3.15	7.77
Clinical Social Work Journal	87	15	.148	.254	.237	1.3	14.9	21.7	.60	1.33
Families in Society: The Journal of Contemporary Human Services	173	29	.212	.272	.345	1.6	8.7	34.4	1.55	3.93
Family Relations	191	31	.900	.978	.662	2.1	6.7	40.2	3.06	6.32
Indian Journal of Social Work	127	21	.110	.019	.011	1.4	11.0	16.3	.10	.57
Journal of Social Policy	57	9	.366	.571	.405	1.3	24.7	59.8	1.00	2.00

Journal of Social Work Education	105	18	.483	.373	.447	1.6	10.1	21.4	2.89	5.94
Research on Social Work Practice	92	16	.261	.349	.313	1.8	14.2	25.7	2.63	5.19
Smith College Studies in Social Work	35	6	.040	.346	.120	1.2	15.5	24.7	.50	2.00
Social Service Review	85	15	.758	.403	1.050	1.3	19.9	51.1	2.13	5.60
Social Work	213	36	1.078	.814	.792	1.6	7.5	35.1	3.97	9.33
Social Work in Health Care	104	18	.118	.099	.173	2.1	15.4	26.2	1.72	4.44
Social Work Research / Social Work Research Abstracts	56	9	.771	.528	.487	2.0	7.1	29.3	2.67	8.56
Total sample	1924	323				1.72	12.06	30.91	3.07	7.26

Note: Sampling fraction of 16.892% was applied to n of articles in sampling frame for publication year and then this product was added for each cell to obtain row total.

[1] Short term impact = M article citation total for publication year and four subsequent years

[2] Long term impact = M article citation total for publication year and ten subsequent years

Table 2.

Prediction of both short and long term impact (n = 323).

Impact and variables	Estimate	Standard Error	df	t	p
Short term impact as criterion					
Intercept	-.5813	1.0030	16	-.58	.5703
Impact Factor Score (IFS)	3.2289	.9071	302	3.56	.0004
Page Length (M)	.06057	.05422	302	1.12	.2649
References (N)	.02636	.01073	302	2.46	.0145
Authors (N)	.06694	.2552	302	.26	.7933
Long term impact as criterion					
Intercept	-1.5534	2.3089	16	-.67	.5107
Impact Factor Score (IFS)	8.1423	2.0994	302	3.88	.0001
Page Length (M)	.1879	.01246	302	1.51	.1326
References (N)	.04918	.02455	302	2.0	.0460
Authors (N)	.03928	.5838	302	.07	.9464

Table 3.

Proportion of variance explained by model using multilevel analysis with stepwise addition of predictors (n = 323).

Impact and variables	R_1^2 [1]	Deviance[2]	*BIC* [3]
Short term impact as criterion		1852.1 (null model)	1860.6
Impact Factor Score (IFS)	.1077	1845.0**	1856.3
References (N)	.1194	1836.2**	1850.3
Page Length (M)	.1213	1834.9	1851.9
Authors (N)	.1225	1834.9	1854.7
Long term impact as criterion		2387.1 (null model)	2395.6
Impact Factor Score (IFS)	.1146	2378.8**	2390.2
References (N)	.1232	2371.8**	2386.0
Page Length (M)	.1273	2369.5	2386.5
Authors (N)	.1276	2369.5	2389.4

[1] R_1^2 = Proportion of first-level variance accounted for; calculation following Snijders and Bosker (1994, 1999).

[2] Deviance = – 2 Log Likelihood (smaller is better); difference in deviance is distributed as a chi-squared variate with degrees of freedom equal to the difference in number of parameters; ** $p < .01$ means statistically significant improvement with respect to the previous model.

[3] *BIC* = Bayesian Information Criterion (smaller is better; Schwarz, 1978).