

# The Hamburg Burnout Inventory (HBI) in Two Large International Online Samples

V.1

**Matthias Burisch<sup>1</sup>**  
**University of Hamburg**  
**burisch@uni-hamburg.de**

## Overview

This technical report will briefly review the history of the *Hamburg Burnout Inventory* (*Hamburger Burnout-Inventar; HBI*), an instrument that was developed out of frustration with previous burnout questionnaires, the *Maslach Burnout Inventory* (*MBI*; Maslach & Jackson, 1986; Maslach, Jackson, & Leiter, 1996) and the *Tedium Measure* (*TM*; Aronson, Pines & Kafry, 1983), later rechristened as the *Burnout Measure*. Although these two instruments still enjoy the status of a „gold standard“ and dominate more than 90% of published empirical burnout research (Schaufeli & Enzmann, 1998, p. 71; Rösing, 2003, p. 69-75) their shortcomings are quite obvious. To mention only a few, the MBI's validity is questionable at best (Burisch, 1984b, unpublished) while the TM, both highly reliable and valid, is even more undifferentiated than the MBI, and its name is probably a misnomer (Enzmann, 1996). Moreover, the original version of the MBI was applicable only to people in service jobs, since many of the items referred to „recipients“, something which has now been corrected with the publication of a „general“ version.

In contrast, the HBI measures burnout with ten very short scales and one additional item. It has been shown to possess adequate reliability and validity vis-à-vis peer ratings. Recently, it has gained popularity due to its availability in a self-scoring and instant-feedback version on *SwissBurnout's* website ([www.swissburnout.ch](http://www.swissburnout.ch)). Between early 2006 and June 2007, about 40.000 individuals have left their (anonymous) traces there. This report will provide a brief background to the instrument and draw some tentative substantive conclusions, particularly from the second of two batches of online data.

## History of the instrument

The *Hamburg Burnout Inventory* (*Hamburger Burnout-Inventar; HBI*) was developed during the late 1980s and early 1990s in collaboration with several graduate students. Originally, it included about 30 constructs and more than 200 items. Gradually, using a mixture of the deductive and the inductive approaches to scale development (Burisch, 1984a), specifically using both common sense and information from empirical data, the scope narrowed. Many of the original constructs proved to be undistinguishable from each other, whereas others collapsed under the evidence provided by item analyses. What remained after

---

<sup>1</sup> Thanks are due Bianka Giesa, Felix Frühauf and Kirsten Steinhoff, who worked on the HBI in its early stage, and Maren Hagge, who conducted another validation study in 2005. I am also grateful to Alexander Harbaugh for editing my raw translation into English (and for much more!) and Catherine Vasey and Alexander Harbaugh for preparing the French version. Lew Goldberg sacrificed precious beach time in order to make this manuscript more readable; his contribution (and much more!) is highly appreciated.

the initial stage — not very well documented since two of the projects were never finished — were 39 items in 10 scales. A little later, item 40 („I am in the middle of a crisis from which I find no way out“) was added. It serves as a 1-item capsule measurement of burnout and is not scored on any of the scales.

Compared to standard personality inventories, both German language and English, an average scale length of just 3.9 items may appear too short. However, the well-known relation between scale length and reliability may not hold for scale length and validity in all cases. To obtain maximally valid scales, three to eight items may be optimal (Burisch, 1997).

For various reasons a couple of items and two full scales may be added in the near future.

The HBI has been employed in a few theses and dissertations, but to my knowledge the only published account of this is Frick and Filipp (1997).

In the spring of 2006, the HBI was posted on *SwissBurnout*'s website; soon after it was made self-scoring. Its major advantages over other measures seem to be threefold: (a) With ten scales and an additional item, it provides a much more differentiated picture than the MBI or the TM, although many of its scales intercorrelate substantially. (b) Some validity information is available, and that looks encouraging. (c) There are rough norms for the German version of the HBI, while those are lacking for both the MBI and the TM.

Between March and November of 2006 more than 17000 users have completed the German language HBI on the *SwissBurnout* website. These data, omitting incomplete records, comprise SWB Sample 1 (N = 16,273).

Late in 2006, some technical improvements were implemented at the website, and English and French versions were introduced. Feedback was provided only if the respondent had answered *all* items. Moreover, more differentiated demographic questions were introduced. Records turned in between Dec 28, 2006 and April 23, 2007 comprise SWB Sample 2 (N = 15,939).

### **Caveat**

There is no guarantee that the two SWB Samples may not contain multiple entries from individuals who „took their pulse“ on a weekly basis. Nor can we exclude the possibility of people who just wanted to „play“ with the device, wanting to know just how high or low scores one can obtain. This is why an additional norming system, based on sample SWB 1, was drawn. For a more detailed discussion, see the section on *Substantive Findings*.

## **Technical Information**

### **Reliability**

For easy reference, Table 1 summarizes all available reliability values for the HBI's scales. These include information from three data bases, all based on paper-and-pencil questionnaires, prior to the online version, namely the thesis by Frühauf (1990), an unfinished thesis project by Steinhoff (about 1991), and a thesis by Hagge (2005).

Table 1  
Scales, Scale Lengths, Reliability and Validity Estimates for the Hamburg Burnout Inventory

Scale	m	Cronbach's Alpha			Mean Alpha			Validity S			Validity H			Mean $r_{tc}$
		F	S	H	SWB 1	SWB 2	$r_k$	$r_{tc}$	$r_{tc-corr}$	$r_k$	$r_{tc}$	$r_{tc-corr}$		
01 Emotional Exhaustion (EE)	5	83	87	86	91	91	61	46	59	60	39	50	43	
02 Personal Accomplishment, reversed (PA)	3	74	63	67	82	71	54	30	41	35	27	46	29	
03 Distance (DIST)	4	71	68	65	76	75	60	42	54	74	47	55	45	
04 Depressive Reaction to Stress (DEP)	3	70	62	77	68	70	63	35	44	68	23	28	29	
05 Helplessness (HELPL)	4	83	73	71	88	87	55	40	53	64	39	49	39	
06 Inner Void (VOID)	4	80	74	73	88	87	65	22	27	69	9	11	15	
07 Tedium (TD)	5	87	89	88	89	91	67	46	56	73	33	39	40	
08 Inability to Unwind (INUN)	3	71	71	73	84	85	61	23	30	68	39	47	31	
09 Overtaxing oneself (OTAX)	5	80	72	81	87	85	61	38	49	50	28	40	33	
10 Aggressive Reaction to Stress (AGG)	3	75	56	74	80	79	50	28	40	29	17	32	23	
<b>Means</b>		<b>3.9</b>	<b>77</b>	<b>72</b>	<b>77</b>	<b>83</b>	<b>60</b>	<b>36</b>	<b>49</b>	<b>61</b>	<b>32</b>	<b>43</b>	<b>34</b>	

Legend: m = scale length. Data sets: F = Frühauf (1990). S = Steinhoff (unpubl.), H = Hagge (2005). SWB 1 = 16273 German SWB on-line tests = Batch 1. SWB 2 = subsample of 14123 German tests, a subsample of Batch 2.  $r_k$  = coefficient of inter-rater agreement.  $r_{tc}$  = validity coefficient.  $r_{tc-corr}$  = validity coefficient corrected for attenuation due to criterion unreliability.

Frühauf (1990) administered an early HBI version to 313 adults, including 75 students, many of whom worked part-time as well, and 238 working men and women. Those volunteers also filled in the *Tedium Measure* and the *Freiburger Persönlichkeits-Inventar* (FPI-R; Fahrenberg, Selg, & Hampel, 1984) for comparison. The first 39 items of the present HBI version were selected on the basis of that study.

In Steinhoff (1991; unpubl.) a sample of 182 adults completed the 40-item HBI version (CRISIS item added in the meantime) and also named two peers who would independently and anonymously rate her or him on 9-point rating scales. She also administered the short form of the *Freiburger Persönlichkeits-Inventar* (FPI-K; Fahrenberg & Selg, 1970) which contains eight 7-item scales.

A study by Hagge (2005) followed the same format as that of Steinhoff, with 77 subjects. For benchmarking purposes, she included the *NEO-FFI* (Borkenau & Ostendorf, 1993) and the *Oldenburg Burnout Inventory* (Ebbinghaus, 1986; Demerouti 1999), which comprises two scales, DISTANCE and EXHAUSTION, each eight items long.

Table 1 also includes reliability values from SWB samples 1 (N = 16,273) and 2 (German language sub-

16,273) and 2 (German language subsample only; N = 14,123), which will be described below. Columns F, S, H, SWB 1, and SWB 2 of Table 1 contain alpha coefficients calculated from Frühauf's, Steinhoff's, Hagge's and the two SWB samples or subsamples, respectively. Thus, Table 1 documents the HBI's internal consistency from five independent samples.

Retest reliability has not been studied so far.

### **Validity**

The middle (Steinhoff sample) and right hand (Hagge sample) blocks of Table 1 contain a coefficient of interrater agreement ( $r_k$ ), which is simultaneously an estimate of the average rating's reliability (Winer, 1962, p. 127). The correlations between scale scores (self ratings) and peer ratings,  $r_{tc}$ , are also displayed in Table 1, as are those coefficients divided by the square root of  $r_k$  (correction for attenuation). The latter coefficients,  $r_{tc-korr}$ , estimate the validity coefficients that would have resulted had the criterion (the peer ratings) been measured error-free.

### **Discussion**

As can be gleaned from the table, there is considerable — though far from perfect — agreement among the *reliability* estimates from the various samples. Across the samples, TD (mean  $r_{tt} = .89$ ) and EE (mean  $r_{tt} = .88$ ) are the most reliable scales, while DEP (.69), DIST (.71), and PA (.71) are the least reliable. Samples SWB 1 and 2 generally yield the highest estimates, probably due to their greater heterogeneity. In view of the shortness of the scales — the first column, headed *m*, contains scale lengths — these figures are remarkably high.

Moreover, in sample SWB 2 almost all items correlate highest (part-whole corrected coefficients) with their own scales. Two items from the DEP scale correlate somewhat higher with scale HELPL, one of the VOID item correlates higher with scale EE, and one AGG item correlates higher with scales DEP, HELPLESS, and VOID. Three of these four anomalies are also present in SWB sample 1, where the picture is generally less favorable. Several options to correct this situation will be considered: (a) Leave things as is, because content considerations justify that, (b) delete the items in question, (c) replace them with others, (d) score them in another scale, or (e) merge scales that correlate substantially (e.g., scales DEP and HELPL).

*Validities* range from excellent to unsatisfactory (i.e., a coefficient of .09 for VOID in the Hagge sample, which rises to .22, when partialling out gender).

To put these findings into perspective, the mean validity coefficient of the *FPI-K*, an established though now obsolete inventory containing eight scales of seven items each, was only .29 in the Steinhoff sample, in contrast to .36 for the HBI. The HBI fared somewhat worse in Hagge's study where the comparison was with the *NEO-FFI* (five scales of twelve items each) and the *OLBI* (two scales of eight items each). While the *NEO-FFI* scored an average validity coefficient of .40 (corrected .45) and the *OLBI* of .38 (corrected .47), the HBI only reached .32 (corrected .43).

**Table 2**  
**Gender, Occupational Status, and Language Used for the HBI**

Status	GERMAN		ENGLISH		FRENCH		% of FR
	male	female	male	female	male	female	
empl. with exec. functions	1000 (13%)	441 (7%)	80 (18%)	39 (14%)	71 (13%)	43 (8%)	11
empl. with line resp.	2386 (31%)	1413 (22%)	149 (33%)	67 (24%)	151 (28%)	82 (15%)	21
empl. without line resp.	2578 (33%)	2808 (44%)	76 (17%)	45 (16%)	192 (36%)	290 (53%)	44
independent	883 (11%)	519 (8%)	58 (13%)	39 (14%)	55 (10%)	35 (6%)	8
other	710 (9%)	913 (14%)	61 (14%)	65 (23w%)	58 (11%)	76 (14%)	12
without employment	172 (2%)	300 (5%)	27 (6%)	23 (8%)	11 (2%)	23 (4%)	3
<b>Sum</b>	<b>7729 (100%)</b>	<b>6394 (100%)</b>	<b>451 (100%)</b>	<b>278 (100%)</b>	<b>538 (100%)</b>	<b>549 (100%)</b>	
<b>% of Language</b>	<b>55</b>	<b>45</b>	<b>62</b>	<b>38</b>	<b>50</b>	<b>50</b>	<b>100</b>

However, many HBI constructs are probably hard to observe and rate from the outside. Thus, although the  $r_k$  value for VOID in the above example is the third highest at .69, reflecting high agreement among raters, these ratings may be of limited value. „Inner Void“ may be an experience people are wary to communicate even to their closest associates because of its ego-threatening impact.

## Description of SWB Sample 2

Respondents in SWB Sample 2 (with few exceptions) provided the following demographic information: gender, year of birth, nationality (in 13 categories), profession (to be typed in; categorization is under way), occupational status (6 categories), and workplace (i. e., Swiss canton vs. „outside Switzerland“; the latter category making up 71%). The language used (German = DE, English = EN, French = FR) and the access date were stored automatically.

A breakdown of the sample by language, gender, and professional status is included in Table 2.

The three language-specific subsamples do not differ much on their gender composition; the FR subsample was practically perfectly balanced in gender, whereas speakers of DE were more likely (55%) and the EN subsample even more likely (62%) to be male.

Professional status was quite similar for DE and FR, whereas in the EN subsample no less than 46% were in the upper echelons of employees with executive functions or line responsibility (DE: 37%; FR 32%) and 13% of it was independent (DE: 10%; FR: 8%). At the same time, 7% of the EN subsample were unemployed, compared to only 3% of the DE and FR subsamples. The average age (computed simply as the difference between 2007 and year of birth)

**Table 3**  
**SWB 2 in terms of geography**

Asia and Oceania	116
Austria	1046
Eastern Europe	76
England	108
France	324
Germany	9511
Italy	119
Latin America	48
Mediterranean countries	164
North America	373
Other countries the EEC	310
Sub-Saharan Africa	35
Switzerland	3703
missing	6
<b>Total</b>	<b>15939</b>

was lowest for EN (33.8) and highest for DE (40.6); the French mean was 38.5. Thus, the EN subsample can be expected to score somewhat differently in terms of burnout.

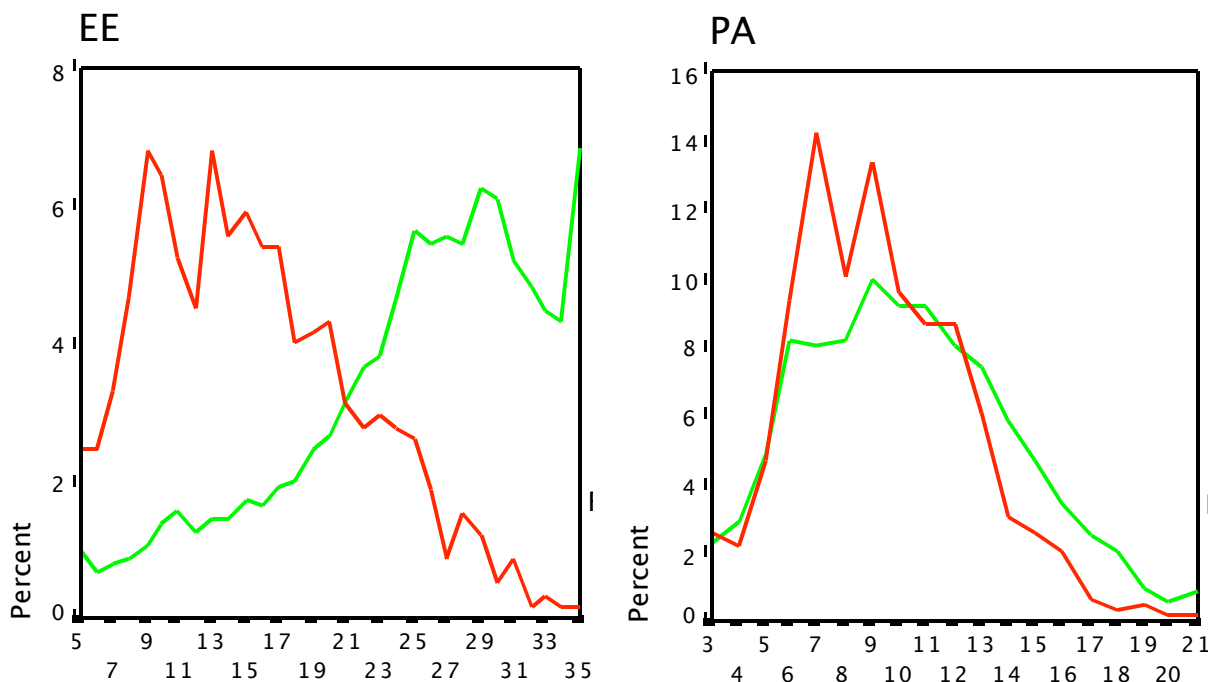
Responses came literally from all over the world, as Table 3 shows, with Germany, Switzerland, and Austria contributing the most participants.

## Substantive findings

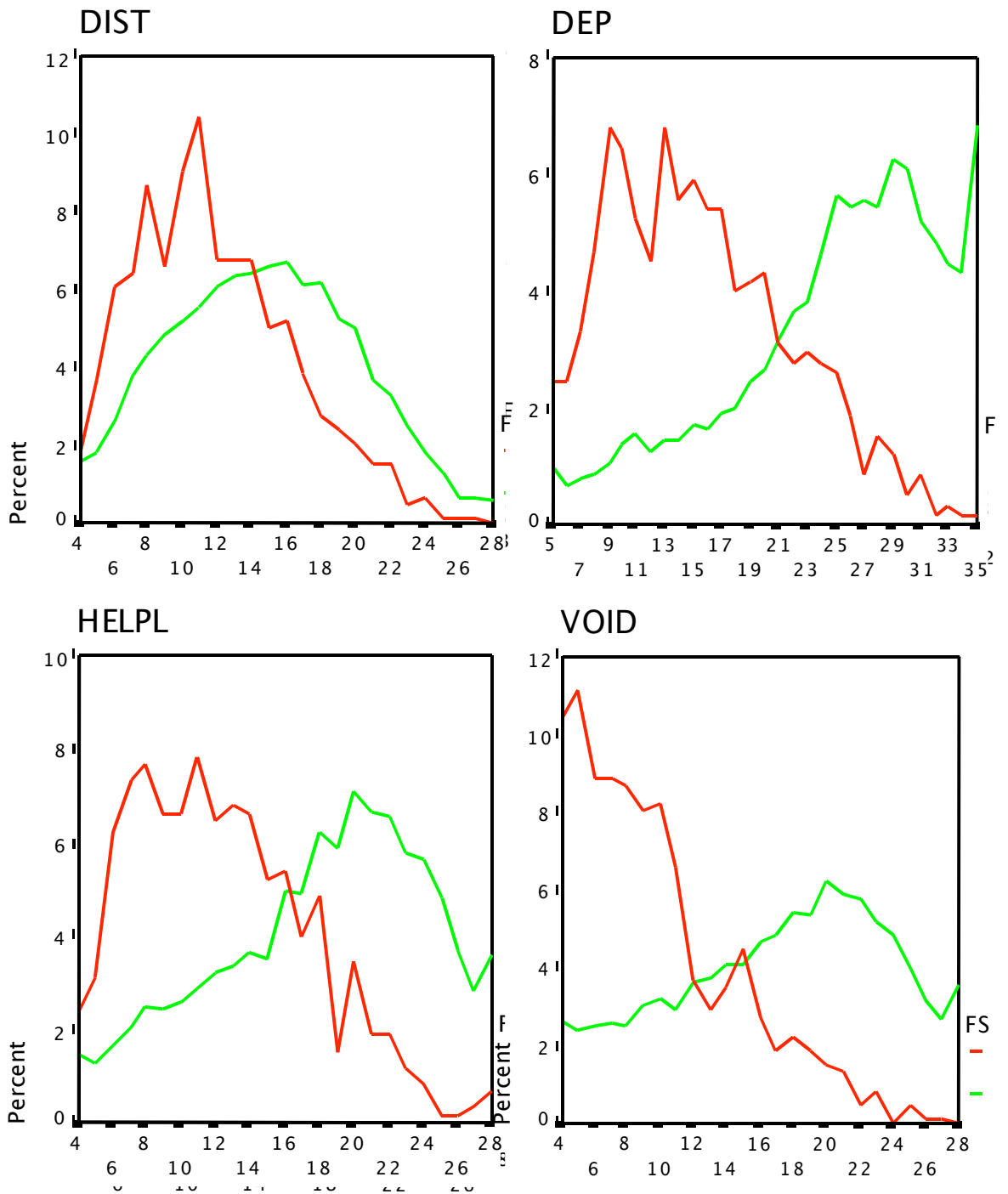
### Comparing the online samples with paper-and-pencil samples

Probably the most striking finding from SWB samples 1 and 2 is that quite persistently those online samples scored dramatically higher on burnout scales than did earlier samples, using paper versions of the HBI (see Figure 1).

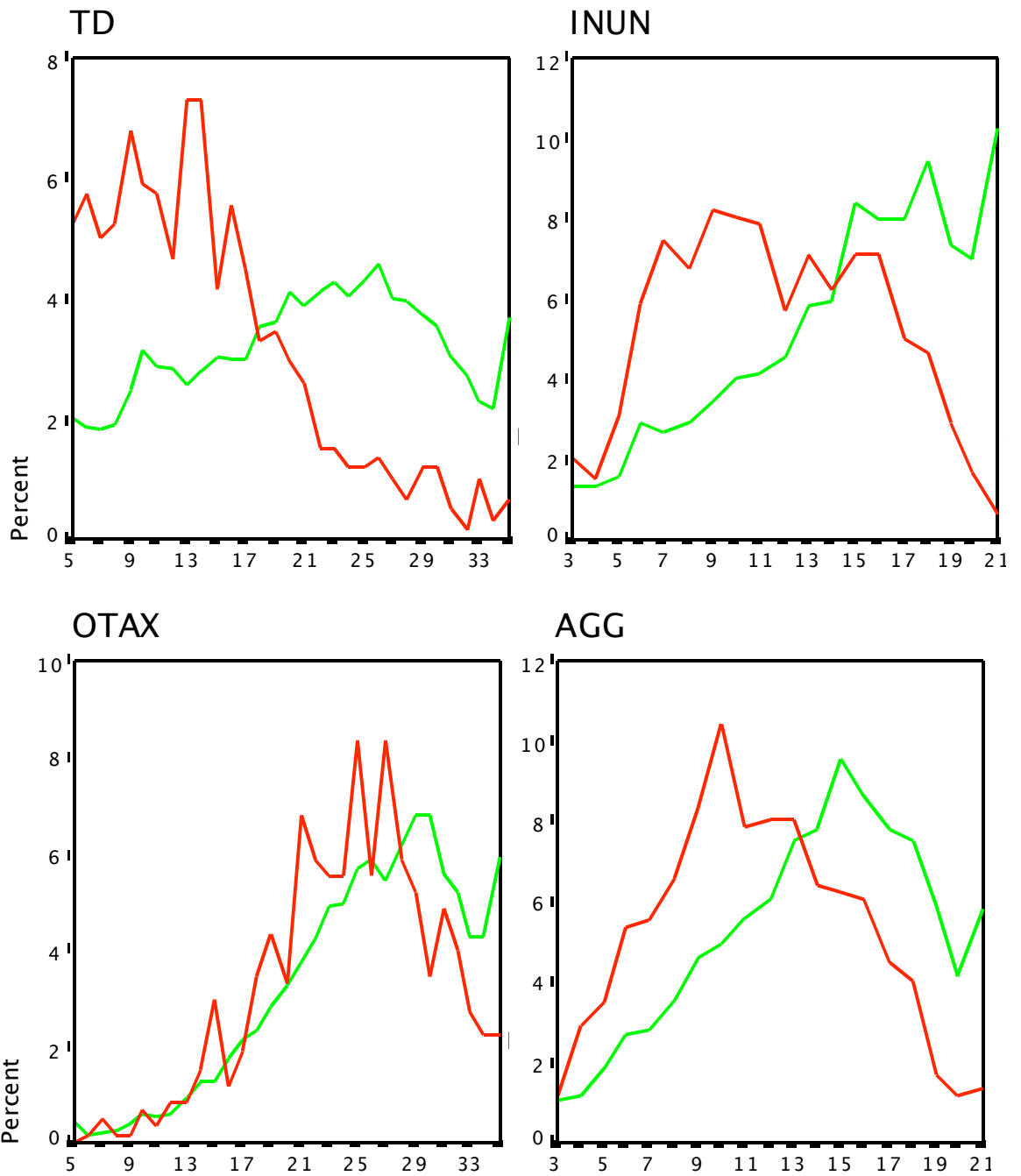
**Figure 1**  
**Comparing Samples Frühauf, Steinhoff, & Hagge with Sample SWB 2 (German only) on all HBI scales**



*Legend:* Red curve = combined sample Frühauf, Steinhoff & Hagge (N = 572; all paper-and-pencil). Green curve: Sample SWB 2, DE only (N = 14123, all online).



*Legend:* Red curve = combined sample Frühauf, Steinhoff & Hagge (N = 572; all paper-pencil). Green curve: Sample SWB 2, DE only (N = 14123, all online).



*Legend:* Red curve = combined sample Frühauf, Steinhoff & Hagge (N = 572; all paper-pencil). Green curve: Sample SWB 2, DE only (N = 14123, all online).

With the exception of scale PA, the *green* curve (sample SWB 2, German version only; N = 14123) indicates much higher means than the *red* curve (samples Frühauf, Steinhoff, and Hagge combined; N = 572). These differences are among the most pronounced ones this writer has ever seen in several decades of empirical research. To exemplify, consider Item 40 (the CRISIS item). Whereas earlier on, only 3% of respondents answered 7 („fully agree“) to item 40 („I am in the middle of a crisis of which I find no way out“) — the lowest percentage of the seven response categories —, a full 21.4% of the German SWB 2 subsample did so, making this the most frequent response!



Note also the more or less marked upward spike at the right hand end of the green curves above. This is an example of a ceiling effect on the scales. It means that a sizable proportion of SWB's online HBI users would have been prepared to describe their lot in even more negative ways, had test items been provided to do so.

*Why?* From hindsight, this may seem unsurprising. After all, users of *SwissBurnout*'s website, even if they got there inadvertently (i.e., not by having typed *burnout* into *Google*), must have possessed enough curiosity and invested the ten minutes to fill in the questionnaire. It is plausible that most of them did *not* do so out of sheer curiosity. In contrast, participants in the studies by Frühauf, Steinhoff, and Hagge (see above) were approached by research assistants who had been warned not to mention the „B word“, in order not to wake any sleeping dogs. Thus, although the representativeness of these early samples may be questioned — they were acquaintances or acquaintances of acquaintances of psychology students living in Northern Germany, and students were overrepresented as subjects — they most probably came closer to the „statistical norm“ in Germany. At any rate, they were not systematically preselected in terms of burnout.

Before we settle for this explanation, let us examine two competing ones. One is that the HBI item content might somehow have „aged“ between the early nineties and 2007. (This would more easily explain a *decline* in burnout scores, though.) Another is that living and working conditions in pertinent areas have sufficiently deteriorated over the past 15 years to be reflected in rising scores.

Fortunately, the data collected by Hagge in 2004 and 2005 are at hand to throw some light on these alternative explanations. Hagge's (2005) German-only sample of 77 adults (53% female; mean age = 39.7 yr.) was biased in terms of education: No less than 68% had graduated from high school („Abitur“) and 39% even held some academic degree. However, no students were included and socio-economic composition resembled that of SWB sample 2 (where pertinent information is available) more closely than the early samples. Does this contemporary sample exhibit signs of *more* burnout than in the early samples?

As Table 4 shows, the answer is a quite unequivocal *no*. There were only three significant differences, and two of these pointed out *less* burnout in the present decade than in the former. There are some mean differences to the contrary, namely *more* Emotional Exhaustion, *more* Inner Void, *more* Inability to Unwind, *more* Overtaxing Oneself in the 2000s. But there are also differences in the opposite direction: *Less* Distancing, *less* Helplessness, *less* Tedium. And all of them are insignificant, i.e., too small not to be explained by chance (sampling error).

**Table 4**  
**Comparing Early Samples Frühauf & Steinhoff with 2005 Sample Hagge**

	Samples Frühauf & Steinhoff		Sample Hagge		<i>p</i>
	Mean	SD	Mean	SD	
<b>01 Emotional Exhaustion</b>	15.4	6.4	16.3	6.6	n.s.
<b>02 Personal Accomplishment (rev.)</b>	9.0	3.2	10.9	2.7	.00
<b>03 Distance</b>	11.9	4.6	11.4	4.6	n.s.
<b>04 Depressive Reaction to Stress</b>	11.0	3.8	9.7	3.6	.01
<b>05 Helplessness</b>	12.5	5.2	12.0	4.4	n.s.
<b>06 Inner Void</b>	9.8	5.0	9.9	4.6	n.s.
<b>07 Tedium</b>	14.5	6.9	12.9	6.0	n.s.
<b>08 Inability to Unwind</b>	11.5	4.4	12.3	3.9	n.s.
<b>09 Overtaxing</b>	24.3	5.8	25.4	5.8	n.s.
<b>10 Aggressive Reaction to Stress</b>	11.6	4.2	10.5	3.9	.05

*Legend:* Combined sample Frühauf & Steinhoff (N = 495; all paper-pencil). Sample Hagge, (N = 77; all paper-pencil). *p* = type I error probability (two-sided).

How about SWB sample 1(all German)? Frequency distributions (not shown here) are strikingly similar to the SWB 2 German subsample, with the exception of having slightly smaller means in all cases. (This latter effect is consistently significant — to be expected with samples that large — but not very strong.)

We may thus conclude with some level of confidence that the dramatically high burnout scores of SWB website users reflect a „real“ self-selection phenomenon: People who visit the website suspect they are in a burnout process. To verify this, they take the test. A portion of them get confirmation of their hunch.

Either that situation tends to get more serious or/and more „hard-core users“ are attracted to the SWB website. (Many caveats!)

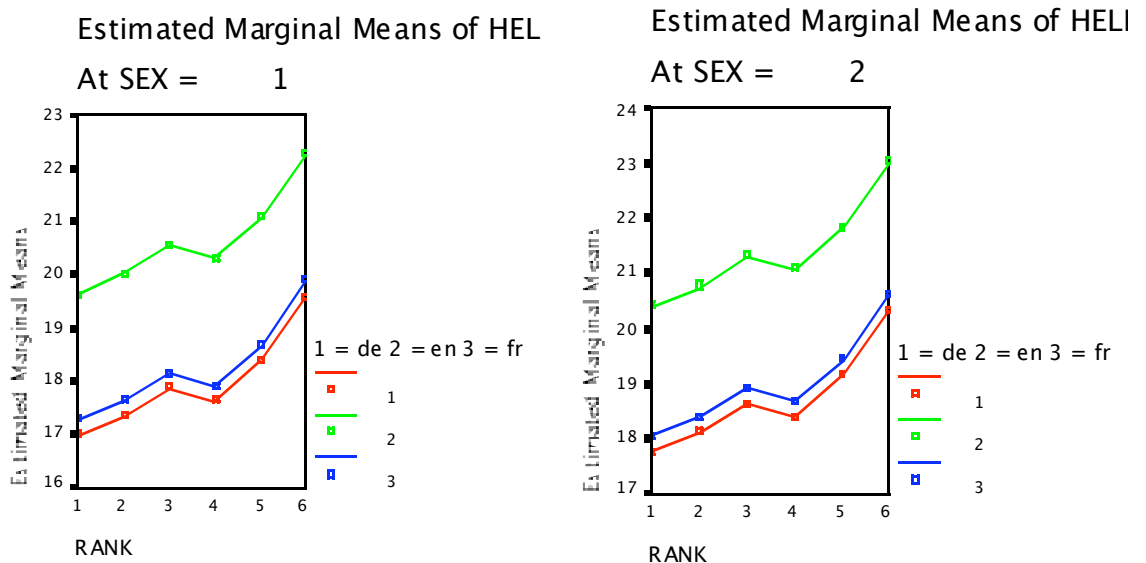
### **Comparing the German, English, and French subsamples of SWB Sample 2**

Only SWB Sample 2 contains data from all three language-specific versions of the HBI, whereas SWB Sample 1 was all German, as were the samples collected by Frühauf, Steinhoff, and Hagge. Thus, SWB 2 provides us with the first chance to examine at national or cultural differences in terms of burnout. Unfortunately, but inevitably, those are inextricably confounded with linguistic differences.

The most conspicuous finding is that EN language means are almost everywhere highest for HBI scales and the CRISIS item (the exception being scale VOID, where DEs score is

the highest). In contrast, DE and FR means are generally pretty close, with DE means being the second highest for six scales while FR means are the second highest for three scales. This finding, although not reflecting a strong effect, looks pretty consistent at first glance. It holds for both sexes and all six status categories and persists after controlling for age,.

Fig. 2 provides a typical example, the dependent variable being scale Helplessness. It is also typical for the effect of status, to be discussed below.



**Figure 2**  
**HELPLESSNESS as a Function of Language, Status, and Sex**

*Legend:* Left panel men, right panel women. Green curve = English, blue curve = French, red curve = German. Rank = status group (cf. Tab. 2)

Does the somewhat special position of the EN subgroup exist also at the item level? For 30 out of 40 HBI items, it does, again controlling for age, sex and status.

The effects of language on HBI scales and the CRISIS item are highly significant, but numerically very weak. Effect size coefficients eta-squared range from a low of .006 (for scales Void and the CRISIS item) to .035 (for scale PA). What is more impressive is the consistency across scales and the lack of interactions with sex and status.

Possible explanations, not mutually exclusive, include: (a) There is a genuine effect, i.e., users from English-speaking parts of the world tend to score higher on burnout; (b) the majority of EN respondents were from North America (probably mostly the US) and England, where living and working conditions may be more conducive to burnout than elsewhere<sup>2</sup>; (c) the translation into English is non-equivalent to the DE and FR version.

In order to weigh the merits of explanation (b), separate ANCOVAs (with age as a covariate) were run, excluding all respondents who had indicated they were from North America

<sup>2</sup>I am indebted to Beate Schulze, Zurich, for providing that hunch.

(N = 373) or England (N = 108 ). Although that left less than half of the EN subsample, lines in the graphs tended to run parallel as before (with the exception of the CRISIS item, where complex interactions took place). And, although the differences tended to diminish somewhat, the EN curves again ran highest for six of the ten HBI scales (the remainder highlighting the DE subsample).

So, although there seems to be something to the North America/England hypothesis, explanations (a) and (c) cannot be ruled out.

### **Gender Differences**

In accord with previous research (Rösing, 2003, p. 94-95), there was not much of a difference in the way men and women described themselves in terms of burnout. Eta squared ranged from zero (for scale Void and the CRISIS item) to .015 (for scale Dist). Females scored somewhat higher on seven of the ten HBI scales; males scored somewhat higher on the remainder (namely, scales PA, DIST, and TD). This held for all three languages and for all six status groups (age controlled). For the CRISIS item, there were interactions; thus, no clear picture emerged.

All of this, of course, may still be a selection effect. Use of the Internet to this date differs for the sexes, and it probably differs differentially in various nations and language groups. Thus, the above differences, very small to begin with, may be explained that way.

### **Age Differences**

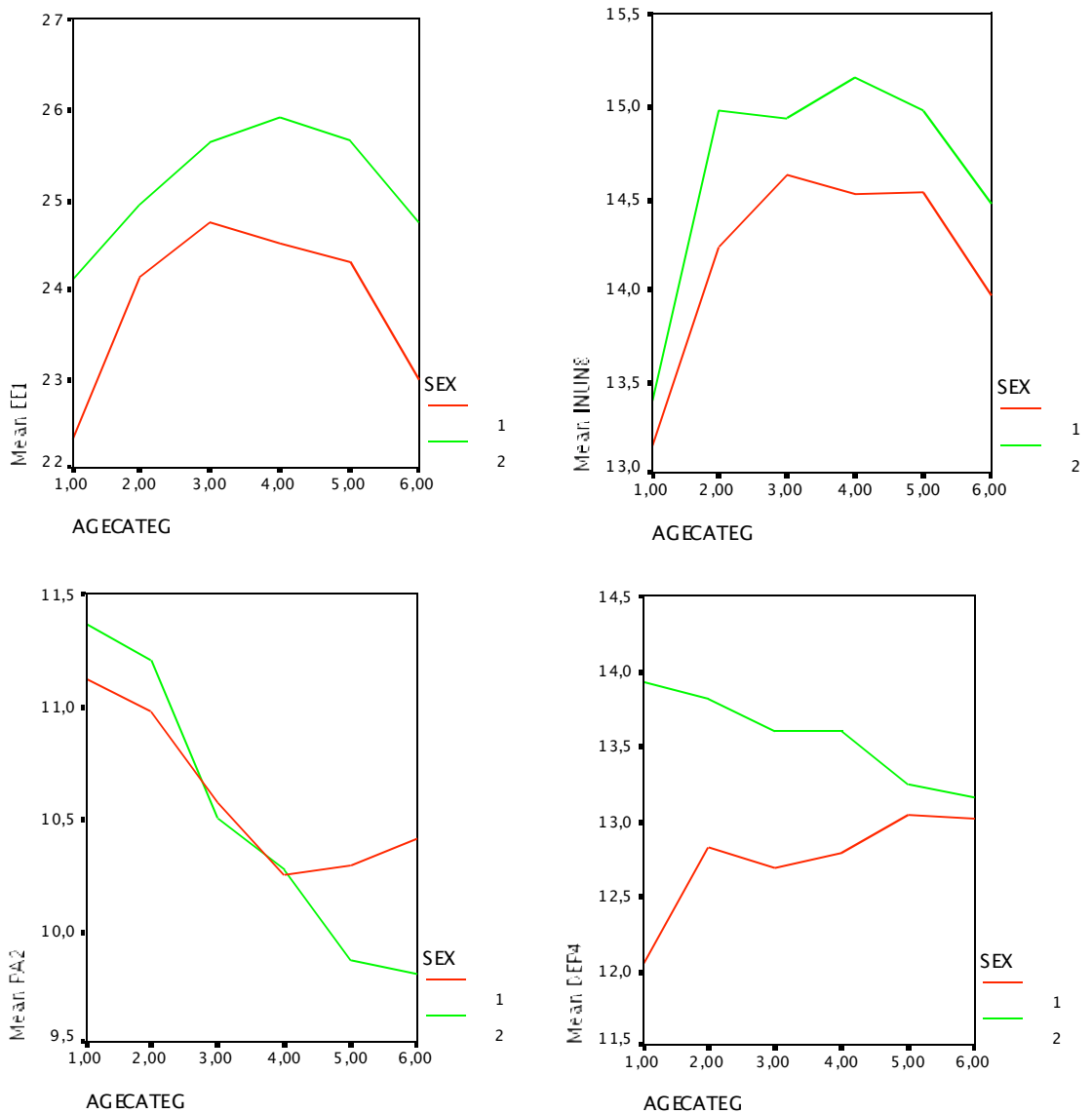
Product-moment correlations between HBI measures and age were minimal, the maximum being -.08 for PA. However, grouping age into decades (<20 = 1; <30 = 2; ..... <70 = 6, omitting the 22 respondents over 69) and plotting means against age categories showed some interesting nonlinear trends, of two types. For scales EE, DIST, HELPL, VOID, INUN, OTAX, AGG, and the CRISIS item, the curves were inversely U-shaped, with maxima in the twenties, thirties, or forties. See Fig. 3 for two examples (top panels). The other type, a more or less monotonic decrease with age, was found for PA and TD (see Fig. 3, lower left panel). For Scale DEP, the female curve was monotonic decreasing, the male curve more or less monotonic increasing (see Fig. 3, lower right panel). Note that the distances between maxima and minima are small, though. Again, the consistency of those trends across scales is more impressive than the size of the effects.

### **Differences Between Professional Status Groups**

Although the same reservations apply here — executives who take the time to fill in the HBI on SWB's website may differ from their unemployed counterparts in more than *one* aspect — the sheer similarity of the plots for status (separately for language and sex, and controlled for age) is stunning (for an example, cf. Fig. 2). With two exceptions, namely scales Inability to Unwind and Overtaxing Oneself, groups 6 („unemployed“) or/and 5 („other“) scored highest on the ten HBI scales and the CRISIS item. Of the nine cases in question, eight saw the Unemployed at the top.

How about minima? Well, that picture looks even clearer: Status group 1 („employee with executive functions“) scores lowest on eight of the ten HBI scales and the CRISIS item.

**Figure 3**  
**Age Trend Examples for Four HBI Scales**



*Legend:* Top left panel: scale EE; top right panel: scale Inun; bottom left panel: Scale PA; bottom right panel: scaleDep. Green curve = Female, red curve = Male. Abscissa: Age decades from teens to sixties.

The exceptions: Scale Inability to Unwind sees status groups „Employee with Line Responsibility“ and „Other“ at the top, and „Employee without Line Responsibility“ at the bottom. With scale Overtaxing Oneself, we see „Independents“ up front, while „Employees without Line Responsibility“ seem to suffer least.

Although admittedly after the fact, that makes some sense at first blush.

Some attention to the „Other“group may be warranted. After all, they scored second highest on nine out of eleven burnout indicators and highest on one (EE). Who are those Others?

I had expected to find students and homemakers, but that was definitely not the whole story. It turned out that the majority of the 1883 respondents who had assigned themselves to the *Other* category would have better fitted one of the defined categories. There were public servants („Beamte“), teachers, apprentices, consultants, firefighters and any number of jobs which in all likelihood were either employees without line responsibility or independent. But how to explain the conspicuously high burnout means?

After inspecting the job codes respondents had typed in (which included „blabla“ and „xxx“, of course), four clusters seemed sufficiently frequent to use them in an analysis: *students* (including pupils, doctoral students, and apprentices; N = 540), *housewives* (all who mentioned they were housewives or homemakers or mothers; N = 104), *teachers* (from Kindergarten teachers to school principals; all „education“; N = 343), and *public servants* („Beamte“; including many police officers; N = 93).

Thus, those four subcategories of „Other“ were compared with the rest of the „Others“ and with the vast remainder of SWB Sample 2 which includes categories 1-4 and 6 (cf. the list in Tab. 2).

The results supported two hunches this writer has long held: Two categories of burnout victims go largely unnoticed, namely housewives and students (in the wide sense). Housewives (by definition all female; there was only a handful of housemen) scored highest in their gender category on scales Emotional Exhaustion, Depressive Reaction, Helplessness, Void, Tedium, Aggressive Reaction, and the CRISIS item, often markedly. Students (both male and female) scored highest on scales Personal Accomplishment and Overtaxing One-self. Moreover, male students obtained the highest means among other males for scales Depression, Helplessness, Tedium, and the CRISIS item.

Also ran and got to the top of their gender field: male teachers for scale Emotional Exhaustion, and teachers of both genders for Inability to Unwind. Again, this does not seem too surprising.

Of course, we cannot rule out the possibility that it was just the most desparate housewives who took the HBI, whereas all the others live happily and will do so ever after. To study questions as these, other — much more costly — research approaches are indicated.

## Conclusions

The HBI appears to be a promising instrument for assessing individual and group levels of burnout. In view of the very short scales, reliabilities are adequate and validities mostly adequate.

This study explored for the most part one large international online sample. What emerged, however, cannot provide more than hunches, given the non-random character of the sample and its sub-samples.

## References

- Aronson, E., Pines, A. M., & Kafry, D. (1983). *Ausgebrannt*. Stuttgart: Klett-Cotta.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae*. Göttingen: Hogrefe.
- Burisch, M. (1984a). Approaches to personality test construction: A comparison of merits. *American Psychologist*, 39, 214-227.
- Burisch, M. (1984b). *The Maslach Burnout Inventory and the Tedium Measure: Reliability and validity in a German sample*. Unpubl. ms.
- Burisch, M. (1997). Test length and validity revisited. *European Journal of Personality*, 303-315
- Demerouti, E. (1999). *Burnout. Eine Folge konkreter Arbeitsbedingungen bei Dienstleistungs- und Produktionstätigkeiten*. Berlin: Peter Lang
- Ebbinghaus, M. (1986). *Erfassung von Burnout. Entwicklung und Überprüfung eines Meßinstrumentes für die Anwendung in verschiedenen Berufsbereichen*. Unpubl. thesis, U. Oldenburg, Germany.
- Enzmann, D. (1996). *Gestresst, erschöpft oder ausgebrannt?* Munich: Profil
- Fahrenberg, J. & Selg, H. (1973). *Das Freiburger Persönlichkeitsinventar*. Göttingen: Hogrefe (2nd ed.).
- Fahrenberg, J., Selg, H. & Hampel, R. (1984). *Das Freiburger Persönlichkeitsinventar FPI. Revidierte Form FPI-R und teilweise geänderte Fassung FPI-A1*. Göttingen: Hogrefe.
- Frick, U. & Filipp, G. (1997). *Gesundheitsberufe im Land Salzburg. Berufs- und Lebenssituation*. Salzburg, Austria: Amt der Salzburger Landesregierung.
- Frühauf, F. (1990). *Entwicklung eines deutschen Burnout-Inventars*. Unpublished thesis, University of Hamburg.
- Hagge, M. (2005). *Vergleich zweier deutscher Burnout-Inventare*. Unpublished thesis, University of Hamburg.
- Maslach, C. & Jackson, S.E. (1986). *Maslach Burnout Inventory Manual (2d. ed.)*. Palo Alto, CA: Consulting Psychologists Press.
- Maslach, C., Jackson, S.E. & Leiter, M.P. (1996). *Maslach Burnout Inventory Manual (3d. ed.)*. Palo Alto, CA: Consulting Psychologists Press.

Rösing, I. (2003). *Ist die Burnout-Forschung ausgebrannt?* Heidelberg: Asanger.

Schaufeli, W.B. & Enzmann, D. (1998). *The burnout companion to study & practice*. Chichester: Taylor & Francis..

Winer, B. J. (1962). *Statistical principles in experimental design*. New York: McGraw-Hill.



