Psychological Test and Assessment Modeling, Volume 54, 2012 (2)

Contents

Psychological Test and Assessment Modeling (Psychol Test Assess Model)

formerly Psychologische Beiträge/Psychology Science *Quarterly*

Indexed in **Directory of Open Access Journals, Infoconnex** (full text)**, PsychInfo, Psyndex, Scopus**

Self-evaluated impact factor 2010: 0.475 (according to Kubinger, Heuberger & Poinstingl, 2010)

Editor:
Klaus D. Kubinger
Head of Division for Assessment and Applied Psychometrics
Department of Developmental Psychology & Psychological Assessment
Faculty of Psychology
University of Vienna
Liebiggasse 5
A-1010 Vienna (Austria)
klaus.kubinger@univie.ac.at

Scientific Office Manager:
Bettina Hagenmüller
p-t-a-m@univie.ac.at

PABST SCIENCE PUBLISHERS
Eichengrund 28
D-49525 Lengerich
Tel. +49 (0) 5484 – 308
Fax +49 (0) 5484 – 550
Postgiro Kto. Köln 208157-503
Internet: www.pabst-publishers.com
www.psychologie-aktuell.com

Publishers Office: Armin Vahrenhorst
Producer: Armin Vahrenhorst
vahrenhorst@pabst-publishers.com

Subscription price per year:
(4 issues) 60,- EUR/USD, incl. postage and VAT, single copy 15,- EUR/USD, incl. postage and VAT.

ISSN 2190-0493 (Print)
ISSN 2190-0507 (Internet)
www.psychological-test-and-assessment-modeling.com

Psychological Test and Assessment Modeling, Volume 54, 2012 (2), 103-127

Examining playfulness in adults: Testing its correlates with personality, positive psychological functioning, goal aspirations, and multi-methodically assessed ingenuity

René T. Proyer[1]

Abstract

The prime aim of this set of studies was to test the disposition to play (playfulness) in adults in its relation with various measures of personality but also ability (self-estimated but also psychometrically measured ingenuity). Study 1 (n = 180) shows that adults playfulness relates primarily to extraversion, lower conscientiousness, and higher endorsements of culture; joy of being laughed at (gelotophilia) and Agreeableness were also predictive in a regression analysis; Study 2 (n = 264) shows that playfulness relates primarily to a high expectation of intrinsic and a low expectation of extrinsic goals as well as greater intrinsic and lower extrinsic importance of goals (for expressive and fun-variants of playfulness); Study 3 (n = 212) shows that playfulness relates to greater self-perception of one's degree of ingenuity and psychometric ingenuity correlated primarily with greater spontaneous and creative variants of playfulness (in about the same range for origence and fluidity of the productions). Overall, the findings were in line with the expectations and could stimulate further studies of playfulness in adults.

Key words: adult playfulness; divergent thinking; life goals; life satisfaction; well-being

[1] *Correspondence concerning this article should be addressed to:* Dr. René Proyer, Department of Psychology, University of Zurich, Binzmühlestrasse 14/7, 8050 Zurich, Switzerland; email: r.proyer@psychologie.uzh.ch

A set of three studies has been designed for contributing to the basic understanding of playfulness in adults. This is an understudied field in psychology and related disciplines – most certainly in comparison to research on play and playfulness in children. The emergence of *positive psychology* has stimulated research in variables that may contribute to positive psychological functioning and well-being and it is argued that playfulness may have such a function in adults.

Although there are efforts for structuring the field, (e.g., Apter, 1982; Barnett, 1990, 2007; Bundy, 1993; Guitard et al., 2005; Proyer, in press) a consensual theory of adult playfulness is missing. In the present studies, playfulness is understood as the predisposition to engage in playful activities and interactions (Barnett, 1991ab). Barnett (2007) defines playfulness as: "[…] the predisposition to frame (or reframe) a situation in such a way as to provide oneself (and possibly others) with amusement, humor, and/or entertainment. Individuals who have such a heightened predisposition are typically funny, humorous, spontaneous, unpredictable, impulsive, active, energetic, adventurous, sociable, outgoing, cheerful, and happy, and are likely to manifest playful behavior by joking, teasing, clowning, and acting silly" (p. 955).

In comparison with studies on children (e.g., Barnett, 1990, 1991ab; Lieberman, 1977), only a comparatively few number of studies has dealt with playfulness in *adults*. Of these, robust relations between playfulness and greater inclination to flow-experiences (Csikszentmihalyi, 1975), intrinsic motivation (Amabile et al., 1994), instrumental and expressive traits (Bozionelos & Bozionelos, 1999), quality of life (Proyer et al., 2010), higher intellectual and emotional strengths of character but lower strengths of restraint (Proyer & Ruch, 2011), creativity and spontaneity (Barnett, 2007; Glynn & Webster, 1992, 1993), stress coping (Qian & Yarnal, 2011), positive attitudes towards the workplace (Glynn & Webster 1992, 1993; Yu et al., 2007), job satisfaction and performance, innovative behavior (Yu et al., 2007), and academic achievement (Proyer, 2011) have been reported. Thus, there is evidence that playfulness as a personality characteristic may play an important role amongst adults.

Assessing adult playfulness. There is a lack of consensus on how to measure playfulness in adults. Different measures exist and one frequently used instrument is Glynn and Webster's (1992) *Adult Playfulness Scale,* which allows testing five different facets of playfulness. The dimensions are *spontaneous* (e.g., impulsive, carefree), *expressive* (e.g., bouncy, open), *fun* (e.g., excitable, bright), *creative* (e.g., imaginative, passive), and *silly* (e.g., childlike, whimsical). An advantage of this instrument is that it allows scoring for a total score and five subscales (spontaneous, expressive, fun, creative, silly) that help describe the correlates of different variants of playfulness. The APS has been used in all of the studies described here. Additionally, the *Short Measure of Adult Playfulness* (SMAP; Proyer, 2012) has been used for the assessment of the global, cognitive aspect of playfulness. Higher scores in the SMAP indicate an easy onset and high intensity of playful experiences along with the frequent display of playful activities.

Three studies were conducted to test (a) personality correlates of playfulness in adults (in terms of the Big Five); (b) its relation to positive psychological functioning (life satisfaction and orientations to happiness); (c) its relations with humor and laughter in terms of three dispositions towards ridicule and being laughed at); (d) intrinsic and extrinsic life goals; and (e) the self-perception of one's degree of ingenuity and psychometric (objectively tested) ingenuity (creativity). Study 1 covers topics (a) to (c); Study 2 deals with topic (d) and Study (3) addresses (e). The APS has been used in study 1 to 3. Since the SMAP has not been available at the time study 1 has been planned and conducted, it has only been used in studies 2 and 3.

Personality correlates of adult playfulness. Studies with children (Barnett, 1991ab) but also work conducted with adults (e.g., Glynn & Webster, 1992, 1993; Proyer, 2012) allow deriving hypotheses on the personality correlates for playfulness in adults. Barnett's (2007; see also Barnett, 2011) definition of playfulness and the characteristics given there to describe a playful adult suggest a picture of a person that is rather extraverted, emotionally stable, and agreeable. Further expectations derived from literature are higher openness towards new experiences (or: culture) and lowered conscientiousness among playful adults. The latter assumptions are supported, for example, by data from Glynn and Webster (1993) who found a correlation of $r(548) = -.36$ to *orderly personality* and of .26 to *innovative attitudes* (all $p = .000$) in a sample of highly intelligent adults.

Playfulness and positive psychological functioning. It has been argued that play and playfulness in adults may facilitate the experience of positive emotions (Fredrickson, 1998, 2001) or positive states such as flow-experiences (Csikszentmihalyi, 1975). Among the indicators of playfulness in young adults, Barnett (2007) lists cheerful and happy. A positive relation between playfulness and life satisfaction was expected. Proyer, Ruch, and Müller (2010) found a positive relation between playfulness for various indicators of *quality of life* in a sample of elderly people. However, as life satisfaction is only one facet of the subjective *well being* a closer look at different facets is warranted. For example, playfulness has previously not been studied in relation to different *orientations towards a good life*. Peterson, Park, and Seligman (2005) describe three distinct routes to happiness: (a) the life of *pleasure* (hedonism); (b) the life of *engagement* (related to flow-experiences); and (c) the life of *meaning* (eudemonia). One might assume that having a playful attitude towards life facilitates seeking pleasures and engaging oneself in activities that enable the experience of flow (the engaged life). While, using ones strengths and talents for a greater good (the *meaningful* life) should not necessarily be related to playfulness. One might also assume that engagement also relates to greater activity and an energetic stance towards daily life, as listed by Barnett (2007) as indicators of playfulness.

Playfulness and dealing with humor and laughter. McGhee (1996) considers humor to be a special variant of playfulness – the play with ideas. In study 1, it was tested whether the way people deal with humor and laughter contributes to the expression of playfulness. Ruch and Proyer (2008, 2009) describe three dispositions in dealing with humor and laughter of others; namely, (a) the fear of being laughed at (*gelotophobia*); (b) the joy of being laughed at (*gelotophilia*); and (c) the joy of laughing at others (*katagelasticism*). Gelotophobes fear being laughed at and appearing ridiculous to others. Although exhibiting playfulness per se does not indicate the risk of being laughed at, one might argue that specific variants of playfulness (e.g., word plays, joking around, etc.) may enhance the risk of being laughed at. It was expected that greater playfulness relates to lower inclinations of fearing to be laughed at. Gelotophiles actively seek and establish situations in which they can make others laugh at them, they do not refrain from telling embarrassing stories or incidents that happened to them for making others laugh. Katagelasticists actively seek and establish situations, in which they can laugh at others, they do not feel bad when doing so but think that those who do not enjoy being laughed at should just fight back. Gelotophilia is associated with playful behavior such as joking, teasing, clowning, and acting silly. This fits well to earlier ideas as, for example, Lieberman (1977) mentions the ability of *making fun of oneself* as a part of social-emotional playfulness. Therefore, positive relations are expected with primarily gelotophilia. However, one might assume that playful people would not refrain from playfully teasing others – and sometimes might overdo it. Thus, positive but numerically lower relations with katagelasticism are also expected.

Intrinsic and extrinsic life goals and playfulness. To the best knowledge of the author, adult playfulness has not yet been tested in its relation with a classification of intrinsic and extrinsic life goals. Kasser and Ryan (1993, 1996) suggest, "extrinsic goals, such as financial success, are those that depend on the contingent reactions of others. Further, they are typically engaged in as means to some other end. Conversely, intrinsic goals, such as self-acceptance, are expressive of desires congruent with actualizing and growth tendencies natural to humans. As such, intrinsic goals are likely to satisfy basic and inherent psychological needs" (1996, p. 280). From a self-determination theory-perspective these needs are autonomy, relatedness, competence, and growth (Deci & Ryan, 1985). Intrinsic goals (and attainment to them) were also shown to benefit well being (e.g., Niemiec, Ryan, & Deci, 2009). Furthermore, Kasser and Ryan (1993, 1996) distinguish between (a) the importance for the person that a specific goal will happen and (b) the likelihood that this will happen.

Already Dewey (1913) described the intrinsic nature of play when he stated that play encompasses activities "which are not consciously performed for the sake of any result beyond themselves; activities which are enjoyable in their own execution without reverence to ulterior purpose" (p. 725). As playfulness is seen as intrinsically motivated (see Amabile et al., 1994; Berlyne, 1960; Bundy, 1993; van der Kooij, 1989), it was expected to demonstrate robust positive relations with intrinsic life goals. Furthermore, it was expected that greater playfulness relates to valuing the importance as well as the likelihood of intrinsic goals to occur. Extrinsic goals should be of lesser importance for playful adults. Nevertheless, it was expected that the higher perceived likelihood of extrinsic should be related with playfulness. Furthermore, it was hypothetisized that exhibiting fun-variants of playful behavior may also be directed towards a specific aim; for example, for being perceived in a special way – e.g., as being eccentric, extravagant, interesting, or attractive. Thus, there may be an extrinsic gain in this kind of behavior. Exhibiting silly-variants of playfulness is expected to be unrelated from life-goals. This variant is most likely related to short-term intrinsic satisfaction but does not necessarily need to relate to any long-term aspirations. The intrinsic nature of playfulness should, as already mentioned, make flow-experiences more likely to occur and enable the experience of positive emotions.

Adult playfulness and divergent thinking. Recently, Proyer (2011) found playfulness in young adults to be widely *unrelated* from self-estimated and psychometric intelligence. In his sample of psychology undergraduates, playfulness correlated with better academic performance and the students' free choice of doing extra-work that was not necessary for passing an exam. However, in this study, only measures of *convergent* thinking were used (verbal, numerical and figural intelligence and memory) and it was argued that playfulness in adults should also be tested in relation to self-estimated and psychometrically measured *divergent* thinking. This call has been based on both theoretical assumptions (see e.g., Chapman, 1978; March, 1976; Sutton-Smith, 1967) but also on empirical studies (again most of these have been conducted with children). For example, Barnett and Kleiber (1982, 1984) found relations of playfulness in children with divergent thinking with gender being a mediating variable (see also Dansky & Silverman, 1973; Lieberman, 1965, 1977; Rossman & Horn, 1972; Taylor & Rogers, 2001; Truhon, 1983).

In this study, psychometrically measured and self-estimated ingenuity were tested and, additionally, the aspects of the *frequency* (number of the productions) and the *origence* (uniqueness of the productions) of verbal, numeric, and figural productions were considered. Playfulness in adults was expected to relate positively to psychometrically measured and self-estimated ingenuity. Especially, facets of playfulness that are directly linked to ingenuity (i.e., its creative and spontaneous variants) should demonstrate positive relations.

Main aims of the present study. It has been stressed out that there is a lack of research in playfulness in adults. The overarching aim of the three studies was adding to the knowledge in the field by studying adult playfulness in its relation with a diverse range of other variables. Study 1 was aimed at describing personality correlates of adults playfulness along with correlates of three dispositions towards ridicule and being laughed at, and several indicators of positive psychological functioning (i.e., life satisfaction and orientations to happiness). In study 2, the relation of life aspirations (the importance and likelihood of intrinsic and extrinsic goals) with adult playfulness was tested. Finally, study 3 was aimed at testing whether greater playfulness in adults relates to greater psychometrically measured performance in a test for ingenuity but also with self-estimates of the own ingenuity.

Study 1

Method

Sample

The sample consisted of 180 adults (84 males and 96 females) from 18 to 81 years (M = 37.0, SD = 15.3). More than three quarters (78.9 %) were employed and 16.7 % were not working (e.g., retired or unemployed; 4.4 % left the answer blank). The largest group (38.9 %) had full vocational training, 13.3 % were students, and 23.3 % held a degree from a university.

Instruments

The *Adult Playfulness Scale* (APS; Glynn &Webster, 1992) is a 32-item adjective list of which 25 items are being scored on a 7-point scale. In this study a total score and additionally, five subscales were computed; namely, *spontaneous* (e.g., spontaneous vs. disciplined, impulsive vs. diligent), *expressive* (e.g., bouncy vs. staid, open vs. reserved), *fun* (e.g., bright vs. dull, excitable vs. serene), *creative* (e.g., imaginative vs. unimaginative, active vs. passive), and *silly* (e.g., childlike vs. mature, whimsical vs. practical). Glynn and Webster report alpha-coefficients between .73 and .83 for the five scales and results from a principal component analysis (rotated to the Varimax-criterion) in which the five extracted factors explained 57.5 % of the variance. Furthermore, they provide data on convergent and predictive validity, which has been supported in further studies (e.g., Amabile et al., 1994; Bozionelos & Bozionelos, 1997, 1999; Fix & Schaefer, 2005). As in Proyer (in press, 2011, 2012) the German version of the scale has been used. In this study, alpha-coefficients were .88 for the total score and yielded a mean of .68 for the five subscales.

The *Inventory of minimal redundant scales* (MRS-45; Ostendorf, 1990; Ostendorf & Angleitner, 1992) is a bipolar list of 45 adjectives for the assessment of *extraversion* (e.g., impulsive vs. restraint), *agreeableness* (e.g., affirmative vs. oppositional), *conscientiousness* (e.g., diligent vs. lazy), *emotional stability* (e.g., robust vs. vulnerable), and *culture* (e.g., inventive vs. conventional). Answers are given on a six-point scale (*very – quite – rather* for each of the poles). The authors report high internal consistencies and provide support for its validity. The scale is frequently used in the German language area for an economic assessment of the dimensions of the five-factor model (e.g., Dormann & Kaiser, 2002; Elfering et al., 2000; Hülsheger & Maier, 2010). Alpha-coefficients in this sample were between .80 (A) and .91 (ES) with a median of .86.

The *PhoPhiKat-45* (Ruch & Proyer, 2009a) is a 45-item questionnaire for the assessment of *gelotophobia* ("When they laugh in my presence I get suspicious"), *gelotophilia* ("When I am with other people, I enjoy making jokes at my own expense to make the others laugh"), and *katagelasticism* ("I enjoy exposing others and I am happy when they get laughed at"; 15 items each). Answers are given on a four-point scale (1 = *strongly disagree*, 4 = *strongly agree*). Ruch and Proyer report high reliability coefficients (alphas ≥ .84) and retest-reliabilities (≥ .77 and ≥ .73 for a three and six months). The PhoPhiKat-45 has been used widely in research and results generally support its validity (see e.g., Proyer, Platt, & Ruch, 2010; Renner & Heydasch, 2010; Samson & Meyer, 2010). The alpha-coefficients in this sample were .88 (gelotophobia) and .89 (gelotophilia and katagelasticism).

The *Orientation to Happiness* scale (OTH; Peterson et al., 2005) is a questionnaire for the subjective assessment of *life of pleasure* (a sample item is "Life is too short to postpone the pleasures it can provide"), *life of engagement* ("I am always very absorbed in what I do"), and *life of meaning* ("I have a responsibility to make the world a better place"). Each scale consists of six items. It utilizes a 5-point answer format (1 = *very much unlike me*, 5 = *very much like me*). The German version of the OTH was used (Ruch et al., 2010). The OTH and its German-language variant have been used in a broad range of studies and proved to be a valid and reliable measure (e.g., Chen et al., 2010; Peterson et al., 2007; Proyer, Annen et al., in press; Proyer, Ruch & Chen, 2012; Vella-Brodrick et al., 2009). The alpha-coefficients in this sample were .68 (E), .75 (P), and .77 (M), respectively.

The *Satisfaction with Life Scale* (SWLS; Diener et al., 1985) is a five-item measure for assessing satisfaction with life. A sample item is "The conditions of my life are excellent". It uses a 7-point answer form (1 = *strongly disagree*, 7 = *strongly agree*). We used a German translation of the scale that had proved its usefulness and good psychometric properties in previous studies (e.g., Peterson et al., 2007; Proyer, Ruch, & Chen, 2012; Ruch et al., 2010). The alpha-coefficient in this sample was .81.

Procedure

Students attending a course on psychometrics collected the data. They were asked to distribute fifteen questionnaires each among adults (with an about equal number of males and females). The data collection was part of the course requirements. Participants were not paid for their services and were informed that they should complete the questionnaire for helping with a scientific study.

Results

An inspection of skewness and kurtosis of all scales that entered the study indicated that they were normally distributed. Means and standard deviations were comparable to data collected in earlier studies with the respective instruments. Pearson correlation coefficients were computed between the total score on the *Adult Playfulness Scale* as well as its five dimensions with measures of personality, orientations to happiness, satisfaction with life, dispositions towards ridicule and being laughed at as well as demographics (sex, age, educational level, and working status). Table 1 contains the correlation coefficients.

Table 1 shows that playfulness as operationalized in the APS could well be located in the framework of the big five. Greater playfulness was associated with emotionally stability, openness to new experiences, extraversion, and low conscientiousness. Out of the three orientations towards happiness, they endorsed primarily the pleasurable life but also the life of engagement. Playfulness was unrelated to an overall estimation of the satisfaction with life. Furthermore, greater playfulness was related with lower fear being laughed at. However, it was related to the enjoyment of being laughed at and joy in laughing *at* others. Demographics were widely unrelated with playfulness but playful people tended to be younger.

These relations were also stable in parts among the five dimensions of playfulness but additionally, they yielded different relations that should be highlighted. For example, those who expressed fun and low silliness in their playfulness were also higher in agreeableness. Those who were higher in the creative aspects of playfulness endorsed culture

Table 1:
Correlations Among (Facets of) Playfulness and Measures of Personality, Orientations to Happiness, Satisfaction with Life, and Dimensions towards Being Laughed at and Ridicule

	Spontaneous	**Expressive**	**Fun**	**Creative**	**Silly**	**Total**
Personality						
ES	.19*	.07	.27*	.37*	-.11	.20*
A	-.03	.02	.21*	.15	-.22*	.02
Cu	.25*	.27*	.34*	.65*	.07	.40*
E	.44*	.59*	.41*	.36*	.18*	.53*
Co	-.42*	-.16*	-.15	.23*	-.52*	-.30*
Happiness						
Pleasure	.32*	.31*	.16	.13	.36*	.34*
Engagement	.23*	.21*	.11	.23*	.12	.24*
Meaning	.09	.10	-.04	.14	.11	.11
Life satisfaction	.13	.10	.31*	.18*	-.10	.15
Laughter						
Gelotophobia	-.25*	-.10	-.34*	-.22*	.01	-.24*
Gelotophilia	.44*	.27*	.23*	.04	.39*	.38*
Katagelasticism	.33*	.20*	-.01	-.06	.40*	.25*
Demographics						
Sex	-.06	.10	.10	.05	-.08	.03
Age	-.29*	-.22*	-.14	.03	-.43*	-.31*
Education	.05	.11	.15	.18*	-.05	.10
Working	-.03	-.06	.10	.03	-.09	-.02

Note. n = 163-180. Personality = MRS-45 (ES = Emotional Stability, A = Agreeableness, Cu = Culture, E = Extraversion, Co = Conscientiousness); Happiness = OTH (Life of Pleasure, Life of Engagement, Life of Meaning); Life satisfaction = SWLS; Laughter = PhoPhiKat-45; sex (1 = male, 2 = female), education (1 = no school to 8 = doing a or having a Phd), working (1 = yes, 2 = no).

* $p < .05$.

strongly (r^2 = .42) and were higher in emotional stability than others but they also yielded (in contrast to the other dimensions) a positive relation to conscientiousness. Those who liked spontaneous and silly expressions of playfulness were lowest in conscientiousness (r^2 between .18 and .27). Higher satisfaction with life was reported from those who pursued fun and creative playfulness. Enjoying to be laughed at by others and to laugh at others was primarily found among those who prefer *spontaneous* and *silly* variants of playfulness. Finally, the latter and *expressive* forms of playfulness were primarily found among younger participants.

Predicting playfulness. A multiple regression was computed (not reported in full detail here) with the total score of the APS as criterion and age along with the other scales as predictors. In a first step, age was entered in the regression and in a second step, personality, the dispositions towards ridicule and being laughed at, the orientations to happiness, and satisfaction with life were added (*stepwise*) to the equation. The analysis yielded a multiple correlation coefficient of .57 (F[6, 153] = 32.96, p = .000). In the final model, age did not contribute to the prediction. Extraversion entered the equation first (ΔR^2 = .25; final model: β = .40, p = .000), followed by low conscientiousness (ΔR^2 = .10; β = -.42, p = .000). Then culture (ΔR^2 = .09; β = .34, p = .000), gelotophilia (ΔR^2 = .02; β = .17, p = .004), and agreeableness entered (ΔR^2 = .01; β = .13, p = .044). When doing the regressions separately for each of the predictors (for an estimate of the individual contribution of these predictors), there was a multiple correlation coefficient of R^2 = .59 for personality (F[8, 147] = 24.71, p = .000) with extraversion (β = .46) entering first followed by conscientiousness (β = -.46), culture (β = .33) and agreeableness (β = .13; all coefficients for the final model). The three dispositions towards ridicule and being laughed at yielded an R^2 = .26 (F[6, 150] = 8.52, p = .000) with gelotophobia (β = -.35) entering the equation first followed by katagelasticism (β = .29). Finally, the three orientations to happiness yielded a R^2 of .21 (F[5, 147] = 7.35, p = .000). In this analysis, younger age contributed significantly to the prediction of playfulness (β = -.25) followed by the pleasurable life (β = .32)[2].

Discussion

The results of the study converged well with the predictions derived from Barnett's (2007) definition of playfulness. Playfulness was most strongly related to extraversion, low conscientiousness, and higher endorsements to culture. Also, gelotophilia and agreeableness were predictive in a regression analysis – yet with a low incremental contribution to the prediction. Hence, they seem to be of lower practical relevance. Also, emotional stability (ES) seemed to be of lesser importance, though correlations point in a direction of higher ES in playful adults. Neither demographics nor fearing to be laughed at (gelotophobia) or enjoying laughing at others predicted playfulness.

[2] The pattern of correlations given in Table 1 suggested that each of the single dimensions of the APS related somewhat differently to the variables tested in this study. Therefore, regressions were conducted with each of the five scales of the APS as criterion and the same specifications as described above. The results are not shown in detail here but suggest that mainly conscientiousness along with extraversion and culture turned out to be potent predictors; *spontaneous playfulness* (R^2 = .51, F[6, 153] = 13.51, p = .000; most important predictors in the final model: Culture [β = .51] and extraversion [β = .34]); *expressive playfulness* (R^2 = .46, F[9, 140] = 15.55, p = .000); most important predictors: extraversion [β = .60] and emotional stability [β = .26]); *fun-oriented playfulness* (R^2 = .40, F[7, 151] = 15.39, p = .000; most important predictors: Conscientiousness [β = -.40], extraversion, agreeableness, culture and life satisfaction with beta-weights between .25 and .29); *Creative playfulness* (R^2 = .53, F[3, 151] = 54.63, p = .000; predictors were age [β = .17], culture [β = .67] and extraversion [β = -.17]); and *silly*-variants of playfulness (R^2 = .48, F[4, 153] = 33.75, p = .000; predictors were age [β = .20], conscientiousness [β = -.40], life of pleasure [β = .25], and katagelasticism [β = -.19]).

At the level of bivariate correlations, the pleasurable life as well as an engaged life demonstrated robust relations with adult playfulness. There was an overlap of 21 % in the variance of the three orientations to happiness with playfulness. These findings are in accordance with the predictions. Hence, hedonism as well as engagement (related to flow-experiences) are associated with playfulness in adults. This seems crucial, as it could be a hint towards a positive impact of playfulness in everyday life. Pending further studies one might discuss whether fostering playfulness in adults (cf. McGhee, 1996, 2010) could be a strategy for boosting flow-related experiences, which, in turn, could have a positive impact on a person's subjective well-being. However, this is at the level of speculations at the moment.

Silly variants of playfulness (childlike, whimsical, frivolous, unpredictable) were primarily predicted by low conscientiousness, the pleasurable life (hedonism) and lower enjoyment of laughing at others. Based on these data, it can only be speculated whether a less ordered environment allows for (silly) playful behavior or whether playfulness leads to less conscientious behavior (or what kind of interaction there may be). However, it seems worth following this line of research. Especially, it would be interesting to know whether this lower expression of conscientiousness in playful adults generalizes into all areas of their life (e.g., leisure time, work time, relationships etc.) or whether this is restricted to specific areas and tasks (e.g., those that are less challenging or pursued just for fun). In study 2, the relations of adult playfulness with intrinsic and extrinsic life goals were tested.

Study 2

Method

Sample. The sample consisted of 268 adults. Two were 17 years old and the others were between 18 and 65 (M = 29.0, SD = 9.1). Slightly more than one quarter were males (n = 69; 25.7 %). More than a third (n = 94; 35.1 %) held a degree from university or were currently students, while 48.9 % (n = 131) had a degree from school that would allow them to study. About one fifth (n = 55; 20.5 %) reported being married.

Instruments

The *Short Measure of Adult Playfulness* (SMAP; Proyer, 2012) is a five-item questionnaire for the assessment of playfulness in adults. It was developed for providing a global, cognitive self-description of playfulness. A sample item is "I am a playful person". Higher scores in the SMAP indicate an easy onset and high intensity of playful experiences along with the frequent display of playful activities. The five items are positively keyed. Answers are given on a 4-point answer format (1= *strongly disagree*, 4 = *strongly agree*). Proyer reports best fit for a one-dimensional solution of the data (in exploratory and confirmatory factor analyses) and high internal consistencies (≥ .80 in three different samples). Also, data on the convergent validity (three other indicators of playfulness), divergent validity (with a seriousness scale), and a more experimental task (ratings for workplaces and pieces of art) support the overall validity of the instrument. The alpha-coefficient in this study was .86.

As in Study1, the *Adult Playfulness Scale* (APS; Glynn & Webster, 1992) has been used. Alpha-coefficients in this sample were .74 (*spontaneous*); .66 (*expressive*); .65 (*fun*); .66 (*creative*); and .69 (*silly*).

The *Aspirations Index* (AI; Kasser & Ryan, 1993, 1996; in the German version by Klusmann, Trautwein, & Lüdtke, 2005) assesses the *importance* ("How important is this to you?") and the *likelihood* ("How likely is it that this will happen in your future?") with 35 items each. Answers are given on a 4-point answer scale (1 = *not at all important*/*very unlikely*, 4 = *very important*/*very likely*). The AI measures *intrinsic* (i.e., personal growth, affiliation, community contribution, physical fitness) and *extrinsic* (i.e., financial success, attractive appearance, fame) aspirations. A sample item is "To have good friends whom you can count on" (affiliation). Total importance and likelihood scores were computed by averaging ratings across the seven domains. The German Version by Klusmann et al. (2005) demonstrated high internal consistencies (mean Cronbach Alpha = .79) and high four-week test-retest correlations (mean = .79 for the importance and .78 for the likelihood). The AI is widely used in research supporting its validity (e.g., Auerbach, et.al., 2010; Kasser, 1996; Kasser & Sheldon, 2000). The alpha coefficients in this sample were satisfactory, with a median of .82 that ranged from .64 to .87.

Procedure. Data for this study has been collected in an online study. The website was hosted by the institution in which the study has been conducted. The study was advertised by means of leaflets handed out at public transport stations and via mailing lists. Additionally, it was posted in several online forums of general interest. Participants were not paid for their services but received a feedback upon request after completing all questionnaires.

Results

When computing bivariate correlations between adult playfulness and life aspirations (not reported here in detail), primarily intrinsic goals demonstrated relations (e.g., greater likelihood of personal growth correlated with all variants of playfulness except for silly playfulness; r^2 were between .02 and .20). The fun-variants of playfulness also correlated robustly with the expectation for affiliation (r^2 = .19) and for physical fitness (r^2 = .12). Overall, these analyses argue for a relation of intrinsic goals to playfulness but also to a relation towards a positive expectancy that (intrinsic and extrinsic) goals will materialize. These analyses, however, might be biased because of a generally high or low overall conviction that goals are important or that it is likely that goals can be achieved. Thus, the analysis of the relation between playfulness and goal aspirations was based on a different approach. Regression analyses were performed with playfulness (general playfulness but also its variants) as criteria. In these analyses, playfulness was regressed onto the overall importance or likelihood of aspirations at Step 1; i.e., total scores were computed for the endorsement towards the importance and likelihood of goals (overall importance / likelihood). At Step 2, the semipartial for the importance of intrinsic aspirations was tested by entering either the importance or likelihood of intrinsic or extrinsic goals into the equation. Kasser and Ryan (1993, 1996) argue for this strategy of analyzing the data as it controls "variance due to having generally high importance or likelihood ratings" (1996, p. 283). Table 2 contains the standardized regression coefficients.

Table 2 shows that facets of adult playfulness were robustly associated with intrinsic and extrinsic goals. Overall playfulness (SMAP) existed independently from goals but there were significant relations with the likelihoods of achieving goals. Expressive (e.g., bouncy, open) and fun (e.g., bright, excitable) variants of playfulness were related with greater overall importance and likelihood of goal achievement. Both, the importance but also the likelihood of extrinsic goals were negatively associated with these two forms of playfulness. This was particularly evident for the likelihood of extrinsic aspirations and fun-variants of playfulness. Thus, pursuing fun-variants of playfulness yielded relations with intrinsic aspirations only. High likelihoods for intrinsic aspirations were associated with both forms of playfulness while only fun-related playfulness demonstrated significant relations towards greater intrinsic importance of aspirations. Greater spontaneous and creative playfulness demonstrated significant relations with greater likelihood of aspirations – and creative playfulness demonstrated associations towards a greater likelihood of intrinsic and a lower likelihood of extrinsic goals. Silly and spontaneous playfulness existed independently from life goals.

Table 2:
Regression of Intrinsic and Extrinsic Goals on (Facets of) Adult Playfulness (n = 263)

	SMAP	**Adult Playfulness Scale**				
	Playfulness	***Spontaneous***	***Expressive***	***Fun***	***Creative***	***Silly***
Importance						
Step1						
Overall	.07	.02	.16*	.16*	.06	.02
Step 2						
Intrinsic	-.03	-.02	.15	.31*	.11	-.12
Extrinsic	.04	.03	-.24*	-.48*	-.17	.18
Likelihood						
Step1						
Overall	.15*	.14*	.14*	.37*	.29*	.06
Step 2						
Intrinsic	.10	.12	.21*	.59*	.17	-.02
Extrinsic	-.13	-.16	-.29*	-.81*	-.24*	.03

Note. Standardized regression coefficients. SMAP = Short Measure for Adult Playfulness.

*p < .05.

Discussion

This study shows that playfulness in adults is associated with intrinsic goals. At the level of bivariate correlations, the perceived likelihood for personal growth correlated with all variants of playfulness and a global, cognitive evaluation of playfulness. The same was true for aspirations towards good relations, which yielded robust relations with expressive, fun, and creative variants of playfulness. It can be speculated that people who experience themselves as playful, report higher inclination towards the pursuit of intrinsic goals.

However, not only intrinsic goals demonstrated positive relations with playfulness. There were also robust relations between extrinsic aspirations towards fame and the likelihood of attractiveness and greater creative and fun-variants of playfulness. However, these seemed to have been of less relevance at the content level. This variant of playfulness yielded the strongest relations with life goals. The silly-forms existed unrelated from intrinsic and extrinsic aspirations in all of their components. It is argued that silly-playfulness is exhibited for immediate pleasure and fun but not for satisfying intrinsic (or extrinsic) needs.

One of the main findings of this study is that greater playfulness relates robustly to a greater expectation of the likelihood that aspirations can be achieved – again, except for silly-variants. Expressive and fun-variants were also predicted, beyond the level of a general tendency to think of aspirations as being important, by higher intrinsic and lower extrinsic likelihood. Greater expressive and fun-related playfulness can also be described by higher intrinsic and lower extrinsic importance of aspirations (if controlled for the overall contribution of the importance ratings). Low extrinsic likelihood yielded the greatest regression coefficient in all analyses. Proyer (2011) found that greater playfulness in students was positively related to better academic performance but also to the willingness of doing more for an exam than would have been necessary for passing the exam (i.e., doing extra reading). This finding may be explained by an interest in personal growth that facilitates the acquisition of further knowledge. Intrinsic aspirations have been linked to positive evaluations of and views on education among students (Henderson-King, & Mitchell, 2011; Vansteenkiste, et. al,, 2008). In study 3, playfulness was related to psychometrically measured and self-estimated ingenuity.

Study 3

Method

Sample

The sample consisted of 212 undergraduate students; 34 were males, 139 were women, and 39 did not indicate their gender. Their mean age was 24.8 years (SD = 5.2) and ranged from 20 to 53 years; 40 did not provide information on their age.

Instruments

As in Study 2, the *Short Measure of Adult Playfulness* has been used[3]; the alpha-coefficient in this study was .69. Also, the *Adult Playfulness Scale* has been used and alpha coefficients were .70, .65, .63, .64, and .71 for *spontaneous*, *expressive*, *fun*, *creative* and *silly*-variants of playfulness.

The *Berlin Intelligence Structure Test Version 4* (BIS; Jäger et al., 1997) is an intelligence test-battery based on Jäger's (1982) *Berlin Intelligence model*. This model comprises different levels of generality; on the top is a g-factor for general intelligence. Then *contents* (verbal, figural, and numeric) and *operations* (speed, memory, ingenuity/creativity, and reasoning) are on the next level. Participants in this study completed three subtests of the "ingenuity" factor that represent verbal (v), numeric (n), and figural (f) ingenuity; i.e. (1) *possible uses of objects* (v), where participants have to generate as many different possible uses for an object within a time-span of two minutes; (2) the task of the participants in the subtest *inventing telephone numbers* (n) is to create as many different sequences of phone numbers (based on specific principles) as possible within a time span of one minute and forty seconds; and (3) drawing real objects out of proposed geometrical figures within two minutes and thirty seconds (*combining symbols*; f). For each of these subtests, two scores can be derived; i.e., one that reflects the *fluidity* of the productions (number of productions), and one that represents the *origence* of the productions (how many different or unique solutions were created). The BIS-4 is widely used in research and practice in the German speaking area (e.g., Bucik & Neubauer, 1996).

Measure for self-estimated ingenuity. Participants rated their ingenuity on a scale from *lowest* (0) to *highest* (100) based on a short definition of ingenuity. This approach has already been successfully used in earlier studies (Proyer, 2011; Proyer & Ruch, 2009).

Procedure. Participants were undergraduate students of psychology. The procedure was identical to the one in Proyer (2011). The students completed the instruments in group settings (fifteen to twenty at a time) as part of the requirements of a lecture. For protecting their anonymity, they were allowed to omit information on age and gender. Three higher-grade students with a special training in test administration conducted the testing sessions. Participants first completed the SMAP, the APS and the self-estimate of ingenuity. Then, they started with the tests on the BIS-4 ingenuity. Afterwards, they completed a battery of other tests, which were unrelated to the present study. The whole procedure took about two hours of testing time (including breaks).

Results

Preliminary analyses. In comparison to the norm of people of similar age, the participants demonstrated average ingenuity with respect to the origence of the productions (figural: mean standard scores = 102; verbal: mean standard scores = 98) but were below

[3] As the final version of the SMAP was not yet available when this data was collected, the version used here consisted only of four out of the five items of the final version of the scale.

average to average in numeric ingenuity (mean standard scores = 88). The median of the self-estimates of ingenuity was above average; i.e., 66.14 on a scale from 0 to 100. This score significantly exceeded the midpoint of the scale (50) that indicated *average* ingenuity ($t(208) = 11.06$, $p = .000$).

Scores for origence and fluidity in the BIS-4 correlated between .41 (numeric) and .79 (figural ingenuity). The two total scores converged well with .71 (all $p = .000$; $n = 210$) but were considered sufficiently distinct for conducting separate analyses. Self-estimated ingenuity and psychometric ingenuity correlated with .06 and .12 for the total scores (fluidity/origence); and with .05/.13, -.04/-.02, and .13/.17 with verbal, numeric, and figural ingenuity (all *n.s.*, except for figural ingenuity, origence, $p = 0.013$; $n = 210$-212).

The relationship between self-estimated and psychometrically measured ingenuity. Correlations between the various indicators of playfulness and ingenuity were computed. Also, partial correlations controlling for demographics and the method used (either self-estimates or the psychometric test) were computed. Finally, correlations of playfulness with difference scores between self-assessment and the psychometric tests were inspected. With respect to the latter, positive correlations would indicate that greater playfulness relates to higher self-estimated than psychometrically tested ingenuity. All correlation coefficients are given in Table 3.

Table 3:
The Relationship between Psychometric and Self-estimated Ingenuity and Adult Playfulness

	SMAP	**Adult Playfulness Scale**				
	Playfulness	***SPO***	***EXP***	***FUN***	***CRE***	***SIL***
Pearson Correlations						
Self-estimates	.26*	.30*	.30*	.21*	.55*	.15*
Fluidity						
Verbal	.13	.18*	.12	.02	.29*	.05
Numeric	.13	.24*	.03	.00	.13	.20*
Figural	.03	.07	.07	.00	.09	-.02
Total	.13	.22*	.10	.01	.22*	.11
Origence						
Verbal	.11	.06	.09	.04	.26*	.03
Numeric	-.03	.15*	-.08	.02	.09	.09
Figural	.07	.12	.10	.00	.11	.03
Total	.07	.16*	.04	.03	.22*	.07
Partial (D)						
Self-estimates	.27*	.26*	.29*	.18*	.51*	.14
Fluidity						
Verbal	.15	.16*	.16*	.06	.28*	.04

	SMAP	Adult Playfulness Scale				
	Playfulness	***SPO***	***EXP***	***FUN***	***CRE***	***SIL***
Numeric	.17*	.26*	.05	.06	.10	.16*
Figural	.05	.14	.14	.04	.09	.03
Total	.16*	.25*	.15	.07	.21*	.11
Origence						
Verbal	.12	-.02	.15	.07	.27*	.03
Numeric	-.08	.11	-.09	.00	.03	.09
Figural	.10	.18*	.16*	.07	.13	.06
Total	.06	.13	.10	.06	.21*	.09
Partial (Method)						
Psychometric-F (S)	.09	.17*	.03	-.03	.13	.08
Psychometric-O (S)	.03	.12	-.01	.00	.15*	.05
Self-estimated (F)	.23*	.26*	.29*	.21**	.54**	.12
Self-estimated (O)	.25*	.27*	.30*	.20**	.54**	.13
Difference scores						
Difference score (F)	.10	.06	.16*	.15*	.26*	.03
Difference score (O)	.14*	.10	.20**	.13	.26**	.06

Note. n = 208-212 (n = 162 for partial correlations with demographics; n = 205 for partial correlations with method; n = 209-212 for difference scores); SMAP = Short Measure of Adult Playfulness; SPO = spontaneous; EXP = expressive; CRE = creative; SIL = silly; verbal, numeric, figural, and total ingenuity (psychometrically measured); Partial (D) = partial correlations controlled for age and gender; Psychometric (S) = correlations between psychometric ingenuity and playfulness controlled for self-estimated ingenuity; Self-estimated (F) = correlations between self-estimated ingenuity and playfulness controlled for fluid aspects of ingenuity; Self-estimated (O) = correlations between self-estimated ingenuity and playfulness controlled for the origence in ingenuity; Difference score (F) = self-estimated ingenuity minus psychometric ingenuity (fluidity; all standardized); Difference score (O) = self-estimated ingenuity minus psychometric ingenuity (origence; standardized); Difference score (FO) = fluidity in ingenuity minus origence in ingenuity (standardized).

* $p < .05$.

Table 3 shows that greater playfulness was associated with higher self-estimated ingenuity with a median of .30 across all facets of the APS. The relation that was numerically highest was found for the creative variants of playfulness (r^2 = .30). Partial correlations indicated that demographics did not contribute strongly to this relation.

The results were less clear for the psychometrically measured ingenuity. Especially spontaneous and creative aspects of playfulness yielded positive correlations with both, fluidity and origence of ingenuity. The numerically highest relation was found between verbal fluidity and creative variants of playfulness (r^2 = .08). A visual inspection of the scatter graph between creative playfulness and psychometric ingenuity indicated that this variant of playfulness could be a necessary but not a sufficient condition for ingenuity (origence) to occur. Those scoring low in creative playfulness were low in ingenuity (there were no outlying cases) but in the group of highly playful participants, there were both, inventive and non-inventive participants. This was less so for the fluidity of ingenious productions. There were outliers in the upper range of playfulness (towards greater fluidity) but no high scores for those low in playfulness. Figural ingenuity existed irrespective of playfulness. Those exhibiting silly variants of playfulness demonstrated a tendency towards greater numeric fluidity. Demographics did not have a large impact on these relations.

When the self-estimates were controlled for psychometric intelligence (separately for origence and fluidity), the results were stable. Thus, the ability as measured by a psychometric test did not have an impact on the outcomes. However, controlling psychometrically measured ingenuity (origence, fluidity) for the self-estimates yielded results in the same direction but the coefficients decreased numerically. Computing a difference score (self-estimated ingenuity minus psychometric ingenuity; again separately for fluidity and origence) indicated that playfulness was higher among those who *over*-estimated their ingenuity (i.e., higher self-estimated than psychometric ingenuity). Although this was not found for the overall assessment of playfulness, it can (mainly) be seen among the creative (r^2 = .07) as well as the expressive-variants of playfulness (though, numerically lower)

Multiple correlation coefficients for all five scales of the Adult Playfulness Scale with the total score for ingenuity (origence/fluidity) were R = .29/R = .34, and R = .18/R = .14, R = .29/R = .34, and R = .28/.36 for figural, numeric, and verbal ingenuity, respectively. Thus, between 2 and 13 % of the variance of psychometric ingenuity could be accounted for by different variants of playfulness (R = .59 for self-estimated ingenuity).

If data were split at the median of the SMAP, there were no significant differences between the groups of high and low playful participants and psychometric ingenuity (origence and fluidity). However, the results indicated that variances were not homogeneous between the two groups for total origence (Levene's test: F = 6.35, p = .012) and figural origence (F = 4.43, p = .036). Variances were homogenous for all fluidity scores except for figural fluidity (F = 4.81, p = .029). In all of these cases, variances were *lower* for the playful participants (above median) indicating more homogenous performances in this group. When analyzing the self-estimated ingenuity in the same way, significant mean score differences and homogeneous variances were found (t(207) = 3.55, p = .000). Mean scores for the group higher in playfulness (M = 68.90, SD = 18.67) exceeded those of the nonplayful ones (M = 59.73, SD = 18.57; d = 0.49).

Discussion

This study provides the first empirical data on self-estimated ingenuity and adult playfulness and allows for comparisons with psychometrically measured ingenuity. The results fit the expectations very well. Overall playfulness relates to higher self-estimates of ingenuity. This was particularly evident for the *creative* (i.e., creative, imaginative, and active) but also the *expressive* (i.e., bouncy, open, animated) and *spontaneous* (i.e., impulsive, free-spirited, adventurous) variants of playfulness. Thus, playful people seem to be aware of their potential for creative and new productions. One might speculate that this perception is shaped by prior experiences when being able to come up with new and genuine solutions for problems or when excelling in creativity as well as a result of feedback from others. For example, having a reputation in a work group as being able to come up with thoughts that have hitherto not been considered. As part of a future study, the implementation of observer ratings for playfulness and creative performances is suggested, as this information would contribute strongly to the further refinement of these findings.

While the relations to the self-estimates could be interpreted straightforward, results were mixed for psychometrically measured ingenuity. The overall playfulness existed widely independently from ingenuity. However, higher expressions of creative and spontaneous variants of playfulness went along with higher tested ingenuity – numerically higher for fluidity and spontaneous playfulness. The latter finding might be explained by arguing that in order to exhibit spontaneous acts of playfulness, the mere number of productions seems to be of primary importance compared to the quality of the productions. This reflects the impulsive character of this aspect of playfulness. Also the playful adults seemed to be more homogenous in their genuine performances in ingenuity (origence) than the nonplayful adults. More research is needed on the situational conditions that allow playfulness to occur as well as conditions that hinder the exhibition of playfulness in the daily life. Overall, specific aspects of playfulness not only seem to be higher in the self-estimates but also in psychometric test scores. The next step in the exploration of these relations would be to examine whether this could be observed in real life situations as well. For example, whether those adults perceived as playful can really produce more creative productions compared to those perceived as nonplayful.

Results also point towards a direction of greater creative but also expressive playfulness among those who *over*-estimated their ingenuity (i.e., higher self-estimates than performance in the psychometric test). Thus, at least for these facets, the self-perception seems to exceed the actual performance. Pending further replication, however, this finding should not be over-interpreted. Overall, there seems to be a stable relation between seeing oneself as being inventive and creative and also being playful. Again, observer ratings would help in disentangling potential biases in the self-perception. Furthermore, it needs to be mentioned that student samples are not the ideal samples for studying playfulness as earlier studies found restricted variances due to the positive relation of playfulness to younger age (see Proyer, 2011, 2012).

General discussion

Based on this set of studies, the prototypic playful adult can be described as being extraverted, low in conscientiousness, open, gelotophilic, agreeable, following intrinsic life goals with extrinsic goals being of low importance (and their likelihood to occur is also lowly valued), an endorsement of a pleasurable and engaged life, and having both a high self-perception of the own ability to be genuine with also a trend towards greater psychometrically measured ingenuity (frequency and origence). Results point towards a potential of playfulness for eliciting positive emotions in adults and for facilitating the occurrence of flow-experiences. While overall playfulness existed independently from life satisfaction, its fun-variants (e.g., bright, excitable) shared about 10 % of variance with a global evaluation of the cognitive aspects of well-being. It can be speculated that playfulness relates to other indicators of well-being as well; e.g., quality of life (see Proyer et al., 2010), or perhaps, psychological and physical well-being. It needs to be acknowledged, however, that this general discussion is based on a set of three different studies with each of them awaiting further replication.

In a follow-up study, it will be tested whether a sense of mastery, an experience of personal growth, and related variables relate to playfulness. Additionally, experimental studies are needed for testing whether interventions aimed at fostering playfulness may be causally related to enhanced well-being and what facets of well-being may be affected. Findings of the present study are interpreted as first hints towards a beneficial effects of playfulness to well-being in adults that should deserve further attention.

This study tested playfulness in general but also at the level of different facets. These were partially heterogeneous in their relation with external variables. The question of the dimensionality of adult playfulness is discussed as controversial in the literature (see Proyer, in press). Thus, one of the main tasks for future research seems to be to identify facets of adult playfulness and test their predictive validity for specific behaviors. It has been argued earlier that there may be specific facets (e.g., those related to a “darker side” of playfulness or of facets that are more strongly oriented towards intellectual interests) that have not yet been described in full detail (Proyer, in press).

The question arises on what benefits are to be expected from these relations. Of course, more studies are needed to further elaborate on the role of playfulness for creativity but also for adjacent fields such as learning. Kolb and Kolb (2010) describe a case study in which a playful activity in a team created a “ludic learning space” that was reported to promote learning. Proyer (2011) discussed several reasons why academic performance was positively related to academic success. Amongst others, he argued that playful students might have a different way of not only approaching an exam but also preparing for the exam. It is not yet known whether this is manifested through the creation of a ludic environment or whether personality traits play a bigger role – nevertheless, there seems to be a potential for play and playfulness in learning- and work-related settings. It is argued that playfulness may have potential for increasing passion and joy or love of learning in the learners (cf. Russ & Christian, 2011; Proyer & Ruch, 2011). Thus, this research may have practical implications (e.g., in pedagogy or in other applied settings) for conditions that facilitate learning, creative productions in general, and academic achievements.

Although expectations were mainly fulfilled regarding the correlates of adult playfulness, more research is needed on the concept itself. For example, its measurement needs to be further developed. The *Adult Playfulness Scale* (Glynn & Webster, 1992) has been criticized for its theoretical background (play is seen as the opposite of work) and for psychometric reasons (Barnett, 2007; Krueger, 1995). Thus, new measures are needed that are based on a theoretical framework. The *Short Measure of Adult Playfulness* (Proyer,

2011) is seen as a first step towards new instruments. It enables a short evaluation of cognitive aspects of playfulness and it is hoped that it can help stimulating further research in the area. Additionally, a new multi-facetted instrument is needed for being able to differentiate among different variants of playfulness. Barnett's (2007) approach to use focus groups of undergraduate students for collecting characteristics of playfulness in adults can be seen as a first step towards the development of such a theory. In combination with other approaches (e.g., a linguistic analysis of what words are being used in relation with playfulness; a psychometric study involving current measures for playfulness; etc.) and earlier theoretical accounts (e.g., Lieberman, 1977), a new theory could be developed and tested.

Author Note

The author wishes to thank Fabian Gander and Sara Wellenzohn for their help in collecting data for Study 2 and Jennifer Hofmann, Lisa Kaufmann, Katharina Klohe, and Annette Weber for their help with the preparation of the testing material, data collection, and data input for Study 3. Tracey Platt and Katharina Klohe proofread the manuscript.

References

Amabile, T. M., Hill, K. G., Hennessey, B. A., & Tighe, E. M. (1994). The Work Preference Inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology, 66*, 950-967. doi:10.1037//0022-3514.66.5.950

Apter, M. J. (1982). *The experience of motivation: The theory of psychological reversals*. London, UK: Academic Press.

Auerbach, R. P., McWhinnie, C. M., Goldfinger, M., Abela, J. R. Z., Zhu, X., & Yao, S. (2010). The cost of materialism in a collectivistic culture: Predicting risky behavior engagement in Chinese adolescents. *Journal of Clinical Child & Adolescent Psychology, 39,* 117-127. doi:10.1080/15374410903401179

Barnett, L. A. (1990). Playfulness: Definition, design, and measurement. *Play and Culture, 3*, 319-336.

Barnett, L. A. (1991a). The playful child: Measurement of a disposition to play. *Play and Culture, 4*, 51-74.

Barnett, L. A. (1991b). Characterizing playfulness: correlates with individual attributes and personality traits. *Play and Culture, 4*, 371-393.

Barnett, L. A. (2007). The nature of playfulness in young adults. *Personality and Individual Differences, 43*, 949-958. doi:10.1016/j.paid.2007.02.018

Barnett, L. A. (2011). How do playful people play? Gendered and racial leisure perspectives, motives and preferences of college students. *Leisure Sciences, 33*, 382-401. doi:10.1080/01490400.2011.606777

Barnett, L. A., & Kleiber, D. A. (1982). Concomitants of playfulness in early childhood: Cognitive abilities and gender. *Journal of Genetic Psychology, 141*, 115-127.

Barnett, L. A., & Kleiber, D. A. (1984). Playfulness and the early play environment. *Journal of Genetic Psychology, 144*, 153-164. doi:10.1080/00221325.1984.9923421

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York, NY: McGraw Hill.

Bozionelos, N., & Bozionelos, G. (1999). Playfulness: its relationship with instrumental and expressive traits. *Personality and Individual Differences, 4,* 749-760. doi:10.1016/S0191-8869(98)00207-4

Bucik, V., & Neubauer, A. C. (1996). Bimodality in the Berlin Model of Intelligence Structure (BIS): A replication study. *Personality and Individual Differences, 21*, 987-1005. doi:10.1016/S0191-8869(96)00129-8

Bundy, A. C. (1993). Assessment of play and leisure: Delineation of the problem. *American Journal of Occupational Therapy, 47*, 217-22.

Chapman, J. A. (1978). Playfulness and the development of divergent thinking abilities. *Child: Care, Health and Development, 4*, 371-383. doi:10.1111/j.1365-2214.1978.tb00096.x

Chen, L. H., Tsai, Y. M., & Chen, M. Y. (2010). Psychometric analysis of the Orientations to Happiness Questionnaire in Taiwanese undergraduate students. *Social Indicators Research, 98*, 239-249. doi:10.1007/s11205-009-9473-2

Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *Journal of Humanistic Psychology, 15*, 41-63.

Dansky, J. L., & Silverman, I. W. (1973). Effects of play on associative fluency in preschool-aged children. *Developmental Psychology, 9*, 38-43. doi:10.1037/h0035076

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Dewey, J. (1913). Play. In P. Monroe (Ed.), *A cyclopedia of education* (pp. 725-727). New York, NY: Macmillan.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment, 49*, 71-75. doi:10.1207/s15327752jpa4901_13

Dormann, C., & Kaiser, D. M. (2002). Job conditions and customer satisfaction. *European Journal of Work and Organizational Psychology*, *11*, 257-283.doi:10.1080/13594320244000166

Elfering, A., Semmer, N. K., & Kälin, W. (2000). Stability and change in job satisfaction at the transition from vocational training into “Real Work”. *Swiss Journal of Psychology, 59*, 256-271. doi:10.1024//1421-0185.59.4.256

Fix, G. A., & Schaefer, C. (2005). Note on psychometric properties of playfulness scales with adolescents. *Psychological Reports, 96*, 993-994. doi:10.2466/PR0.96.4.993-994

Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology, 2*, 300-319. doi:10.1037//1089-2680.2.3.300

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology – the broaden-and-build theory of positive emotions. *American Psychologist, 56*, 218-226. doi:10.1037//0003-066X.56.3.218

Glynn, M. A., & Webster, J. (1992). The Adult Playfulness Scale: An initial assessment. *Psychological Reports, 71*, 83-103. doi:10.2466/PR0.71.5.83-103

Glynn, M. A., & Webster, J. (1993). Refining the nomological net of the Adult Playfulness Scale: Personality, motivational, and attitudinal correlates for highly intelligent adults. *Psychological Reports, 72*, 1023-1026.

Guitard, P., Ferland, F., & Dutil, É. (2005). Toward a better understanding of playfulness in adults. *OTJR: Occupation, Participation and Health, 25*, 9-22.

Henderson-King, D., & Mitchell, A. M. (2011). Do materialism, intrinsic aspirations, and meaning in life predict students' meanings of education? *Social Psychology of Education, 14*, 119-134. doi:10.1007/s11218-010-9133-z

Hülsheger, U. R., & Maier, G. W. (2010). The careless or the conscientious: Who profits most from goal progress? *Journal of Vocational Behavior,* doi:10.1016/j.jvb.2010.04.001

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells [Multimodal classification of cognitive abilities: An experimentally controlled development of a descriptive model of intelligence structure]. *Diagnostica, 28*, 195-225.

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. BIS-Test, Form 4* [Berlin Intelligence-Structure-Test, BIS, Form 4]. Göttingen, Germany: Hogrefe.

Kasser, T. (1996). Aspirations and well-being in a prison setting. *Journal of Applied Social Psychology, 26*, 1367-1377. doi:10.1111/j.1559-1816.1996.tb00076.x

Kasser, T., & Ryan, R. M. (1993). A dark side of the American dream: Correlates of financial success as a central life aspiration. *Journal of Personality and Social Psychology, 65*, 410-422. doi:10.1037//0022-3514.65.2.410

Kasser, T., & Ryan, R. M. (1996). Further examining the American dream: Differential correlates of intrinsic and extrinsic goals. *Personality and Social Psychology Bulletin, 22*, 280-287. doi:10.1177/0146167296223006

Kasser, T., & Sheldon, K. M. (2000). Of wealth and death: Materialism, mortality salience, and consumption behavior. *Psychological Science, 11*, 352–355. doi:10.1111/1467-9280.00269

Klusmann, U., Trautwein, U., & Lüdtke, O. (2005). Intrinsische und extrinsische Lebensziele. Reliabilität und Validität einer deutschen Fassung des Aspirations Index [Intrinsic and extrinsic personal goals: Reliability and validity of a German translation of the Aspirations Index]. *Diagnostica, 51*, 40-51. doi:10.1026/0012-1924.51.1.40

Kolb, A. Y., & Kolb, D. A. (2010). Learning to play, playing to learn: A case study of a ludic learning space. *Journal of Organizational Case Management, 23*, 26-50. doi:10.1108/09534811011017199

Krueger, A. (1995). The Adult Playfulness Scale: A review. *Psychology*, *32*, 36-38.

Lieberman, N. J. (1965). Playfulness and divergent thinking: Investigation of their relationship at the kindergarten level. *The Journal of Genetic Psychology, 107*, 219-224.

Lieberman, N. J. (1977). *Playfulness: Its relationship to imagination and creativity*. New York, NY: Academic Press.

March, J. G. (1976). The technology of foolishness. In J. G. March, & J. P. Olsen (Eds.), *Ambiguity and choice in organizations* (pp. 69-81). Bergen, Norway: Universitetsforlaget.

McGhee, P. E. (1996). *Health, healing and the amuse system: Humor as survival training*. Dubuque, IO: Kendall/Hunt Publishing Company.

McGhee, P. E. (2010). *Humor: The lighter path to resilience and health*. Bloomington, IN: AuthorHouse.

Niemiec, C. P., Ryan, R. M., & Deci, E. L. (2009). The path taken: Consequences of attaining intrinsic and extrinsic aspirations in post-college life. *Journal of Research in Personality, 43*, 291-306. doi:10.1016/j.jrp.2008.09.001

Ostendorf, F. (1990). *Sprache und Persönlichkeitsstruktur. Zur Validität des Fünf-Faktoren-Modells der Persönlichkeit* [Language and personality. On the validity of the Five-factor model of personality]. Regensburg, Germany: Roderer.

Ostendorf, F., & Angleitner, A. (1992). On the generality and comprehensiveness of the Five-Factor model of personality: Evidence for five robust factors in questionnaire data. In G. V. Caprara & G. L. van Heck (Eds.), *Modern personality psychology: Critical reviews and new directions* (pp. 73-109). New York, NY: Harvester Wheatsheaf.

Peterson, C., Park, N., & Seligman, M. E. P. (2005). Orientations to happiness and life satisfaction: The full life vs. the empty life. *Journal of Happiness Studies, 6*, 25-41. doi:10.1007/s10902-004-1278-z

Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Washington, DC: American Psychological Association.

Proyer, R. T. (in press). A psycho-linguistic study on adult playfulness: Its hierarchical structure and theoretical considerations. *Journal of Adult Development*. doi:10.1007/s10804-012-9141-7

Proyer, R. T. (2011). Being playful and smart? The relations of adult playfulness with psychometric and self-estimated intelligence and academic performance. *Learning and Individual Differences*, 21, 463-467. doi:10.1016/j.lindif.2011.02.003

Proyer, R. T. (2012). *Development and initial assessment of a short measure for adult playfulness: The SMAP*. Manuscript submitted for publication.

Proyer, R. T., Annen, H., Eggimann, N., Schneider, A., & Ruch, W. (in press). Assessing the "good life" in a military context: How does life -and work-satisfaction relate to orientations to happiness and career-success among Swiss professional officers? *Social Indicators Research*, doi:10.1007/s11205-011-9823-8

Proyer, R. T., Platt, T., & Ruch, W. (2010). Self-conscious emotions and ridicule: Shameful gelotophobes and guilt free katagelasticists. *Personality and Individual Differences, 49*, 54-58. doi:10.1016/j.paid.2010.03.007

Proyer, R. T., & Ruch, W. (2011). The virtuousness of adult playfulness: The relation of playfulness with strengths of character. *Psychology of Well-Being: Theory, Research and Practice*. doi:10.1186/2211-1522-1-4

Proyer, R. T., Ruch, W., & Chen, G. (2012). Positive psychology and the fear of being laughed at: Gelotophobia and its relations to orientations to happiness and life satisfaction in Austria, China, and Switzerland. *Humor: International Journal of Humor Research, 25*, 23-40. doi:10.1515/humor-2012-0002

Proyer, R. T., Ruch, W., & Müller, L. (2010). Sense of humor among the elderly: Findings with the German version of the SHS. *Zeitschrift für Gerontologie und Geriatrie, 43*, 19-24. doi:10.1007/s00391-009-0082-0

Qian, X. L., & Yarnal, C. (2011). The role of playfulness in the leisure stress-coping process among emerging adults: An SEM analysis. *Leisure/Loisir, 35*, 191-209. doi:10.1080/14927713.2011.578398

Renner, K.-H., & Heydasch, T. (2010). Performing humor: On the relations between self-presentation styles, gelotophobia, gelotophilia, and katagelasticism. *Psychological Test and Assessment Modeling, 52*, 171-190.

Rossman, B. B., & Horn, J. L. (1972). Cognitive, motivational, and temperamental indicants of creativity and intelligence. *Journal of Educational Measurement, 9*, 265-286. doi:10.1111/j.1745-3984.1972.tb00959.x

Ruch, W., Harzer, C., Proyer, R. T., Park, N., & Peterson, C. (2010). Ways to happiness in German-speaking countries: The adaptation of the German version of the Orientations to Happiness Questionnaire in paper-pencil and internet samples. *European Journal of Psychological Assessment, 26*, 227-234. doi:10.1027/1015-5759/a000030

Ruch, W., & Proyer, R. T. (2008). The fear of being laughed at: Individual and group differences in gelotophobia. *Humor: International Journal of Humor Research, 21*, 47-67. doi:10.1515/HUMOR.2008.002

Ruch, W., & Proyer, R. T. (2009). Extending the study of gelotophobia: On gelotophiles and katagelasticists. *Humor: International Journal of Humor Research, 22*, 183-212. doi: 10.1515/HUMR.2009.00

Ruch, W., Proyer, R. T., Harzer, C, Park, N., Peterson, C., & Seligman, M. E. P. (2010). Adaptation and Validation of the German Version of the Values in Action Inventory of Strengths (VIA-IS) and the Development of a Peer-Rating Form. *Journal of Individual Differences, 31*, 138-149. doi:10.1027/1614-0001/a000022

Russ, S. W., & Christian, K. M. (2011). Play. In M. A. Runco, & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (2nd ed., Vol. II, pp. 238-243). Amsterdam, NL: Elsevier.

Samson, A. C., & Meyer, Y. (2010). Perception of aggressive humor in relation to gelotophobia, gelotophilia, and katagelasticism. *Test and Assessment Modeling, 52*, 217-230.

Sutton-Smith, B. (1967). The role of play in cognitive development. *Young Child, 22*, 159-160.

Taylor, S. I., & Rogers, C. S. (2001). The relationship between playfulness and creativity of Japanese preschool children. *International Journal of Early Childhood, 33*, 43-49. doi:10.1007/BF03174447

Tegano, D. W. (1990). Relationship of tolerance ambiguity and playfulness to creativity. *Psychological Reports, 66*, 1047-1056. doi:10.2466/PR0.66.3.1047-1056

Truhon, S. A. (1983). Playfulness, play, and creativity: A path analytic model. *Journal of Genetic Psychology, 143*, 19-28.

van der Kooij, R. (1989). Research on children's play. *Play & Culture, 2*, 20-34.

Vansteenkiste, M., Timmermans, T., Lens, W., Soenens, B., & Van den Broeck, A. (2008). Does extrinsic goal framing enhance extrinsic goal-oriented individuals' learning and performance? An experimental test of the match perspective versus self-determination theory. *Journal of Educational Psychology, 100*, 387-397. doi:10.1037/0022-0663.100.2.387

Vella-Brodrick, D. A., Park, N., & Peterson, C. (2009). Three ways to be happy: Pleasure, engagement, and meaning: Findings from Australian and US samples. *Social Indicators Research, 90*, 165-179. doi:10.1007/s11205-008-9251-6

Yu, P., Wu, J.-J., Chen, I., & Lin, Y.-T. (2007). Is playfulness a benefit to work? Empirical evidence of professionals in Taiwan. *International Journal of Technology Measurement, 39*, 412-429. doi:10.1504/IJTM.2007.013503

Psychological Test and Assessment Modeling, Volume 54, 2012 (2), 128-148

The case of dependency of responses and response times: A modeling approach based on standard latent trait models

Jochen Ranger[1] *& Tuulia Ortner*[2]

Abstract

When modeling responses and response times in tests with latent trait models, the assumption of conditional independence between responses and response times might be too strong in the case that both data are gained from reactions to the same item. In order to account for the possible dependency of responses and response times from the same item, a generalization of the model of Linden2007 is proposed. The basic idea consists in the assumption of a latent continuous response that underlies the observed binary response. This latent response is assumed to be correlated with the corresponding response time. The main advantage of this approach consists in the fact that the marginal models for responses and response times follow well known, standard latent trait models. Model estimation can be accomplished by marginal maximum likelihood estimation. The adequacy of the estimation approach is demonstrated in a small scale simulation study. An empirical data application illustrates the practicability of the approach in practice.

Key words: item response theory, response time, log-normal distribution, conditional independence

[1] *Correspondence concerning this article should be addressed to:* Jochen Ranger, PhD, Martin-Luther-University Halle-Wittenberg, Institute for Psychology, Universitätsplatz 10, 06108 Halle, Germany; email: jochen.ranger@psych.uni-halle.de

[2] University of Salzburg

1. An approach to account for the dependency of responses and response times in tests

Due to the computerized application of tests, item response times are widely available today. Therefore, it is not surprising that studies on the meaning or utility of response times in tests have come into the focus of psychological research (Schnipke & Scrams, 2002; van der Linden, 2009). One field of research addresses the question, whether it is possible to extend latent trait models to both, the responses and the response times of individual test takers. Such models are attractive as they provide the opportunity to include response times into the measurement of individual characteristics (Ferrando & Lorenzo-Sevas, 2007; van der Linden, 2008; Ranger & Ortner, 2011). However, care must be exercised when formulating latent trait models for responses and response times. Standard latent trait models might not be appropriate for all fields of application. This is the case when responses and response times gained from the same item are more closely related than responses and response times from different items. As will be shown later, ignoring this extra association in the data can lead to noticeable distortions of certain parameter estimates. Referring to these considerations, the following manuscript proposes an approach to account and test for the possible dependency between responses and response times in the same item.

2. Approaches to modeling responses and response times in tests

In general, latent trait models intend to model dependencies between observable quantities (Bartholomew & Knott, 1999). These models are based on the assumption of latent traits that are supposed to represent the entirety of all systematic influencing factors of the observable quantities. In the framework of item response models, this reasoning leads to the local independence assumption, which states that responses from different items are independent when conditioning on the underlying latent traits.

When modeling the joint distribution of responses and response times in tests, it is tempting to assume local independence as well. This comprises four aspects of local independence: The local independence of the responses, the local independence of the response times, the local independence of the responses and response times from different items and the local independence of the response and response time in the same item. While conditional independence is reasonable for data from different items, it is rather controversial when it comes to the response and response time in the same item (Thissen, 1983). As both quantities can be seen as the result of the same response process, they might share common influences that have not been accounted for when using the latent traits as conditioning variables.

The possible causes for a violation of the fourth facet of local independence are numerous. In achievement tests it is well known that individuals can increase their speed of responding at the cost of response accuracy, a phenomenon called speed accuracy trade-off. Generally speaking, a speed accuracy trade-off can be seen as a negative relation between the level of ability and work pace, at which the individual is able to operate. As Linden (2009) has pointed out, the existence of a speed accuracy trade-off does not contradict a latent trait model with constant latent traits, as long as individuals choose a level for ability and work pace before beginning the test and maintain this level throughout working. However, it seems possible that individuals do not rely totally on a stable choice during the test, but unsystematically fluctuate slightly around their chosen level over different items. This is a potential source of local dependency between item responses and response times. Additional sources of local dependency are random fluctuations of attention (Pieters & van der Ven, 1982). With reference to personality scales, research has revealed the average response latencies increase if a single item possesses an emotional evocative character that is independent from the trait measured (Temple & Geisinger, 1990; Tyron & Mulloy, 1993). It seems probable that such specific arousal evoked by single items also affects the test taker's response process for this particular item. Therefore, when modeling responses and response times, the used model should allow for violations of the local independence assumption in data from the same item.

2.1 A classification of responses and response time models

In the next paragraphs, different approaches to the joint distribution of responses and response times will be discussed. The following starting point will be used. Let the joint distribution of the responses and the response times in a test depend on two latent traits, namely ability θ and work pace ω, and some item parameters γ_g. The item parameters of the item response model will be denoted as $\beta_g(\gamma_g)$ and the item parameters of the response time model as $\alpha_g(\gamma_g)$, however the dependency on γ_g will only be stated when necessary. When assuming conditional independence of observations from different items, the joint distribution of the responses $\boldsymbol{x}=[x_1,\ldots,x_G]$ and response times $\boldsymbol{t}=[t_1\ldots t_G]$ can be stated most generally for a test taker as

$$\mathrm{f}(\boldsymbol{x},\boldsymbol{t}\,|\,\theta,\omega;\gamma)=\prod_{g=1}^{G}\mathrm{f}(x_g,t_g\,|\,\theta,\omega;\gamma_g), \tag{1}$$

where G denotes the number of the items, x_g is the response to item g, t_g is the corresponding response time and vector $\gamma=[\gamma_1,\cdots,\gamma_G]$ represents the parameters of the different items.

Different approaches can be chosen in order to specify the joint distribution $\mathrm{f}(x_g,t_g\,|\,\theta,\omega;\gamma_g)$ of the response and the response time in a single item (Bloxom, 1985). In the simplest case, the assumption of conditional independence can be applied to single items. In this case one can factor $\mathrm{f}(x_g,t_g\,|\,\theta,\omega;\gamma_g)$ as $\mathrm{f}(x_g\,|\,\theta,\omega;\beta_g(\gamma_g))\mathrm{f}(t_g\,|\,\theta,\omega;\alpha_g(\gamma_g))$, such that standard latent trait models can be used for the responses and the response times. Of course, further simplifications like $\mathrm{f}(x_g\,|\,\theta;\beta_g(\gamma_g))\mathrm{f}(t_g\,|\,\omega;\alpha_g(\gamma_g))$ might be more reasonable if one assumes that the responses and the response times depend on different latent traits. This approach has been advocated by Thissen (1983) and Linden (2007).

Alternatively, one can factor $\mathrm{f}(x_g, t_g \mid \theta, \omega; \gamma_g)$ as $\mathrm{f}(x_g \mid \theta, \omega; \beta_g(\gamma_g))\mathrm{f}(t_g \mid x_g, \theta, \omega; \alpha_g(\gamma_g))$. Again simplifications like $\mathrm{f}(x_g \mid \theta; \beta_g(\gamma_g))\mathrm{f}(t_g \mid x_g, \omega; \alpha_g(\gamma_g))$ might be more reasonable. This approach has been investigated by Linden2009 and revealed excellent power to detect even minor violations of conditional independence.

And finally, one can factor $\mathrm{f}(x_g, t_g \mid \theta, \omega; \gamma_g)$ as $\mathrm{f}(x_g \mid t_g, \theta, \omega; \beta_g(\gamma_g))\mathrm{f}(t_g \mid \theta, \omega; \alpha_g(\gamma_g))$ with the possible simplification of $\mathrm{f}(x_g \mid t_g, \theta; \beta_g(\gamma_g))\mathrm{f}(t_g \mid \omega; \alpha_g(\gamma_g))$. This approach has been proposed by Breukelen1991 and Verhelst1997.

The question addressing the adequate strategy for response time modeling should be answered empirically in each case, depending on the characteristics of the data. Nevertheless, some of the proposed models might be more preferable as a first choice from a theoretical point of view. The available publications show that responses from tests can be modeled with standard item response models when ignoring response time. This implies that the marginal response distribution

$$\mathrm{f}(x_g \mid \theta, \omega; \beta_g(\gamma_g)) = \int \mathrm{f}(x_g, t_g \mid \theta, \omega; \gamma_g)\,\mathrm{d}t_g \tag{2}$$

should follow (a potentially bidimensional version of) a standard item response model. Likewise, response times in tests have been modeled with standard latent traits models for a long time (Scheiblechner, 1979; van der Linden, 2006). Therefore, the marginal response time distribution

$$\mathrm{f}(t_g \mid \theta, \omega; \alpha_g(\gamma_g)) = \sum_{x_g} \mathrm{f}(x_g, t_g \mid \theta, \omega; \gamma_g) \tag{3}$$

should also be a standard response time model. So, when setting up a new model, it would be desirable that the corresponding marginal distributions of responses and response times are known latent trait models, as this is what we would expect from empirical findings.

In fact, a similar claim has already been made by Ip (2002) for item response models that account for the dependency between responses. Referring to the different approaches described above, only the model of Verhelst (1997) fulfills this claim. The model of Verhelst (1997) however assumes exponentially distributed response times. The exponential distribution implies a constant hazard rate and possesses the memoryless property and therefore might be a rather unrealistic model for data sets. As a possible solution, we propose an alternative model that is based on the log-normal distribution. The log-normal distribution is known to fit real data remarkably well (van der Linden, 2009). In the following paragraphs we will introduce this model that can be regarded as a generalization of the approach of Linden (2007), with the slight modification that we use the two-parameter probit model whereas Linden (2007) used the three-parameter probit model for the responses.

3. A model for the joint distribution of responses and response times

The model for the joint distribution of responses and response times in a test is introduced in two steps. First, it is described how responses and response times are distributed in a single item when conditioning on the latent traits θ and ω. At this level of the model, no assumption of conditional independence will be made. Second, the joint distribution of the responses and response times from different items will be derived.

3.1 The distribution of responses and response times in a single item

A standard model for binary responses in tests is the two-parameter probit model (Lord & Novick, 1968, p. 365). This model can be derived from the assumption that the binary response to an item rests on a continuous but unobservable response (Baker, 1992, p. 8). Let θ be the ability level of an individual and let the unobserved latent response z_g to item g depend on θ only. More specifically, it has to be assumed that conditionally on ability θ the latent response z_g is normally distributed with expected value

$$\mathrm{E}(z_g \mid \theta; \beta_{0g}, \beta_{1g}) = \beta_{0g} + \beta_{1g}\theta \tag{4}$$

and variance $\sigma^2_{z_g} = 1$. The quantities β_{0g} and β_{1g} are item parameters, β_{0g} reflecting the difficulty of an item and β_{1g} being the item discrimination. Whenever the latent response z_g exceeds the threshold zero, the observable item response is positive, otherwise it is negative. Or more formally, $x_g = 1$ when $z_g \geq 0$ and $x_g = 0$ when $z_g < 0$. In this case, the distribution of the observed response x_g can be derived as a binomial distribution with success probability

$$\mathrm{P}(x_g = 1 \mid \theta; \beta_{0g}, \beta_{1g}) = \int_0^\infty \mathrm{f}(z_g \mid \theta; \beta_{0g}, \beta_{1g}) \mathrm{d}z_g = \Phi\left(\beta_{0g} + \beta_{1g}\theta\right), \tag{5}$$

where $\Phi(x)$ denotes the distribution function of the standard normal distribution.

The second component of the model describes the distribution of the response times and is based on the log-normal distribution. Log-normal models have been used successfully for response times in tests (van der Linden, 2009). Such a response time model follows from the assumption that conditionally on work pace ω the logarithm of the response time $t_g' = \log(t_g)$ is normally distributed with expected value

$$\mathrm{E}(t'_g \mid \omega; \alpha_{0g}, \alpha_{1g}) = \alpha_{0g} + \alpha_{1g}\omega \tag{6}$$

and the variance $\sigma^2_{t'_g} = \alpha_{2g}$ independent of the test taker's characteristics. Again, α_{0g} and α_{1g} are item parameters, α_{0g} reflecting the general response time level of an item and α_{1g} accounting for the strength of the relationship between work pace and the response time.

Within this framework, the assumption of conditional independence within an item can easily be abandoned by allowing for a correlation between the latent response z_g and the log response time t'_g. In this case, the distribution of z_g and t'_g follows a bivariate normal distribution with expected values according to Equation (4) and Equation (6) and correlation ρ_g, which accounts for the dependency of responses and response times in a single item. As a consequence, the joint distribution of the observable response x_g and the log response time t'_g is

$$\mathrm{f}(x_g, t'_g \mid \theta, \omega; \alpha_g, \beta_g, \rho_g) = \int_{-\infty}^{\infty} \mathrm{I}(z_g, x_g)\mathrm{f}(z_g, t'_g; \mu(\theta, \omega, \alpha_g, \beta_g), \Sigma(\alpha_g, \rho_g))\mathrm{d}z_g. \qquad (7)$$

In Equation (7), function $\mathrm{I}(z_g, x_g)$ is an indicator function with $\mathrm{I}(z_g, 1) = 1$ when $z_g > 0$, $\mathrm{I}(z_g, 0) = 1$ when $z_g < 0$ and zero elsewhere. Function $\mathrm{f}(z, t'; \mu, \Sigma)$ is a bivariate normal distribution with mean vector $\mu = \mu(\theta, \omega, \alpha_g, \beta_g)$ given by Equation (4) and Equation (6) and covariance matrix $\Sigma = \Sigma(\alpha_g, \rho_g)$ with diagonal elements $\Sigma_{11} = 1$, $\boldsymbol{S}_{22} = \alpha_{2g}$ and off-diagonal elements $\rho_g \sqrt{\alpha_{2g}}$. Equation (7) can easily be generalized to polytomous items and the graded response model by slightly modifying the indicator function.

The proposed model in Equation (7) is a variant of the model of Linden2007. In the model of Linden2007, the probability of a correct solution is given by a three parameter logistic model $\mathrm{P}(x_g = 1) = c_g + (1 - c_g)\Phi(\beta_{1g}(\theta - \beta_{0g}))$ and the response times are distributed according to a log-normal distribution with $\mathrm{E}(t_{g'} \mid \omega; \alpha_{0g}) = \alpha_{0g} - \omega$. Contrary to the present model (see Equation (6)), the model of Linden2007 contains a restriction of the different α_{1g} parameters to the same value. The present model avoids this assumption as such constraints are unusual in factor analysis, but it always is possible to implement it in case it is justified by the data set. The mayor difference between the two models is the assumption of conditional independence between the response and the response time in the same item, which is made by Linden2007 but not in the present model.

There are several reasons to follow the approach proposed in this manuscript. The structure of Equation (7) accounts for the fact that responses and response times can often be modeled separately by unidimensional standard latent trait models. The applicability of unidimensional models to responses and response times clearly excludes the existence of additional, neglected common traits that have not been taken into consideration and that are responsible for remaining associations between responses and response times. However, even though one can exclude the presence of traits that influence all items, one still can assume specific factors that influence the response and response time in just one item, thereby causing a correlation between z_g and t'_g. The presence of such specific factors leaves the validity of the unidimensional response model and the unidimensional response time model unaffected. The assumption of specific factors resembles models for testlets, where the dependency of items based on the same content is similarly modeled by assuming a testlet specific factor (Wainer, Bradlow, & Wang, 2007; Li, Bolt, & Fu, 2006). Contrary to the present model however, the specific influences in the testlet model affect items measuring the same trait.

The psychological interpretation of the item specific factor depends on the kind of the test and the context of testing. As outlined in the introduction, it could represent random fluctuations in the speed accuracy level of an individual, account for the effects of isolated random guessing or be the consequence of the specific arousal evoked by a single item.

3.2 The distribution of responses and response times in a test

Responses and response times in a single item are based at least in part on the same cognitive process. As a consequence, the assumption of conditional independence seems not realistic. Reactions to different items however do not share the same response process. Therefore, the assumption of conditional independence is plausible for responses and response times from different items. Let $\mathrm{f}(x_g, t'_g \mid \theta, \omega; \alpha_g, \beta_g, \rho_g)$ be the distribution of the response and the response time of a test taker in item g. According to the conditional independence assumption, the joint distribution of the latent traits and the responses and response times in the G items of the test can be stated as

$$\mathrm{f}(\boldsymbol{x}, \boldsymbol{t}', \theta, \omega; \gamma, \rho_{\theta\omega}) = \prod_{G} \mathrm{f}(x_g, t'_g \mid \theta, \omega; \alpha_g, \beta_g, \rho_g) \mathrm{f}(\theta, \omega; \rho_{\theta\omega}). \tag{8}$$

In Equation (8), the distribution $\mathrm{f}(\theta, \omega; \rho_{\theta\omega})$ is the distribution of the latent traits in the population of the potential test takers. As in the original model of Linden2007, this is a bivariate normal distribution with zero means, unit variances and coefficient of correlation $\rho_{\theta\omega}$. In this aspect the model resembles an oblique factor model.

4. Estimating item parameters

Having observed the responses and response times of N test takers, the unknown item parameters can be estimated according to the marginal maximum likelihood approach or the limited information approach. In limited information estimation one first estimates the tetrachoric correlation matrix between the responses, the correlation matrix between the response times and the biserial correlation matrix between responses and response times. Using these correlation matrices, the model parameters can be estimated with standard software for structural equation models by allowing for correlated residuals in responses and response times from the same item. However, as limited information estimates are not efficient, marginal maximum likelihood estimation is preferred in this manuscript. We therefore propose an algorithm that can generally be used for response and response time modeling and might be useful for other response time models as well.

Marginal maximum likelihood estimates can be found by an application of the expectation maximization (EM) algorithm (Rubin, 1976; McLachlan & Krishnan, 1997. It is well known that the unknown latent traits can be considered as missing data. This idea is the basis for the application of the EM algorithm to the estimation of item parameters in the two parameter logistic model (Bock & Aitkin, 1981) or to the estimation of the loadings in the linear factor model (Rubin & Thayer, 1982). However, to estimate the parameters of the proposed model, it is advantageous to introduce another type of missing data. The item characteristic curve of the two-parameter probit model can be justified by the assumption of a latent continuous response z_g which underlies the observed binary response x_g, see the explanations above. In fact, the exact value of the latent response z_g cannot be observed as it only is known whether this variable exceeds the threshold zero or not. Therefore, this latent response can also be considered as missing data. Although not immediately apparent, the estimation of the item parameters can be simplified by pretending that the latent responses to the different items are known. In fact, this approach is similar to the technique of data augmentation, which has been applied in Markov Chain Monte Carlo estimation of item response models (Albert, 1992).

Let $\boldsymbol{z}_i = [z_{i1},\ldots,z_{iG}]$ be the latent responses and $\boldsymbol{t}'_i = [\log(t_{i1}),\ldots,\log(t_{iG})]$ be the log response times of the i-th individual from altogether N test takers. Simplifying the notation slightly as $\mu_{ig} = \mu(\theta_i,\omega_i,\alpha_g,\beta_g)$ and $\Sigma_g = \Sigma(\alpha_g,\rho_g)$ and using the nomenclature $\boldsymbol{y}_{ig} = [z_{ig},\log(t_{ig})]'$ and $\lambda_i = [\theta_i,\omega_i]'$, the relevant kernel of the complete log-likelihood function can be written as

$$\mathrm{LL} = \sum_{i=1}^{N}\left[\sum_{g=1}^{G}\log\left[\frac{1}{|\Sigma_g|^{1/2}}\right] - \frac{1}{2}[(\boldsymbol{y}_{ig}-\mu_{ig})'\Sigma_g^{-1}(\boldsymbol{y}_{ig}-\mu_{ig})] + \log\left[\frac{1}{|\Sigma_{\theta\omega}|^{1/2}}\right] - \frac{1}{2}[\lambda_i'\Sigma_{\theta\omega}^{-1}\lambda_i]\right]. \quad (9)$$

where $\Sigma_{\theta\omega}$ denotes the variance covariance matrix of the latent traits. The complete log-likelihood function is a function of the unknown item parameters and of sufficient statistics of the missing data, that is, the latent traits and the latent responses. A more accessible version of the complete log-likelihood function as well as a list of the unknown sufficient statistics is given in the Appendix. In case of known latent observations, the maximization of the complete log-likelihood function would be straightforward. However, as the latent variables and likewise the sufficient statistics are not known, the complete log-likelihood function can not directly be used to estimate the item parameters.

One possible solution to this problem consists in the iterated replacement of the unknown sufficient statistics by preliminary values. These values are determined as follows. First, provisional item parameters have to be chosen for the items. With these item parameters it is possible to calculate the conditional expectation of the unobserved sufficient statistics when conditioning on the observed data, that is, on the responses $\boldsymbol{x}_i$ and log response times $\boldsymbol{t}'_i$ of the test takers. For example, the conditional expectation of $\sum_{i=1}^{N} z_{ig}$ can be calculated as

$$\sum_{i=1}^{N}\mathrm{E}(z_{ig}\,|\,\boldsymbol{x}_i,\boldsymbol{t}'_i) = \sum_{i=1}^{N}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z_{ig}\ \mathrm{f}(z_{ig}\,|\,\theta,\omega,\boldsymbol{x}_i,\boldsymbol{t}'_i)\mathrm{f}(\theta,\omega\,|\,\boldsymbol{x}_i,\boldsymbol{t}'_i)\mathrm{d}z_{ig}\,\mathrm{d}\theta\mathrm{d}\omega. \quad (10)$$

Although not explicitly stressed, the different distributions in Equation (10) depend on the provisional item parameters. The inner most integral over z_{ig} is the expectation of a truncated normal distribution that can be stated in closed form. Therefore the triple integral simplifies to a twofold integration problem. The integral over $f(\theta,\omega|\boldsymbol{x}_i,\boldsymbol{t}'_i)$ can be approximated with Gauss Hermite Quadrature (Stroud, 1971). A more thorough description of the algorithm is given in the Appendix.

After all the unobserved quantities in the complete log-likelihood function have been replaced by their conditional expectation, the resulting equation is a function of the item parameters alone that can be maximized easily. Maximization for the item parameters is not computationally intensive because the time-demanding calculations have been made when calculating the conditional expectations. Having found the maximum, one can use the corresponding item parameter estimates as new provisional values for determining the updated conditional expectations of the unknown sufficient statistics. This sequence of calculating expected statistics and maximization is consecutively iterated until parameter estimates converge.

5. Simulation study

In order to test the practicability of the proposed approach, we performed a simulation study. With this simulation two intentions were pursued. First, to demonstrate the applicability of the estimation method. And second, to investigated whether a maximum likelihood ratio test comparing the proposed model with a model assuming independence has proper Type I error rates and power.

5.1 Estimation of item parameter

Model estimation was demonstrated with a test of 20 items for samples of 500 and 1000 subjects. This range was supposed to cover the sample sizes reported in empirical applications. Ability and work pace were sampled from a standard bivariate normal distribution. Thereby, a correlation of $\rho_{\theta\omega} = 0.3$ was assumed between ability and work pace. Such correlations between ability and work pace have been reported for achievement tests (van der Linden, 2009). Responses and response times were generated according to the proposed model. The employed item parameters are given in Table 1. The chosen item parameters resembled more or less values that had been found in previous studies.

Four different levels of correlation between the responses and response times were considered, ranging from $\rho_g = 0.0$ for the first five items to $\rho_g = 0.3$ for the last five items. This increase was thought to be a realistic pattern as correlations between responses and response times might increase during the test due to effects of test speededness.

Altogether 500 datasets were generated for every sample size. Preliminary item parameters for the response model were estimated by fitting a probit model to the responses alone. Additionally, preliminary item parameters for the response time model were estimated by factor analyzing the logarithmized response times. These estimates were used as starting values for the EM algorithm. The starting values for the correlations between the responses and the response times were set to zero.

Table 1:
True item parameter of the simulated items

Item	β_0	β_1	α_0	α_1	α_2	ρ_g
1	1.00	1.00	3.00	0.40	0.36	0.00
2	1.00	1.00	3.00	0.40	0.36	0.00
3	1.00	1.00	3.00	0.40	0.36	0.00
4	1.00	1.00	3.00	0.40	0.36	0.00
5	0.50	1.00	3.50	0.40	0.36	0.00
6	0.50	1.00	3.50	0.40	0.36	0.10
7	0.50	1.00	3.50	0.40	0.36	0.10
8	0.50	1.00	3.50	0.40	0.36	0.10
9	0.00	1.00	4.00	0.40	0.36	0.10
10	0.00	1.00	4.00	0.40	0.36	0.10
11	0.00	1.00	4.00	0.40	0.36	0.20
12	0.00	1.00	4.00	0.40	0.36	0.20
13	-0.50	1.00	4.50	0.40	0.36	0.20
14	-0.50	1.00	4.50	0.40	0.36	0.20
15	-0.50	1.00	4.50	0.40	0.36	0.20
16	-0.50	1.00	4.50	0.40	0.36	0.30
17	-1.00	1.00	5.00	0.40	0.36	0.30
18	-1.00	1.00	5.00	0.40	0.36	0.30
19	-1.00	1.00	5.00	0.40	0.36	0.30
20	-1.00	1.00	5.00	0.40	0.36	0.30

The EM algorithm was implemented in R (R Development Core Team, 2009). The integrals in the E-Step were approximated with Gauss Hermite Quadratur and 20 nodes per dimension. The expected log-likelihood function was maximized with the package optim. Note that although the true discrimination coefficients of the items were the same, their estimates were not restricted to the same value. The EM algorithm was ended when item parameter values did not change for more than 0.0008. The code can be obtained from the authors on request.

Altogether, the EM algorithm worked well as it converged in every sample. On the whole, the true item parameters could be recovered well without bias. The mean and the standard deviation of the estimates are given in Figure 1 for the item correlation parameter ρ_g. Results for the remaining parameters can be obtained from the authors.

Additionally, the item parameters were estimated according to the limited information approach. As not all programs for structural equation modeling can handle mixtures of

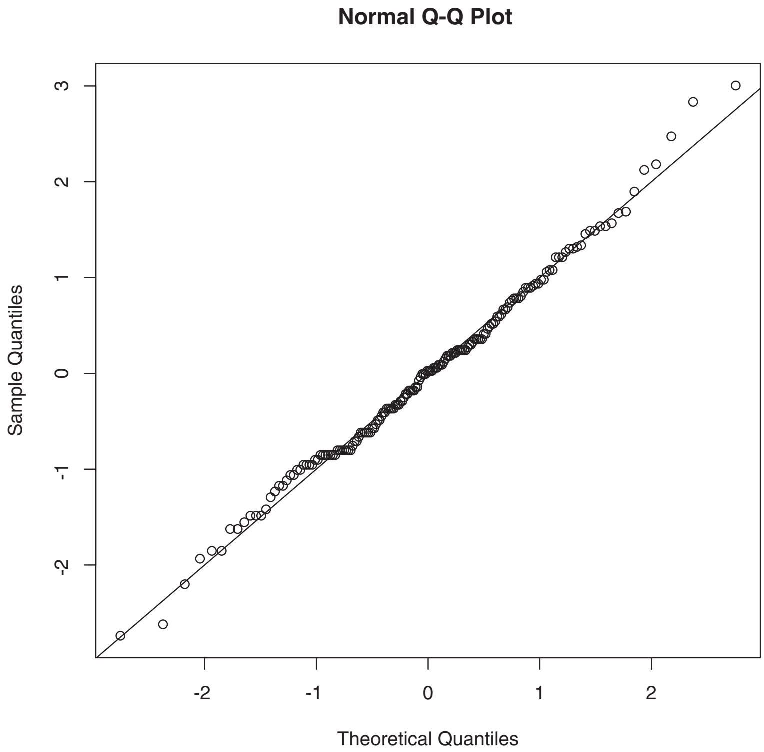


Figure 1:
Estimation results for correlation parameter ρ_g

continuous and discrete responses, the response times were dichotomized and the item parameters were estimated according to Muthen1978. However, instead of the weighted least squares approach proposed by Muthen1978, the item parameters were estimated with unweighted least squares. Item parameter estimates were unbiased, but not as efficient as the corresponding maximum likelihood estimates. The standard deviations of the correlation coefficients for example were about twice as large as the corresponding standard deviations of the maximum likelihood estimates. Therefore, as it is well known, maximum likelihood estimation is the first choice when one is interested in precise estimates.

5.2 Testing for independence

The proposed approach offers a framework for testing whether the responses and response times in the same item are independent or not. A first test of this hypothesis is the likelihood ratio test that compares the proposed model with a version where ρ_g is set to zero. As second test may serve a z-test that evaluates whether the correlation parameters ρ_g deviate from zero. This test employs Wald's second partial deviations for variance estimation. Both tests were evaluated with respect to power and Type I error rate in a simulation study.

Altogether, three scenarios were investigated. In all scenarios, the responses and response time were generated for 20 items. In the first scenario, the Type I error rate of the tests was investigated. The data was generated according to the item parameters given in Table 1 with the exception that there was no correlation between the responses and response times in a single item. In the second and third scenario, the power of the tests was the quantity of interest. In the second scenario, the data was generated according to the proposed model using the parameter values in Table 1. Local independence is violated as the correlations ρ_g are not zero any more. In the third scenario, a different violation of the conditional independence assumption was considered. This time, the data was generated according to the model of Linden2009. In this model, the response times are distributed log-normally with expected value according to Equation (6) and the modification that the intercept term α_{0g} is different for positive and negative responses. As a consequence, the response times are distributed differently for positive and negative responses. The motivation for this approach is the observation that wrong answers sometimes take longer than right answers (Thissen, 1983). Contrary to the original version of the model of Linden2009, the two-parameter probit model was used for the responses instead of the three-parameter probit model.

For every scenario, 500 simulation samples with a size of 500 and 1000 subjects were generated. Two different models were calibrated with marginal maximum likelihood estimation: The proposed model with the correlation parameters ρ_g estimated freely and the restricted version with all correlations ρ_g set to zero. The validity of this restriction was tested with the proposed likelihood ratio test. The model was also calibrated with limited information estimation. The results were used for a z-test of the hypothesis that $\rho_g = 0$ for all items. Empirical rejection rates are given in Table 2 for the three scenarios and the two tests. Note that the first two lines of Table 1 (Scenario 1) contain the empirical Type I error rate of the tests whereas the remaining lines contain the power.

As can be seen, both the likelihood ratio test and the z-test adhere to the nominal Type I error rate well. The likelihood ratio test however seems to have a slightly reduced nominal Type I error rate in small samples. The power of both tests is excellent. With sample sizes that are usually used for item response models both tests can detect small deviations from the independence assumption with a high probability. This is especially remarkable for the z-test that is based on dichotomous responses such that some information of the response times is lost. Interestingly, both tests also can verify model violations when the true model is the model of Linden (2009). However, in this case the likelihood ratio test is clearly superior to the z-test. Therefore, although the implementation of the z-test in standard software for structural equation models offers a quick check for model violations, the proposed maximum likelihood ratio test might be preferable.

Table 2:
Empirical Type I error rates and empirical power

Scenario	Sample	Test: LR-Test			Test: z-Test		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
1 - Independence	500	0.074	0.032	0.008	0.104	0.066	0.010
	1000	0.104	0.046	0.010	0.126	0.052	0.012
2 - Correlation	500	1.000	1.000	1.000	1.000	1.000	0.996
	1000	1.000	1.000	1.000	1.000	1.000	1.000
3 - Intercept	500	0.974	0.960	0.888	0.782	0.666	0.428
	1000	1.000	1.000	1.000	0.974	0.960	0.866

5.3 Consequences of model violations

Testing the local independence assumption is only necessary when parameter estimates are highly distorted in the case of unaccounted dependency. Therefore, in a further simulation study the effects of a misspecified model were investigated. Thereby, response patterns were generated for 10000 subjects and six different tests. The first two tests consisted of 20 items with item parameters as given in Table 1. However, for the first test the correlation between the response and the response time in the same item was set to $\rho_g = 0.50$ for all items, whereas for the second test this correlation was set to $\rho_g = 0.25$. The third and fourth test were generated by using every second item of the first and second test, thus reducing test length from 20 items to 10 items. The last two tests were generated similarly by choosing only every fourth item of the first two tests. In all conditions the correlation between ability and work pace $\rho_{\theta\omega}$ was set to zero.

Having generated the responses and response times, the proposed model was fit to the data with the correlation parameters ρ_g restricted to zero, thus ignoring the extra association between the responses and response times in the same item. Despite fitting the wrong model, the item parameters of the item response and the response time model could be recovered without serious bias. The parameter estimates $\widehat{\alpha}$ and $\widehat{\beta}$ of the independence model differed maximally by 0.03 from the corresponding estimates of the full model. However, the estimates of the correlation between ability and work pace were distorted, ranging from 0.018 up to 0.148. The exact results are given in Table 3.

Two effects can be noted. First, the misspecification of the model affects mostly the correlation of the latent traits. This is due to the structure of the model. As the responses and the response times depend on different latent traits, the only way to allow for an association between the responses and response times in the reduced model consists in the admission of a positive correlation between the latent traits. Therefore, it is expectable that the effect of the misspecification is mostly reflected in a distortion of

Table 3:
Estimated $\widehat{\rho}_{\theta\omega}$ in a misspecified model depending on the length of the test and the correlation between the response and the response time

Items	5		10		20	
ρ_g	0.250	0.500	0.250	0.500	0.250	0.500
$\widehat{\rho}_{\theta\omega}$	0.075	0.148	0.036	0.075	0.018	0.035

$\widehat{\rho}_{\theta\omega}$. Note that a special feature of the proposed model is the fact that the responses and response times follow standard latent trait models when considered separately. Second, the effect diminishes with a growing number of items. This is due to the fact that the number of misspecified associations grows more slowly than the number of correctly specified associations. In a test of G items, only $2G$ associations, the associations between responses and response times in the same items, are misspecified, while the remaining $2G \times 2G - 2G$ associations are correctly specified. This is similar to the concept of essential independence, which is present when the average covariance tends to zero (Junker, 1991). However, as Junker (1991) pointed out, even though the effects on consistency might not be large, they can be considerable on the standard errors of estimates.

From a practitioner's perspective, the correlation between ability and work pace is a key quantity. It can be shown that the accuracy of ability estimates can be improved by jointly considering responses and response times (van der Linden, Klein Entink, & Fox, 2010). The actual gain however depends on the amount of correlation between ability and work pace in the population of the test takers. Therefore, without checking the independence assumption one risks the overestimation of the benefits of response time modeling.

6. Empirical data application

To investigate the applicability of the proposed approach to real data, the model was used for data from an application of the German pre-version of the Eysenck Personality Profiler. The German Eysenck Personality Profiler has been published with reduced amount of items by Eysenck (1998). The Eysenck Personality Profiler measures 21 traits of personality which are consistent with the three major dimensions of personality as defined by Eysenck. In line with the original form of the questionnaire, three response options were offered including the 'don't know' option. Responses were dichotomized by scoring 'don't know' answers as rejections. Data was collected by Ortner2008 and consisted of 171 men called up for military service. If they agreed (about 80%), they were tested after the standardized psychological testing conducted by the Psychological Service of the Austrian Armed Forces. Persons were only included if they were evaluated as being motivated by the conductor and if no language problems were known. To reduce faking, the conductor pointed out that all results are handled anonymously and are not evaluated to determine the military appropriateness of the persons. Nevertheless one individual had to be excluded due to unusual short response times. Here only results for the anxious scale will be presented. Although originally the scale consists of 15 items one item had to be excluded as it was rejected by almost all subjects.

First, the response times were logarithmized. Normal Q-Q plots revealed that this transformation was capable of normalizing the data, see Figure 2 for an example.

Then, the responses and log response times were analyzed separately. The motivation behind this step was the generation of starting values for the EM algorithm and the assessment of model fit for the marginal models. Responses were analyzed with R and

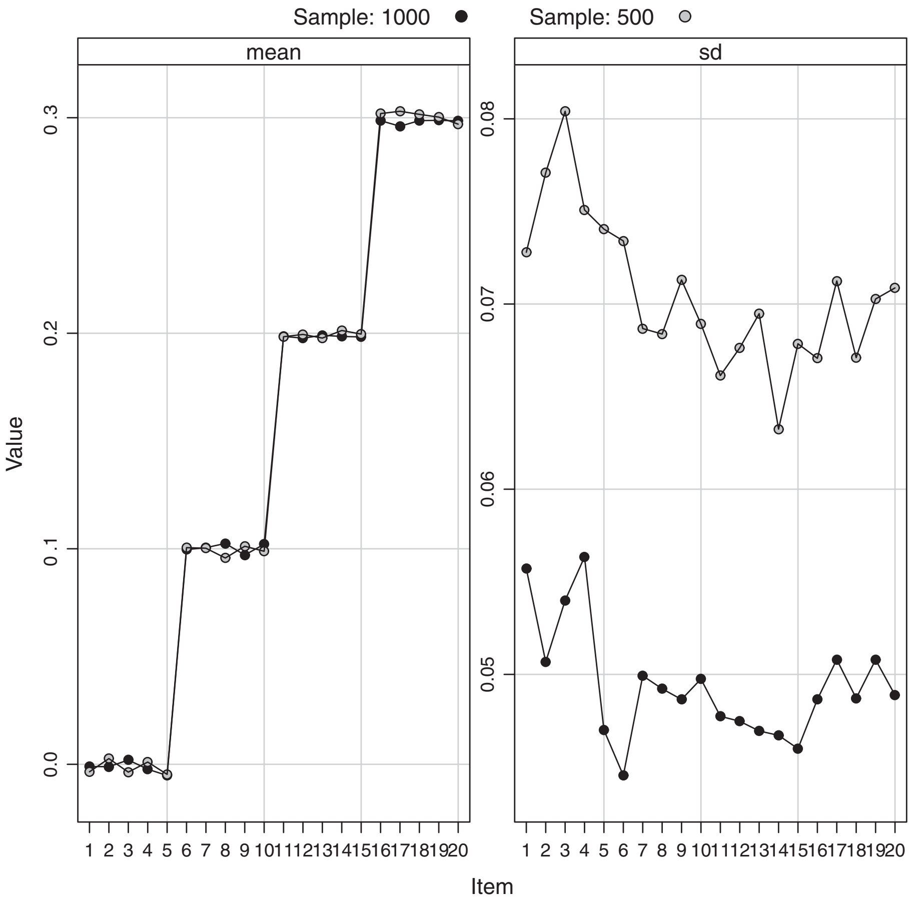


Figure 2:
Normal Q-Q plot: Fit of the log-normal distribution to response times of item 13 of the anxious scale

the ltm package (Rizopoulos, 2006). As the probit model is not implemented in the ltm package, a one-parameter logit model was used instead. When appropriately transforming the parameters, logit models are virtually identical to probit models. Assessing model fit via a parametric bootstrap test based on Pearson's chi-squared statistic did not show any evidence for model violations ($p = 0.44$). Therefore, the one-parameter probit model was used for the subsequent analysis because parsimonious models are preferable, especially in samples of moderate size. Replacing the two-parameter probit model with the one-parameter probit model is a change with little implications for the proposed estimation approach. Then, the log response times were analyzed by maximum likelihood factor analysis. A one dimensional model was sufficient to account for the dependence between the log response times ($p = 0.41$). Results of the factor analysis did not change when extreme observations were truncated, such that no truncation of the data seemed necessary. Finally, standardized residuals were calculated and plotted against the estimated factor scores (Bollen & Arminger, 1991). These plots did not show any systematic violation of the assumption of linearity and variance homogeneity. Summing up, it seemed reasonable to use the one-parameter probit model and the linear factor model for the marginal distribution of responses and log response times.

In the next step, the item parameters of the joint distribution of responses and log response times were estimated using the proposed EM algorithm. The resulting parameter estimates are given in Table 4. Ability and work pace were correlated with $\widehat{\rho}_{\theta\omega} = 0.246$. When the original model of Linden (2007) was used, almost the same estimates resulted. The estimate of the correlation between the latent traits slightly increased to $\widehat{\rho}_{\theta\omega} = 0.262$. The remaining estimates (the coefficients of correlation ρ_g excluded) were identical up to the second decimal place. This indicates that no extra association of the responses and response times in the same item has to be considered.

In order to test for $\rho_g = 0$, a likelihood ratio test was used. In this test the effect of restricting every parameter $\widehat{\rho}_g$ to zero was evaluated. This test yielded a non-significant test statistic of $\chi^2 = 11.49$ ($\mathrm{df} = 14$, $p = 0.65$). Therefore, the results indicate that the assumption of conditional independence can be extended to observations from the same item.

Finally, the independence model, the model with correlated responses and response times and a version of the model of Linden (2009) based on the one-parameter probit model were compared with respect to AIC. This comparison yielded values of 4628.50 for the independence model, of 4645.00 for the proposed model and of 4646.61 for the model of Linden2009. Again findings suggest the independence model can describe the data best. The actual model and the model of Linden (2009) revealed similar model fit with a slight advantage for the actual model.

7. Discussion

Due to the popularity of computer administered tests, interest in item response times and their possible applications is growing. Whereas not for all applications joint models for responses and response times are needed, some applications depend on them crucially. This is always the case when response times are incorporated into the estimation of the unknown trait level, see for example Linden (2010).

When jointly modeling responses and response times Achilles' heel is the question whether responses and response times in the same item can be considered as independent when conditioning on the latent traits. Although there is evidence for the independence (van der Linden & Glas, 2010) this assumption could be too strong for some tests. As it is well known that ignoring association in the data can distort confidence intervals (Ip, 2002) and parameter estimates (Wang, Cheng, & Wilson, 2005), it is a wise choice to check this assumption and to account for the dependency when it exists.

In the present article a new method was proposed that can account for the dependency between responses and response times in the same item. This can be done by only slightly generalizing the model of Linden (2007). The first advantage of this approach consists in the fact that the resulting marginal models are standard latent trait models. This is especially advantageous as the marginal response and response time distributions have been analyzed routinely with these models. A second advantage is the possibility to implement the actual approach in standard software for structural equation models. Although in this case, suboptimal limited information estimation has to be used, the results might be good enough for practical applications. Only when the exact amount of extra correlation between responses and response times has to be determined or one is interested in very high power, more complex estimation routines are recommended.

An empirical data application demonstrated the usefulness of the proposed approach in practice. Although most applications of response time models can be found in the field of achievement tests, in this study the applicability of the model to a personality test was shown. Therefore, the presented findings might have more implications than the mere checkout of a new model. In fact, it was shown that the model of Linden (2007) can be used for data from personality tests, such that it is not limited to the field of achievement tests. And second, equally to results from Linden (2009) findings indicate that the assumption of conditional independence in tests might not be totally unjustified in some cases. These findings might increase the popularity of response time modeling in the future.

8. Appendix

8.1 Sufficient statistics of the complete log-likelihood function

The relevant kernel of the complete log-likelihood function is given in Equation (9). Using Equation (4) and Equation (6), Equation (9) can be written as

$$\begin{aligned} \mathrm{LL} = & \sum_{i=1}^{N}\sum_{g=1}^{G}[\log\left[(\alpha_{2g}(1-\rho_g^2))^{-\frac{1}{2}}\right] - \frac{1}{2(1-\rho_g^2)}[\frac{(z_{ig}-\beta_{0g}-\beta_{1g}\theta_i)^2}{1} \\ & + \frac{(t'_{ig}-\alpha_{0g}-\alpha_{1g}\omega_i)^2}{\alpha_{2g}} - 2\rho_g \frac{(z_{ig}-\beta_{0g}-\beta_{1g}\theta_i)(t'_{ig}-\alpha_{0g}-\alpha_{1g}\omega_i)}{\sqrt{\alpha_{2g}}}]] \\ & + \sum_{i=1}^{N}[\log\left[(1-\rho_{\theta\omega}^2)^{-\frac{1}{2}}\right] - \frac{1}{2(1-\rho_{\theta\omega}^2)}(\theta_i^2+\omega_i^2-2\rho_{\theta\omega}\theta_i\omega_i)]. \end{aligned} \tag{11}$$

Expanding Equation (11) and summing over the N test takers reveals that the complete log-likelihood function is a function of the following unobserved sufficient statistics:

$$\sum_{i=1}^{N} z_{ig} \qquad \sum_{i=1}^{N} z_{ig}^2$$

$$\sum_{i=1}^{N} \theta_i \qquad \sum_{i=1}^{N} \theta_i^2$$

$$\sum_{i=1}^{N} \omega_i \qquad \sum_{i=1}^{N} \omega_i^2$$

$$\sum_{i=1}^{N} z_{ig} t_{ig} \qquad \sum_{i=1}^{N} \theta_i \omega_i$$

$$\sum_{i=1}^{N} \theta_i z_{ig} \qquad \sum_{i=1}^{N} \theta_i t'_{ig}$$

$$\sum_{i=1}^{N} \omega_i z_{ig} \qquad \sum_{i=1}^{N} \omega_i t'_{ig}$$

8.2 Calculation of conditional expectation of sufficient statistics

The complete log-likelihood function depends on unknown sufficient statistics, which are replaced by their conditional expectation during the E-Step. First, provisional values λ^*

have to be chosen for the unknown item parameters. Given these preliminary values for the item parameters, the conditional expectation of $\sum_{i=1}^{N} z_{ig}$ is

$$\sum_{i=1}^{N} \mathrm{E}(z_{ig} \mid \boldsymbol{x}_i, \boldsymbol{t}'_i; \gamma^*) = \sum_{i=1}^{N} \iiint z_{ig} \, \mathrm{f}(z_{ig} \mid \theta, \omega, \boldsymbol{x}_i, \boldsymbol{t}'_i; \gamma^*) \mathrm{f}(\theta, \omega \mid \boldsymbol{x}_i, \boldsymbol{t}'_i; \gamma^*) \mathrm{d}z_{ig} \, \mathrm{d}\theta \mathrm{d}\omega. \quad (12)$$

The inner most integral over z_{ig} can be given in closed form. Conditional on the latent traits, the latent response in item g is independent of the responses and the response times from different items, such that $\mathrm{f}(z_{ig} \mid \theta, \omega, \boldsymbol{x}_i, \boldsymbol{t}'_i; \gamma^*)$ can be simplified to $\mathrm{f}(z_{ig} \mid \theta, \omega, x_{ig}, t'_{ig}; \gamma^*_g)$. Conditional on θ and ω, the joint distribution of z_{ig} and t'_{ig} is a bivariate normal distribution, see Equation (7) for details. Therefore, the conditional distribution $\mathrm{f}(z_{ig} \mid \theta, \omega, t'_{ig}; \gamma^*_g)$ is a normal distribution, with expected value

$$\mathrm{E}\left(z_{ig} \mid \theta, \omega, t'_{ig}; \gamma^*_g\right) = (\beta^*_{0g} + \beta^*_{1g}\theta) + \frac{\rho^*_g}{\sqrt{\alpha^*_{2g}}} \cdot (t'_{ig} - (\alpha^*_{0g} + \alpha^*_{1g}\omega)) \quad (13)$$

and conditional variance $1 - \rho_g^{*2}$. Conditioning finally on the observed response x_{ig} yields a truncated normal distribution with corresponding expected value and variance. The expectation of a truncated normal distribution can be given in closed form. Let $\mathrm{E}(z_{ig} \mid \theta, \omega, t'_{ig}, x_{ig}; \gamma^*_g)$ be the expectation of the truncated normal distribution implied by Equation (13). Using this expectation in case of z_{ig}, one can simplify Equation (12) to

$$\sum_{i=1}^{N} \mathrm{E}(z_{ig} \mid \boldsymbol{x}_i, \boldsymbol{t}'_i; \gamma^*) = \sum_{i=1}^{N} \iint \mathrm{E}(z_{ig} \mid \theta, \omega, t'_{ig}, x_{ig}; \gamma^*_g) \mathrm{f}(\theta, \omega \mid \boldsymbol{x}_i, \boldsymbol{t}'_i; \gamma^*) \mathrm{d}\theta \mathrm{d}\omega. \quad (14)$$

The solution of the two-fold integral can not be given in closed form. However, it can be approximated by Gauss Hermite quadrature. Using cartesian quadrature rules, Equation (14) can be approximated by

$$\sum_{i=1}^{N} \mathrm{E}(z_{ig} \mid \boldsymbol{x}_i, \boldsymbol{t}'_i; \gamma^*) = \sum_{i=1}^{N} \sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \frac{\mathrm{E}(z_{ig} \mid \theta_{q_1}, \omega_{q_2}, t'_{ig}, x_{ig}; \gamma^*_g) \mathrm{f}(\boldsymbol{x}_i, \boldsymbol{t}'_i \mid \theta_{q_1}, \omega_{q_2}; \gamma^*) w_{q_1} w_{q_2}}{\sum_{q_1=1}^{Q_1} \sum_{q_2=1}^{Q_2} \mathrm{f}(\boldsymbol{x}_i, \boldsymbol{t}'_i \mid \theta_{q_1}, \omega_{q_2}; \gamma^*) w_{q_1} w_{q_2}} \quad (15)$$

where summation is over quadrature points θ_{q_1} and ω_{q_2} with corresponding weights w_{q_1} and w_{q_2}. The remaining sufficient statistics are calculated alike.

Acknowledgement

We would like to thank the editor and two referees for their helpful and constructive comments, which led to many improvements.

References

Albert, J. H. (1992). Estimation of Normal Ogive Item Response Curves Using Gibbs Sampling. *Journal of Educational Statistics, 17,* 251-269.

Baker, F. B. (1992). *Item Response Theory: Parameter Estimation Techniques*. New York: Marcel Dekker.

Bartholomew, D., & Knott, M. (1999). *Latent variable models and factor analysis*. London: Arnold.

Bloxom, B. (1985). Considerations in psychometric modeling of response time. *Psychometrika, 50,* 383-397.

Bock, R. D., & Aitkin, M. (1981). Marginal Maximum Likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46, 443-459.

Bollen, K. A., & Arminger, G. (1991). Observational residuals in factor analysis and structural equation models. *Sociological* Methodology, 21, 235-262.

Ferrando, P. J., & Lorenzo-Sevas, U. (2007). A measurement model for Likert responses that incorporates response time. *Multivariate Behavioral Research, 42,* 675-706.

Ip, E. H. (2002). Locally dependent latent trait model and the dutch identity revisited. *Psychometrika, 67,* 367-386.

Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika, 56,* 255-278.

Li, Y., Bolt, D. M., & Fu, J. (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement, 30,* 2-21.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics, 31,* 181-204.

van der Linden, W. J. (2009). Conceptual issues in Response-time modeling. *Journal of Educational Measurement, 46,* 247-272.

van der Linden, W. J. (2008). Using response times for item selection in adaptive tests. *Journal of Educational and Behavioral Statistics, 31,* 5-20.

van der Linden, W. J., & Glas, C. A. W. (2010). Statistical tests of conditional independence between responses and/or response times on test items. *Psychometrika, 75,* 120-139.

van der Linden, W. J., Klein Entink, R. H., & Fox, J. P. (2010). IRT Parameter Estimation With Response Times as Collateral Information. *Applied Psychological Measurement, 34,* 327-347.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Addison-Wesley, Reading.

McLachlan, G. J., & Krishnan, T. (1997). *The EM algorithm and extensions*. New York: Wiley.

Pieters, J. P. M., & van der Ven, A. H. G. S. (1982). Precision, Speed, and Distraction in Time-Limit Tests. *Applied Psychological Measurement, 6,* 93-109.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ranger, J., & Ortner, T. M. (2011). Assessing Personality Traits through Response Latencies using item response theory. *Educational and Psychological Measurement, 71,* 389-406.

Rizopoulos, D. (1976). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17,* 1-25.

Rubin, D. B. Inference and missing data. *Biometrika, 63,* 581-592.

Rubin, D.B., & Thayer, D. (1982). EM Algorithms for ML Factor Analysis. *Psychometrika, 47,* 69-76.

Scheiblechner, H. (1979). Specifically objective stochastic latency mechanisms. *Journal of Mathematical Psychology, 19,* 19-38.

Schnipke, D. L., & Scrams, D. J. (2002). Exploring issues of examinee behavior: Insights gained from response-time analysis. In Mills, C. N., Potenza, M. T., Fremer, J. J, & Ward, W. C. (eds.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 237-267). Mahwah: Lawrence Erlbaum.

Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals*. Englewood Cliffs: Prentice-Hall.

Temple, D. E., & Geisinger, K. F. (1990). Response latency to computer-administered inventory items as an indicator of emotional arousal. *Journal of Personality Assessment, 54,* 289-297.

Thissen, D. (1983). Timed testing: An approach using item response theory. In Weiss, D. J. (ed.), *New horizons in testing: Latent trait test theory and computerized adaptive testing* (pp. 179-203). New York: Academic Press.

Tyron, W. W., & Mulloy, J. M. (1993). Further Validation of Computer-Assessed Response Time to Emotionally Evocative Stimuli. *Journal of Personality Assessment*, *61,* 231-236.

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. New York: Cambridge University Press.

Wang, W. C., Cheng, Y. Y., & Wilson, M. (2005). Local item dependence for items across tests connected by common stimuli. *Educational and Psychological Measurement*, *65,* 5-27.

Psychological Test and Assessment Modeling, Volume 54, 2012 (2), 149-150

Configural Frequency Analysis (CFA) and other non-parametrical statistical methods: Introduction to Special Topic (Part I)

Mark Stemmler[1] *& Alexander von Eye (Guest Editors)*

„Called to rest from a hobby called science" is the epitaph of Gustav A. Lienert (GAL). He was called to rest more than ten years ago. However, the Configural Frequency Analysis (CFA) community did not rest. Since 2001 they have met several times to keep the ideas of GAL alive. For instance, the members of this community performed simulation studies on the behavior of CFA statistical tests, they undertook studies concerning the dependency structure of CFA tests, they worked on the reformulation of existing models of CFA, they developed new models for CFA, and they improved computer programs of CFA (a CFA-package is now also available in R).

The CFA community met first in 2003 at the University of Kassel organized by Erwin Lautsch, a symposium funded by the German Research Council (i.e., Deutsche Forschungs-gemeinschaft (DFG)). Two years later, a second conference was organized also by Erwin Lautsch, again in Kassel. This symposium was financially supported by the Lienert Foundation (i.e., Lienert-Stiftung). The works presented at these two conferences were published in Special Issues in the journal *Psychology Science* under title of "Charting the future of Configural Frequency Analysis I + II: Developments and applications". The guest editors were *Erwin Lautsch* and *Alexander von Eye*. The Editor-in-Chief of the journal, an outlet of *Wolfgang Pabst Publishers*, was *Klaus Kubinger*.

In 2006, *Mark Stemmler* organized a "Gustav A. Lienert Memorial Symposium" at the 45th Congress of the German Psychological Association (i.e., Deutsche Gesellschaft für Psychologie (DGPs)) in Nuremberg. The organizer of this congress was *Friedrich Lösel*. Again, the presented works were published by the Pabst Science Publisher, thanks to *Petra Netter* and her support through the Lienert Foundation. Guest editors were *Erwin Lautsch, Dirk Martinke* and *Mark Stemmler*.

---

[1] *Correspondence concerning this article should be addressed to:* Mark Stemmler, PhD, University of Erlangen-Nuremberg, Institute for Psychology, Nägelsbachstr. 49c, 91052 Erlangen, Germany; email: mark.stemmler@psy.phil.uni-erlangen.de

In December of 2011, the ten-year Gustav A. Lienert Memorial Symposium took place. *Mark Stemmler*, the organizer, was able to draw the attention of a renowned international group, from the US, Mexico, Sweden, Luxembourg and Austria, to take part in this meeting. This effort was honored by the DFG, which supported this meeting financially (GZ: STE 126-1). This Special Topic presents the first set of papers articles which are based on the papers presented during this symposium. The first article is by *Alexander von Eye* and *Eun-Young Mun*. The authors present a new methodological approach within Configural Frequency Analysis (CFA) to assess interindividual differences in intraindividual change. They propose new models of CFA to explore and test hypotheses concerning person-specific local associations in longitudinal data. The article by *Clemens B. Fell, Alexander von Eye, Gabriel Schui* and *Günther Krampen* examines the relationship between membership in the German Psychological Association (i.e., Deutsche Gesellschaft für Psychologie (DGPs)) and number of citations in English journal articles, excluding self-citations. They authors found out that membership in several subsections of the DGPs is preferable over membership in one or no subsection, with regard to international visibility. The third manuscript by *Matthias J. Müller* and *Petra Netter* uses CFA in order to show different specific physiological reaction patterns in response to stressful events. With the help of a two-sample CFA the authors showed that cortical increase (C+) and testosterone decrease (T-) characterizes uncontrollable stressful situations and the opposite pattern in controllable stress. Finally, *Mark Stemmler* and *Friedrich Lösel*, apply Prediction-Configural Frequency Analysis (P-CFA) to demonstrate the stability of a proactive pattern of aggressiveness (i.e., high externalizing and low internalizing problems) in boys from kindergarten to secondary school. In addition, this behavior pattern was related to self-reported delinquent behavior in mid-adolescence.

With are grateful to the financial support of the DFG in order to organize for the organization of the international meeting. We hope to repeat this sort of symposia symposium regularly in the future. We are also indebted to *Wolfgang Pabst* who benevolently accompanied all of the CFA or GAL- related publications in the past. Last but not least, we are indebted to *Klaus Kubinger,* the Editor-in-Chief of the newly founded journal *Psychological Test and Assessment Modeling* for providing us with the space of two issues with contributions based on the person-centered approach which was co-invented by GAL.

Psychological Test and Assessment Modeling, Volume 54, 2012 (2), 151-167

Interindividual differences in intraindividual change in categorical variables

Alexander von Eye[1] *& Eun-Young Mun*[2]

Abstract

In this article, we proceed from the assumption that constancy and change in development are not necessarily universal. This deviation from the general assumption of universal developmental patterns is embedded in the theory of person-oriented research. In addition, we propose that constancy and change can reflect local associations instead of associations that cover the entire range of admissible scores. Models of Configural Frequency Analysis are proposed to explore and test hypotheses concerning person-specific local associations in repeated observation data. Three models are considered for lagged data. These models differ in the reasons that are assumed as causes for local associations. The first model reflects variable associations of any kind. The second model reflects case-specific variable associations. The third reflects differences between cases. In an example, data from a study on the development of alcoholics are used. The data in this example reflect case-specific associations in the development of drinking behavior over a span of two years. In the discussion, the person- and the variable-oriented elements of longitudinal research are addressed. In addition, assumptions concerning the independence of longitudinal data are made explicit.

Key words: ... Author: please add ...

[1] *Correspondence concerning this article should be addressed to:* Alexander von Eye, PhD, University of Vienna, Faculty of Psychology, Liebiggasse 5, 1010 Vienna; email: voneye@msu.edu

[2] Rutgers, The State University of New Jersey

Interindividual Differences in Intraindividual Change in Categorical Variables

Developmental change processes and patterns of developmental change are not universal. Individuals differ in all characteristics of change, for example, timing, speed, duration of change process, amount of change, or qualitative characteristics of change. In standard statistical analysis, researchers often proceed from the assumption that change is universal and that differences from average change reflect measurement error or imperfections of a model.

In this article, we operate under just the opposite assumption. We propose that individuals do differ in change characteristics. Individuals can be grouped together only if these differences are no greater than random. We propose a new statistical method for the analysis of interindividual differences in intraindividual, developmental change. The new method is a variant of Configural Frequency Analysis (CFA; Lienert & Krauth, 1975; von Eye & Gutiérrez-Peña, 2004; von Eye, Mair, & Mun, 2010).

This article is structured as follows. First, we establish the background of this work in the contexts of person-oriented research and CFA. We then present the new approach to comparing individuals in their change patterns. Third, we present an empirical data example from research on alcohol use disorders. Finally, we discuss the present approach in substantive and statistical contexts.

Before we delve into the theoretical and the technical elements of the method to be presented here, however, we briefly illustrate the meaning of the expression "interindividual differences in intraindividual change" (Baltes, Reese, & Nesselroade, 1977). We present two examples. The first illustrates lack of change, on average. In this example, seven artificial trajectories are depicted. Each case is observed on three occasions. The cases start from different levels of behavior. From Time 1 to Time 2, there is no change. Every case stays where they are. In contrast, from T2 to T3, six of the seven cases change, only the case whose trajectory begins with the value of 4 stays unchanged. On average, the sample shows the scale value of 6 for each of the three observation points. However, only one out of 7, that is, 14.3%, shows consistently the same score over the entire observation period. This is depicted in Figure 1. The bold line indicates the average trajectory in the data.

From this example, we draw two conclusions. First, average values often fail to describe the activity in a population. For example, using data from a study on the development of alcohol use disorders, von Eye and Bergman (2003) showed that autocorrelations of averaged scores may fail to describe a single individual in a population. The second conclusion we draw is that differences in development, that is, interindividual differences in intraindividual constancy and change may disappear when change is described based on average scores. Both of these conclusions are a key to the person-oriented research perspective outlined in the following section.

The second example uses artificial data. It describes six cases, also observed over three points in time. Each of these cases displays change. Three of the six cases show a change in the linear trend such that an increase from T1 to T2 is followed be a decrease from T2 to

T3. The other three cases show an increase from T1 to T2 and an accelerated increase from T2 to T3. This is depicted in Figure 2. The bold line indicates the averaged increase.

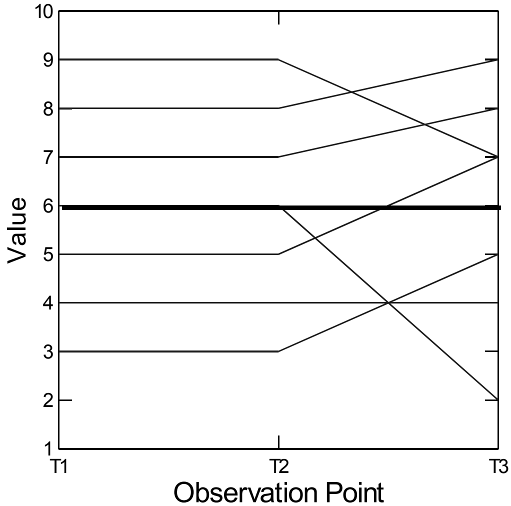


Figure 1:
Interindividual differences in intraindividual change when there is, on average, constancy

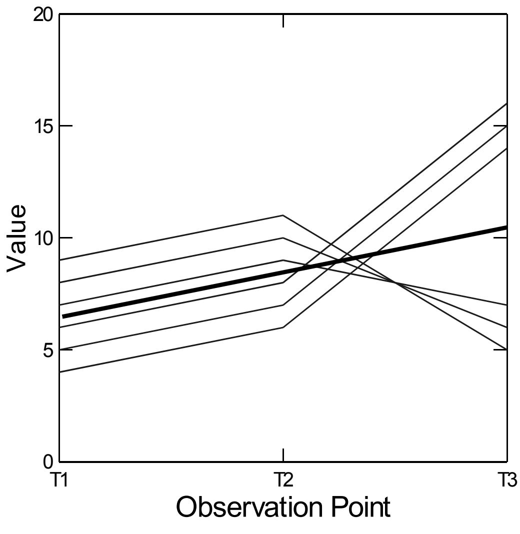


Figure 2:
Interindividual differences in intraindividual change when there is, on average, a linear trend

Figure 2 illustrates again that the average trajectory can fail to describe any of the cases in the sample. Whereas the average trajectory suggests a consistent linear trend over the three observation points, each of the cases shows a break in linear trend. The conclusion from this example is that the stability in development that is suggested by the unchanged average linear trend misses the developmental activity simulated in these data. The conclusions from both examples are captured by the tenets of person-oriented research. Two of these tenets are discussed in the next section.

Person-oriented Developmental Research

Person-oriented research was introduced by Bergman and Magnusson (1997; cf. von Eye & Bergman, 2003; Bergman, von Eye, & Magnusson, 2006; von Eye, Bergman, & Hsieh, 2013). Interwoven with holistic research, the main tenets that are of interest for the present work are

1. Functioning, process, and development of behavior are, at least in part, specific and unique to the individual; and
2. developmental processes occur in a lawful way and can be described as *patterns* of the involved factors; development can be described by constancy and change in these patterns; the meaning of the involved factors is determined by the factors' interactions with other factors.

Other tenets concern the holistic nature of development, the number of factors that need to be taken into consideration, the number of meaningfully different patterns, and the conditions that must be fulfilled for meaningful comparisons (see von Eye & Bergman, 2003; von Eye, 2010).

The first of the two tenets considered here proposes that, in principle, each individual can exhibit developmental characteristics that are unique and make the individual different from all other individuals. This does not imply that, as was illustrated in the examples in Figures 1 and 2, every individual differs from all other individuals in all respects. Similarly, this does not imply that every individual necessarily differs from all others in one or more aspects. However, differences can exist and they can be meaningful. Therefore, we intend to take them seriously, and we assign individuals to the same group only when we can be sure that they are homogeneous within the group.

The second of these tenets is related to Bergman and Magnusson's (1997) holistic perspective. It proposes that change is multifaceted and multidimensional, and that all facets and dimensions need to be taken into account. In other words, it is not sufficient to describe change in just one variable. Change (or lack of change) in multiple variables needs to be described simultaneously. Change in multiple variables constitutes *patterns of change*, and these patterns, once established, are the unit of analysis (cf. Bergman, Nurmi, & von Eye, 2012). Patterns can be described by lists of categories in categorical variables or, as was illustrated in the examples in Figures 1 and 2, by trend parameters that represent a trajectory. Interindividual differences in intraindividual change are reflected in different categories or in differences in trend parameters. In the following paragraphs, we describe change in categorical variables from a CFA perspective.

Analyzing Change in Categorical Variables with Configural Frequency Analysis

Change in categorical variables implies moving from one category to another. Interindividual differences in such change imply that intraindividual development originates in the same category but goes on to different categories, over time. This is exemplified in Table 1.

In Table 1, Individual A transitions from Category a to Category b over time from Time *i* to Time *j*. Individual B, in contrast, transitions from Category a to Category c. The same table can be used to illustrate the multivariate change, that is, the change in patterns addressed in the second tenet of person-oriented research. Suppose the category labels of the variable that spans the turnover tables in Table 1 represent patterns instead of single categories. Then, the transition from a to b is a transition from one multivariate pattern to a different one. Accordingly, the transition from a to c describes the transition from the same original pattern to a third pattern.

In Configural Frequency Analysis (CFA), a pattern is termed a *configuration*. Longitudinal CFA asks questions concerning the characteristics of transitions. In general, CFA asks questions concerning individual configurations. If a configuration is observed more often than expected, it is said to constitute a *CFA type*. If a configuration is observed less often than expected, it is said to constitute a *CFA antitype*. Based on von Eye & Gutiérrez-Peña (2004; cf. von Eye, Mair, & Mun, 2010), the statistical null hypothesis for a type/antitype decision can be formulated as follows.

Table 1:
Turnover Tables for Two Individuals

		Individual A			Individual B		
		Time j			Time j		
	Categories				Categories		
Time i		a	b	c	a	b	c
	a		x				x
	b						
	c						

Consider a two- or higher-dimensional cross-classification with R cells (configurations). For Configuration r, a test is performed under the null hypothesis H_0: $E[m_r] = \widehat{m_r}$, where m_r is the observed frequency of Configuration r, $\widehat{m_r}$ is the corresponding expected frequency, and E[.] indicates the expectancy. This null hypothesis proposes that Configuration r does not constitute a type or an antitype. If, however, Configuration r constitutes a CFA type, the null hypothesis is rejected because (using the binomial test for an example)

$$B_{N,\pi_r}(m_r - 1) \geq 1 - \alpha ,$$

where π_r indicates the probability of Configuration r. In other words, the null hypothesis is rejected because the cell contains more cases than expected. If Configuration r constitutes a CFA antitype, the null hypothesis is rejected because (again using the binomial test)

$$B_{N,\pi_r}(m_r) \leq \alpha .$$

This indicates that the null hypothesis is rejected because Cell r contains fewer cases than expected.

In longitudinal research, CFA identifies *types of constancy* and *types of changes*. Accordingly, there are *antitypes of constancy* and *antitypes of change*. A type or an antitype of constancy suggests that a particular temporal pattern that indicates no change was observed at a different rate than expected. A type or antitype of change suggests that a particular temporal pattern that indicates change was observed at a different rate than expected. Here, the terms constancy and change can refer to any parameter of series of measures.

The decision as to whether a configuration constitutes a type, an antitype, or is not suspicious is made with reference to the *CFA base model*. This model contains *all effects that* are *not of interest* to the researcher (von Eye, 2004). If this model is rejected, the effects the researcher is interested in must exist. The original CFA base model was that of variable independence (Lienert, 1968). This model takes all main effects into account, but no interactions of any order. Therefore, if the researcher is interested in interactions, the original base model is suitable. Most CFA base models are log-linear models (but other models have been discussed; see, e.g., von Eye, 2002). Consider, for example, the four variables A, B, C, and D. The log-linear base model of variable independence for these four variables is

$$\log \hat{m} = \lambda + \lambda^A + \lambda^B + \lambda^C + \lambda^D .$$

From a person-oriented research perspective, it is important to realize that the results of CFA are not expressed in terms of variable relationships. As can be concluded based on the CFA null hypothesis, the results of CFA are lists of type- or antitype-constituting configurations. Each of these reflects *local relationships*, that is, relationships that apply to patterns of variable categories, not necessarily to entire variables with all their categories (Havránek & Lienert, 1984).

A large number of CFA base models have been proposed for the analysis of longitudinal data (von Eye, 2002; von Eye, Mair, & Mun, 2010; von Eye, Mun, & Bogat, 2008, 2009). To introduce CFA models for longitudinal data, we present two models, the second of which being already suitable for the analysis of interindividual differences in intraindividual change. Later, in the next section, we show how lag analysis can be used for this purpose.

Consider the two variables X and Y, each observed twice to result in the four measures X1, X2, Y1, and Y2. One suitable base model for the longitudinal analysis of these four measures is

$$\log \hat{m} = \lambda + \lambda^{X1} + \lambda^{X2} + \lambda^{Y1} + \lambda^{Y2} + \lambda^{X1,Y1} + \lambda^{X2,Y2}.$$

This model proposes that, at Time 1, the two measures X1 and Y1 are associated, and that, at Time 2, they are associated again. This model can be rejected only if diachronous, that is, cross-time relationships between X and Y exist. In other words, this model can be rejected only if one or more of the following interactions exist: [X1, Y2], [X2, Y1], [X1, X2, Y1], [X1, X2, Y2], [X1, Y1, Y2], [X2, Y1, Y2], and [X1, X2, Y1, Y2]. Each of these terms reflects a particular diachronous interaction. Von Eye and Mair (2007, 2008) have proposed methods to determine which variable relationships cause types and antitypes in longitudinal CFA.

So far, the base models for original and longitudinal CFA did not distinguish between subgroups or even individuals. The model of longitudinal CFA just discussed can be respecified as conditional on the subjects under study. Specifically, let the first individual be labeled A and the second individual B. Then, the base model of longitudinal CFA is

$$\log \hat{m} = \lambda + \lambda^{A} + \lambda^{X1|A} + \lambda^{X2|A} + \lambda^{X1,Y1|A} + \lambda^{X2,Y2|A} + \lambda^{B} + \lambda^{X1|B} + \lambda^{X2|B} + \lambda^{X1,Y1|B} + \lambda^{X2,Y2|B}.$$

This base model can be rejected for any of the following three reasons:

1. The interactions [X1, Y2], [X2, Y1], [X1, X2, Y1], [X1, X2, Y2], [X1, Y1, Y2], [X2, Y1, Y2], and [X1, X2, Y1, Y2] exist for Individual A;
2. The interactions [X1, Y2], [X2, Y1], [X1, X2, Y1], [X1, X2, Y2], [X1, Y1, Y2], [X2, Y1, Y2], and [X1, X2, Y1, Y2] exist for Individual B; and
3. synchronous or diachronous interactions exist that link Individual A with Individual B.

Types and antitypes that result from this base model may be hard to interpret. Therefore, researchers may wish to consider the following options. First, one can specify a base model that is saturated within both individuals. The resulting base model would be

$$\log \hat{m} = \lambda + \lambda^{A} + \lambda^{X1|A} + \lambda^{X2|A} + \lambda^{X1,Y1|A} + \lambda^{X2,Y2|A} + \lambda^{X1,Y2|A} + \lambda^{X2,Y1|A}$$
$$+\lambda^{X1,X2,Y1|A} + \lambda^{X1,X2,Y2|A} + \lambda^{X1,Y1,Y2|A} + \lambda^{X2,Y1,Y2|A} + \lambda^{X1,X2,Y1,Y2|A}$$
$$+\lambda^{B} + \lambda^{X1|B} + \lambda^{X2|B} + \lambda^{X1,Y1|B} + \lambda^{X2,Y2|B} + \lambda^{X1,Y2|B} + \lambda^{X2,Y1|B}$$
$$+\lambda^{X1,X2,Y1|B} + \lambda^{X1,X2,Y2|B} + \lambda^{X1,Y1,Y2|B} + \lambda^{X2,Y1,Y2|B} + \lambda^{X1,X2,Y1,Y2|B}$$

This model can be rejected only if interactions between Individuals A and B exist. These interactions can be either synchronous or diachronous. A second option would make this model more complex because all synchronous interactions between measures from the two individuals need to be included. Finally, a 2-group model can be considered in which one attempts to discriminate between the two individuals. This model would be

$$\log \hat{m} = \lambda + \lambda^{X1} + \lambda^{X2} + \lambda^{Y1} + \lambda^{Y2} + \lambda^{X1,Y1} + \lambda^{X2,Y2}$$
$$+\lambda^{X1,X2,Y1} + \lambda^{X1,X2,Y2} + \lambda^{X1,Y1,Y2} + \lambda^{X2,Y1,Y2} + \lambda^{X1,X2,Y1,Y2} + \lambda^{I},$$

where the last term distinguishes between the two individuals to be compared. This model is saturated in the four measures used for discrimination. It can be rejected only if any of the interactions between the individuals and the measures exists. Even if an interaction that discriminates between the two individuals is synchronous, it can be interpreted as developmental because the resulting statement would be that the two individuals differ at a particular point in time.

Configural Lag Analysis

Configural lag analysis (CLA; von Eye, Mair, & Mun, 2010) allows one to analyze *intensive longitudinal data* (Walls & Schafer, 2006). This type of data involves large numbers of repetitions and typically is created, in particular in psychological research, for relatively small numbers of cases (Nesselroade & Molenaar, 2010). To introduce the concept of a *lag*, let observations be made over T occasions, with $T > 2$. Then, an observation from a point in time $t + k$, occurs with a k time units lag (for $t \leq T - k$, and $k > 0$). Accordingly, negative lags can be defined, with $k < 0$. An observation that takes place at a t - k time point occurred with a negative lag of k time units, that is, k time units before the observation at time t.

For data analysis, the string of observed measures is shifted up (for negative lags) or down (for positive lags), by k steps. This is illustrated for positive lags in Table 2 (cf. Table 11.3 in von Eye et al., 2010).

Two strings of measures, shifted by a lag of size k, can be crossed to create an I x I cross-classification, where I is the number of categories of the observed variable. This is illustrated in Table 3 (see von Eye et al., 2010; Table 11.4).

Table 2:
Longitudinal Measures with Lag 1, Lag 2, and Lag 3

Time	Original measures	Measures with Lag 1	Measures with Lag 2	Measures with Lag 3
1	x_1	-	-	-
2	x_2	x_1	-	-
.	.	x_2	x_1	-
.	.	.	x_2	x_1
.	.	.	.	x_2
		.	.	.
			.	.
				.
n -1	$x_{n\text{-}1}$	$x_{n\text{-}2}$	$x_{n\text{-}3}$	$x_{n\text{-}4}$
n	x_n	$x_{n\text{-}1}$	$x_{n\text{-}2}$	$x_{n\text{-}3}$

Table 3:
Cross-classification of a String of Scores, for a Lag of Size k

	Lag k Measures		
Original Measures	$I = 1$	$I = 2$	$I = 3$
$I = 1$	m_{11}	m_{12}	m_{12}
$I = 2$	m_{21}	m_{22}	m_{23}
$I = 3$	m_{31}	m_{32}	m_{33}

Similarly, strings from multiple lags can be cross-classified. For $k = 1$, the number of entries in a cross-classification of the type in Table 3 is reduced by 1. For $k = m$, the number of entries is reduced by m. This applies accordingly when more than one lag is considered simultaneously. In addition, the simultaneous analysis of lagged information with time-invariant information such as Gender is straightforward. The interpretation of the entries in cross-classifications like the one exemplified in Table 3 is without problems. Entry ij (for $i, j = 1, ..., I$) in this cross-classification indicates the frequency with which an observation of Category i at time t was preceded by an observation of Category j, at Time t - k.

Base models for CLA can be specified using the same criteria as for standard CFA. Consider, for example, the case in which two strings of data are available, for the two individuals A and B. The second string results from a shift by k time units. Let the original observations be labeled with O, and the lagged observations with K. Then, the base model of original CFA which proposes variable independence is

$$\log \hat{m} = \lambda + \lambda^{ID} + \lambda^{O} + \lambda^{K} \,,$$

where *ID* indicates the variable that distinguishes the two individuals. Types and antitypes from this model suggest local associations between the variables *ID*, *O*, and *K*. More pertinent to longitudinal research are the following two models. First, we ask whether the two respondents differ in their lag structure. The base model for this question includes the association between *O* and *K*, but proposes independence of *ID*, *O*, and *K*, or

$$\log \hat{m} = \lambda + \lambda^{ID} + \lambda^{O} + \lambda^{K} + \lambda^{O,K}$$

This model can be rejected only if any of the interactions [*ID*, *O*], [*ID*, *K*], and [*ID*, *O*, *K*] exist. Types and antitypes from this base model suggest that the respondents differ either in their development over time (two-way interactions) or in the development of their patterns of constancy and change, over a lag of *K* (three-way interaction). In its structure, this base model is identical to the base model of 2-group CFA (cf. von Eye, 2002).

Taking also into account that the two respondents may differ in their temporal pattern of behavior both in the original and the lagged string, the model

$$\log \hat{m} = \lambda + \lambda^{ID} + \lambda^{O} + \lambda^{K} + \lambda^{O,K} + \lambda^{ID,O} + \lambda^{ID,K}$$

may be considered. If this model is rejected, only the interaction [*ID*, *O*, *K*] can exist. Types and antitypes from this base model suggest that the two respondents differ in their pattern of constancy and change over a lag of size *K*.

Data Example

In this section, we illustrate CLA using data from a project on the development of alcohol use disorders (Perrine, Mundt, Searles, & Lester, 1995) in adulthood. A sample of self-diagnosed alcoholic males provided information about their drinking the day before, and their subjective ratings of mood, health, or quality of the day every morning. Here, we ask, whether the drinking pattern of Respondent 3000 differs from the drinking pattern of Respondent 3004. We use a lag of seven to assess drinking constancy and change in a weekly rhythm. The interesting aspect of lagged configural analysis of seven days is that the corresponding days in each week are used. Respondent 3000 provided data for 735 consecutive days. Respondent 3004 did the same for 742 consecutive days. Here, we focus on the consumption of beer.

In preliminary analyses, we found that these two respondents display quite different drinking behavior (see von Eye & Bergman, 2003). The drinking pattern of Respondent 3000 is erratic in the sense that there is no strong autocorrelation structure that can be linked to weekdays, months, or any other calendar patterns. Respondent 3004 is quite the opposite. His drinking is predictable over very long stretches of time. Not one of his autocorrelations is non-significant. Even for a lag of k = 50, the autocorrelation is stronger than 0.50. In contrast, not a single autocorrelation of Respondent 3000 reaches

0.30, not even for the shorter intervals. The two respondents also differed in the number of beers they consumed. Over the entire observation span, Respondent 3000 consumed an average number of 1.1 beers per day, with a range from 0 and 9 beers. Respondent 3004 consumed, on average, 5.1 beers per day, with a range from 0 to 14. Both respondents consumed liquor in addition (not analyzed here). Figure 3 displays the autocorrelation patterns of the two respondents.

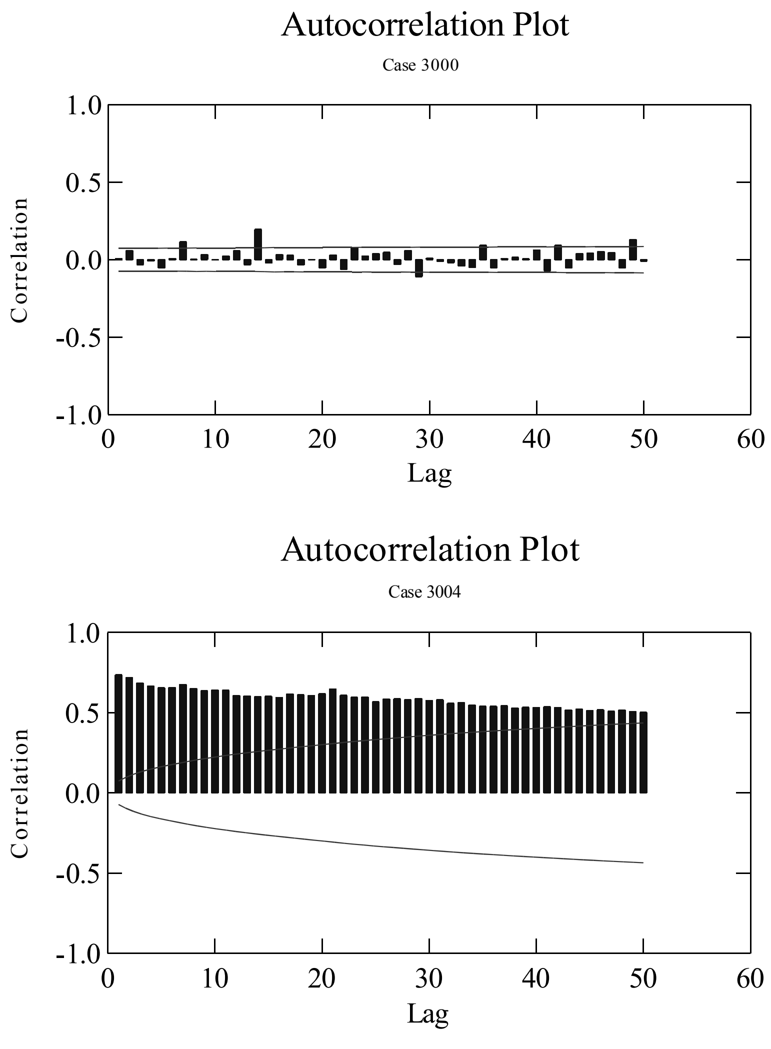


Figure 3:
Autocorrelation patterns of the beers consumed by Respondent 3000 and Respondent 3004

For the following log-linear and configural analyses, we winsorized the frequencies of beer use. This was done to prevent the cross-classifications from becoming overly sparse. Respondent 3000 never had more than 9 beers on any given day. Therefore, we created a category that represents “9 or more beers on a single day.” The number-of-beers-consumed variable (*B*) is, therefore, no longer ratio but ordinal scale. It has 10 categories, ranging from 0 through 9. The following analyses examine the 2 (*ID*) x 10 (*B*) x 10 (*B7*) cross-classification given in the appendix, where *B7* indicates beer consumption with a lag of seven days. We consider three log-linear models. The first is the main effect model of variable independence,

$$\log \hat{m} = \lambda + \lambda^{ID} + \lambda^{B} + B^{B7} .$$

Types and antitypes from this model suggest local associations between the variables *ID*, *B*, and *B7*. The second model considered here asks whether the two respondents differ in their lag structure. The base model for this question includes the association between *B* and *B7*, but proposes independence of *ID*, *B*, and *B7*, that is,

$$\log m = \lambda + \lambda^{ID} + \lambda^{B} + \lambda^{B7} + \lambda^{B,B7} .$$

This model can be rejected only if any of the interactions [*ID*, *B*], [*ID*, *B7*], and [*ID*, *B*, *B7*] exists. Types and antitypes from this base model suggest that the two respondents differ either in their longitudinal pattern of beer drinking over time (two-way interactions) or in the development of their longitudinal patterns of constancy and change in beer drinking, over a lag of $K = 7$ (three-way interaction).

The third model takes into account that the two respondents may differ in their temporal pattern of beer drinking both in the original series of observations and the lagged observations. This is the model

$$\log m = \lambda + \lambda^{ID} + \lambda^{B} + \lambda^{B7} + \lambda^{ID,B} + \lambda^{ID,B7} + \lambda^{B,B7}$$

If this model is rejected, only the interaction [*ID*, *B*, *B7*] can exist. Types and antitypes from this base model suggest that the two respondents differ in their pattern of constancy and change in beer drinking over a lag of size *K* for a span of over two years of daily observations.

The first of these three models comes with a goodness-of-fit $G^2 = 2217.75$ ($df = 180$; $p < 0.01$). This large value indicates that there are strong relationships in the three-way table. From a configural perspective, we ask where, in the table, the biggest deviations can be found. In this example, the biggest deviations, in units of standardized deviates, suggest weekday-specific stability. The biggest deviation constitutes a type. It is found in Cell 8 8, for Respondent 3004. For 32 of the 742 observation days as well as for the corresponding day one week later, Respondent 3004 had indicated that he consumed 8 beers. Under the main effect base model of variable independence, 4.15 days had been expected ($z = 13.67$; $p < \alpha^*$). This type suggests drinking behavior that is stable for the same day in consecutive weeks. The second biggest deviation, also indicating a type, was also found for Respondent 3004, for Cell 8 9. This configuration indicates drinking behavior that is close to stable. Eight beers on one weekday are followed by 9 or more beers on the same day of the next week. Similarly, the strongest deviation for Respondent 3000 also indicates a stability type. It is found for Configuration1 1. For 169 of the 735 observation days and the corresponding same weekday one week later, Respondent 3000 indicated that he had not consumed any beer. Under the main effect base model, 76.74 days had been expected.

In general, most of the (near-) stability types for Respondent 3000 were found for corresponding days on which this respondent consumed no beer or only small numbers of beers. For Respondent 3004, the opposite was true. (Near-) stability types were found for the corresponding days on which he consumed 5 or more beers.

Only a few configurations constituted antitypes. These were all configurations that describe corresponding days for Respondent 3000 on which he leaped from drinking nothing to drinking 7 or more beers. The same applies for Respondent 3004.

From the perspective of studying interindividual differences in intraindividual change, more interesting is the direct comparison of the beer drinking patterns of the two respondents. The second of the above three base models allows one to perform such a comparison. The goodness-of-fit $G^2 = 1033.71$ ($df = 99$; $p < 0.01$) for this model suggests that this model is significantly better than the original base model ($\Delta G^2 = 1184.04$; $\Delta df = 81$; $p < 0.01$). It should be noted, however, that both of the programs we used to analyze the frequency table in the appendix (Lem and SYSTAT 13) indicated convergence problems. Here, we interpret the solution provided by Lem (Vermunt 1993), because even the Delta option, invoked with $\Delta = 0.05$, did not improve the solution provided by SYSTAT (cf. the discussion of computational issues in log-linear modeling in von Eye & Mun, 2012). The issue was that a number of cell frequencies were estimated to be near zero, which caused the program to have problems estimating parameters for these cells.

Still, model fit is poor and we tentatively inspect the differences between the observed and the expected cell frequencies and the corresponding standardized residuals to identify the largest differences between the two respondents. We find that, as for the original base model, the largest discrepancies come in the form of types, and they suggest stable drinking behavior over the observation period of more than two years. Specifically, Respondent 3000 shows stability mostly in the domain of no drinking or drinking only small numbers of beers. In the domain of drinking large numbers of beers, we find stability antitypes. The strongest of these suggests that this respondent drinks 7 beers on a given day as well as the corresponding day one week later at a rate far lower than expected. In fact, this pattern was not reported at all (see the table in the appendix). The same applies to (not) consuming 9 beers or more.

In contrast, Respondent 3004 shows an antitype that is constituted by Cell Configuration 1 1. This antitype suggests that this respondent reported significantly fewer corresponding non-drinking days than expected (84 versus 127.02). Clearly this difference goes in the opposite direction as for Respondent 3000, and it is significantly stronger for respondent 3000. The other discrepancies are similar to the ones found with the original base model. Respondent 3004 is more stable than Respondent 3000 when it comes to reporting the consumption of large numbers of beers (7 and more), and Respondent 3000 is more stable than his counterpart when it comes to reporting the consumption of small numbers of beers, or none. No additional antitypes emerged from this base model.

Considering that the development of beer drinking may be specific to the respondent (third base model) yields a different picture. The overall goodness-of-fit for this model was $G^2 = 92.33$ ($df = 81$; $p = 0.18$). This result suggests that this model is not only significantly better than the first base model ($\Delta G^2 = 2125.42$; $\Delta df = 99$; $p < 0.01$), but it can also stand for itself and describes the data well. The three-way interaction between the three variables that span the cross-classification in the appendix is not needed to explain the frequencies. In the context of hierarchical log-linear models, this base model would be saturated anyway, and no types or antitypes could possibly result.

We conclude that the differences between Respondent 3000 and Respondent 3004 can be satisfactorily explained by the respondent-specific constancy and change in alcohol consumption. The differences in the one week lags are not needed to describe the interindividual differences in intraindividual change between these two respondents.

Discussion

From the perspective person-oriented research, the study of interindividual differences in intraindividual change is most interesting and important. There exist, unfortunately, only small numbers of empirical projects that create data that allow one to undertake such comparisons. In most cases, intensive longitudinal data as described by Walls and Schafer (2006) are needed for such comparisons. Most of the small number of published works on interindividual differences in intraindividual change is either theoretical (Baltes, Reese, & Nesselroade, 1977) or methodological (Nesselroade & Molenaar, 2010, von Eye et al., 2010). Empirical papers exist in which P-technique and related methods are applied mostly to physiological and mood data (for an overview, see Jones & Nesselroade, 1990).

Interestingly, P-technique and related methods carry strong variable-oriented elements even when they are applied in person-oriented research. These elements are that statements are made under the assumption that variable relationships exist and are valid for the entire range of admissible scores. The approach presented in this article shows that this assumption may not always hold. We found that the model of all two-way interactions explains the data well. However, the two-way interactions are not carried by all levels of drinking. Specifically, we found that the differences between the two respondents are most evident at the extreme levels, that is, for zero or small numbers of beers consumed (Respondent 3000) and for seven or more beers (Respondent 3004). In between these numbers, the observed frequencies do not differ from those estimated, even for the base model of variable independence. We conclude that *local associations* as defined by Havránek and Lienert (1984) can be identified, in particular, when a configural approach is adopted.

Two issues of concern should not be overlooked. The first may be specific to the data used for the example in this article, and the second is more general in nature. The first issue concerns the nature of data that can make it hard for computer programs to estimate models. In the present data example, both the column and the row that are constituted by having consumed eight beers are empty, for Respondent 3000. As soon as interactions are estimated that involve this respondent, this pattern causes problems, and generally available, even commercially available software report problems with convergence. For example, Lem reported, for the third of the above base models, eight (nearly) boundary or non-identified (log-linear) parameters and 29 zero estimated frequencies. In contrast, SYSTAT reports that the solution to this problem is not unique and gives an overall goodness-of-fit result of $G^2 = 3247.48$ ($df = 164$; $p < 0.01$). This result is dramatically different than the result given by Lem, and this difference cannot be explained by rounding. Instead, these differences reflect the way the programs handle the problem of estimating frequencies to be zero. We, therefore, recommend that researchers recalculate their results using different programs, to make sure that published results are data-specific instead of reflecting software peculiarities.

The second issue is statistical. It concerns with the assumptions made about the independence of data in frequency tables such as the one in the appendix. This issue is important because it concerns most of the longitudinal models that are estimated for log-linear models or CFA. Assuming independence between the observations of the same individual in longitudinal studies is rather common (it simplifies things), but may have some undesirable consequences. Liang and Zeger (1986) consider independence a *working assumption*, but also consider some alternatives. To the best of our knowledge, the modern approach to dealing with the lack of independence is to assume that independence holds, but *conditional on a latent process* (or latent random variable). This approach seems to work but may be difficult to implement. Further work is required in data analysis under this assumption.

References

Baltes, P. B., Reese, H. W., & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods*. Monterey, CA: Brooks/Cole.

Bergman, L. R., & Magnusson, D. (1997). A person oriented approach to research on developmental psychopathology. *Developmental Psychopathology*, *9*, 291-319.

Bergman, L. R., von Eye, A., & Magnusson, D. (2006). Person-oriented research strategies in developmental psychopathology. In D. Cicchetti & D. J. Cohen (Eds.), *Developmental psychopathology*, 2nd ed. (pp. 850-888). London, UK: Wiley.

Havránek, T., & Lienert, G. A. (1984). Local and regional versus global contingency testing. *Biometrical Journal, 26*, 483-494.

Jones, C. J., & Nesselroade, J. R. (1990). Multivariate, replicated, single-subject designs and P-technique factor analysis: A selective review of intraindividual change studies. *Experimental Aging Research*, *16*, 171-183.

Liang, K.-L., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*, 13-22.

Lienert, G. A. (1968). *Die "Konfigurationsfrequenzanalyse" als Klassifikationsmethode in der klinischen Psychologie* [*Configural Frequency Analysis - a classification method for clinical psychology*]. Paper presented at the 26. Kongress der Deutschen Gesellschaft für Psychologie in Tübingen 1968.

Lienert, G. A., & Krauth, J. (1975). Configural Frequency Analysis as a statistical tool for defining types. *Educational and Psychological Measurement*, *35*, 231-238.

Nesselroade, J. R., & Molenaar, P. C. M. (2010). Analyzing intra-person variation: Hybridizing the ACE model with P-technique factor analysis and the idiographic filter. *Behavior Genetics*, *40*, 775-783.

Perrine, M. W., Mundt, J. C., Searles, J. S., & Lester, L. S. (1995). Validation of daily self-report consumption using interactive voice response (IVR) technology. *Journal of Studies on Alcohol and Drugs*, *56*, 487-490.

Vermunt, J. K. (1997). *Lem: A general program for the analysis of categorical data*. Tilburg University: unpublished program manual.

von Eye, A. (2002). *Configural Frequency Analysis - Methods, models, and applications*. Mahwah, NJ: Lawrence Erlbaum.

von Eye, A. (2010). Developing the person-oriented approach - Theory and methods of analysis. *Development and Psychopathology*, *22*, 277-285.

von Eye, A., & Bergman, L.R. (2003). Research strategies in developmental psychopathology: Dimensional identity and the person-oriented approach. *Development and Psychopathology*, *15*, 553-580.

von Eye, A., Bergman, L. & Hsieh, C.-A. (2013). Person-oriented approaches. In Overton, W. F. & Molenaar, P.C.M. (eds.), *Handbook of child Psychology – Theory and methods*. New York: Wiley (in press).

von Eye, A., & Gutiérrez Peña, E. (2004). Configural Frequency Analysis - the search for extreme cells. *Journal of Applied Statistics*, *31*, 981-997.

von Eye, A., & Mair, P. (2007). A functional approach to Configural Frequency Analysis. *Austrian Journal of Statistics*, *37*, 161-173.

von Eye, A., & Mair, P. (2008). Functional Configural Frequency Analysis: Explaining types and antitypes. *Bulletin de la Société des Sciences Médicales, Luxembourg*, *144*, 35-52.

von Eye, A., Mair, P., & Mun, E.-Y. (2010). *Advances in Configural Frequency Analysis*. New York: Guilford Press.

von Eye, A., & Mun, E.-Y. (2012). *Log-linear modeling – Concepts, interpretation, and applications*. New York: Wiley. (in preparation)

von Eye, A., Mun, E.-Y., & Bogat, G. A. (2008). Temporal patterns of variable relationships in person-oriented research - Longitudinal models of Configural Frequency Analysis. *Developmental Psychology*, *44*, 437-445.

von Eye, A., Mun, E. Y., & Bogat, G. A. (2009). Temporal patterns of variable relationships in person-oriented research - Prediction and auto-association models of Configural Frequency Analysis. *Applied Developmental Science*, *13*, 172-187.

Walls, T. A., & Schafer, J. L. (Eds.)(2006). *Models for intensive longitudinal data*. New York: Oxford University Press.

Appendix

Cross-classification Respondent x Number-of-beers-consumed x Number-of-beers-consumed on the corresponding day in the following week

Observed Frequencies

IDNUMBER	BEER1	BEER7									
		0	1	2	3	4	5	6	7	8	9
3,000	0	169	65	50	28	15	7	4	1	0	0
	1	81	37	25	18	3	1	0	0	0	0
	2	44	35	17	11	8	1	1	0	0	1
	3	22	17	18	4	2	0	0	0	0	0
	4	18	7	2	0	0	0	1	0	0	0
	5	2	2	4	0	0	0	0	0	0	0
	6	2	1	1	2	0	0	0	0	0	0
	7	0	0	1	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0	0	0
	9	0	1	0	0	0	0	0	0	0	0
3,004	0	84	15	16	9	5	2	2	1	1	1
	1	21	4	3	2	0	0	0	1	0	1
	2	11	6	2	2	1	2	4	4	3	1
	3	9	3	1	3	1	0	6	2	2	1
	4	2	1	2	3	3	5	15	3	3	5
	5	3	1	2	4	7	10	14	5	14	4
	6	1	0	3	3	10	11	29	21	20	17
	7	1	2	2	0	5	11	12	19	17	16
	8	2	0	2	2	4	10	18	16	32	24
	9	3	1	1	0	6	13	15	13	18	17

Psychological Test and Assessment Modeling, Volume 54, 2012 (2), 168-178

Scientometric analyses of the international visibility of German psychology researchers and their range of specialization

Clemens B. Fell[1], *Alexander von Eye*[2],
Gabriel Schui[3] *& Günter Krampen*[4]

Abstract

With reference to the role of networking, accelerated by current developments within large parts of the scientific community, the assumption is examined that the range of specialization of scientists in terms of membership in professional sections of scientific societies is related to the international impact of their publications. The sample consists of 2,788 German psychologists enrolled in the German Psychological Society (Deutsche Gesellschaft für Psychologie, DGPs). A log-linear model suggests that the citation pattern of DGPs members with no citations of their papers published in 2000 or 2005 respectively in the time intervals 2000-2004 or 2005-2009 generally differs from that of their colleagues across four ranges of specialization categories. Configural Frequency Analysis led to the identification of distinct subgroups of scientific specialization and international visibility, i.e., citations by others. Specifically, for those individuals who enjoy international visibility, one key to success seems to be multiple professional specializations with reference to different subdisciplines of psychology.

Key words: Professional specialization, Internationality, International impact, Citation analysis, Configural Frequency Analysis (CFA)

---

[1] Saarland University, Saarbrücken, Germany

[2] Michigan State University, East Lansing, MI, USA, and University of Vienna, Vienna, Austria

[3] *Correspondence concerning this article should be addressed to:* Gabriel Schui, PhD, Leibniz Institute for Psychology Information (ZPID), 54286 Trier, Germany; email: gabriel.schui@zpid.de

[4] Leibniz Institute for Psychology Information (ZPID), University of Trier, Trier, Germany

Introduction

Gigerenzer et al. (1999) proposed seven recommendations for improving the international visibility and reception of German psychology. These recommendations have given renewed rise to a fervid discussion: the language dispute in German psychology. In the context of this discussion, the internationalization that results from English language publishing and citations of German psychologists' publications continuously increases (Krampen, Schui, & Fell, 2010). However, internationalization does not take place homogeneously in German psychology because psychology is a heterogeneous discipline (Krampen et al., 2010): Particular subfields more than others address reports about their gain of knowledge to a more international audience which makes them internationally more visible than others. More visible implies that publications and citations by scholars in the same subfields are mostly available in English. In this paper, we ask whether there exists any factor that is unrelated to the content of subdisciplines but related to the international visibility and citations of German psychology. We hypothesize that networking within and between different professional specializations may be a crucial but, up to now, empirically unexplored influence. In the present study, we consider membership in one (i.e., single specialization) or more subgroups (i.e, multiple specialization) of the German Psychological Society (*Deutsche Gesellschaft für Psychologie, DGPs*) to be such a factor.

Many authors emphasize the importance of scientific societies for the success of their members (Anderson & Schultz, 2003; Bickel, 2007; Schimank, 1988). Bickel (2007, p. 91) sums up: "Professional societies form a living matrix where minds meet and engage and where trusted colleagues pool their knowledge, helping each other to glimpse and plumb larger forces at work, to see connections among events, and to imagine the future." Anderson and Schultz (2003) emphasize that professional societies advance both their disciplines and their individual members. In member support, the authors distinguish between instrumental, supportive, normative and voluntary interests of their members. Schimank (1988) focuses on instrumental functions and differentiates: 1) furthering scientific communication, 2) supporting careers and representing collective interests, 3) providing a meeting place and a research result trafficking spot, 4) advising in science policy decisions. Especially in consideration of the changing scientific work environment, scientific societies are important – and they may become ever more important for the scientific success of their members. Authors widely agree about the quintessential changes in scientific knowledge production and the necessity to appropriately adapt to these changes (Houghton, 2005; Houghton, Steele, & Henty, 2004; Rowlands & Fieldhouse, 2007; Thagard, 2005): There is an increasing diversity in the work conditions of scientists, and the portion of inter-/multi-/transdisciplinary and problem-oriented research is also increasing. In addition, one can find an emphasis on collaborative work and (more diverse/informal) communication and an increase in the size of research teams.

Nowadays, most scientists are involved in wider collegial networks. We concur with Houghton et al.'s (2004, p. 240) conclusion that there are two main reasons for collaboration: The first reason reflects technical or material reasons concerning costly equipment and facilities. The second reason reflects demands of "a wider range of specialist skills, with collaboration a response to complexity, increasing interdisciplinarity and an increasingly problem-oriented approach to research" exist. We also concur with Thagard (2005) that the more distinct the professional profile of a scientist is, the more effective his networking will be, and the more suitable he/she is for problem-focused, multidisciplinary and transdisciplinary research in collaborative teams, the more successful in terms of impact and visibility he/she will probably be. Considering the important functions of scientific societies, the current changes in the scientific environment, the societies' influence on success and visibility of every single member becomes obvious.

Comparable to the American Psychological Association (APA) in the U.S., the German DGPs is "[…] an incorporated association of qualified psychologists engaged in research and teaching. Its goal is to advance and expand scientific psychology" (German Psychological Association, 2008). Comparable to the APA divisions, the DGPs is subdivided in 15 professional sections, largely corresponding to the common psychological subfields (e.g., General Psychology, Clinical Psychology, Work and Organizational Psychology, Differential Psychology, Developmental Psychology). Every DGPs member is encouraged to join, in addition to his/her general membership, one or more of these sections (section membership comes with an annual fee). The choice of joining one or more sections depends on and reflects (e.g., in his/her public DGPs online profile) a scholar's scientific main foci. Thus, this sections' memberships are an indicator of single versus multiple specialization in one, two, three or even more subdisciplines of psychological research.

Summing up, networking within scientific associations is very helpful to scientists. The ongoing changes in the scientific environment motivate scholars to join professional societies. In addition, professional specialization in terms of a membership in one or more professional sections of a scientific association may be of importance. A clear scientific profile facilitates collaboration and communication in only one or in multiple and diverse research groups. One may assume that the international visibility of scientists in terms of citations in international journals relates to their networking in the context of single versus multiple professional specialization. In brief, the present scientometric study explores whether professional specialization of German psychologists in terms of memberships in one or more professional sections of the DGPs is related to their international impact in terms of the citations of their publications in English language journals over a span of 10 years.

Method

The number of DGPs members examined in this study is 2,788. This includes every member of the DGPs as of January 2010. The DGPs has 15 sections, and a large number of members of the DGPs is enrolled in multiple sections (M = 1.41, SD = 0.96, Med = 1, *Range* [0;15]).

"Most attempts to measure the 'quality' of papers or the recognition of scientists use either citations received or the impact factor of the journal as indicators" (Larivière & Gingras, 2010, p. 424). In this study, we calculate the international impact for every

DGPs member as the number of citations in an international journal indexed in the *Social Sciences Citation Index (SSCI)* during two consecutive time periods, excluding self-citations. International journals are defined as English language journals. The two citation periods cover the years 2000-2004 of the DGPs members' publications published in 2000 (= C0) and 2005-2009 of those published in 2005 (= C5). So, two citation numbers resulted for each member, one for each time period of five years.

This approach intentionally differs from the Journal Impact Factor (JIF), which is calculated for journals and which is – sometimes – added up to a personal cumulative JIF of a particular scientist. Using the JIF for evaluating scientists is far from indisputable. In his highly cited paper, Seglen (1997) describes a number of problems of the JIF. Our approach allows one to deal with some of the biggest problems: The sum of citations of an article during a certain time period is more closely related to the article itself than the JIF, which is a descriptor of the journal. The JIF does not care about self-citations; in contrast we explicitly cull them. The SSCI, data basis for JIF calculation, has an English language bias. Overall, only 7.2 % to 7.4 % (in the examined subject area 'psychology') of the indexed publications were not published in English. In the study at hand the object of research is the international visibility of individual psychologists. So the subject matter is unaffected by this problem, because only citations in Anglophone publications are examined here. In addition, according to the JIF calculation, every article has a time span of one to three years after its publication during which it can make a contribution to the corresponding JIF. However, several authors (e.g., Glänzel & Schoepflin, 1995; Vanclay, 2009) assume that the ageing of publications varies depending on the particular journal or field. The JIF fails to reflect these variations. Vanclay (2009) illustrates this point:

> The JIF does not deal evenly with 'Hares' (journals to which citations accrue quickly over a confined period) and 'Tortoises' (journals to which citations accrue slowly over an extended period), because the 2-year sample represents a much larger proportion of total citations for the Hares. (p. 4)

And whereas Tortoises show quite consistent citation trends, the Hares do not (Vanclay, 2009). Considering the scientific heterogeneity of psychology (Krampen et al., 2010) it is assumed that both exist, Tortoises as well as Hares. For that reason, we use 5-year periods for both citation measuring intervals.

The two variables *Citation Record for the years 2000-2004 (C0),* and *Citation Record for the years 2005-2009 (C5)* were created as follows. First, a frequency variable was defined that consisted of the counts of citations (excluding self-citations) for each individual in the subject area, i.e., journal subject type, 'psychology' in the SSCI, for each of the two 5-year periods. Then, in order to prevent the cross-classification of citation counts and section belonging from becoming too sparse, the upper end of the distribution was aggregated into one category. This category contains all those incidences in which an individual was cited 5 or more times during the observation period. Thus, both C0 and C5 now have 6 categories. The first category indicates that an individual was not cited at all. The second contains those DGPs members who were cited once, and the 6th category contains those DGPs members who were cited five or more times by others.

Considering that the DGPs has 15 sections, and also considering that a large number of members of the DGPs is enrolled in multiple sections, crossing sections with C0 and C5 would have resulted in a very large, unwieldy, and sparse table. Therefore, membership patterns and sections were aggregated based on the *number of section affiliations (CS)* indicated by the DGPs members. The aggregation resulted in four superordinate sections. The first contains those 342 DGPs members who do not belong to any of the sections (it is possible to be a member without joining a section). The second contains those 1,317 DGPs members who belong to one section (single specialization in research). The third contains those 862 DGPs members who belong to two sections (double specialization in research), and the fourth contains those 267 DGPs members who belong to three or more sections (multiple specialization in research).

The following analyses use the three variables: Classification of Subdisciplines of Psychology (CS), Citation Record for the years 2000-2004 of the publications published in 2000 (C0), and Citation Record for the years 2005-2009 of the publications published in 2005 (C5). The classification is based on the section system of the DGPs. Crossed, the three variables under study form a 4 x 6 x 6 frequency table.

Results

In the following sections, we present results from two approaches to the analysis of the SSCI citation data. First, we present a log-linear model that explains the frequency distribution in the cross-classification of the variables Classification of Subdisciplines of Psychology (CS), Citation Record for the years 2000-2004 (C0), and Citation Record for the years 2005-2009 (C5). Second, we perform a Configural Frequency Analysis of the same data. The goals of these analyses are: 1) Explanation of the association structure of CS, C0, and C5 patterns at the level of variables; and 2) Identification of those patterns of the development of citation records that stand out as particularly likely or particularly unlikely.

The hierarchical log-linear model that provides a satisfactory explanation of the 4 x 6 x 6 (CS x C0 x C5) cross-classifications contains all three two-way interactions. The model comes with a goodness-of-fit LR-χ^2 = 93.69 [illegible]h suggests non-significant model-data discrepancies, that is, good fit (df = 75; p =.07). The model involves 1 parameter estimate for the intercept, 13 main effect parameters, and 55 interaction parameters. Of these 55 interaction parameters, 12 are significant. The majority of these indicate that the citation pattern for DGPs members with no citations differs from the citation pattern for members with 1 or more citations, across all categories of the aggregated number of subdisciplines scale.

This log-linear model is equivalent to a logit model in which C0 and C5 are predicted from CS. We, therefore, conclude that, in general, a membership pattern that reflects the number of subdivisions a DGP member is enrolled in is a good predictor of the member's citation success. We now follow up this result by highlighting the particular patterns that carry this predictive relationship.

Instead of interpreting each parameter in detail, we perform a Configural Frequency Analysis (CFA; Lienert & Krauth, 1975; von Eye & Gutiérrez Peña, 2004; von Eye, Mair, & Mun, 2010). This technique of data analysis allows one to identify those cells in a cross-classification that stand out as observed significantly more often (CFA types) or less often (CFA antitypes) than expected. CFA types and antitypes reflect the interactions needed to explain a table as a whole. For the present purposes, we perform a first order CFA. This CFA model estimates the expected cell frequencies based on the main effects of the variables that span the table. The types and antitypes explain the interactions needed for the above log-linear model. For the cell-wise CFA tests, we use the z-test. To protect α, we use the Holland-Copenhaver (1987) procedure. Table 1 displays a summary of results from CFA, and shows that CFA resulted in 16 types and 9 antitypes.

Pattern 1

Pattern 1 describes those DGPs members, who are not members of any professional section (12.27 %, n = 342). Type 1 1 1 describes those 191 (55.85 %) of those members whose 2000 and 2005 papers experience no citation in an English psychology-related publication during either 5 years observation period. Far more individuals display Pattern 1 1 1 than one would expect by chance. Therefore, Configuration 1 1 1 is said to constitute a type. One may speculate whether these individuals are still active as scholars at all. They could be retirees or practitioners since at least 2000. Accordingly, less than half of the DGPs members without section affiliation (44.15 %, n = 151) achieve at least one English citation during either period. Far fewer members than one would expect by chance, one could call them the ‘whizz kids’, are in lowest C0- and in the highest C5-citation category (Antitype 1 1 6, 0.88 %, n = 3).

Pattern 2

Pattern 2 describes those DGPs members, who participate in exactly one professional section, i.e. they are involved in one research area with a single specialization (47.24 %, n = 1,317). It is the largest group of the sample. Here also, there are remarkably many individuals who experience no international citation during either observation period (Type 2 1 1, 42.90 %, n = 565), but their percentage is smaller than within the previous group. Then there are 29.69 % (n = 391) of the one-section members whose C5 citation amount is increased compared to C0 (vs. 18.13 %, n = 62 no-section members) and 7.74 % (n = 102; vs. 6.43 %, n = 22 no section-members) with stable, high citation numbers. So, altogether, there are more ‘climbers’ and ‘keepers’ and, for some patterns, there are even more than one would expect by chance (Types 2 4 6, 2 5 5, 2 6 6). Significantly fewer scientists than expected experience their international citation breakthrough between 2000 and 2005 (Antitypes 2 1 5, 2 1 6, 1.75 %, n = 23), but, again, these numbers are higher than in the previous no-section-group. Another fact worth noting is that remarkably fewer one-section members than one would expect by chance experience dramatically lower citation numbers in C5 than in C0 (Antitypes 2 3 1, 2 4 1, 2 6 1).

Table 1:
Types and Antitypes of the Configural Analysis of the Citation Development Data from the Social Science Citation Index

Configuration CS C0 C5	m	$\hat{m}$	*z*	*p*	
1 1 1	191	106.77	8.15	0.0000	Type
1 1 6	3	17.04	-3.40	0.0003	Antitype
1 6 6	6	1.57	3.53	0.0002	Type
2 1 1	565	411.17	7.59	0.0000	Type
2 1 5	11	31.67	-3.67	0.0001	Antitype
2 1 6	12	65.61	-6.62	0.0000	Antitype
2 3 1	36	65.29	-3.62	0.0001	Antitype
2 4 1	14	36.71	-3.75	0.0001	Antitype
2 4 6	16	5.86	4.19	0.0000	Type
2 5 5	6	1.63	3.42	0.0003	Type
2 6 1	8	37.94	-4.86	0.0000	Antitype
2 6 6	27	6.05	8.51	0.0000	Type
3 1 6	12	42.94	-4.72	0.0000	Antitype
3 3 1	20	42.73	-3.48	0.0003	Antitype
3 4 6	13	3.83	4.68	0.0000	Type
3 5 4	7	1.60	4.26	0.0000	Type
3 5 5	6	1.07	4.77	0.0000	Type
3 6 1	0	24.83	-4.98	0.0000	Antitype
3 6 5	10	1.91	5.85	0.0000	Type
3 6 6	29	3.96	12.58	0.0000	Type
4 2 4	9	2.68	3.86	0.0001	Type
4 5 3	5	0.92	4.26	0.0000	Type
4 5 6	4	0.69	4.00	0.0000	Type
4 6 3	6	1.65	3.40	0.0003	Type
4 6 6	11	1.23	8.82	0.0000	Type

Note. *m* = observed cell frequencies; $\hat{m}$ = expected cell frequencies; *z* = *z*-Score; *p* = significance level.

Pattern 3

Pattern 3 describes those 862 (30.92 %) psychologists who are members of two professional DGPs sections, i.e. they are involved in two research areas with a double specialization. For this pattern, dramatic crashes (Antitypes 3 3 1, 3 6 1) as well as a comet-like rises (Antitype 3 1 6) exist less often than one would expect by chance. In contrast, modest changes in both directions are significantly noticeable (Types 3 4 6, 3 5 4, 3 5 5, 3 6 5). In comparison with the previous patterns, we note that the percentage of not cited scientists (34.11 %, *n* = 294) decreases and the rates of keepers (7.89 %, *n* = 68), climbers (31.15 %, *n* = 303) and whizz kids (3.13 %, *n* = 27) are higher.

Pattern 4

A look at those 267 (9.58 %) DGPs members who are members of three or more professional sections (Pattern 4 = multiple specializations) reveals considerable non-random differences between C0 and C5: increases (Types 4 2 4, 4 5 6) and decreases (Types 4 5 3, 4 6 3). It is interesting to note that these members do not experience zero citations in significant numbers. Again, compared to all previous patterns, the percentages of uncited scientists (28.46 %, *n* = 76) are lower and those of keepers (11.99 %, *n* = 32), climbers (35.21 %, *n* = 94) and whizz kids (4.49 %, *n* = 12) are higher.

In all 2000 and 2005 publications (if available at all) 40.39 % (*n* = 1,126) of the 2,788 analyzed DGPs members were never cited in an international psychology-related journal indexed in the SSCI during the observed time span of 10 years. Of these, the majority are members of one DGPs section (*n* = 565; 50.18 %) which is, in absolute numbers, the largest group showing single specialization in research. In percent, pattern 1 contains most of the uncited scientists.

Configuration '. 6 6' (2.62 %, *n* = 73) constitutes a type for each of the four categories of CS, which indicates, that across all membership categories, there are significantly more DGPs members than expected who are cited five or more times in an English publication in the two observation periods. Two characteristics of these four types are worth noting. First, the numbers of cases in these cells are much smaller than the numbers of cases who are never cited. Second, of those four types, the one for the members who do not belong to any subgroup (1 6 6) describes the smallest group (both in percent and frequency), whereas the one for the members who belong to three or more subgroups contains most of the highly cited scientists (in percent).

Moving to the antitypes, we note that configuration '. 1 6' (*n* = 35) constitutes (configuration 4 1 6 representing the only exception) antitypes for each of the section categories. These are individuals who experience an increase from zero to five or more citations from the first to the second observation period. In comparison with the sizes of the CS-groups, the relatively largest portion of whizz kids scrimmage in 4 1 6 (3.00 %, *n* = 8), or, in percent: the smaller the number of section memberships, the fewer whizz kids.

Discussion and conclusions

In view of the increase in international visibility (decreasing numbers of uncited scientists, increasing numbers of keepers, climbers and whizz kids) of pattern 4 . ., one might argue that, at least, the frequencies in the last pattern indicate evidence of the benefits of professional diversification. But, in fact, only 55 (20.60 %) of the 267 scientists in pattern 4 . . belong to even more than three sections of the DGPs. So, one may assume that the majority of DGPs members of the fourth group do not exhibit an indiscriminate diversification, but rather their multiple professional focusing. Ascribing the large number of uncited scientists during the observed time span of 10 years to retirees may not be an entirely satisfactory explication. Fewer than 10 % of all DGPs members are at an age in which they could have been retirees for the whole 10 year span of observation.

The cell frequencies of the types 1 6 6, 2 6 6, 3 6 6, and 4 6 6 indicate that among the psychologists organized in the DGPs there are just a few who are well established in terms of international visibility. In absolute numbers, most of them belong to one or two professional sections, but the relative frequencies within configuration '. 6 6' increase consistently across the stages of CS, i.e. increasing number of subdivision memberships, suggesting multiple specialization.

Two conclusions might be drawn from these results: (1) There is indeed only a small number of individuals who enjoy solid international visibility; (2) For these individuals, one key to success seems to be multiple professional specialization within psychological research.

We note that configuration '. 1 6' constitutes antitypes for each of the section categories (4 1 6 is the only exception). The group of those members who report fewer section memberships contains fewer whizz kids (in percent) than the groups of members who report more section memberships. But who are these up-and-coming scientists – the so called whizz kids? One may speculate whether they are perhaps scientific youngsters at the beginning of their careers who experience first citation success after the first five observation years. In accordance with Schneller and Schneider (2004) in Germany, graduate students of psychology are 29.9 years old on average (*SD* = 5.61). In the DGPs 115, members are younger than 30 years. Krampen et al. (2010) point out that the internationalization of German psychology in terms of publication and reception has continuously been rising for about ten years, first of all within younger psychologists. So, there could be about 35 budding scientists in the sample: no international citations of their 2000-publications (if available at all) in 2000-2004 and five or more citations of their 2005-publications during the following five years of observation. This very successful group of scientists has managed to establish themselves in the international psychology community in a short period of time.

Altogether, our results confirm the assumption that section membership within the DGPs, in terms of networking and multiple professional profile sharpening, improves the international visibility of German psychologists, as measured by the citation numbers of their publications in the international community. Of course, this is not a proof of a causal relationship. One could also conclude that the more successful scientists turn their attention to more particular topics, and, thus join the corresponding sections. Either way, there is a relation between professional focusing on one hand and international visibility as well as international scientific success on the other. From our perspective, the former explanation seems considerably more traceable. To conclude, we find it worth noting that the 16 types and 9 antitypes suggest that psychology is a very heterogeneous field in regard to the multiple subdisciplines and the psychologists themselves in terms of publication and reception of scholarly work (Krampen, Fell, & Schui, 2011).

Four aspects are worthy of further examination: First, we do not know whether DGPs members generally differ in international visibility from those psychologists who do not belong to the DGPs or other learned societies. However, by far the most of the German psychology researchers are members of the DGPs. Second, we do not know whether the citation advantage of the individuals belonging to multiple DPGs sections is rooted in the possibility that they particularly cite each other, maybe because they are more familiar with the papers of their colleagues. Third, we do not know whether the citation patterns found here are subfield-specific and, thus, reflect the subfields' well-analyzed specific internationalization. Last, but not least cross-subdisciplinary research within psychology and diversity of specialization was operationalized by the number of individual memberships to different sections of a scientific association, but not by scientometric indicators (see, e.g., Bordons, Morillo, & Gomez, 2004; Porter & Chubin, 1985). Thus, the construct validity of this indicator must be tested in the context of scientometric indicators.

References

Anderson, M. S., & Shultz, J. B. (2003). The role of scientific associations in promoting research integrity and deterring research misconduct. *Science and Engineering Ethics,* 9, 269-272. doi:10.1007/s11948-003-0013-1

Bickel, J. (2007). The role of professional societies in career development in academic medicine. *Academic Psychiatry, 31*, 91-94. Retrieved from http://ap.psychiatryonline.org/article.aspx?articleid=50523

Bordons, M., Morillo, F., & Gomez, I. (2004). Analysis of cross-disciplinary research through bibliometric tools. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 437-456). Dordrecht, The Netherlands: Kluwer.

German Psychological Society (2008, March 27). Aims and responsibilities: What is the German Psychological Society (DGPs)? Retrieved March 1, 2012 from http://www.dgps.de/en/dgps/aufgaben/

Gigerenzer, G., Rösler, F., Spada, H., Amelang, M., Bierhoff, H. W., Ferstl, R., … Vaitl, D. (1999). Internationalisierung der psychologischen Forschung in Deutschland, Österreich und der Schweiz: Sieben Empfehlungen [Internationalizing psychological research in Germany, Austria, and Switzerland: Seven recommendations]. *Psychologische Rundschau, 50*, 101-105. doi:10.1026//0033-3042.50.2.101

Glänzel, W., & Schoepflin, U. (1995). A bibliometric study on ageing and reception processes of scientific literature. *Journal of Information Science, 21*, 37-53. doi:10.1177/016555159502100104

Holland, B. S., & Copenhaver, M. D. (1987). An improved sequentially rejective bonferroni test procedure. *Biometrics, 43*, 417-423. Retrieved from http://www.jstor.org/stable/2531823

Houghton, J. W. (2005) Changing research practices and research infrastructure development. *Higher Education Management and Policy, 17*, 63-82. Retrieved from http://www.oecd.org/dataoecd/53/62/42347801.pdf

Houghton, J. W., Steele, C., & Henty, M. (2004). Research practices and scholarly communication in the digital environment. *Learned Publishing, 17*, 231-249. doi:10.1087/0953151 04323159667

Krampen, G., Fell, C., & Schui, G. (2011). Psychologists' research activities and professional information-seeking behavior: Empirical analyses with reference to the Theory of the Intellectual and Social Organization of the Sciences. *Journal of Information Science, 37*, 439-450. doi:10.1177/0165551511412148

Krampen, G., Schui, G., & Fell, C. (2010). ZPID-Monitor 2008 zur Internationalität der Psychologie aus dem deutschsprachigen Bereich: Der Kurzbericht [2008 ZPID Monitor on the internationality of psychology in the German-speaking countries: A brief report]. *Psychologische Rundschau, 61*, 210-212. doi:10.1026/0033-3042/a000050

Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew Effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology, 61*, 424-427. doi:10.1002/asi.21232

Lienert, G. A., & Krauth, J. (1975). Configural frequency analysis as a statistical tool for defining types. *Educational and Psychological Measurement, 35*, 231-238. doi:10.1177/ 001316447503500201

Porter, A. L., & Chubin, D. E. (1985). An indicator of cross-disciplinary research. *Scientometrics, 8*, 161-176. doi:10.1007/BF02016934

Rowlands, I., & Fieldhouse, M. (2007). *Information behaviour of the researcher of the future: A British Library/JISC study* (Work package I: Trends in scholarly information behaviour). Retrieved from JISC website: http://www.jisc.ac.uk/media/documents/programmes/ reppres/ggworkpackagei.pdf

Schimank, U. (1988). Scientific associations in the German research system - Results of an empirical study. *Knowledge, Technology & Policy, 1*, 69-85. doi:10.1007/BF02687214

Schneller, K., & Schneider, W. (2004). *Bundesweite Befragung der Absolventinnen und Absolventen des Jahres 2003 im Studiengang Psychologie* [Nationwide survey of the 2003 vintage of the German graduate students of psychology]. Retrieved from German Psychological Society website: http://www.dgps.de/studium/danach/_download/bericht. pdf

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal,* 314, 498-502. doi:10.1136/bmj.314.7079.497

Thagard, P. (2005). How to be a successful scientist. In M. E. Gorman, R. D. Tweney, D. C. Gooding, & A. P. Kincannon (Eds.), *Scientific and technological thinking* (pp. 159-171). Mawah, NJ: Erlbaum.

Vanclay, J. K. (2009). Bias in the journal impact factor. *Scientometrics, 78*, 3-12. doi:10.1007/s11192-008-1778-4

Von Eye, A., & Gutiérrez Peña, (2004). Configural frequency analysis: The search for extreme cells. *Journal of Applied Statistics, 31*, 981-997. doi:10.1080/0266476042000 270545

Von Eye, A., Mair, P., & Mun, E.-Y. (2010). *Advances in configural frequency analysis*. New York, NY: Guilford Press.

Psychological Test and Assessment Modeling, Volume 54, 2012 (2), 179-194

Endocrine response patterns after uncontrollable experimental stress: an application of CFA

Matthias J. Müller[1] *& Petra Netter*[2]

Abstract

According to biological stress theories cortisol increases (C+) and testosterone decreases (T-) characterize uncontrollable (UC) stress and the opposite pattern is observed in controllable (CON) stress. The influence of CON/UC on hormone responses to a mental (d2) and a physical (E) short-term stressor was tested by two-sample configural frequency analysis in a cross-over design on 74 healthy males assigned to either CON or UC conditions. Areas under the response curves of saliva C and T were computed and dichotomized (+/-). The evaluation of bivariate response patterns (C/T) revealed that the combination C+T- was significantly more prevalent after UC than after CON with both stressors. The pattern C-T+ constituted a significant discrimination type between CON and UC across both stressors.

Key words: configural frequency analysis; discrimination type; hormone response pattern; cortisol; testosterone

[1] *Correspondence concerning this article should be addressed to:* Matthias J. Müller, MD, Vitos Clinical Centre Giessen-Marburg, Licher Str. 106, 35394 Giessen, Germany; email: mjmueller@gmx.de

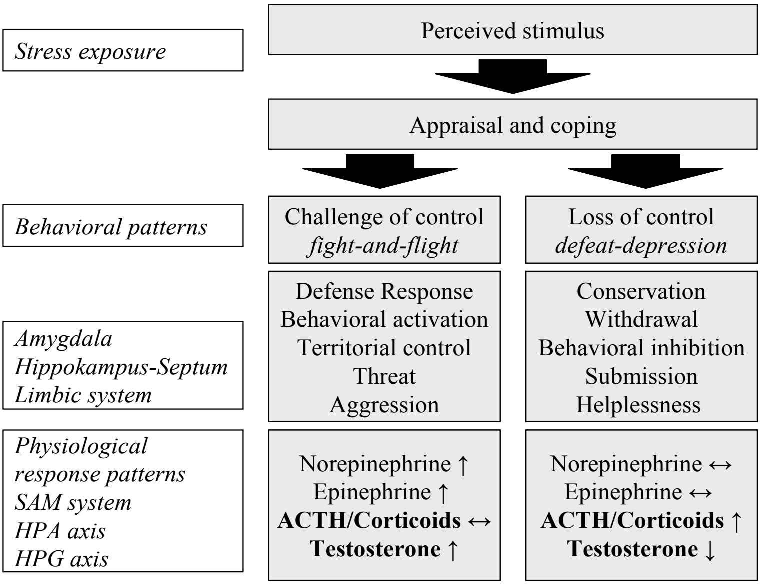
[2] Department of Psychology, University of Giessen, Giessen, Germany

Background

This paper is designed to demonstrate that Configural Frequency Analysis (CFA) (Krauth & Lienert, 1973; Lautsch & von Weber, 1995; Lienert, 1969; von Eye, 1990, 2001, 2002) is a suitable device for the statistical identification of physiological response patterns elicited by different stress conditions, thus contributing to assessment as well as to stress research.

Uncontrollability and endocrine stress response

After the development of Learned Helplessness Theory by Seligman (1975) a number of experiments tried to identify aspects of stressors which contribute to behavioral and physiological stress reactions. Two of the salient components were unpredictability and uncontrollability (Dess, Limwick, Patterson, Overmier, & Levine, 1983). In particular the role of uncontrollability in psychological and endocrine stress responses attracted researchers from different fields of psychophysiology and neurochemistry (Christianson,



SAM, sympatho-adreno-medullary system; HPA, hypothalamo-pituitary–adrenal axis; HPG, hypothalomo-pituitary-gonadal axis; ACTH, adrenocorticotropic hormone; ↑, increase; ↓, decrease; ↔, no change

Figure 1:
Specific response patterns for controllable and uncontrollable stressors (Henry, 1982, 1986, 1997; Henry & Stephens, 1977)

Ragole, Amat, Greenwood, Strong, Paul, Fleshner, Watkins, Maier, 2010; Dickerson & Kemeny, 2004; Diener, Struve, Balz, Kuehner, & Flor, 2009; Foley & Kirschbaum, 2010; Rozeske, Evans, Frank, Watkins, Lowry, & Maier, 2011).

Many of them tried to test the psychoneuroendocrinological model of controllable and uncontrollable stress developed by J.P. Henry (Henry, 1982, 1986, 1997; Henry & Stephens, 1977) depicted in Figure 1 which gave rise to the concept that sustained perceived uncontrollability and helplessness are risk factors for affective disorders and the accompanying physiological changes may cause cardiovascular and metabolic disorders (Björntop, 1997; Bose, Oliván, & Laferrère, 2009; Chrousos 2000; Ruige, 2011).

The psychoendocrinological hypothesis derived from this model states that uncontrollable stress leads to activation of the hypothalamo-pituitary-adrenal (HPA) axis with the consequence of elevated cortisol levels (C↑) and inhibition of the gonadal axis leading to a decrease of testosterone (T↓) in men, a pattern which is the opposite in controllable stress conditions (C↓, T↑). So the endocrinological response pattern of this model hypothetically links controllable stress (CON) to a response of cortisol decrease (C↓) and of testosterone increase (T↑), and uncontrollable stress (UC) to a response of cortisol increase (C↑) and testosterone decrease (T↓) in men.

Objective

So far, no data are available as to the validity of this hypothesis for different types of short term laboratory stressors, e.g., physical and mental stressors, in humans. Accordingly, following questions were investigated: (1) Do short term experimental laboratory stressors elicit cortisol increases and testosterone decreases in uncontrollable and cortisol decreases and testosterone increases in controllable conditions? (2) Do mental and physiological stressors elicit the same response patterns in healthy men?

Transformed into the hypothesis of bivariate response patterns, question (1) would read: CON leads to C↑T↓, and UC is followed by C↓T↑. This hypothesis requires adequate statistical analysis techniques. Configuration Frequency Analysis (CFA) developed by Lienert (Krauth & Lienert, 1973; Lienert, 1969; von Eye, 1990, 2002) seems highly appropriate for this purpose.

Method

Participants

Seventy-four healthy male students (age 18 - 40 years, M = 25, SD = 3) were included. Exclusion criteria were checked by a trained physician and comprised severe physical diseases, psychiatric disorders, continuous medication in particular with substances which could have an impact on cortisol and testosterone levels.

Stressors

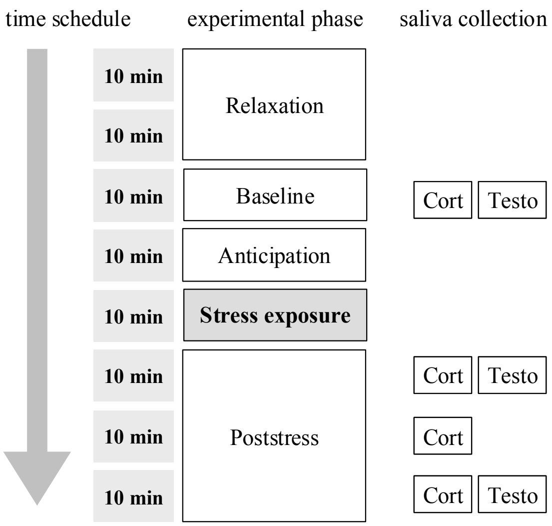
A modified version of the letter cancellation attention test d2 (Brickenkamp, 1994) was used as mental stressor and applied for 10 minutes according to the original instructions.

Electrical skin stimuli (E) were applied as physical stressor. The stimuli were tested for just painfully experienced individual thresholds in a preceding trial and all participants were exposed to 30 stimuli with a mean inter-stimulus-interval of about 20-sec (10 min duration of stress exposure; for details of the procedure c.f. Müller, 2011).

Controllability was achieved by contingent vs. non-contingent feedback in the d2 test and by self vs. experimenter administered electrical stimuli in the physical stress condition (E).

Experimental Design

A two-group balanced cross-over design was used, i.e., two experiments, applied one week apart, comprising one day with the physical stressor (E) and one day with the mental stressor (d2) presented in randomized order and lasting 80 minutes each. Groups were defined as "uncontrollable" (UC) with both stressors presented in the uncontrollable condition (n = 37) and "controllable" (CON) with both stressors presented in the controllable condition (n = 37).



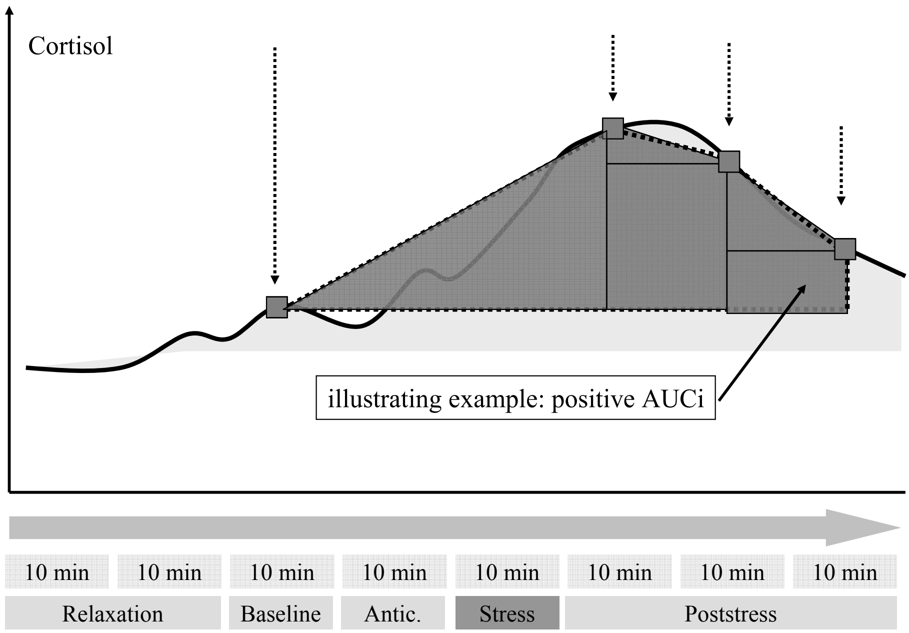
Cort, Cortisol; Testo, Testosterone; Cort and Testo indicate the time of saliva sampling for obtaining levels of the two hormones in question

Figure 2:
Procedure of testing on experimental days

Testing was started at 2 p.m. in order to hit the phase of a fairly balanced steady state of both hormones since they show circadian rhythmicity with a decelerating decline in the afternoon. The procedure on each day of testing is depicted in Figure 2.

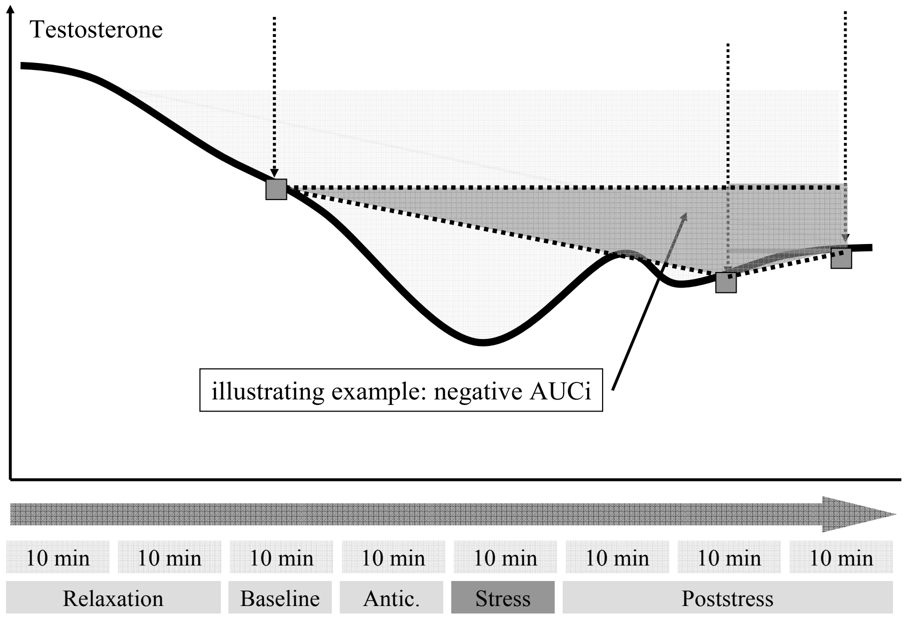
Measures

Endocrine parameters: Free saliva *cortisol (C)* and free saliva *testosterone (T)* concentrations were taken at baseline and after stress exposure (c.f. Fig. 2). Hormone concentration analysis was achieved by ELISA technique resulting in coefficients of variation of $CV < 11\%$ each. Areas under the hormone response curve *(AUCs)* of endocrine parameters were calculated with reference to baseline measures according to the literature (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003). AUCs were dichotomized (+/-) with "AUC-" = decrease or no change and "AUC+"= increase compared to baseline. Figures 3 and 4 demonstrate how dichotomized scores for positive and negative AUC results were obtained.



AUCi, area under the response curve with respect to increase compared to baseline; illustrating example of a positive cortisol response ("+")

Figure 3:
Definition of a positive AUC in the case of a fictive cortisol response curve



AUCi, area under the response curve with respect to increase compared to baseline; illustrating example of a negative testosterone response (“-“)

Figure 4:
Definition of a negative AUC in a case of a fictive testosterone response curve

Statistical analysis

Descriptive statistics (means, standard deviations, proportions) were calculated for hormone data at baseline and for possible confounders (age, body mass index [BMI], and smoking status [yes/no]). Differences between groups were analyzed by means of Mann-Whitney *U*-tests for ordinal data and by χ^2 test in case of categorical data. Areas under the response curves (AUCs) were descriptively shown as cumulative frequency distributions, Mann-Whitney *U*-tests were used to explore group differences in AUC values.

Differences between corresponding dichotomized AUC values (“+” and “-“) of the experimental groups (CON and UC) were tested by conventional χ^2 tests. Conventional χ^2 tests yielded single χ^2 values (df = 1) for the probability of frequencies of “+” and “–“ values in groups CON and UC separately for cortisol (C) and testosterone (T) in each of the two stressors. Additionally, χ^2 tests were computed for the global comparison of the distribution of bivariate response patterns (C+/- and T+/-) in both groups (CON and UC) (df = 3) in each of the two stressors. Results of conventional χ^2 tests were reported as frequencies of endocrine responses, χ^2, df, and corresponding two-tailed p-values.

Two-sample configuration frequency analysis (CFA)

As outlined above, the questions raised required a procedure allowing to test for the validation of a combined hypothesis concerning the pattern of two variables (C/T) in two independent samples (CON/UC). Two-sample CFA is suitable to compare two samples with respect to a series of configurations of states of random variables (Krauth & Lienert, 1973; Stemmler & Bingham, 2003; von Eye, 1990, 2002).

Basically, two-sample CFA is used to explore whether the frequency distributions of the studied configurations are homogenous when compared across two samples. Therefore, the null hypothesis of no differences is locally tested for each configuration under a base model with the following features (von Eye, 2002, p. 174):

a) the model is saturated in the variables used to compared two groups, and
b) the model assumes independence between the grouping variable(s) and the comparison ("discriminant") variables.

In two-sample CFA, significant "types" can emerge only if a relationship between the "discriminant" variables and the "grouping" variable exists. A "type" suggests that in one of the two samples a particular configuration was observed more often than expected based on the above mentioned model. Such configurations are called "discrimination types". If in one group a particular configuration was observed more often than in the other group, this "discrimination type" always constitutes both a "type" (observed more often than expected in one group) and an "antitype" (observed less often than expected in the other group).

To compare the $2^2 = 4$ bivariate patterns of C+/- and T+/- (i.e., C+T+, C+T-, C-T+, and C-T-) between CON and UC and to detect potential discrimination types, the data were subjected to two-sample CFAs, separately for the mental and physical stress condition. Finally, a two-sample CFA comparing the frequency of $2^4 = 16$ patterns (4 endocrine response patterns, 2 stressor types) between CON and UC was computed to explore a hypothetical endocrine discrimination type (CON vs. UC) for both stressor types. CFA results are shown as frequencies of endocrine response patterns, χ^2 statistics ($df = 1$), and two-tailed p-values. For an estimation of the strength of association and effect sizes, the correlation coefficient φ, the squared coefficient $\varphi^2 * 100$ for the percentage of variability accountable by the relationship, the odds ratio θ, and the binomial effect size (*BES*) are reported. Odds-ratio (Bland & Altman, 2000) and correlation coefficients are standard indicators of the degree of non-independence in two-sample CFA (von Eye, 2002). The binomial effect size (*BES*; Rosenthal & Rubin, 1982) is a measure of effect size in 2×2 tables and is defined as $BES = N_{11} / (N_{11} + N_{12}) - N_{21} / (N_{21} + N_{22})$ for $N_{i.}$ and $N_{.j} > 0$, with a range of $-1 \leq BES \leq +1$. In two-sample CFA, the *BES* is an estimation of the discrimination effect and indicates "the proportionate surplus of cases in one group over the other" (von Eye, 2002, p. 187).

The level of statistical significance was set at $\alpha = .05$; in case of multiple tests (two-sample CFA) alpha levels were Bonferroni-adjusted. Conventional χ^2 tests and Mann-

Whitney *U*-tests were calculated with SPSS 15.0 (2006), two-sample CFAs were computed by the CFA program by von Eye (2001).

Results

To prove that randomization between the “controllable” and “uncontrollable” groups was successful, Table 1 gives means and standard deviations of baseline values of the two hormones in each stress condition as well as means and standard deviations of possible confounders (age, body mass index and smoking) in the two groups. Significance tests yielded no significant group difference in any of the variables indicating that randomization was successful.

Distribution of endocrine AUCs

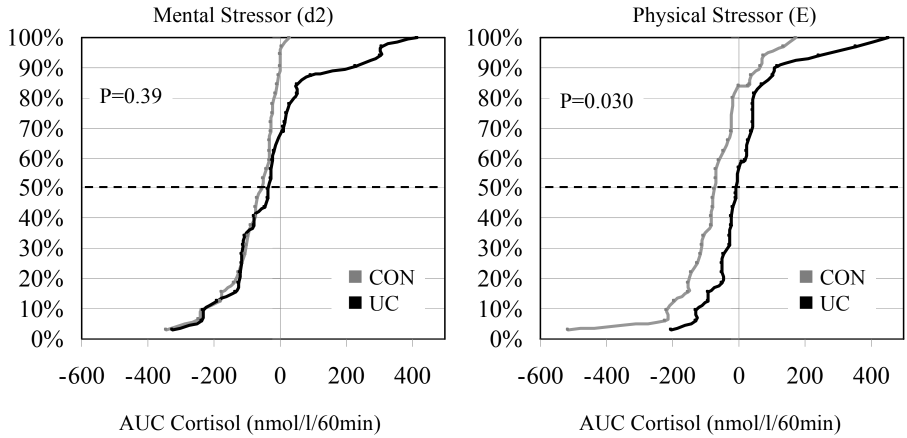
The cumulative AUC distribution curves for cortisol values obtained with the mental and physical stressor in the “controllable” and “uncontrollable” group are depicted in Figure 5; the respective testosterone AUC distribution curves are shown in Figure 6.

According to Figures 5 and 6 the group with uncontrollable conditions exhibited larger positive cortisol AUC responses and lower negative AUC responses than the group with controllable conditions. The non-parametric group differences (CON vs. UC) in AUC

Table 1:
Means, standard deviations, and statistical tests for group differences between the “controllable” and “uncontrollable” stress group for possible confounders and baseline hormone levels

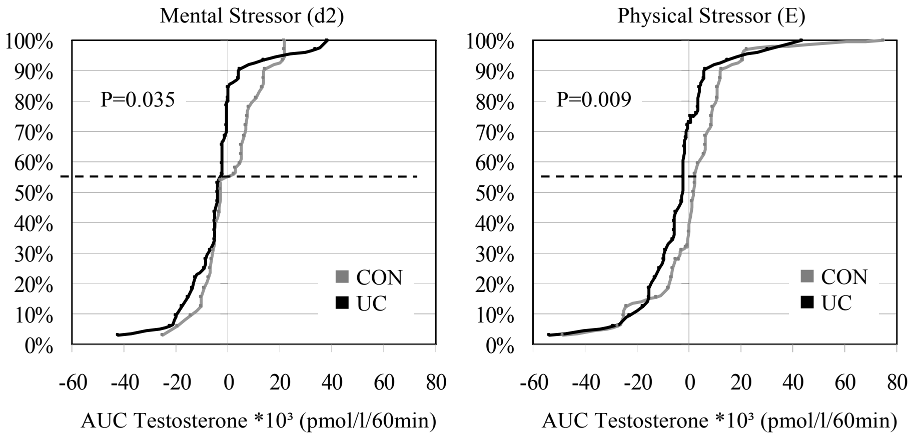
	Total	CON	UC	group difference[a,b]		*p*-value
N	74	37	37	-		-
Age (years)	24.8 ± 2.9	23.6 ± 2.2	25.8 ± 3.1	z = 1.23		.20[a]
BMI (kg/m²)	22.2 ± 1.8	22.4 ± 1.7	22.1 ± 1.9	z = .59		.55[a]
Smoking (% yes)	47 %	53 %	43 %	χ^2 = 1.004	df = 1	.32[b]
Baseline salivary hormone levels						
Cortisol (d2)[c]	5.4 ± 3.3	5.4 ± 3.1	5.3 ± 3.5	z = .26		.80[a]
Cortisol (E)[c]	4.7 ± 2.9	5.1 ± 3.2	4.3 ± 2.6	z = 1.20		.23[a]
Testosterone (d2)[d]	961 ± 542	913 ± 392	1009 ± 661	z = .18		.86[a]
Testosterone (E)[d]	1009 ± 732	1030 ± 718	989 ± 757	z = .53		.60[a]

BMI, body mass index; d2, mental stressor; E, physical stressor; CON, controllable condition, UC, uncontrollable condition; [a], Mann-Whitney *U*-test; [b], χ^2 test; [c], nmol/l; [d], pmol/l



Mann-Whitney *U*-test for differences between the “controllable” (CON) and “uncontrollable” (UC) group, n = 37 each

Figure 5:
Cumulative presentation of areas under the curve (AUC < 0: decrease, AUC > 0: increase) in the mental (left) and physical stress condition (right) for cortisol responses



Mann-Whitney *U*-test for differences between the “controllable” (CON) and “uncontrollable” (UC) group, n = 37 each

Figure 6:
Cumulative presentation of areas under the curve (AUC < 0: decrease, AUC > 0: increase) in the mental (left) and physical stress condition (right) for testosterone responses

values was more pronounced for testosterone but was only significant on the 5 % level for cortisol in the physical condition whereas it did not reach significance for cortisol responses to the mental stressor. It must be emphasized that the considerable number of negative AUC values is due to the circadian decline of both hormones.

Group comparisons by conventional χ^2 tests

The results of comparisons between the experimental groups (CON vs. UC) with respect to endocrine responses by using conventional cross-classification and χ^2 tests are presented in Table 2.

Single χ^2 tests for endocrine responses obtained with both types of stressors yielded significantly higher proportions of increases (+) in cortisol in the UC group (d2: 30 %; E: 43%) than in the CON group (d2: 5 %, E: 16 %) (see also Figure 5). On the other hand, significantly more subjects in the UC group showed a decrease (-) in testosterone AUC (d2: 78 %; E: 73 %) compared to the CON group (d2: 54 %; E: 32 %) (c.f. Figure 6).

Table 2:
Dichotomized cortisol (C) and testosterone (T) AUC responses in the “controllable” and “ uncontrollable” group: results of χ^2 tests

Stressor type	C	T	Controllable *n*	Controllable %	Uncontrollable *n*	Uncontrollable %	χ^2	*df*	*p*-value (2-tailed)
Mental (d2)	-		35	94.6	26	70.3	7.559	1	.012
	+		2	5.4	11	29.7			
		-	20	54.1	29	78.4	4.893	1	.048
		+	17	45.9	8	21.6			
Physical (E)	-		31	83.8	21	56.8	6.469	1	.021
	+		6	16.2	16	43.2			
		-	12	32.4	27	73.0	12.198	1	.001
		+	25	67.6	10	27.0			
Mental (d2)	-	-	19	51.4	19	51.4	10.885	3	.012
	-	+	16	43.2	7	18.9			
	+	-	1	2.7	10	27.0			
	+	+	1	2.7	1	2.7			
Physical (E)	-	-	10	27.0	16	43.2	17.573	3	.001
	-	+	21	56.8	5	13.5			
	+	-	2	5.4	11	29.7			
	+	+	4	10.8	5	13.5			

+/-, positive or negative area under the response curve (AUCs) after stress exposure with respect to baseline; C, cortisol; T, testosterone

The numerical group differences between controllable and uncontrollable conditions was slightly more pronounced for cortisol than for testosterone in the mental stress condition and clearly more pronounced for testosterone than for cortisol in the physical stress condition.

Global χ^2 tests to compare the distribution of combined cortisol/testosterone (C/T) response patterns revealed significant effects in favor of group discriminating patterns with both stressor types. After mental stress, in the UC group n = 10 (27 %) showed the response pattern C+T- whereas only n = 1 subject (3 %) in the CON group showed this pattern. Following physical stress (E) the endocrine response pattern C+T- was detected in n = 11 (30 %) subjects of the UC group and in n = 2 (5 %) of the CON group. Inverse response patterns (C+T-) occurred in n = 16 (43 %) of CON, vs. n = 7 (19 %) of UC after mental stress (d2) whereas after physical stress (E) this response pattern was revealed in n = 21 (57 %) subjects of the CON and in n = 5 (14 %) of the UC group.

Group comparisons by two-sample Configuration Frequency Analysis (CFA)

To test for discrimination types which represent our hypothesis of high cortisol and low testosterone in uncontrollable and the reverse pattern in controllable stress, two-sample CFAs were computed separately for the two stressors. Table 3 shows the results of both analyses.

Table 3:
Patterns of combined dichotomized cortisol (C) and testosterone (T) AUC responses in the "controllable" and " uncontrollable" group: results of two-sample CFAs

Stressor type	C	T	CON		UC		χ^2	Statistics					
			n	%	*n*	%	(df = 1)	p-value[a]	*OR* θ	*BES*	φ	φ^2 %	*DT*
Mental (d2)	-	-	19	51.4	19	51.4	0	1	1	0	0	0	no
	-	+	16	43.2	7	18.9	5.110	.0238	3.27	.24	.26	6.9	no
	+	-	1	2.7	10	27.0	8.649	.00327[a]	13.33	-.24	-.34	11.7	yes
	+	+	1	2.7	1	2.7	0	1	1	0	0	0	no
Physical (E)	-	-	10	27.0	16	43.2	2.135	.144	2.06	-.16	-.17	2.9	no
	-	+	21	56.8	5	13.5	15.179	.000098[a]	8.40	.43	.45	20.5	yes
	+	-	2	5.4	11	29.7	7.559	.00597[a]	7.40	-.24	-.32	10.2	yes
	+	+	4	10.8	5	13.5	.126	.722	1.29	-.03	-.041	0.2	no

C, cortisol; T, testosterone; +/-, increase/decrease or no change of hormone responses (AUC) from baseline; CON, controllable conditions; UC, uncontrollable conditions; p-value, two-tailed p-value of χ^2 statistic; *OR* θ, odds ratio; φ, correlation coefficient; φ^2 %, squared coefficient φ * 100; *BES*, binomial effect size estimation; *DT*, discrimination type according to Krauth & Lienert (1973); [a] Bonferroni-corrected significant p-values if $p < \alpha/4 = .0125$

Two-sample CFAs resulted in clearly significant differences between groups CON and UC in the mental as well as in the physical stress condition. After mental stress, the response pattern C+T- was found to significantly discriminate between CON and UC (discrimination type) with moderate effect size (e.g., φ = -.34). After physical stress, two endocrine response patterns differed significantly between CON and UC. Again, the response pattern C+T- was significantly associated with the UC condition (φ = -.32), but the inverse response pattern (C-T+) showed an even stronger association with the C condition (φ = .45). Thus, both patterns have to be considered as significant discrimination types.

Testing combined response patterns produced by mental and physical stress

Since the two stressors elicited similar patterns for the controllable as well as for the uncontrollable condition, it was expected that a two-sample CFA based on the patterns of cortisol (C) and testosterone (T) responses combined for the two stressors (d2/E) would result in a general discrimination type even more suitable to separate the CON from the UC group.

The results obtained by a two-sample CFA for all possible C/T response patterns combined for both stressors are shown in Table 4. The global test of the model revealed χ^2 = 25.18, df = 15, p = .0476.

The results revealed that the endocrine response pattern C-T+ after both stressors (d2 and E) represents a significant discrimination type between controllable and uncontrollable stress with sufficient effect size (φ = .39). After Bonferroni correction the response pattern C+T-C+T- failed to reach significance. This would indicate that increases in testosterone (T+) and no change or a decrease in cortisol (C-) form a reliable characteristic pattern for responses to the attention test d2 as well as for the physical pain of electric stimulation, provided the participant is convinced he has control over stimulus presentation.

Discussion

Summary

It was expected that according to theoretical frameworks of psychoneuroendocrinological stress research (Henry, 1982, 1986, 1997; Henry & Stephens, 1977) derived from animal models uncontrollability of stress would be accompanied by an increase in cortisol (C+) and a decrease in testosterone (T-) in men as opposed to controllable stress which should elicit the opposite pattern (C-T+). It was hypothesized that (1) particular endocrine response patterns are associated with either uncontrollable (C+T-) or controllable (C-T+) stress conditions, and (2) that these response patterns are occurring intra-individually consistently across different types of stressors.

Table 4:
Combined response patterns of cortisol (C) and testosterone(T) obtained for mental (d2) and physical (E) stress: results of two-sample CFA

d2		E		CON		UC		χ^2	Statistics					
C	T	C	T	*n*	%	*N*	%	($df = 1$)	*p*-value[a]	*OR* θ	*BES*	φ	φ^2 %	*DT*
-	-	-	-	8	21.6	11	29.7	0.474	.491	1.53	-.08	-.093	0.9	no
-	-	-	+	8	21.6	2	5.4	3.600	.058	4.83	.16	.237	5.6	no
-	-	+	-	1	2.7	4	10.8	1.800	.180	4.36	-.08	-.162	2.6	no
-	-	+	+	2	5.4	2	5.4	0	1	1	0	0	0	no
-	+	-	-	2	5.4	4	10.8	0.667	.414	2.12	-.05	-.099	1.0	no
-	+	-	+	12	32.4	1	2.7	9.308	.00228[a]	17.3	.30	.391	15.3	yes
-	+	+	-	1	2.7	2	5.4	0.333	.563	2.06	-.03	-.069	0.5	no
-	+	+	+	1	2.7	0	0	1.000	.317		.03	.117	1.4	no
+	-	-	-	0	0	1	2.7	1.000	.317		-.03	-.117	1.4	no
+	-	-	+	0	0	1	2.7	1.000	.317		-.03	-.117	1.4	no
+	-	+	-	0	0	5	13.5	5.000	.025		-.135	-.269	7.2	no
+	-	+	+	1	2.7	3	8.1	1.000	.317	3.18	-.05	-.120	1.4	no
+	+	-	-	0	0	0	0	0	1		0		0	no
+	+	-	+	1	2.7	1	2.7	0	1	1	0	0	0	no
+	+	+	-	0	0	0	0	0	1		0		0	no
+	+	+	+	0	0	0	0	0	1	1.53	0	-.093	0	no

Stressor type: d2, mental stressor; E, physical stressor; C, cortisol; T, testosterone; +/-, increase/decrease or no change of hormone responses (AUC) from baseline; CON, controllable conditions; UC, uncontrollable conditions; *p*-value, two-tailed *p*-value of χ^2 statistic; *OR* θ, odds ratio; *BES*, binomial effect size estimation; φ, correlation coefficient; φ^2 %, squared coefficient φ * 100; *DT*, discrimination type according to Krauth & Lienert (1973); [a] Bonferroni-corrected significant *p*-values if $p < \alpha/16 = .003125$

The influence of controllability or uncontrollability (CON/UC) on endocrine short-term stress responses was studied by means of a 10 minutes mental test (d2) and a 10 minutes physical (E) stressor applied in a cross-over design to 74 healthy male students randomly assigned either to CON or UC conditions. Saliva cortisol and testosterone concentrations were assessed at baseline and after stress exposure. The areas under the endocrine response curves (AUCs) were calculated and dichotomized (+/-) with respect to baseline. Inspection of hormone levels revealed that all endocrine concentrations of cortisol and testosterone were within normal ranges. Conventional χ^2 tests were used to explore the frequency of univariate and bivariate hormone responses after mental and physical stress in controllable and uncontrollable conditions. Two-sample CFA was applied to test the hypotheses that (1) bivariate response patterns C+/T- and C-/T+ could discriminate between CON and UC conditions; and (2) that these response patterns are observed consistently across both types of stressors.

The evaluation of univariate response patterns of cortisol and testosterone (C+/- or T+/-) indicated that an increase in cortisol (C+) or a decrease in testosterone (T-) responses was observed more often with uncontrollable stressors as compared to controllable stressors. Analyses of bivariate response patterns (C/T) revealed that the combination C+T- was more prevalent in uncontrollable conditions (d2: 27 %, E: 30 %) than in controllable conditions (d2: 3 %; E: 5 %). However, the non-specific decrease in cortisol and testosterone (C-T-) was the most prevalent pattern observed in 51 % (d2) and 35 % (E) of participants of both groups (CON and UC combined) and has to be attributed to the circadian decline of both hormones.

Two-sample Configural Frequency Analysis (CFA)

After mental stress (d2), the response pattern C+T- constituted a significant discrimination type sensu Lienert (Krauth & Lienert, 1973). After physical stress, both theoretically specified endocrine response patterns C+T- and C-T+ could significantly discriminate between controllable and uncontrollable conditions. Moreover, the response pattern C-T+ constituted a significant discrimination type between CON and UC across both stressors (d2: C-T+ and E: C-T+) with a higher frequency in CON (12/37 = 32 %) than in UC (1/37 = 3 %). Thus, the results corroborate the general hypothesis put forward in question (1) and give a tentatively positive answer to question (2), asking if the endocrine response patterns discriminating controllable and uncontrollable conditions are valid for different types of stressors.

Conclusions

For a mental as well as for a physical stressor the patterns of cortisol and testosterone responses expected according to psychoneuroendocrinological theory derived from animal models could be corroborated. However, when the type of stressors was included into CFA analysis results only partially confirmed the hypothesis: The pattern of a cortisol decrease and a testosterone increase was observed more often after controllable stress, but the reverse pattern for uncontrollability was evidently less consistent across types of stressors.

The experiment demonstrates that controllability and the lack of controllability (uncontrollability) have a clear impact on endocrine response patterns under both types of stressors applied as short term stimuli in a laboratory setting to young healthy men. CFA has been demonstrated to be highly suitable for analyzing the distribution of dichotomized response patterns in samples with $n < 100$.

References

Björntorp, P. (1997). Stress and cardiovascular disease. *Acta Physiologica Scandinavica Supplement, 640,* 144-148.

Bland, J. M., & Altman, D. G. (2000). Statistics notes. The odds ratio. *British Medical Journal, 320,* 1468.

Bose, M., Oliván, B., & Laferrère, B. (2009). Stress and obesity: the role of the hypothalamic-pituitary-adrenal axis in metabolic disease. *Current Opinion in Endocrinology, Diabetes, and Obesity, 16,* 340-346.

Brickenkamp, R. (1994). *Test d2: Aufmerksamkeits-Belastungstest*. 8. Auflage. Göttingen: Hogrefe.

Christianson, J. P., Ragole, T., Amat, J., Greenwood, B. N., Strong, P. V., Paul, E. D., Fleshner, M., Watkins, L. R., & Maier, S. F. (2010). 5-hydroxytryptamine 2C receptors in the basolateral amygdala are involved in the expression of anxiety after uncontrollable traumatic stress. *Biological Psychiatry, 67,* 339-345.

Chrousos, G. P. (2000). The HPA axis and the stress response. *Endocrinology Research, 26*, 513-514.

Dess, N. K., Limwick, D., Patterson, J., Overmier, J. B., & Levine, S. (1983). Immediate and proactive effects of controllability and predictability on plasma cortisol responses to shocks in dogs. *Behavioral Neuroscience, 97*, 1005-1016.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*, 355-391.

Diener, C., Struve, M., Balz, N., Kuehner, C., & Flor, H. (2009). Exposure to uncontrollable stress and the postimperative negative variation (PINV): prior control matters. *Biological Psychology, 80*, 189-195.

Foley, P., & Kirschbaum, C. (2010). Human hypothalamus-pituitary-adrenal axis responses to acute psychosocial stress in laboratory settings. *Neuroscience and Biobehavioral Reviews, 35*, 91-96.

Henry, J. P. (1982). The relation of social to biological processes in disease. *Social Science and Medicine, 16,* 369-380.

Henry, J. P. (1986). Neuroendocrine patterns of emotional response. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, Research, and Experiences, Vol. 3* (pp. 37-60). San Diego: Academic Press.

Henry, J. P. (1997). Psychological and physiological responses to stress: the right hemisphere and the hypothalamo-pituitary-adrenal axis, an inquiry into problems of human bonding. *Acta Physiologica Scandinavica Supplement, 640*, 10-25.

Henry, J. P., & Stephens, P. M. (1977). *Stress, health, and the social environment*. Berlin: Springer.

Krauth, J. & Lienert, G. A. (1973). *KFA Die Konfigurationsfrequenzanalyse und ihre Anwendung in Psychologie und Medizin.* Freiburg: Alber.

Lautsch, E., & von Weber, S. (1995). *Methoden und Anwendungen der Konfigurationsfrequenzanalye (KFA)*. Weinheim: Psychologie Verlags Union.

Lienert, G. A. (1969). Die Konfigurationsfrequenzanalyse als Klassifikationsmittel in der klinischen Psychologie. In M. Irle (Hrsg.), *Bericht über den 26. Kongreß der Deutschen Gesellschaft für Psychologie 1968 in Tübingen*. Göttingen: Hogrefe.

Müller, M. J. (2011). Helplessness and perceived pain intensity: relations to cortisol concentrations after electrocutaneous stimulation in healthy young men. *Biopsychosocial Medicine, 5*, 8.

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology, 28*, 916-931

Rosenthal, R., & Rubin, D. B. (1982). A simple general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74,* 166-169.

Rozeske, R. R., Evans, A. K., Frank, M. G., Watkins, L. R., Lowry, C. A., & Maier, S. F. (2011). Uncontrollable but not controllable stress desensitizes 5-HT1A receptors in the dorsal raphe nucleus. *Journal of Neuroscience, 31*, 1407-1415.

Ruige, J. B. (2011). Does low testosterone affect adaptive properties of adipose tissue in obese men? *Archives of Physiology and Biochemistry, 117*, 18-22.

Seligman, M. E. P. (1975). *Helplessness. On depression, development and death*. San Francisco: Freeman and Company.

SPSS Inc. (2006). *SPSS 15.0 for Windows*. Chicago, Illinois: SPSS Inc.

Stemmler, M., & Bingham, C. R. (2003). Log-linear modeling and two-sample CFA in the search of discrimination types. *Psychology Science, 45,* 421-429.

Von Eye, A. (1990). *Introduction to configural frequency analysis. The search for types and antitypes in cross-classification.* Cambridge: Cambridge University Press.

Von Eye, A. (2001). Configural Frequency Analysis - Version 2000. A Program for 32 Bit Windows Operating Systems. *Methods of Psychological Research Online, 6*, 129-139.

Von Eye, A. (2002). *Configural frequency analysis: methods, models, and applications*. Mahwah, Illinois: Lawrence Erlbaum Associates, Inc. Publishers.

Psychological Test and Assessment Modeling, Volume 54, 2012 (2), 195-207

The stability of externalizing behavior in boys from preschool age to adolescence: A person-oriented analysis[1]

Mark Stemmler[2] *& Friedrich Lösel*[3]

Abstract

The continuity of externalizing behaviors such as aggression, delinquency and hyperactivity has been noted by many researchers. There is also increasing knowledge on different developmental subtypes of problem behavior. In previous person-oriented analyses we found two types of externalizing problems in boys (Stemmler et al., 2005, 2008; Stemmler & Lösel, 2010). One pattern contained externalizing problems only, whereas the other type showed both externalizing and internalizing problems. The present study addressed these two groups in an extended prospective longitudinal design. It was investigated whether the groups remained stable over time and whether the two types of antisociality were related to offending in adolescence. The sample consisted of 295 boys from the Erlangen-Nuremberg Development and Prevention Study (Lösel et al., 2009). Social behavior was rated by mothers, kindergarten educators, and school teachers; offending was self-reported by the adolescents. The time lag between the first and last data assessment was more than eight years.

Approximately nine percent of the boys revealed stable externalizing behavior problems over the entire assessment period. Criminal behavior correlated positively with externalizing problems and negatively with internalizing problems. In a person-oriented prediction-Configural Frequency Analysis (P-CFA; von Eye, 2002) the 'externalizing only' pattern could be replicated and suggested high stability over time. Moreover, this pattern was clearly related to self-reported delinquent behavior. In contrast to our previous studies with shorter follow up periods, the 'combined externalizing and internalizing' pattern did not appear as a type. It was also not significantly related to juvenile offending. Potential explanations for these findings are discussed.

Key words: Prediction-Configural Frequency Analysis (P-CFA), proactive and reactive aggression, delinquency, longitudinal research

---

[1] This research was supported by a grant from the German Federal Ministry of Family Affairs, Seniors, Women and Youth

[2] *Correspondence concerning this article should be addressed to:* Mark Stemmler, PhD, University of Erlangen-Nuremberg, Institute for Psychology, Nägelsbachstr. 49c, 91052 Erlangen, Germany; email: mark.stemmler@psy.phil.uni-erlangen.de

[3] University of Cambridge and University of Erlangen-Nuremberg

Introduction

The study of the origins and continuity of aggression, delinquency and other forms of externalizing behavior is one of the most important topics of developmental criminology (Boers, Lösel & Remschmidt, 2009). Externalizing behavior is not only a frequent and serious problem in cross-sectional studies on prevalence, but a substantial number of children with early externalizing symptoms are at risk to enter a pathway of relatively persistent antisocial behavior (e.g., Loeber & Farrington, 1998; Lösel & Bender, 2003). Therefore the persistence and aggravation of externalizing problems over time is not only an important issue of basic research, but also highly relevant for approaches to prevention (Farrington & Welsh, 2007; Lösel & Bender, 2012).

Studies on the development of antisociality suggest that one must differentiate between various forms of problem behavior. Loeber and Hay (1997), for example, distinguished between 1) overt antisociality (hitting, bullying, fighting, and cruelty to animals; leading to later assault or rape), 2) covert antisociality (shoplifting, lying, vandalism, and firesetting; leading to burglary, fraud, or serious theft), and 3) authority conflict (stubborn behavior, defiance, or disobedience; leading to truancy, running away, or staying out late). Similar to the first two of these types, other authors differentiated between physical aggression versus relational (social, indirect) aggression or proactive versus reactive aggression (Dodge, Lochmann, Harnish, Bates & Pettit, 1997; Fontaine, 2007; Vitaro, Barker, Boivin, Brendgen & Tremblay, 2006). Although these forms of aggression show substantial overlap, they seem to have partially different origins and developmental trajectories (Vitaro & Brendgen, 2012).

The differentiation of subgroups and developmental pathways of antisocial behavior requires person-oriented research strategies (Magnusson & Allen, 1993). In previous studies we applied Configural Frequency Analysis (CFA; Lienert & Krauth, 1975; von Eye, 2002) to investigate different patterns of externalizing problems in the Erlangen-Nuremberg Development and Prevention Study (Stemmler, Lösel, Beelmann, Jaursch & Zenkert, 2005; Stemmler, Lösel, Beelmann & Jaursch, 2008; Stemmler & Lösel, 2010). We found two types of externalizing problems that supported the above-mentioned differentiations. One pattern contained externalizing problems only, whereas the other type showed both externalizing and internalizing problems. The first group could be seen as primarily proactive, instrumental and unemotional in their exhibition of aggression, whereas the second group seemed to be more reactive and emotionally driven.

Our previous analyses addressed the time period from preschool to elementary school and did not allow conclusions on the longer-term stability of the two developmental patterns. Therefore, the present study has been carried out. We are now using CFA to investigate the stability of externalizing behavior patterns in boys from kindergarten to secondary school. CFA is a statistical tool for the analysis of multi-way contingency tables of categorical variables (von Eye, Mair & Mun, 2010). Each CFA is based on an underlying null model, usually the model assuming independence among the variables involved. In order to find so-called *types*, that is higher frequencies than expected under the null model, or *antitypes*, that is lesser frequencies than expected under the null model, the null hypothesis needs to be rejected (Stemmler & Bingham, 2003). Whereas ordinary correlations reveal only bivariate or multivariate relationships, pattern-oriented techniques such as CFA and log-linear models are based on a multinomial statistical model, allowing the detection of interactions of a higher order.

In this paper prediction-CFA (P-CFA) will be applied (von Eye, 2002). In P-CFA the variables A and B are the predictors for the criterion variable C. The respective log-linear base model is a model which is saturated within each set of predictors and criteria. It looks like the following:

$$\ln e_{ijk} = \lambda 0 + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} \quad (1)$$

ln e_{ijk} is the natural logarithm of the expected frequencies, λ_0 is the intercept, λ_i is the parameter of variable A and λ_j is the parameter of variable B. The lamba parameters can be interpreted similarly to beta weights in a regression equation. For each P-CFA or log-linear model a goodness-of-fit between the null model and the observed model is calculated using the likelihood ratio (LR) and/or the Pearson chi-square statistic (von Eye, 2002). If the null model of the P-CFA in equation (1) does not apply or results in a significant LR or chi-square statistic, interactions between any predictors and the criterion variable are prevalent. The advantage of a P-CFA over the ordinary first order CFA is, that the resulting types or antitypes contain information of a directed relationship (Stemmler et al., 2008). Common types or configurations refer to a class of persons sharing the same characteristics (Stemmler, 2001).

Method

Sample

The data were taken from the Erlangen-Nuremberg Prevention and Development Study (Lösel, Beelmann, Stemmler & Jaursch, 2006; Lösel, Stemmler, Jaursch & Beelmann, 2009). The original sample of the core study consisted of 675 kindergarten children (336 boys, 339 girls) from 609 families. The project is a longitudinal study that started at preschool age and is now containing seven waves of data collection. The sample was nearly representative to young families living in Erlangen and Nuremberg (Franconia). According to an index of the socioeconomic status (SES; Geißler, 1994) which included income, education, profession, and housing conditions, 13.3% of the families were lower class, 32.3% were lower middle class, 30.6 % middle class, 15.4% upper middle class, and 3.0 % upper class. Approximately 86% of the parents were married at *Time 1*. The retention rates varied over time; in the most recent wave (nearly 10 years after the first one) circa 90% of the original sample participated. The analyses in this paper focused on all boys with available data at three measurement points: At *Time 1* the average age of the boys was M = 5.81 years (SD = 5.55 months). At *Time 2* the mean age was M = 9.22 years (SD = 9.43 months). Finally, at *Time 3* the boys were nearly 14 years old (M = 13.92; SD = 9.90 months). Therefore, the time lag between the first and last time-point was more than eight years. A total of N = 295 boys fulfilled the inclusion criteria. In some statistical analyses the sample size varied a little due to missing data for some variables.

Measures

Child's social behavior. The children's social behavior in kindergarten and at school was assessed by our German adaptations of the Social Behavior Questionnaire (SBQ; Tremblay et al. 1987; Tremblay et al., 1992). The SBQ is available in several versions. Here, kindergarten educators', school teachers', and mothers' ratings were used (Lösel, Beelmann & Stemmler, 2002). The mother's ratings were used when the children were in secondary school. The content and format of the teacher's SBQ versions are identical and consist of 46 items. The mother's version has two additional items. The teacher's version item 'stealing things' is divided for the mothers' version into 'stealing things at home' and 'stealing things outside home'. Each item is rated on a 3-point scale ranging from *'0' = never/not true* to *'2' = almost always/true most of the time*. Two scales of the SBQ were used. The *Externalizing Problems* scale is a second order scale consisting of four primary scales: *Physical Aggression*, *Hyperactivity/Attention Problems*, *Destroying Things/Delinquency* and *Indirect Aggression*. The reliabilities for the different informants were α =.89 (preschool teachers/kindergarten nurses), α =.91 (school teachers), and α =.74 (mothers). The second scale on Emotional Problems/Anxiousness addressed internalizing behavior problems scale; α =.75 (kindergarten); α =.78 (school), and α =.63 (mothers).

Offending in adolescence. The adolescent's delinquent behavior was assessed by a German delinquency self-report scale (DBS; Lösel, 1975). The DBS-scale, which was filled in by the youngsters themselves, consists of a total score and various subscales. The *Total Scale* is a summary of all 28 items. Its reliability in various studies varied between α =.77 and .89. In addition, we used the subscales *Property Crimes* (α =.60 - .78) and *Violent Offenses* (α =.56 -.74) for our analyses. Each item is answered according to whether the delinquent act under question was ever been committed, and if yes, how often in the last year.

Results

Table 1 shows the correlations between the kindergarten educators' SBQ ratings (*Time 1*), the elementary school teacher's rating (*Time 2*), the mothers' SBQ ratings of the boys and the boys' self reports in the DBS at *Time 3* (secondary school). The longitudinal correlations for externalizing problems were significant, even from *Time 1* to *Time 3* (r = .24), suggesting some stability in the rank order of problem behavior over more than eight years. The longitudinal correlations for internalizing problems were lower, but still significant: i.e., r = .13 between *Time 1* to *Time 2*: r = .20 from *Time 2* to *Time 3*, and r = .18 between *Time 1* and *Time 3*. The cross-sectional correlations revealed that externalizing problems correlated significantly with internalizing problems (ranging from $r = .12$ at *Time 1* through $r = .28$ at *Time 3*). This suggests some co-occurrence between these two types of problem behaviors. Self-reported offending correlated significantly with externalizing problems across all three time-points (ranging from $r = .20$ at *Time 2* and *Time 3* through $r = .28$ at *Time 1*), but negatively with internalizing behavior (ranging from $r = -.08$ at *Time 3* through $r = -.19$ at *Time 2*). With one exception (i.e., *Time 1* to *Time 3*: $r = .14$), there were no significant longitudinal correlations between externalizing and internalizing behavior.

In order to investigate the long-term stability of problem behavior patterns, three individual characteristics were selected for Prediction-Configural Frequency Analysis (P-CFA): Externalizing behavior in kindergarten, and externalizing and internalizing behavior in secondary school. All variables were dichotomized close to the 75^{th} percentile. Table 2 shows the observed and expected frequencies for the P-CFA as well as the *z*-values based on the standardized residuals, which are basically the normal approximation of the χ^2 –component.

From kindergarten to secondary school nine percent ($n = 26$) of the sample stayed in the upper quarter of the distribution for externalizing problems. The base model for the P-CFA provided a non-satisfactory fit ($LR = 13.21$, $df = 3$, $p = .00$) suggesting an interaction between the *Time 1* predictor (i.e., externalizing) and any of the *Time 3* criterion variables (i.e., externalizing and internalizing). One externalizing pattern turned out to be significant over the eight-year period: The *type* '+ – +' representing high externalizing behavior at *Time 1* and *Time 3*, but no internalizing problems. The 'combined externalizing and internalizing' pattern did not appear as a type.

Table 1:
Longitudinal and cross-sectional correlations for boys' externalizing and internalizing behavior from kindergarten (T_1) via elementary school (T_2) to secondary school (T_3) and the correlation with delinquent behavior (T_3)

		Kindergarten (T1)		Elementary School (T2)		Secondary School (T3)	
		Ext	Int	Ext	Int	Ext	Int
T_1	Ext						
	Int	$.12^{*}$					
T_2	Ext	$.34^{***}$	-.02				
	Int	-.02	$.13^{*}$	$.21^{**}$			
T_3	Ext	$.24^{***}$	$.14^{*}$	$.13^{**}$	.06		
	Int	-.05	$.18^{**}$	-.04	$.20^{*}$	$.28^{***}$	
	DB	$.28^{***}$	$-.10^{+}$	$.20^{**}$	$-.19^{**}$	$.20^{***}$	-.08

Note: $^{+}p < .10$, $^{*}p < .05$, $^{**}p < .01$. The sample sizes for the bivariate correlations varied from $n = 259$ through $n = 298$. T_1 = kindergarten teachers' ratings, T_2 = elementary school teachers' ratings, and T_3 = mothers' ratings; DB = delinquent behavior; DB is children's self-report.

Table 2:
Prediction CFA for externalizing problems in kindergarten ($ExtT_1$) and externalizing and internalizing in secondary ($Ext/IntT_3$)

Cell Index			P-CFA		
$ExtT_1$	$IntT_3$	$ExtT_3$	$f(o)_{ijk}$	$f(e)_{ijk}$	z_{ijk}
-	-	-	148	140.82	0.61
-	-	+	23	32.91	-1.73
-	+	-	31	29.08	0.36
-	+	+	23	22.19	0.17
+	-	-	36	43.18	-1.09
+	-	+	20	10.09	3.11 T
+	+	-	7	8.92	-0.64
+	+	+	6	6.81	-0.64

Note: $N = 294$. T = Type, A = Antitype. z_{ijk} = z-approximation of the chi-square statistic. '-' = below the 75th percentile; '+' = above the 75th percentile. The listed z-value is based on the *standardized residuals* in the SPSS printout.

Table 3 contains the relationship of different behavioral patterns with offending at Time 3. As the sizes of the subsamples with different patterns of problem behavior were rather small, we restricted our analysis to the configurations of *Time 3*. We entered the *Total Scale* of the self-report delinquency scale and the two subscales in oneway-ANOVAs

Table 3:
One-way ANOVAs comparing the four patterns of externalizing and internalizing behavior regarding to adolescent delinquent behavior in secondary school (cross-sectional analyses)

	Behavior Pattern				
	i + e + (n = 30)	i – e + (n = 43)	i- e – (n = 181)	i + e – (n = 35)	
Delinquency Scales	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*F (df)*
Total Score	1.94^{b} (2.11)	3.84^{a} (3.80)	2.18^{b} (2.49)	2.14^{b} (2.71)	4.82^{**} (3,285)
Violent Offenses	0.37^{b} (0.89)	0.93^{a} (1.28)	0.45^{b} (0.87)	$0.51^{a,b}$ (0.95)	3.26^{*} (3,285)
Property Crime	1.08^{b} (1.18)	2.05^{a} (2.00)	1.21^{b} (1.35)	1.11^{b} (1.41)	4.50^{**} (3,285)

Note: $^{**}p < .01$. $N = 208$. Means with different indices were significantly different (post hoc comparisons). For the post hoc comparison the least significant difference (LSD) was applied. Delinquency ratings were based on children's self-report. Behavior ratings were based on mother's report.

with four behavior patterns based on the combination of externalizing and internalizing behavior. The behavior patterns were distributed as follows: ‘i + e +‘ (n = 30), ‘i – e +‘ (n = 43), ‘i – e – ‘(n = 181), ‘i + e –‘ (n = 35). The *Total Scale* and the subscale *Property Crime* revealed significant mean differences among the four behavior patterns involved. The highest values were found in the ‘externalizing only’ pattern ‘i – e +’. This configuration was significantly different from the patterns with no externalizing behavior (i.e., ‘i – e –‘ and ‘i + e –‘) and also showed a clear tendency to enhanced scores in the scale on *Violent Offenses*. The means for the ‘combined externalizing and internalizing’ pattern were not higher than the patterns with no externalizing problems, suggesting that ‘i – e +’ is the only group with a deviant behavior pattern.

Discussion

As we set our cutoff point of externalizing problems around the 75th percentile of the sample’s distribution, 23% and 24% fell in this category at *Time 1* and at *Time 3*. Such point-prevalence rates are above the rates of clinically relevant behavior problems, but only slightly above the broader definition in the nationwide German Child and Youth Health Survey of the Robert Koch Institute (Hölling, Erhart, Ravens-Sieberer & Schlack, 2007). Contrary to the latter survey, the present paper provides data on the stability of externalizing behavior for boys from kindergarten through secondary school. There are about nine percent of the boys who remained in the highest quarter of externalizing behavior for more than eight years. The majority of this subset of boys consists of the ‘externalizing only’ pattern (20 out of 26). This group of approximately 7% of early starting and relatively persistent antisocial youngsters is in accordance with international findings on the prevalence of problem stability (Loeber & Farrington, 1998; Lösel & Bender, 2003).

Stability of externalizing behavior was also found in the longitudinal correlations of our study. The correlations were significant over time, although for each measurement time different raters were used. The latter usually leads to relatively small inter-correlations (Achenbach, 2006; Lösel, Stemmler, Beelmann & Jaursch, 2005). The correlations between different informants in cross-sectional studies are even smaller than the longitudinal correlations of behavior ratings by the same informant (Lösel, 2002). As expected, the broader characteristic of externalizing problems was significantly and substantially related to offending behavior at *Time 3* (correlations between r = .20 and .28). This could be seen as a further indicator of measurement validity as the criminal behavior was assessed via self- report. In contrast to the externalizing problems, internalizing behavior at all three measurement times correlated negatively with juvenile delinquency. This result is line with research suggesting that internalizing problems may protect children from aggression and delinquency (Loeber, Farrington, Stouthamer-Loeber & White, 2008; Lösel & Farrington, 2012; Tremblay, Pihl, Vitaro & Dobkin, 1994). However, some findings also suggest that internalizing problems may increase the risk of continuity/aggravation when youngsters already developed antisocial problems (Donker, 2004; Loeber et al., 2008; Lösel & Farrington, 2012).

The risk aspect of internalizing problems had been suggested in our previous finding of an 'externalizing only' and an 'externalizing plus internalizing' type of problem behavior. However, the long-term data of the present study did not confirm the latter type. Our application of Prediction-Configural Frequency Analysis (P-CFA) was able to detect a three-way interaction between externalizing and internalizing variables which could not be found in a variable-oriented approach (Stemmler et al., 2005). This underlines the importance of a person-oriented research strategy (Magnusson & Allen, 1983). However, we now only found a significant pattern of 'externalizing only' problems and no longer a type of 'combined externalizing and internalizing' problems. The first longitudinal pattern indicates the more proactive, instrumental and overt type of antisociality (Dodge et al., 1997; Fontaine, 2007; Loeber & Hay, 1997; Vitaro & Brendgen, 2012). It goes along with physiological under-arousal in the respective individuals (Raine et al., 1998). In contrast, the reactive type involves more emotions such as angry outbursts, anxiousness, and impulsive reactions to perceived provocation that go along with enhanced physiologically arousal (Scarpa & Kolko, 1994).

We can only speculate why the latter type could not be replicated in our long-term assessment from preschool to adolescence. In our former article (Stemmler & Lösel, 2010) investigating the time from kindergarten to elementary school both types were significant and they encompassed 27 boys (18 boys in pattern 'i– e+'; 9 boys in pattern 'i+ e+'). In the present article both patterns contained (almost) the same number of boys, that is 26 boys (20 boys in pattern 'i– e+'; 6 boys in pattern 'i+ e+'). They are, however, mixed differently, with more boys in the 'externalizing only' group and less boys in the 'combined externalizing and internalizing' group. Some boys from the comorbide type moved to the proactive type because they did not reach the high level of the 75^{th} percent cut-off for internalizing. Due to small numbers, the minor reduction from nine to six boys resulted in a significant loss of power in the 'i– e+' pattern. In addition to this random explanation one must also take into account that internalizing problems are generally less stable over time than antisocial behavior (Robins & Price, 1991). They are also less visible for informants and therefore more difficult to assess (Lösel et al., 2005). This would explain the minor differences between our previous and current findings. One must also take into account changes in social contexts over time. With regard to school bullying, for example, the 'i+ e+' pattern can be interpreted as the group of bully-victims, whereas the 'i– e+' pattern indicates the group of typical proactive bullies (Lösel & Bliesener, 2003; Olweus, 1993). As longitudinal studies have shown, bullying perpetration is relatively stable and predicts antisocial outcomes such as offending and violence (Ttofi, Farrington, Lösel & Loeber, 2011a). Although bullying victimization is a significant predictor of internalizing problems, it seems to be less stable and more weakly related to later behavioral outcomes (Ttofi, Farrington, Lösel & Loeber, 2011b). This has also been observed in the Erlangen-Nuremberg Development and Prevention Study (Lösel & Bender, 2011). The change from primary to secondary schools and also psychological processes of puberty may have contributed to our finding that the combined 'externalizing and internalizing' group became less concise than in our previous shorter follow-up studies.

The disappearance of the second antisocial type is also reflected in the relations between the developmental patterns and offending in adolescence. Only the more 'cold-blooded' behavioral pattern of 'i– e+' showed clearly enhanced criminality. In contrast, the 'i+ e+' pattern revealed similar mean scores of self-reported offending as the large non-deviant 'i– e-' group. Together with the variable-oriented findings in Table 1, this suggests a moderate protective effect of emotional problems such as anxiousness, depressive mood and social withdrawal.

The particular relevance of the 'externalizing only' pattern and its relationship to delinquency is in line with other studies, suggesting the high predictive validity of proactive aggression for later externalizing behavior and delinquency. Vitaro, Gendreau, Tremblay & Olign (1997), for example, investigated proactively and reactively aggressive twelve-years old boys from a low SES. Only the proactive boys revealed high delinquency rates and behavior problems such as conduct problems at the age of 15. Similarly, in a study by Pulkkinen (1996) proactively but not reactively aggressive male adolescents committed crimes during adulthood. Such findings may be due to more peer rejection and less social-cognitive skills in reactively aggressive boys (Dodge, Lochman, Harnish, Bates & Pettit, 1997). Subsequently, reactively aggressive males revealed significantly lesser friends and therefore lesser delinquents peers than their proactive aggressive counterparts with whom criminal acts are usually committed. In our study the 'combined externalizing and internalizing' group also showed the lowest quality of social information processing from all boys in elementary school (Stemmler & Lösel, 2010).

Although the above findings underline the particular criminological relevance of the small group with stable externalizing problems one has to bear in mind the following issues: The average age of the boys at *Time 3* was approximately 14 years and the majority of scores on self-reported offending were rather low. As offending increases during adolescence and reaches its peak at about 17 to 18 years (Farrington, Coid & West, 2009), one may expect larger variance and stronger statistical relationships when the boys become older. Another limitation of our study is related to the cutoff at the 75th percentile of the SBQ distributions. This suggests that the 'externalizing only' subgroup must not have been particularly low in the scale on emotional problems. Therefore, we should only assume some overlap but not identity with the callous-unemotional type of antisociality that has been described by Frick, Cornell, Barry, Bodin and Dane (2003). Further data waves in our study may help to clarify this relation.

References

Achenbach, T. M. (2006). As others see us: clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science, 15,* 94-98.

Boers, K., Lösel, F., & Remschmidt, H. (Eds.) (2009). Developmental and life-course criminology (Special Issue). *Monatsschrift für Kriminologie und Strafrechtsreform [Monthly Journal of the Criminal and Penal Law Reform],* 2-3.

Dodge, K. A., Lochman, J. E., Harnish, J. D., Bates, J. E., & Pettit, G. S. (1997). Reactive and proactive aggression in school children and psychiatrically impaired chronically assaultive youth. *Journal of Abnormal Psychology*, *106*, 37-51.

Donker, A. (2004). *Precursors and prevalence of young-adult and adult delinquency*. Doctoral thesis. The Netherlands Institute for the Study of Crime.

Farrington, D. P., & Welsh, B. C. (2007). *Saving children from a life of crime: Early risk factors and effective interventions*. Oxford: Oxford University Press.

Farrington, D. P., Coid, J. W., & West, D. J. (2009). The development of offending from age 8 to age 50: Recent results from the Cambridge Study in Delinquent Development. *Monatsschrift für Kriminologie und Strafrechtsreform, 59,* 160-173.

Fontaine, R. G. (2007). Disentangling the psychology and law of instrumental and reactive subtypes of aggression. *Psychology, Public Policy and Law, 13*, 143-165.

Frick, P. J., Cornell, A. H., Barry, C. T., Bodin, S. D., & Dane, H. E. (2003). Callous-unemotional traits and conduct problems in the prediction of conduct problems severity, aggression, and self report of delinquency. *Journal of Abnormal Child Psychology, 31*, 457-470.

Geißler, R. (Ed.), (1994). Soziale Schichtung und Lebenschancen in Deutschland *[Social strata and life opportunities in Germany]*. Stuttgart: Enke.

Hölling, H., Erhart. M., Ravens-Sieberer, U., Schlack, R. (2007). Verhaltensauffälligkeiten bei Kindern und Jugendlichen: Erste Ergebnisse aus dem Kinder und Jugendgesundheitssurvey [Behavior problems in children and adolescents: First results from the children and juvenile health survey]. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, *50,* 784-793.

Lienert, G. A., & Krauth, J. (1975). Configural Frequency Analysis as a statistical tool for defining types. *Educational and Psychological Measurement, 35*, 197-209.

Loeber, R., & Farrington, D. P. (Eds.), (1998). *Serious and violent juvenile offenders: Risk factors and successful interventions*. Thousand Oaks, CA: Sage.

Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & White, H. R. (2008). *Violence and serious theft: Development and prediction from childhood to adulthood*. New York: Routledge.

Loeber, R., & Hay, D. (1997). Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology*, *48*, 371-410.

Lösel, F. (1975). *Handlungskontrolle und Jugenddelinquenz [Self-control and juvenile delinquency]*. Stuttgart: Enke.

Lösel, F. (2002). Risk/need assessment and prevention of antisocial development in young people: Basic issues from a perspective of cautionary optimism. In R. Corrado, R. Roesch, S.D. Hart & J. Gierowski (Eds.), *Multiproblem violent youth: A foundation for comparative research needs, interventions and outcomes* (pp. 35-57). Amsterdam: IOS/NATO Book Series.

Lösel, F., Beelmann, A., Stemmler, M., & Jaursch, S. (2006). Prävention von Problemen des Sozialverhaltens im Vorschulalter: Evaluation des Eltern- und Kindertrainings EFFEKT [Prevention of social behavior problems at preschool age: Evaluation of the preant and child training EFFEKT]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 35*, 127-139.

Lösel, F., Beelmann, A., & Stemmler, M. (2002). Skalen zur Messung sozialen Problemverhaltens bei Vorschul- und Grundschulkindern: Die deutschen Versionen des Eyberg Child Behavior Inventory (ECBI) und des Social Behavior Questionnaire (SBQ). *[Scales for measuring problem behavior in preschool- and elementary school children: The German versions of the Eyberg Child Behavior Inventory (ECBI) und des Social Behavior Questionnaire (SBQ)]*. Universität Erlangen-Nürnberg: Institut für Psychologie.

Lösel, F., & Bender, D. (2003). Resilience and protective factors. In D. Farrington & J. Coid (Eds.), *Early prevention of adult antisocial behaviour* (pp. 130-204). New York: Cambridge University Press.

Lösel, F., & Bender, D. (2011). Emotional and antisocial outcomes of bullying and victimization at school: a follow-up from childhood to adolescence. *Journal of Aggression, Conflict and Peace Research, 3*, 89-96.

Lösel, F., & Bender, D. (2012). Child social skills training in the prevention of antisocial development and crime. In D. P. Farrington & B. C. Welsh (Eds.), *The Oxford handbook of crime prevention* (pp. 102-129). Oxford, UK: Oxford University Press.

Lösel, F., & Bliesener, T. (2003). *Aggression und Delinquenz unter Jugendlichen: Untersuchungen von kognitiven und sozialen Bedingungen [Aggression and delinquency among adolescents: Studies on cognitive and socialorigins]*. Neuwied: Luchterhand.

Lösel, F., & Farrington, D. P. (2012). Direct protective and buffering protective factors in the development of youth violence. *American Journal of Preventive Medicine, 43* (2), in press..

Lösel, F., Stemmler, M., Jaursch, S., & Beelmann, A. (2009). Universal prevention of antisocial development: Short- and long-term effects of a child and parent-oriented program. *Monatsschrift für Kriminologie und Strafrechtsreform, 92*, 289-308.

Lösel, F., Stemmler, M., Beelmann, A., & Jaursch, S. (2005). Aggressives Verhalten im Vorschulalter: Eine Untersuchung zum Problem verschiedener Informanten [Aggressive behavior at preschool age: A study on the issue of different informants]. In I. Seiffge-Krenke (Ed.), *Aggressionsentwicklung zwischen Normalität und Pathologie [Development of aggression between normality and pathology]* (pp. 141-167). Göttingen: Vandenhoeck & Ruprecht.

Magnusson, D., & Allen, V. L. (1983). Implications and applications of an interactional perspective for human development. In D. Magnusson & V. L. Allen (Eds.), *Human development: An interactional perspective* (pp. 370-387). New York: Academic Press.

Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Cambridge, MA: Blackwell.

Pulkkinen, L. (1996). Proactive and reactive aggression in early adolescence as precursors to anti- and prosocial behavior in young adults. *Aggressive Behavior, 22*, 241-257.

Raine, A., Reynolds, C., Venables, P. Mednick, S. A., & Farrington, D. P. (1998). Fearlessness, stimulation-seeking, and large body size at age 3 years as early predispositions to childhood aggression at age 11 years. *Archives of General Psychiatry*, *55*, 745-751.

Robins, L. N., & Price, R. K. (1991). Adult disorders predicted by childhood conduct problems: Results from the NIMH Epidemiologic Catchment Area Project. *Psychiatry, 43*, 113-132.

Scarpa, A., & Kolko, D. J. (1994). Salvary testosterone and cortisol in disruptive children: relationship to aggressive, hyperactive, and internalizing behavior. *Journal of the American Academy of Child and Adolescent Psychiatry*, *33*, 1174-1184.

Stemmler, M. (2001). *Das Problem der Identifikation differentieller Entwicklungsverläufe: neuere methodische Ansätze zur Analyse von inter- und intraindividuellen Unterschieden. [The problem of identifying different developmental trajectories: New methodological approaches for the analysis of inter- and intraindividual differences]*. Unveröffentlichte Habilitationsschrift. Universität Erlangen-Nürnberg.

Stemmler, M., & Bingham, C. R. (2003). Log-linear modeling and two-sample CFA in the search of discrimination types. *Psychology Science, 45*, 421-429.

Stemmler, M., Lösel, F., Beelmann, A., Jaursch, S., & Zenkert, B. (2005). Child problem behavior in kindergarten and in primary school: A comparison between prediction configural frequency analysis and multiple regression. *Psychology Science*, *47*, 467-478.

Stemmler, M. Lösel, F., Beelmann, A., & Jaursch, S. (2008). A configural perspective on the stability of externalizing problem behavior in children: Results from the Erlangen-Nuremberg Development and Prevention Study. In M. Stemmler, E. Lautsch & D. Martinke (Eds.), *Configural Frequency Analysis (CFA) and other non-parametrical statistical methods: Gustav A. Lienert memorial Issue* (pp. 70-83). Lengerich: Pabst Science Publishers.

Stemmler, M., & Lösel, F. (2010). Different patterns of boys' externalizing behavior and their relation to risk factors: A longitudinal study of preschool children. *Bulletin de la Société des Sciences Médicales Luxembourg, 10*, 53-62.

Tremblay, R. E., Desmarais-Gervais, L., Ganon, C., & Charlebois, P. (1987). The Preschool Behavior Questionnaire. Stability of its factor structure between cultures, sexes, ages and socioeconomic classes. *International Journal of Behavioral Development, 10*, 467-484.

Tremblay, R. E., Pihl, R. O., Vitaro, F., & Dobkin, P. L. (1994). Predicting early onset of male antisocial behavior from preschool behavior. *Archives of General Psychiatry*, *51*, 732-739.

Tremblay, R. E., Vitaro, F., Gagnon, C., Piché, C., & Royer, N. (1992). A prosocial scale for the Preschool Social Behavior Questionnaire: Concurrent and predictive correlates. *International Journal of Behavioral Development, 15*, 227-245.

Ttofi, M.M., Farrington, D.P., Lösel, F., & Loeber, R. (2011a). The predictive efficiency of school bullying versus later offending: A systematic/meta-analytic review of longitudinal studies. *Criminal Behaviour and Mental Health, 22, 80*-89.

Ttofi, M.M., Farrington, D.P., Lösel, F., & Loeber, R. (2011b). Do the victims of school bullies tend to become depressed later in life? A systematic review and meta-analysis of longitudinal studies. *Journal of Aggression, Conflict and Peace Research, 3,* 63-73.

Vitaro, F., Barker, E. D., Boivin, M., Brendgen, M., & Tremblay, R. E. (2006). Do early difficult temperament and harsh parenting differentially predict reactive and proactive aggression? *Journal of Abnormal Child Psychology, 34*, 685-695.

Vitaro, F., & Brendgen, M. (2012). Subtypes of aggressive behaviors: Etiologies, development, and consequences. In T. Bliesener, A. Beelmann and M. Stemmler (Eds.), *Antisocial behaviour and crime: Contributions of developmental and evaluation research to prevention and intervention* (pp. 17-38). Göttingen & Cambridge, MA: Hogrefe Publishing.

Vitaro, F., Gendreau, P. L., Tremblay, R. E., & Oligny, P. (1997). Reactive and proactive aggression differentially predict later conduct problems. *Journal of Child Psychology and Psychiatry*, *39(3)*, 377-385.

von Eye, A. (2002). *Configural Frequency Analysis: Methods, models and applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

von Eye, A., Mair, P., & Mun, E.-Y. (2010). *Advances in Configural Frequency Analysis*. London: The Guilford Press.