

Psychometric analysis of the Dutch language Facilitative Interpersonal Skills (FIS) video clips

Sabine J. van Thiel,^{1,2} Margot C.W. Joosen,^{1,3} Anne-Linde Joki,⁴ Arno van Dam,^{1,2} Jac J.L. van der Klink,^{1,5} Kim de Jong⁴

¹Tilburg University, Tilburg School of Social and Behavioral Sciences, Tranzo - Academic collaborative center Work & Health, Tilburg, The Netherlands; ²Mental Health Institute GGZ WNB, Bergen op Zoom, The Netherlands; ³Department Human Resource Studies, Tilburg University, Tilburg School of Social and Behavioral Sciences, Tilburg, The Netherlands; ⁴Institute of Psychology, Leiden University, Leiden, The Netherlands; ⁵Optentia, North West University of South Africa, Vanderbijlpark, South Africa

ABSTRACT

With the motivation of investigating the replicability and transferability of the findings employing the Facilitative Interpersonal Skills (FIS) performance task beyond Anglophone countries, a set of Dutch FIS clips have been scripted and recorded. In this study the psychometric properties of the Dutch clips was tested. Furthermore, an additional set of FIS clips portraying a non-challenging client-therapist interaction was tested. 369 psychology students rated the interpersonal impact (IMI-C) and the affect (positive and negative affect schedule) displayed by the hypothetical client. Thirteen out of sixteen FIS clips were located in the same IMI-C quadrant as the US clips, indicating good content validity for all sets of FIS clips. Inter-rater reliability was reasonable for one set of Dutch language FIS clips ($k=0.416$). Visual inspection of quadrants showed the different character of the non-challenging set of FIS clips. The Dutch FIS clips are directly applicable for educational and research purposes.

Correspondence: Sabine van Thiel, Tilburg University, Tilburg School of Social and Behavioral Sciences, Tranzo, Professor Cobbenhagenlaan 125, 5037 DB, Tilburg, The Netherlands.
E-mail: s.j.vanthiel@tilburguniversity.edu

Citation: van Thiel, S.J., Joosen, M.C.W., Joki, A.L., van Dam, A., van der Klink, J.J.L., & de Jong, K. (2021). Psychometric analysis of the Dutch language Facilitative Interpersonal Skills (FIS) video clips. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 24(1), 94-105. doi: 10.4081/ripppo.2021.513

Acknowledgments: the authors would in particular like to thank Prof. Timothy Anderson for answering many questions and thinking along in the process of translation and recording. We like to thank Dr. Miel Vugts who contributed to the analyses of this study. We thank Scott Minnaugh, PhD., Prof. Dr. Antje Guntz, and Dr. Thomas Munder for answering questions and allowing us to use their IMI-C data. We also thank Anton Hafkenscheid for thinking along. In addition, we thank all respondents who participated in the study.

Funding: this work was supported by PSION, *Psychische Interventies en Ondersteuning Nederland*.

Data availability: study data were collected and managed using Qualtrics Survey Software and hosted at Tilburg University. Derived data supporting the findings of this study are available from the corresponding Author on request.

Ethical approval and consent to participate: this study was reviewed and approved by the School of Social and Behavioral Science Ethics Review Board at Tilburg University (reference EC-2018.77). All participants in this study were given the opportunity to (re-)consider participation and submit questions prior to digitally accepting the informed consent form.

Received for publication: 17 December 2020.

Revision received: 11 March 2021.

Accepted for publication: 13 March 2021.

This work is licensed under a Creative Commons Attribution Non-Commercial 4.0 License (CC BY-NC 4.0).

©Copyright: the Author(s), 2021

Licensee PAGEPress, Italy

Research in Psychotherapy:

Psychopathology, Process and Outcome 2021; 24:94-105

doi:10.4081/ripppo.2021.513

Key words: Interpersonal skills; therapist effects; performance-based assessment; common factors.

Introduction

Research has provided compelling evidence that some therapists achieve consistently better outcomes with their clients than other therapists, regardless of the treatment delivered or the characteristics of the clients (Barkham, Lutz, Lambert, & Saxon, 2017; Kim, Wampold, & Bolt, 2006; Okiishi *et al.*, 2006; Wampold, 2015). Baldwin and Imel (2013) show that a typical client who is treated by one of the best 10% of therapists has twice the probability of recovery and half the probability of deterioration compared to a client that is treated by a therapist classified in the bottom 10%. They estimate that 5% to 7% of the outcome variance in therapies might be due to the personality of the therapist. Therefore, there is evidence for the view that some therapist consistently facilitate better client outcomes than others (Johns, Barkham, Kellett, & Saxon, 2019).

Although client outcomes clearly vary among therapists, it is less clear what specific characteristics and actions of therapists account for their differential outcomes. For example, therapists' age, gender, ethnicity, religion, marital status, clinical experience, theoretical orientation and professional degree are not consistently linked to client outcome (Huppert *et al.*, 2001; Wampold & Brown,

2005; Owen, Wampold, Kopta, Rousmaniere, & Miller, 2016; Bjaastad *et al.*, 2018). While there is a lack of consistent findings supporting therapist characteristics, research on therapy process and relational variables is building ground (Hatcher, 2015; Heinonen & Nissen-Lie, 2019). Therapists may differ in the extent to which they possess interpersonal skills that facilitate an environment in which client developments can take place (Wampold & Imel, 2015).

The Facilitative Interpersonal Skills (FIS) performance task was developed to measure interpersonal skills of therapists and subsequently link these skills to client outcomes (Anderson, Ogles, Patterson, Lambert, & Vermeersch, 2009; Anderson, Crowley, Himawan, Holmberg, & Uhlin, 2016). The FIS consists of video recordings of challenging therapy situations recorded with professional actors in a clients' role. These recorded 'video vignettes' (see also Hillen, van Vliet, de Haes, & Smets, 2013) represent different clients with various psychological problems. Participants/therapists watch the video vignettes. Following each case story, participants are asked to respond to the client in the video as if they were the client's therapist. The therapist's responses are recorded by video and thereafter scored on a set of interpersonal skills using 5 point scales, *i.e.* verbal fluency, hope and positive expectations, persuasiveness, emotional expression, empathy, alliance bond capacity, and alliance rupture-repair responsiveness. The sum score on all clips represents the degree of a participant's interpersonal skills, higher scores indicating better interpersonal skills. The FIS measure has predictability of treatment outcomes (Anderson *et al.*, 2009). Specifically, clients of therapists with higher FIS scores experience significant more symptom reduction than clients of lower FIS therapists (Anderson, Crowley *et al.*, 2016; Anderson, McClintock, Himawan, Song, & Patterson, 2016).

FIS appears to be a good instrument to experimentally test the skills and actions of therapists (Wampold, Baldwin, Holtforth, & Imel, 2017). The importance of interpersonal skills of therapists stems from research on alliance and countertransference. Countertransference refers to the therapist's feelings, cognitions, and behaviors that arise in response to dynamics occurring in the counseling relationship (Gelso & Hayes, 2007). Meta-analyses by Hayes, Gelso, Goldberg, and Kivlighan (2018) revealed that successful countertransference management is related to better therapy outcomes. Eubanks *et al.* (2018) have shown that difficulties in the interpersonal process are linked to negative therapy outcomes, such as drop out.

Several FIS practice situations versions are established (see also www.fisresearch.com). In 2016, Jeremy Safran and the New School for Social Research introduced a new version of the video clips (Safran *et al.*, 2016). These clips included superior video quality to the originals and should serve training purposes. FIS was also translated into five

different languages. Munder *et al.* (2019) tested the psychometric properties of the German language version of the FIS task and found high inter-rater agreement and internal consistency, suggesting the FIS is a unidimensional scale.

For this study, Dutch language FIS clips were established and recorded. Three sets of clips were developed; the first set was translated from the original US clips (Anderson *et al.*, 2009), the second set was translated from the new US clips (Safran *et al.*, 2016), and the third set concerned a newly developed set of non-challenging benign clips (Steggles & De Jong, 2018). This additional set of clips forms a controlled stimulus allowing for a test of the nature of the interaction (challenging *versus* non-challenging), in future research. In the process of translating and creating FIS clips for other contexts, it is important that the psychometric qualities of the translated FIS clips are tested. The objective of the current study was therefore to examine the psychometric characteristics of the Dutch language FIS clips with regard to validity and reliability.

The following research questions will be addressed: i) To what extent do the Dutch language FIS clips represent the same interpersonal affect as the American FIS recordings (content validity)?; ii) To what extent do different observers report the same experienced interpersonal affect in response to the Dutch language FIS clips (reliability/inter-rater agreement)?; and iii) Are the non-challenging benign Dutch FIS clips significantly different in affective response compared to the Dutch language challenging FIS clips? It is hypothesized that the interpersonal affect in the Dutch clips resembles the affect of the corresponding American clips. For the non-challenging clips we hypothesize that the interpersonal affect will be more positive, friendly and less distinct compared to the original FIS clips.

Methods

Design

In this cross-sectional study, all participants are exposed to a randomly assigned (between-subjects) mixed set of Dutch-language FIS clips (within-subjects). This study was reviewed and approved by School of Social and Behavioral Science Ethics Review Board at Tilburg University (reference EC-2018.77).

Participants

Undergraduate psychology students were recruited to participate in the study through the online research subject recruitment platform PURS of Tilburg University. In this study, students were deliberately chosen as test subjects because of their affinity, impartiality and unbiasedness in confrontation with the clients. Their assessment is primarily based on the pattern of observable behavior of the clients in the clips and is not influenced by previous experiences

with clients in therapy. Therefore, impartiality was not further investigated or verified. Because it concerns Dutch spoken clips, only native-level Dutch-speaking students of 17 years and older were eligible to participate in the study. As there is no strict theoretical or empirical basis for the calculation of the sample size for the current study (the ratio of subjects to variables) (Rouquette & Falissard, 2011), a target of 380 participants (five participants per item) was set to ensure adequate sample size.

A total of 412 students registered to participate in the study. Participants were given the opportunity to (re-)consider participation and submit questions before digitally accepting the informed consent form. A total of 401 students accepted the informed consent and got access to the survey. Of those 401 who consented and started the survey, we removed 32 because they had completed less than 20% of the questionnaires. Figure 1 shows participant flow throughout the study. The final analyses included 369 participants. The sample's age range was 17-37 [mean (M)=19.65, standard deviation (SD)=2.41] years, with a median age of 19 years. Fifty-four participants identified

as male (14.6%) and 313 (84.8%) as female. Participants were required to report their Student Number for the researchers to verify participation to allocate participation credits. This information was removed, analyses therefore took place anonymously.

Testing procedure

After signing-up for the research study, participants were sent a hyperlink link to one of the three survey sets (see Table 1), where they gained access to the FIS clips and the questionnaires. Prior to commencing the study, participants received information regarding the study's purpose, aim, and procedure. All data was processed in a coded and anonymous fashion. Data were stored in line with confidential regulations and are only accessible to the investigators at Leiden and Tilburg University, who were directly engaged in the project.

For each FIS clip, participants received background information about the client in the FIS clip and then watched it. After each clip, the participant was asked to

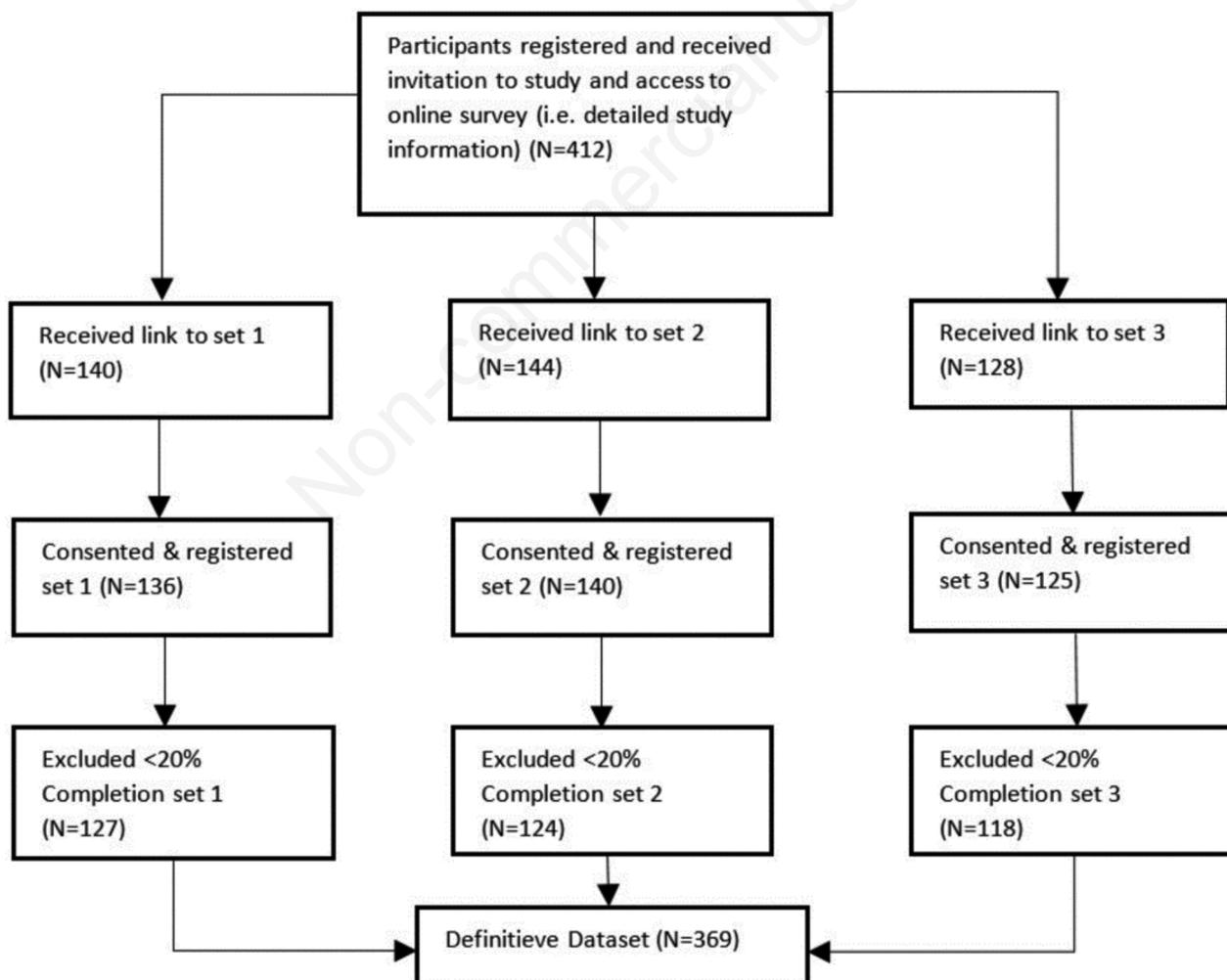


Figure 1. Flowchart of sample inclusion.

rate the client's interpersonal messages on the impact message inventory-circumplex (IMI-C) and rate the clients' experienced emotions on the positive and negative affect schedule (PANAS). Participants had the opportunity to re-watch the client clip and no time limit was set for the task.

Materials

Dutch-language Facilitative Interpersonal Skills clips

Three sets of Dutch FIS video vignettes were developed using semi-professional Dutch actors. The first set is a Dutch translation (further referred to as D-O) of the original challenging clips from Anderson *et al.* (2009) (further referred to as US-O). To illustrate such a challenging situation, this is a section of the dialogue with Suzie: *'No, I am not upset with you, it's just that I keep asking for guidelines or something and... I just don't feel I ever get anything... I don't think there is anything you can do to help me and I don't know what I can do to help myself'*.

The second set is a Dutch translation (further referred to as D-N) of the second version clips from the New School For Social Research (further referred to as US-N). To illustrate a challenging situation, this is a section of the clip of Sean: *'I need you to tell me how I can represent myself in a way that's gonna be where I'm gonna get a job. Because I am paying you every week and if you want me to keep doing that, then I need to have a job'*.

Steggles & De Jong (2018) developed a third set of experimental non-challenging, benign FIS clips (further referred to as D-B) of client therapist interactions. The transcripts for these vignettes were drawn from the same sessions as the original challenging (US-O) FIS clips (Strupp, 1993), but then representing a more common, much less challenging client-therapist manifestation. This is a section of the script of the non-challenging clip of Jack: *'Because if someone were to offend me or, or make me mad, upset me or whatever, I will do everything*

that I can in order to get them back... So I guess, what I'm trying to say... I think I take those things too seriously'.

In the process of translation and development of the FIS clips, the principles set by Anderson, Patterson, McClintock, McCarrick, Song, and The Psychotherapy and Interpersonal Lab Team (2018) were followed. The transcript was translated into Dutch by two independent translators. A synthesis meeting took place, and the back-translation method was used to assure the FIS clips' linguistic equivalency. All three sets of FIS video vignettes were filmed against the same plain background and all clips are of similar duration.

In total, twenty-three FIS clips were recorded of which seven D-O FIS clips, nine D-N FIS clips, and seven D-B FIS clips. To limit task duration, attention load, and for the purpose of randomization, these clips were divided into three survey sets (see Table 1). Each survey-set contains a mix of the original (D-O), second version (D-N) and benign (D-B) FIS clips. To increase ecological validity, a number of clients were given a modified (Dutch) name, *e.g.* Suzie became Suzan, Les became Lex, *etc.* For ease of reading, we will only refer to the associated American names through this article.

The impact message inventory circumplex - Dutch language version

The impact message inventory-circumplex (IMI-C) was used to measure the interpersonal message experienced by the participant as a response to each client in the FIS clip. The IMI-C is a 56-item questionnaire that measures 'distinctive internal reactions, referred to as impact messages' to interpersonal behaviors (Kiesler & Schmidt, 2006; Dutch language version Hafkenscheid & Rouckhout, 2013). The inventory contains words, phrases and statements that describe how the respondent is emotionally engaged or impacted when interacting with another targeted person. The IMI-C subscales cor-

Table 1. Assignment of Facilitative Interpersonal Skills clips to participants.

Experimental Sets		
Survey set 1	Survey set 2	Survey set 3
Bonnie (D-O)	Les (D-N)	Hillary (D-O)
Sean (D-N)	Lauren (D-O)	Jack (D-N)
Jack (D-B)	Suzie (D-B)	Bonnie (D-B)
John (D-O)	Jessica (D-N)	Suzie (D-N)
Lauren (D-N)	Jack (D-O)	Les (D-O)
Les (D-B)	John (D-B)	Lauren (D-B)
Suzie (D-N)	Bonnie (D-N)	John (D-N)
Hillary (D-B)	Sam (D-N)	

D-O, Dutch language original challenging clips; D-N, Dutch language New School challenging clips; D-B, non-challenging benign clips.

respond to the type of impact they represent on the interpersonal circumplex. The interpersonal circumplex is a graph in the form of a circle. The horizontal axis represents the degree of affiliation while the vertical axis the degree of control. Responses are given on a 4-point Likert scale, from 1 'Not at all' to 4 'Very much so'. The Dutch language version of the IMI-C has shown to adequately locate target persons in the two-dimensional interpersonal space using the main dimension, affiliation and control scores (Hafkenscheid & Rouckhout, 2013). The interpersonal experience of the FIS clips was examined by calculating the mean control (CO) and affiliation (AF) score for all FIS clips.

Timothy Anderson, Ohio University, and Scott Mimaugh, New School for Social Research (unpublished data) provided the affiliation and control scores of the original (US-O) and second version (US-N) American language FIS clips for comparison purposes.

The positive and negative affect schedule - Dutch language version

The positive and negative affect schedule (PANAS) (Engelen, De Peuter, Victoir, Van Diest, & Van den Bergh, 2006) is a 20-item scale with 10 positive and 10 negative affective descriptors. After each client vignette, participants assessed to what extent the client in the clip experiences the affective descriptors described. Responses are given on a 5-point Likert scale, from 1 'very little or not at all' to 5 'very much'. Sum scores of two broad domains of affect, termed positive affect and negative affect, were calculated for all FIS clips. Both positive affect and negative affect represent largely independent constructs ranging from low to high levels of emotional experience (Watson, Clark, & Tellegen, 1988). Low positive affect scores reflect 'sadness and lethargy' whereas high positive affect scores reflect 'high energy, full concentration, and pleasurable engagement' (Watson *et al.*, 1988). Low negative affect scores describe 'a state of calmness and serenity' whereas high negative affect scores suggest 'subjective distress and unpleasurable engagement'. Studies have shown that the PANAS is a reliable and valid measure of experienced emotions and mood (Watson *et al.*, 1988; Engelen *et al.*, 2006).

Based on theoretical considerations we expected no differences between D-O clips and D-N clips but we did expect higher scores on positive affect and lower scores on negative affect for D-B clips compared to D-O clips and D-N clips.

Instructed response items

To assess participant engagement and attention, instructed response items were included in the three survey sets (Gummer, Rossmann, & Silber, 2018). These items were inserted in the questionnaire among the regular questions. An item was for example: 'in this item we check

your attention; please click on moderately so'. Each survey set had one instructed response item.

Statistical analyses

Data were analyzed using IBM SPSS statistics V.24. First, the instructed response items were analyzed. Data were retained up until that point of the missed attention-check and subsequent responses were excluded from analyses. Subsequently, in instances where a participant submitted the same response for all questions for a given vignette, all responses for that client vignette were excluded from analyses.

To test content validity quadrants of the interpersonal circumplex of the US and Dutch language clips were compared via visual inspection. Good content validity of a set of FIS clips is met when clips are distributed from a variety of interpersonal situations and are equally distributed around the interpersonal circumplex space (Anderson *et al.*, 2018).

To test the inter-rater reliability we analyzed the level of agreement of the interpersonal experience for each survey set by calculating Fleiss' Kappa (Fleiss, 1971), an adaptation of Cohen's kappa for more than two raters, using the IMI-C scores. Interpretations are based on the guidelines from Altman (1999), adapted from Landis & Koch (1977). Due to the nature of the dataset, three separate tests were conducted: survey set one, survey set two and survey set three (see Table 1 for division of client vignettes). For methodological reasons, only participants with ratings for each vignette per set were included for this analysis (set1 N=116, set2 N=110, set3 N=109). Next, in order to gain more insight into the agreement and sensitivity across raters, distribution of ratings per quadrant (Friendly-Dominant, Friendly-Submissive, Hostile-Submissive and Hostile-Dominant) were provided. Based on the Cohen's Kappa value (Altman, 1999) 61% agreement would be considered as substantial agreement. Yet, having 40% of the evaluations being wrong could give serious quality problems (McHugh, 2012). In this study therefore percentages above 70 are interpreted as acceptable agreement between assessors.

To investigate differences in affective response between the challenging and non-challenging clips, an independent samples t-test was conducted to compare means of control and affiliation between the D-O and D-B clips. An exploratory multiple linear regression analysis, with clips version (D-O, D-N, and D-B) nested within client, was conducted to examine whether clip version predicts scores on negative affect and positive affect. Multiple pair wise tests were performed on the same dataset and therefore the Bonferroni correction is used to reduce the odds of false-positive results (Type I errors), resulting in a corrected P-values of 0.004 (0.05/14 tests). For this analyses the datasets were restructured and merged into one data-set. In addition, scores of D-B clips will be compared with the D-O clips on the interpersonal circumplex through visual inspection.

Results

Content validity: impact message inventory-circumplex scores and comparisons with American language Facilitative Interpersonal Skills clips

The US-O, US-N, D-O, D-N, and D-B clips scattered along the affiliation and control dimensions of the interpersonal circumplex. The US-O *versus* the US-N clips differ on several clients: Hillary does not exist in the US-N

clips. The US-N clips, however, have three additional clients: Sean, Jessica and Sam.

The clients in the D-O set represent a variety of interpersonal stances and are distributed across the interpersonal space, except for Friendly-Dominant. All clients are located in the same quadrant as the US-O FIS clips, although the intensity of the affect within the quadrant varies (see Figure 2A and B). Furthermore, three out of seven US-O clips (Hillary, Lauren and John) were developed after assessment of the first four (Jack, Suzie, Les

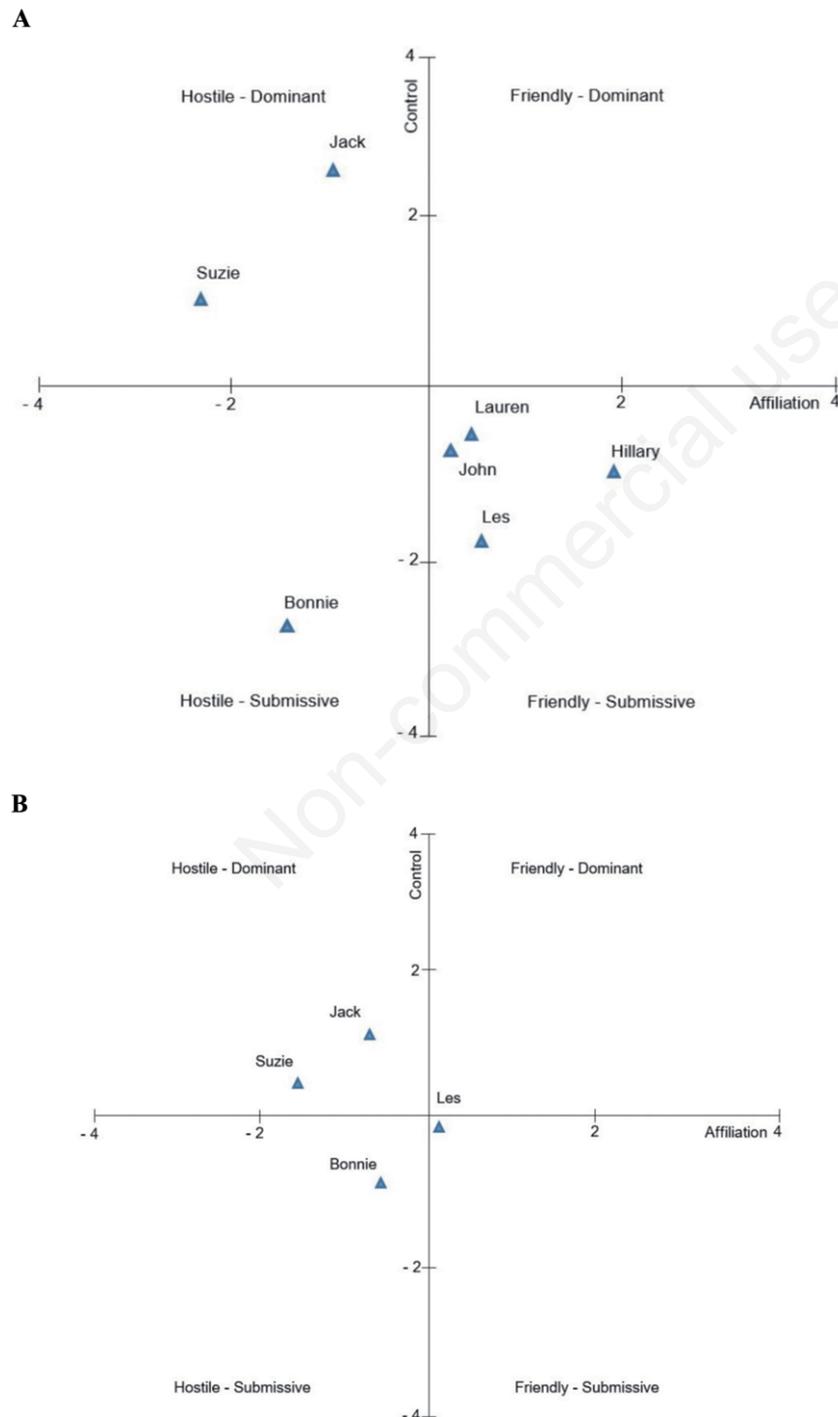


Figure 2. A) Impact message inventory-circumplex (IMI-C) ratings of original version Dutch language (D-O) Facilitative Interpersonal Skills (FIS) clips plotted within interpersonal circumplex space; B) IMI-C ratings of original Anderson (US-O) FIS clips plotted within interpersonal circumplex space. IMI-C scores of Lauren, John and Hillary are not available.

and Bonnie) to have client profiles that were more friendly, but would also provide difficult situations for the average therapist. Our analysis confirms that the three more recently developed client vignettes are indeed rated in the friendly quadrant.

Comparing the IMI-C scores of the D-N and US-N clips, show that six clients are located in the same quad-

rant and three clients are located in a different quadrant (see Figure 3A and B). Lauren is originally located in the friendly-submissive quadrant, Jessica is originally located in the hostile-dominant quadrant, and Sam is originally located in the friendly-dominant quadrant. An independent t-test was conducted to compare the D-O (affiliation: $M = -0.27$, $Mdn = -0.35$, $SD = 1.90$, $range = -4.85, -4.51$,

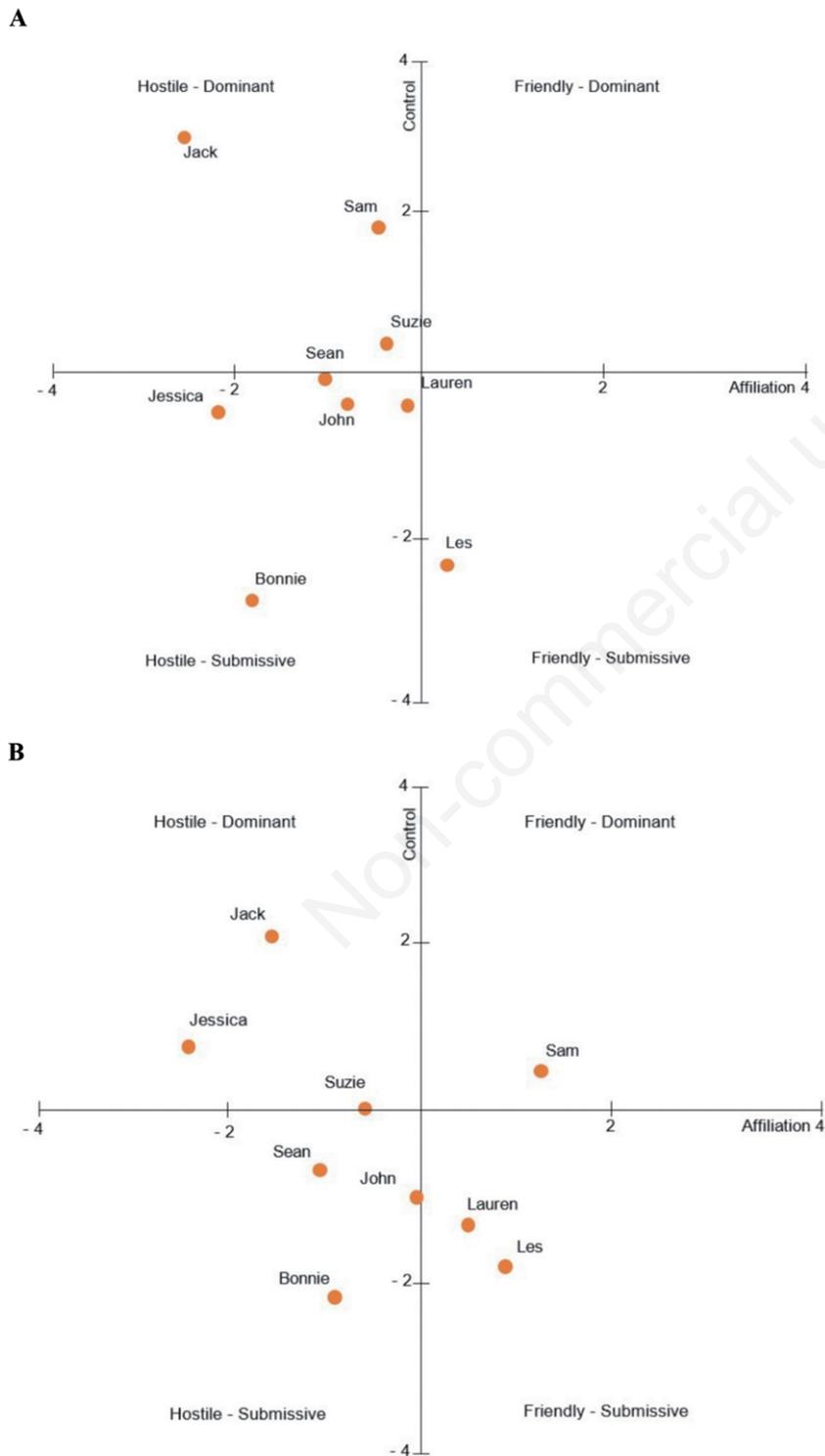


Figure 3. A) Impact message inventory-circumplex (IMI-C) ratings of second version (D-N) Dutch language Facilitative Interpersonal Skills (FIS) clips plotted within interpersonal circumplex space; B) IMI-C ratings of second version New school (US-N) FIS clips plotted within interpersonal circumplex space.

control: $M = -0.50$, $Mdn = -0.69$, $SD = 2.03$, $range = -4.87$, -4.91 and D-N (affiliation: $M = -1.01$, $Mdn = 1.68$, $SD = -0.99$, $range = -5.69$, 3.98 , control: $M = -0.18$, $Mdn = 2.06$, $SD = -0.29$, $range = -5.95$, 5.24) clips on control and affiliation. There is a tendency for the D-N clips to be less friendly; $t(1887) = 8.89$, $P = 0.000$, and less dominant; $t(1887) = 3.368$, $P = 0.001$ associating with the D-O clips.

Reliability of the Dutch language Facilitative Interpersonal Skills clips

Fleiss' Kappa analysis (Fleiss, 1971) was conducted in order to determine whether there was an adequate agreement between the rater's interpersonal judgements across the client vignettes, using the score on the interpersonal complex.

There was moderate agreement between the raters' experiences for the first survey set, $\kappa = 0.416$ [95% confidence interval (CI), 0.410 to 0.422], $P < 0.0005$. This finding indicates that the proportion of agreement between raters is beyond that expected by chance. The degree of agreement

between in survey set 2 and survey set 3 is deemed low: $\kappa = 0.194$ (95% CI, 0.188 to 0.201), $P < 0.0005$ and $\kappa = 0.195$ (95% CI, 0.187 to 0.203), $P < 0.0005$, respectively.

Using the mean ratings on the IMI-C, each client was assigned a global classification in line with one of the four interpersonal circle quadrants: Friendly-Dominant, Friendly-Submissive, Hostile-Submissive, Hostile-Dominant. Looking at the share of scores across all four quadrants per client vignette (see Table 2), we gained a more nuanced insight into the extent of agreement across the raters (see also Conger, 1980).

The scores of the D-O clips show that the clips of Bonnie, Suzie, Jack and Hillary evoke one dominant interpersonal experience across raters. In the case of Les a clear split in interpersonal experience was observed between Friendly-Submissive (64%) and Hostile-Submissive (32%). In the case of John and Lauren a much larger range of interpersonal experiences was observed across the raters. The scores of the D-N clips suggest that the clips of Bonnie, Jack and Jessica evoke one dominant interpersonal experience across raters. In the case of Les there is

Table 2. Distribution of ratings (in percentages) per interpersonal circle quadrants.

Client clip	Version	Friendly-Dominant	Friendly-Submissive	Hostile-Submissive	Hostile-Dominant
Bonnie	D-O	0	11.8	<i>87.4</i>	0.8
	D-N	1.7	8.7	<i>88.7</i>	0.9
	D-B	6.9	52.6	37.1	3.5
John	D-O	5.7	49.6	21.1	23.6
	D-N	0.9	25.9	33.9	39.3
	D-B	20.7	40.5	27.0	11.7
Suzie	D-O	4.2	2.5	13.3	<i>80.0</i>
	D-N	14.2	27.4	15.9	42.5
	D-B	35.3	47.4	7.8	9.5
Lauren	D-O	15.4	48.7	23.9	12.0
	D-N	9.7	37.1	25.8	27.4
	D-B	18.4	9.2	25.7	46.8
Jack	D-O	17.4	0	0	<i>82.6</i>
	D-N	0.8	0	0	<i>97.5</i>
	D-B	76.8	12.8	0.8	9.6
Hillary	D-O	14.4	<i>80.5</i>	2.5	2.5
	D-B	29.6	56.5	8.7	5.2
Les	D-O	0.9	64.3	32.1	2.7
	D-N	1.6	67.2	30.3	0.8
	D-B	32.0	45.1	8.2	14.8
Jessica	D-N	2.6	5.3	<i>70.2</i>	21.9
Sam	D-N	43.8	0.9	0.9	54.5
Sean	D-N	4.0	8.8	46.4	40.8

D-O, Dutch translation of the original challenging clips from Anderson *et al.* (2009); D-N, Dutch translation of the second version clips from the New School For Social Research; D-B, non-challenging, benign clips. Values in italics represent percentages above 70, considered acceptable agreement.

a split in the evoked interpersonal experience between Friendly-Submissive (67%) and Friendly-Dominant (30%). In the case of John, Suzie, Lauren, Les, Sam, and Sean there is a more extensive range of interpersonal experience across the raters. The D-B clips scores show that only the clip of Jack evokes a dominant interpersonal experience across raters (76.8%). In the other clips, the reported interpersonal experience between raters was less centered and more scattered.

Challenging versus non-challenging benign Dutch Facilitative Interpersonal Skills clips

The comparison of D-O clips showed that they differ significantly on affiliation and control (affiliation: $M = -0.27$, $Mdn = -0.35$, $SD = 1.90$, $range = -4.85, -4.51$, control: $M = -0.50$, $Mdn = -0.69$, $SD = 2.03$, $range = -4.87, -4.91$) and D-B (affiliation: $M = 0.63$, $Mdn = 0.66$, $SD = 1.29$, $range = -4.06, 4.31$, control: $M = -0.06$, $Mdn = -0.14$, $SD = 1.41$, $range = -4.31, 5.09$), resulting in the D-B clips to be more friendly; $t(1656) = 11.23$, $P = 0.000$, and less dominant; $t(1656) = 5.05$, $P = 0.000$, compared with the D-O clips. Visual inspection of the D-B clips on the circumplex (Figure 4) shows that these clips are more clustered around the friendly-submissive axis.

Assessment of the exposed positive and negative affect

Participants were asked to rate on the PANAS what affect the clients in the FIS clips displayed. Using multi-level regression analyses, we tested whether scores on positive affect and negative affect for D-N and D-B clips deviate from D-O clips. Table 3 shows regression coeffi-

cients, standard errors, t-values, and P-values for the clips demonstrating significant differences.

In the case of Bonnie, Suzie, Jack, and Les there was a lower negative affect score and a higher positive affect score for the D-B clips compared to the D-O clips. Only in the case of Bonnie and Les, these differences can be considered to be significant. For John and Hillary the negative affect score was higher for the D-B clips compared to the D-O clips, and the positive affect score was lower for the D-B clips compared to the D-O. In both cases these differences were significant. For Lauren both the negative affect and the positive affect score were higher for the D-B clip compared to the D-O clip. These differences are not significant.

Discussion and Conclusions

Therapists differ in their effectiveness with some therapists consistently achieving better outcomes with their clients than others. The FIS instrument, measuring interpersonal skills of therapist, has shown to be predictive of treatment outcomes (Anderson *et al.*, 2009; Anderson, Crowley, *et al.*, 2016). This study aimed to test the psychometric properties of the Dutch language FIS clips.

Main findings

To test content validity, we compared the D-O clips with the US-O FIS clips by Anderson *et al.* (2009). As hypothesized, all reported interpersonal experiences of the D-O clips corresponded to those of the US-O clips (see Figure 2A and B). Furthermore, our research confirms that

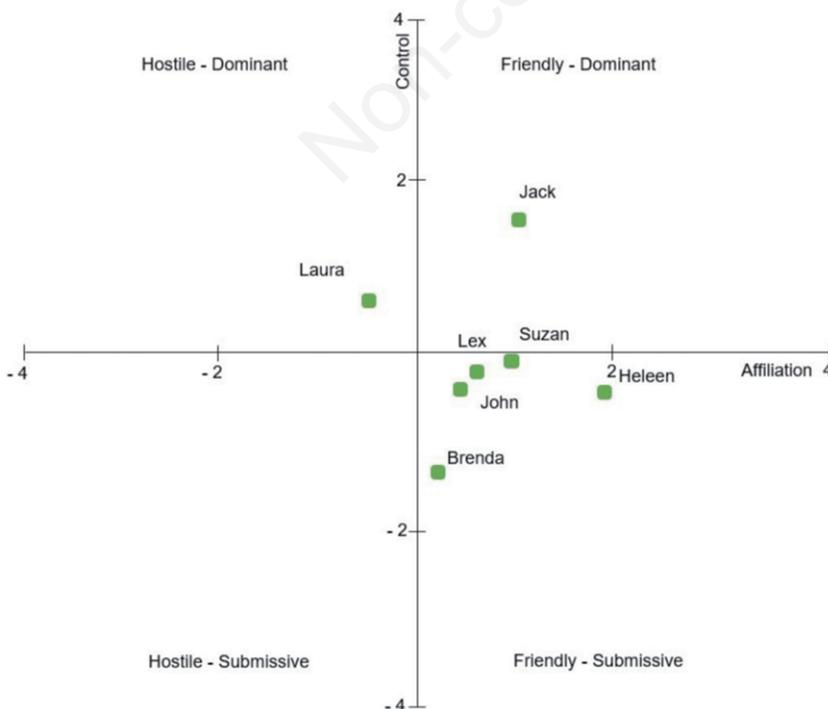


Figure 4. Impact message inventory-circumplex ratings of non-challenging benign Dutch language (D-B) Facilitative Interpersonal Skills clips plotted within interpersonal circumplex space.

the three client vignettes that have most recently been developed to have client profiles that were more friendly, are indeed perceived and experienced by raters as Friendly(-Submissive). Comparing the D-N clips with the US-N clips we see that six out of nine clips are located in the same quadrant of the interpersonal circumplex. Three clips are located in a different quadrant (see Figure 3A and B). There seems to be a meta effect that the D-N clips are more hostile than the US-N clips. In directing and recording these clips, we may have put too much emphasis on the resistance of the clients. Our assessment confirms that the range and distribution of the clients in both the D-O clips and D-N FIS clips represent a sufficiently broad case mix, optimally reflecting clinical complexity and good content validity.

In the majority of the D-O clips (four out of seven), we found a high agreement in the interpersonal experience between the raters; 80%-87% agreement. For the D-N clips, three out of nine clips had an agreement between 70%-97.5%. The interpersonal challenge appears to be less explicit in these clips. For the D-B clips the agreement between raters is the lowest. Only one out of the seven clips had adequate agreement, namely Jack (77%) (see Table 2). Based on these results, the D-O clips can be considered as reliable and can therefore be applied in further research and practice. However, the non-challenging benign clips (D-B) were expected to enhance variability in scoring because the purpose of these clips is to generate little interpersonal challenge and therefore these clients show less pronounced behavior and emotions.

IMI-C scores of the D-O clips and the D-B clips were compared in order to investigate whether this newly developed set forms a set of controlled stimuli, allowing for the effect of the nature of the interaction on interpersonal skills to be isolated and tested in future research. As hy-

pothesized, we found the D-B FIS clips to be reported as less dominant and more friendly. The D-B clips are less distributed and more centered around the friendly-submissive axis in the circumplex (see Figure 4). These results indicate less experienced interpersonal complexity by participants. We also compared the D-O and the D-B clips on positive and negative affect (see Table 3). Four of the seven clips were rated higher on positive affect and lower on negative affect. Two of the seven D-B clips were rated the opposite: lower on positive affect and higher on negative affect. One D-B clip was scored both higher on negative and positive affect. Despite the experimental manipulation (challenging *versus* non-challenging) these varying results can be explained by the content of the clips. In both D-O and D-B clips, positive (*e.g.*, interested, active) and negative emotions (*e.g.*, nervous, guilty) can be displayed by the clients.

Strengths, limitations and suggestions for future research

To our knowledge, this was the first study to measure the psychometric characteristics of FIS clips rather than the psychometric properties of the ratings. Our findings therefore offer a valuable contribution to the knowledge of the FIS measurement. As described in the introduction, the clips were derived from transcripts of real therapy sessions (Strupp, 1993), assuring their relevance to real practice situations (Gould, 1996). This study incorporates a new set of matched benign, non-challenging scripted video vignettes (controls) and to empirically test the discriminability between the non-challenging benign and challenging sets. All clips were translated and directed by researcher practitioners, with first-hand clinical experience. This study can serve as a fruitful basis for future re-

Table 3. Results of multilevel regression analyses for positive and negative affect.

Client clip	Version	Negative affect				Positive affect			
		β	SE	<i>t</i> -value (<i>df</i>)	P	β	SE	<i>t</i> -value (<i>df</i>)	P
Bonnie	D-O	-	-	-	-	-	-	-	-
	D-N	-0.96	0.75	-1.28 (352)	0.20	-0.12	0.62	-0.20 (352)	0.84
	D-B	-3.09	0.61	-4.12 (352)	0.00*	3.19	0.62	5.12 (352)	0.00*
John	D-O	-	-	-	-	-	-	-	-
	D-N	1.50	0.94	1.60 (348)	0.11	-5.84	0.91	-6.51 (348)	0.00*
	D-B	5.63	0.93	6.03 (348)	0.00*	-5.66	0.89	-6.34 (348)	0.00*
Hillary	D-O	-	-	-	-	-	-	-	-
	D-B	3.52	0.86	4.09 (226)	0.00*	-8.32	0.91	-9.17 (226)	0.00*
Les	D-O	-	-	-	-	-	-	-	-
	D-N	2.43	0.88	2.76 (347)	0.01	0.09	0.77	0.12 (347)	0.90
	D-B	-1.97	0.87	-2.26 (347)	0.03	5.42	0.76	7.09 (347)	0.00*

SE, standard error; D-O, Dutch translation of the original challenging clips from Anderson *et al.* (2009); D-N, Dutch translation of the second version clips from the New School For Social Research; D-B, non-challenging, benign clips. *Only significant results are displayed; differences are significant at $P < 0.004$.

search and practice with the Dutch language FIS clips. The results from this study can be used in selecting clips and can accommodate the interpretation of results in follow-up studies.

The following limitations that can have implications for further research and applicability must be acknowledged. First, the sample of raters is largely homogeneous (84.8% female) and consisted of inexperienced undergraduate psychology students. There could consequently be a selection bias. In our statistical analyses, we have not checked for gender differences but other research has underlined differences in interpersonal accuracy between men and women (Hall, Gunnery, & Horgan, 2016). In addition, the personal plea in the challenging clips may be more noticeable by experienced therapists, compared to students, as they carry a joint responsibility for process, relationship and recovery in contact with clients. A second measurement in a smaller and diverse group of experienced therapists could provide an interesting addition to the research results.

Second, variety in the assessment of the interpersonal affect may be determined by other factors than the client in de clip. Research employing the IMI-C has tested the generalizability of impact messages across therapists and has found that while some impact messages are generalizable across therapists (e.g., Dominance & Hostile-Dominant), other categories of impact messages turned out to be poorly generalizable (Hafkenscheid, 2003). Differences in the experienced interpersonal affect among raters could reflect interpretations, filled or motivated by the rater's personal reaction or experience, a concept also referred to as countertransference (Gelso & Hayes, 2007). Also research by Holmqvist and Arnelius (1996) suggest that therapist characteristics play a significant part in the connection between therapists' emotional reactions towards characteristics of the clients.

Third, a range of positive and negative affect ratings was observed for both D-O and D-B clips. This could indicate that the PANAS may not have been the optimal choice to test the research question it was used for. Clients in both the D-O and D-B clips may be perceived to experience a range of positive and negative affect. The D-O clips are characterized by the presence of a challenging interpersonal problem between the therapist and the client, which is not directly being assessed with this instrument.

Finally, in order to limit participation burden, fatigue and drop-out, the decision was made to randomly mix the original, second version and non-challenging benign clips across three experimental sets. Due to this design, we were limited, from a methodological point of view, in some of the empirical evaluations. Future research may consider having the same participant rate all the original and non-challenging benign clips. This would enable to perform a stronger reliability analyses and would give the possibility to compare the interpersonal assessment of the challenging and non-challenging clips within subjects.

Conclusions

In sum, our findings showed that the results on the content validity test for the D-O FIS clips were satisfactory as the interpersonal affect fully matched the US-O clips. We also found adequate agreement across raters in the majority of the D-O clips. This set is therefore best suitable to use in follow-up research. The D-N clips only partly matched the interpersonal affect of the US-N clips. This set also scores less high on the reliability analysis. In a setting where consultation about interpretation is possible, such as training or workshops, these clips can be more applicable. D-N clips with adequate reliability (e.g., Jessica) and can be used as a supplement to the D-O set to obtain larger case mix.

The clustering of the D-B clips around the friendly-submissive axis in combination with significantly higher scores on friendliness and lower scores on dominance are indicative that there is indeed less interpersonal challenge in the benign clips compared to the D-O clips.

This is the first paper to demonstrate how psychometric characteristics of FIS clips were assessed. Research employing the Dutch FIS clips will contribute to extending our understanding of the role of facilitative interpersonal skills in the therapeutic setting. Findings can contribute re-shaping the academic curricula for clinical training. The Dutch FIS clips are directly applicable for educational purposes; both from a point of view of teaching and of assessment of facilitate interpersonal skills. Using the clips in further research, limitations regarding reliability and manipulation must be taken into account.

References

- Altman, D. G. (1999). *Practical statistics for medical research*. New York, NY: Chapman & Hall/CRC Press.
- Anderson, T., Crowley, M. E., Himawan, L., Holmberg, J. K., & Uhlin, B. D. (2016). Therapist facilitative interpersonal skills and training status: a randomized clinical trial on alliance and outcome. *Psychotherapy Research*, 26, 511-529.
- Anderson, T., McClintock, A. S., Himawan, L., Song, X., & Patterson, C. L. (2016). A prospective study of therapist facilitative interpersonal skills as a predictor of treatment outcome. *Journal of Consulting and Clinical Psychology*, 84(1), 57-66.
- Anderson, T., Ogles, B. M., Patterson, C. L., Lambert, M. J., & Vermeersch, D. A. (2009). Therapist effects: Facilitative interpersonal skills as a predictor of therapist success. *Journal of Clinical Psychology*, 65(7), 755-768.
- Anderson, T., Patterson, C., McClintock, A. S., McCarrick, S. M., Song, X., & The Psychotherapy and Interpersonal Lab Team. (2018). *Facilitative Interpersonal Skills Task and Rating Manual*. Unpublished Rating Manual. Athens, OH: Ohio University.
- Baldwin, S. A., & Imel, Z. E. (2013). Therapist effects: Findings and methods. In: M.J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change*, 6 (pp. 258-297). New York, NY: Wiley.
- Barkham, M., Lutz, W., Lambert, M. J., & Saxon, D. (2017).

- Therapist effects, effective therapists, and the law of variability. In L. G. Castonguay & C. E. Hill (Eds.), *How and why are some therapists better than others?: Understanding therapist effects* (pp. 13-36). Washington, DC: American Psychological Association.
- Bjaastad, F.J., Wergeland, H.G. J., Haugland, M. B. S., Gjestad, R., Havik, O. E., Heiervang, E. R., & Öst, L. G. (2018). Do clinical experience, formal cognitive behavioural therapy training, adherence, and competence predict outcome in cognitive behavioural therapy for anxiety disorders in youth? *Clinical psychology & psychotherapy*, 25(6), 865-877.
- Conger, A. J. (1980). Integration and generalization of kappas for multiple raters. *Psychological Bulletin*, 88(2), 322.
- Engelen, U., De Peuter, S., Victoir, A., Van Diest, I., & Van den Bergh, O. (2006). Verdere validering van de Positive and Negative Affect Schedule (PANAS) en vergelijking van twee Nederlandstalige versies. *Gedrag en gezondheid*, 34(2), 61-70.
- Eubanks, C. F., Muran, J. C., & Safran, J. D. (2018). Alliance rupture repair: a meta-analysis. *Psychotherapy*, 55(4), 508.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378.
- Gelso, C. J., & Hayes, J. (2007). *Countertransference and the therapist's inner experience: Perils and possibilities*. New York, NY: Routledge.
- Gould, D. (1996). Using vignettes to collect data for nursing research studies: How valid are the findings? *Journal of Clinical Nursing*, 5, 207-212
- Gummer, T., Roßmann, J., & Silber, H. (2018). Using instructed response items as attention checks in web surveys: properties and implementation. *Sociological Methods & Research*, 0049124118769083.
- Hafkenscheid, A. (2003). Objective countertransference: do patients' interpersonal impacts generalize across therapists?. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, 10(1), 31-40.
- Hafkenscheid, A., & Rouckhout, D. (2013). The impact message inventory-circumplex (IMI-C): A replication study of its circumplex structure in a dutch sample. *Journal of personality assessment*, 95(4), 417-422.
- Hall, J. A., Gunnery, S. D., & Horgan, T. G. (2016). Gender differences in interpersonal accuracy. In J. A. Hall, M. Schmid Mast, & T. V. West (Eds.), *The social psychology of perceiving others accurately* (pp. 309-327). Cambridge, UK: Cambridge University Press.
- Hatcher, R. L. (2015). Interpersonal competencies: responsiveness, technique, and training in psychotherapy. *American Psychologist*, 70(8), 747.
- Hayes, J. A., Gelso, C. J., Goldberg, S., & Kivlighan, D. M. (2018). Countertransference management and effective psychotherapy: meta-analytic findings. *Psychotherapy*, 55, 496-507.
- Heinonen, E., & Nissen-Lie, H. A. (2019). The professional and personal characteristics of effective psychotherapists: a systematic review. *Psychotherapy Research*, 1-16.
- Hillen, M. A., van Vliet, L. M., de Haes, H. C., & Smets, E. M. (2013). Developing and administering scripted video vignettes for experimental research of patient-provider communication. *Patient Education and Counseling*, 91(3), 295-309.
- Holmqvist, R., Armelius, B.A. (1996). Sources of therapists' countertransference feelings. *Psychotherapy Research*, 6(1), 70-78.
- Huppert, J. D., Bufka, L. F., Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2001). Therapists, therapist variables, and cognitive-behavioral therapy outcome in a multicenter trial for panic disorder. *Journal of Consulting and Clinical Psychology*, 69(5), 747.
- Johns, R. G., Barkham, M., Kellest, S., & Saxon, D. (2019). A systematic review of therapist effects: A critical narrative update and refinement to review. *Clinical Psychology Review*, 67, 78-93.
- Kiesler, D. J. & Schmidt, J. A. (2006). *The Impact Message Inventory-Circumplex (IMI-C) Manual: Sampler set, manual, test booklet, scoring key, work sheets*. Redwood City, CA: Mind Garden.
- Kim, D. M., Wampold, B. E., & Bolt, D. M. (2006). Therapist effects in psychotherapy: a random-effects modeling of the National Institute of Mental Health Treatment of Depression Collaborative Research Program data. *Psychotherapy Research*, 16(02), 161-172.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*: *Biochemia medica*, 22(3), 276-282.
- Munder, T., Schlipfenbacher, C., Toussaint, K., Warmuth, M., Anderson, T., & Gumz, A. (2019). Facilitative interpersonal skills performance test: Psychometric analysis of a German language version. *Journal of Clinical Psychology*, 75(12), 2273-2283.
- Okishi, J., Lambert, M.J., Eggett, D., Nielsen, L., Dayton, D.D., & Vermeersch, D.A. (2006). An analysis of therapist treatment effects: Toward providing feedback to individual therapists on their clients' psychotherapy outcome. *Journal of Clinical Psychology*, 62(9), 1157-1172.
- Owen, J., Wampold, B. E., Kopta, M., Rousmaniere, T., & Miller, S. D. (2016). As good as it gets? Therapy outcomes of trainees over time. *Journal of Counseling Psychology*, 63(1), 12.
- Rouquette, A., & Falissard, B. (2011). Sample size requirements for the internal validation of psychiatric scales. *International Journal of Methods in Psychiatric Research*, 20(4), 235-249.
- Steggles, K., & de Jong, K. (2018). *Does therapist emotion regulation moderate their facilitative interpersonal skills?* Paper presented at the 49th International Society for Psychotherapy Research International Annual Meeting, Amsterdam, The Netherlands.
- Strupp, H. H. (1993). The Vanderbilt psychotherapy studies: synopsis. *Journal of Consulting and Clinical Psychology*, 61(3), 431-433.
- Wampold, B. E. (2015). How important are the common factors in psychotherapy? An update. *World Psychiatry*, 14(3), 270-277.
- Wampold, B. E., & Brown, G. S. J. (2005). Estimating variability in outcomes attributable to therapists: a naturalistic study of outcomes in managed care. *Journal of Consulting and Clinical Psychology*, 73(5), 914.
- Wampold, B. E., Baldwin, S. A., Holtforth, M. g., & Imel, Z. E. (2017). What characterizes effective therapists? In L. G. Castonguay & C. E. Hill (Eds.), *How and why are some therapists better than others?: Understanding therapist effects* (pp. 37-53). Washington, DC: American Psychological Association.
- Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate: The evidence for what makes psychotherapy work*. New York, NY: Routledge.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063.