This is the post-print version. This article may not exactly replicate the final version published in the journal. The final peer-reviewed and edited copy of this manuscript can be found at the homepage of the Journal of Career Assessment:

https://journals.sagepub.com/doi/full/10.1177/1069072719865159

Citation:
Opening the Career Counseling Black Box: Behavioral Mechanisms of Empathy and Working Alliance

Klonk, F.E.¹, Will, T.², Ianiro-Dahm, P.M.³, & Kauffeld, S.²

¹Curtin University, Future of Work Institute / Centre for Transformative Work Design
florian.klonek@curtin.edu.au

²Technische Universität Braunschweig

³Hochschule Bonn-Rhein-Sieg, University of Applied Sciences
patrizia.ianirodahm@h-brs.de

Abstract

Many individuals who struggle with career decisions seek professional guidance from career counselors. However, little is known about how to ensure that career guidance sessions are of high quality. Vital factors for a successful counseling process include a positive working alliance and empathy of the counselor. This study investigates how counselors’ behaviors, measured with a behavioral coding instrument, relate to positive working alliances and empathy. We introduce a video-based instrument that measures micro-behaviors and skills during career counseling sessions. Psychometric analyses were conducted on the basis of 53 video recordings. Double-coded sessions (N = 18) showed excellent interrater reliabilities for the coded micro-behaviors.

By using a multi-trait–multi-method approach, we established construct validity with another coding instrument, whereas criterion validity was established with respect to counselors’ empathy and working alliance ratings. We discuss theoretical implications of using video-based methods in career sessions to advance process research, and present practical applications in terms of assessing quality assurance in career interventions.

Keywords: career counseling, empathy, working alliance, video methods, observational analyses
Due to globalization, as well as to the increasing changes in work and forms of employment, the demands on graduates and young adults have increased enormously. Individuals entering the workforce often seek guidance for career decisions from professionals (Gati & Levin, 2014). Professionals offering career guidance are found under many different names, such as career counseling, career consultancy, guidance interviews, helping professionals, transition advisors, and career or life coaches (Chung & Gfroerer, 2003; Grant, Passmore, Cavanagh, & Parker, 2010; McMahon, 2004). They help individuals to make important career decisions, develop career-related goals (e.g., Grant et al., 2010), and improve overall career planning (Brown et al., 2003). More importantly, professional career advisors are estimated to make up to 0.8% of the entire labor force (Grant et al., 2010; OECD, 2003; Whiston, 2003).

Whereas the backgrounds of these professionals are manifold—many having no career-specific training (OECD, 2003)—the question arises as to what these professionals actually do to help their clients to make important life decisions? To answer this, we need fine-grained taxonomies that capture counseling behaviors in-situ, which will allow us to study counseling sessions with precision (Heppner & Heppner, 2003). Behavioral taxonomies can be analyzed based on videotaped counseling sessions (e.g., Multon, Ellis-Kalton, Heppner, & Gysbers, 2003), which offer methodological advances to survey-based approaches because the latter can be biased through enhanced self-perceptions or memory inaccuracies (Miller & Mount, 2001; Miller, Yahne, Moyers, Martinez, & Pirritano, 2004). To illustrate, Miller and Mount (2001) showed that counselors’ self-reports of counseling skills shared very little overlap with objective

---

1 Note: In the present manuscript, we use the term career professionals for any professional who offers a career counseling service (i.e., counselors, coaches, mentors, educators etc.)
observations of counseling behaviors in practice. That is, counselors tended to largely overestimate their own skills when answering survey-based measurements.

With little knowledge about what happens within career counseling sessions, research must open the “black box” of the career counseling process (Heppner & Heppner, 2003, Whiston, 2003). Video-based approaches allow one to repeatedly inspect specific behaviors and interaction sequences (e.g., Klonek, Wunderlich, Spurk, & Kauffeld, 2016; Mittendorff, den Brok, & Beijaard, 2010). The current study aims to advance our understanding of career guidance sessions by introducing a video-based observational instrument.

First, we introduce a video-based instrument specifically adapted to the context of career counseling. This instrument is based on the Advanced Interaction Analysis for Consulting (act4consulting, short “ACT”; Hoppe, 2013) and captures practitioners’ micro-behaviors and skills within the career guidance process. The behavioral categories (i.e., the counselor’s procedural and socioemotional skills) are theoretically based on control and self-determination theory (e.g., Gregory, Beck & Carr, 2011; Markland, Ryan, Tobin, & Rollnick, 2005), and refer to a non-directive and respectful inquiry style of the counselor (Van Quaquebeke & Felps, 2016). Second, we provide detailed reliability analyses for the micro-behaviors that the instrument captures. Third, we demonstrate construct and criterion validity based on technical (e.g., questions, paraphrasing, summaries) and relational outcome (i.e., empathy and working alliance) criteria. We focus on relationship quality criteria of the counselor–client interaction (i.e., empathy, Moyers & Miller, 2013; working alliance, Baron & Morin, 2009; De Haan, Duckworth, Birch, & Jones, 2013) because these are essential criteria for counseling success across a variety of settings, such as psychotherapy, coaching, and career guidance (e.g., McKenna & Davis, 2009).
Reliability of Measures Obtained From Video-Based Instruments

As observational studies are labor-intensive research projects, the reliability of measures that are obtained with video-based instruments is of vital importance (Klonek, Quera, & Kaufffeld, 2015; Multon et al., 2003). Accuracy of behavioral observations is usually assessed by using several observers and comparing their obtained scores with statistical measures of interrater reliability. Unfortunately, reliability of observational instruments is often not reported in career counseling studies (e.g., Multon et al., 2003). Another problem is that many researchers only report overall reliability measures (e.g., Gessnitzer & Kaufffeld, 2015; Mittendorf et al., 2010), which lack information on category-specific reliability. This approach can be criticized for several reasons: First, an overall reliability index does not guarantee high reliability for specific behavioral categories that are captured by a coding instrument (Hewes, 1985). Second, this approach fails to inform other researchers about category-specific reliability (cf., Klonek, Quera, Burba, & Kaufffeld, 2016). For example, Moyers, Martin, Catley, Harris, and Ahluwalia (2003) reported category-specific reliability estimates for all counselors’ micro-behaviors of their observational instrument, and showed that 50% of variables provided poor reliability. This underscores that high reliabilities in observational research should not be taken for granted. Low reliabilities can be circumvented by using a coding manual, appropriate observer trainings (Bakeman & Quera, 2011), and the integration of software support (Klonek et al., 2015). The current study applied these aforementioned quality measures (i.e., coding manual, observer training, and software support). Furthermore, we provide information on code-specific reliabilities.

**H1: Micro-behaviors and counseling skills obtained from a video-based career counseling instrument will provide reliable scores.**
Convergent Validity of Video-Based Assessed Career Counselors’ Micro-Behaviors

Beside reliability, validity is crucial for a new coding system. Convergent validity can be estimated by comparing results obtained from a second instrument that captures similar measures. We selected this second coding instrument based on two criteria: First, there had to be scientific research using the second observational instrument in a career counseling context (Kronek et al. 2016; Rochat & Rossier, 2016). Second, there had to be data available on the second instrument that indicated strong psychometric properties. Using these criteria, we selected the Motivational Interviewing Treatment Integrity (MITI). First, the MITI has been used in career counseling research contexts: For example, in one study it was used to evaluate whether career counselors meet quality benchmarks in Motivational Interviewing (MI), which is a client-centered intervention in which the counselor’s empathy is a key component (Kronek et al., 2016). The MITI allows us to measure both client-centered behaviors (open questions and reflective listening, which are important indicators of counselor empathy, Moyers, Martin, Manuel, Miller, & Ernst, 2010), neutral behaviors (giving information, closed questions), and behaviors that are specific to MI interventions (i.e., MI adherent behaviors, e.g., affirmations, vs. MI non-adherent behaviors, e.g., confrontations). Second, due to the strong empirical research tradition in MI, there is a substantial amount of publications on the MITI, reporting excellent psychometric properties (e.g., Kronek et al., 2015; Moyers, Martin, Manuel, Hendrickson, & Miller, 2005). Finally, the MITI also overlaps conceptually with existing career counseling instruments, such as the “Revised hill counselor verbal response modes category system” (Hill, 1985; Multon et al., 2003). Due to these arguments and the ease of accessibility of the MITI (available at no charge at casaa.unm.edu), we used it to assess convergent validity. Overall, we hypothesize that:
H2: Objectively coded micro-behaviors (i.e., close questions, open questions, paraphrasing, summarizing, appreciation, interruption of the client) will be positively associated with conceptually similar categories of the MITI (i.e., closed question, open question, simple reflection, complex reflection, MI adherent behavior, MI non-adherent behavior).

Criterion Validity of Video-Based Assessed Counseling Skills

With respect to criterion validity, we focus on two outcome variables: empathy and working alliance. Whereas empathy is considered to be a crucial success measure in therapeutic counseling contexts (Moyers & Miller, 2013), working alliance is an essential variable in the context of career counseling (Whiston, Rossier, & Barón, 2016).

In counseling research, counselor empathy can be conceptualized as interpersonal efforts to take the clients’ perspectives by trying to accurately understand their cognitive and emotional experiences (Moyers & Miller, 2013). A variety of therapy studies have indicated that empathy of counselors is an essential predictor of client-related therapy outcomes (e.g., Pantalon, Chawarski, Falcioni, Pakes, & Schottenfeld, 2004; Ritter et al., 2002). For example, Pantalon et al. (2004) showed that rated counselor empathy, based on videotaped sessions, prospectively explained objective measures of their respective clients’ health. Building on this evidence, Moyers and Miller (2013) concluded that low counselor empathy might even be toxic for clients. Whereas psychotherapy research contributed substantially to a better understanding of how counselor skills affect patient outcomes, research in career counseling has less to offer concerning these issues (Whiston & Oliver, 2005). Based on the evidence from therapy research, we assume that counselor empathy in career interventions provide an equally important ingredient for helping clients to deal with career-related problems. Taken together, counselors’ behavioral skills should be related to how empathetically they approach their client, and how
well they deal with sensitive content that could occur in the session (Hirschi, 2017). In general, counselor skills associated with empathy comprise the expression of understanding and caring (e.g., De Haan, 2008, Rogers, 1961). Typical behaviors associated with empathy are, for example, paraphrasing or addressing the client’s feelings (e.g., Rogers, 1961, Korman, Bavelas, & De Jong, 2013). These counselor skills can be termed as functional socioemotional skills. We expect that functional socioemotional counselor skills measured with objective video recordings will contribute towards counselor empathy. Thus, we hypothesize:

\[ H3: \text{There will be a positive relationship between functional socioemotional skills and empathy.} \]

Counselors may also display insensitive behaviors, and thereby harm the relational empathic bond with the client. Instead of active listening, paraphrasing, and supporting the client, the counselor may be distracted, express impatience, or even displaying lacking interest. We consider such adverse behaviors as dysfunctional. Thus, we hypothesize:

\[ H4: \text{There will be a negative relationship between dysfunctional socioemotional skills and empathy.} \]

**The Importance of Working Alliance in Career Counseling**

The working alliance between counselor and client has received increased attention in career counseling research (Whiston et al., 2016). Whiston et al. (2016) reviewed 17 working alliance studies and showed that it positively related with various client outcome measures (e.g., career maturity, job and life satisfaction, or vocational self-efficacy). Despite knowledge about the various consequences of a good working alliance, we know significantly less about the counseling behaviors that positively predict a working alliance (cf., Whiston et al., 2016).
A counseling working alliance characterizes an affective bond with a strong focus on tasks and goals (McKenna & Davis 2009). The affective bond is strongly associated with constructs such as trust (O’Brien & Palmer, 2010), interpersonal comfort, and mutual sympathy (Baron & Morin, 2009). Certain behaviors are generally considered beneficial to the development of an affective bond (e.g., demonstrating empathy, encouragement, or appreciation; e.g., Gregory & Levy, 2011; O’Brien & Palmer, 2010). The affective bond seems to provide the necessary framework in which the client feels secure enough to open up, and to being self-reflective (Greif, 2007; Ianiro, Lehmann-Willenbrock & Kauffeld, 2014). However, the concept of working alliance goes beyond the affective bond between counselor and client: It also relates to the degree of mutual agreement as to the goals to be attained, as well as to the ways and means of attaining them (Baron & Morin, 2009). Thus, the construct working alliance also refers to the quality of the counseling procedure. There are several micro-behaviors that can be related to a high-quality counseling procedure. Research linking control theory and counseling has stressed the importance of a counselors’ goal orientation, the ability to provide structure within the process, and support for the client’s self-reflection (e.g., Grant, 2005; Gregory et al., 2011). To achieve this, counselors need to display procedural skills, such as good time management, suggesting the next procedural step, and staying focused within the session (e.g., Grant, 2007). Another important counseling behavior is the use of open questions because they stimulate self-awareness and require personal responsibility of the client (Passmore & Fillery-Travis, 2011).

Taken together, we term these behaviors associated with the quality of the counseling procedure as functional procedural skills. We expect that functional procedural skills positively contribute to a working alliance. Thus, we hypothesize:
H5: There will be a positive relationship between functional procedural skills and working alliance.

For many counselors, it is often difficult to use open, instead of closed, questions within the session (Miller & Mount, 2001), and to prevent themselves from quickly suggesting solutions for the client (Klonk et al., 2015). However, in non-directive counseling, it is particularly important for counselors to be able to step back and let the client find his or her own solutions (e.g., Markland et al., 2005). We expect that counselor’s behaviors, such as proposing a problem or solution instead of encouraging the client to find the solution for him/herself, are dysfunctional in the counseling process and thereby negatively affect the working alliance.

H6: There will be a negative relationship between dysfunctional procedural skills and working alliance.

Whereas empathy reflects the affective-relational bond between counselors and clients, working alliance is most strongly grounded in goal- and task-related behavior. A coding instrument used for career counseling needs to address both affective goal- and task-related behavior, as well as relational aspects, such as empathy. As our instrument allows for assessing a counselor’s behavior related to goals, and a client’s progress, we expect that—in the context of counseling—it exceeds explanatory power of existing coding instruments that were developed for a specific intervention, and which have a more narrow focus (such as the MITI). Thus, we compared the explanatory variance of our instrument with the MITI and hypothesize:

H7: Counselor skills (measured with ACT) explain working alliance (in terms of goals and tasks) above and beyond client-centered and motivational counselor skills (measured with the MITI).

Method
Sample

The current analyses are based on a dataset of 53 career counseling sessions (with unique counselor–client dyads). Data was collected at two German universities from 2008 to 2013. The counselors were young professionals with a bachelor’s degree in psychology and extensive training in career counseling (approx. 200 hours) over a period of one year that covered theoretical (learning about counseling, literature study) and practical (peer coaching) components. The counselors were, on average, 25 years old ($SD = 3.64$), and the majority was female (92.5%, $N = 49$), which corresponds to a growing female gender distribution of graduates in psychology (Ostertag & McNamara, 1991). The clients’ sample consisted of young professionals or students with a bachelor’s degree before finishing their master’s degree. The mean client age was 25 years ($SD = 3.11$), and they showed a balanced gender distribution (56.6% female clients, $N = 30$).

Procedure

The clients were recruited via advertisement or via the website of the two institutes. They could participate if they had enrolled in one of the two universities that offered the career counseling program. After giving written consent to be videotaped, counselors and clients (who did not know each other before) were randomly assigned to each other. The entire career guidance program covered five well-structured sessions that each took about 90 minutes.

Video-Observation of Counselor Behaviors and Skills

We analyzed the first session because it builds the baseline for establishing a working relationship with the client. Recorded counselor behaviors were coded using the ACT$^2$ instrument (Gessnitzer & Kauffeld, 2015, Hoppe, 2013), which is mutually exclusive and

---

$^2$ The entire coding manual for the video-analytical instrument can be obtained from the second author
exhaustive in that every observational unit is assigned to only one code (Bakeman & Quera, 2011). Videotaped sessions were subdivided into units, each representing the smallest meaningful communicative statement that could be understood by the other member of the interaction. We were interested in micro-behaviors of counselors that are considered important for the development of the affective bond, such as active listening, or demonstrating empathy or encouragement (e.g., Duckworth & De Haan, 2009; O’Broin & Palmer, 2010). Further, we coded micro-behaviors associated with a counselor’s goal- and task-orientation, which support clients’ self-regulation (Gregory, et al. 2011). Counseling approaches assuming clients’ innate tendency for personal growth, and their own expertise over their issue, emphasize the importance of a counselor’s non-directive and respectful inquiring attitude toward the client (e.g., Markland et al., 2005; Van Quaquebeke & Felps, 2016). In terms of respectful inquiry, three components are important: asking questions, question openness, and attentive listening (Van Quaquebeke & Felps, 2016). Markland and colleagues (2005) further described counselors’ behavior that provides structure (e.g., helping clients to develop appropriate goals or giving them feedback) and support a client’s autonomy (e.g., exploring options with the client and avoiding coercion). Following this line of reasoning, we focused on counselors’ micro-behaviors’ that showcase a degree of structure, respectful inquiry, and directivity. We selected functional and dysfunctional socioemotional and procedural micro-behaviors associated with bonding, structure, and respectful inquiry and directivity.

Examples for functional socioemotional skills include establishing contact (e.g., “Hi, was it easy to find your way here?”), bonding (e.g., “In case of emergency, you can call me”), and facilitating behaviors, such as signs of listening (e.g., “Hmm, yes”), addressing client’s feelings (e.g., “You felt excited in this moment”), and appreciation (e.g., “Well done, you are very
ambitious!”). *Dysfunctional socioemotional skills* are, for example, *interruptions of the client* or *disruptive behavior* (e.g., “What you told me is not important to me!”).

*Functional procedural skills* cover behaviors that help to structure sessions, such as *goal orientation* (e.g., “Today, we will talk about your career planning”), *time management* (e.g., “We still have 10 min. left for…”), or *open questions* (e.g., “What would help you to enhance your person–job fit?”). *Dysfunctional procedural skills* include *losing the train of thoughts and running off topic* (e.g., “My cousin also works for this company… Oh, I need to call him later”). Moreover, directive behaviors, such as *suggesting* or *explaining* solutions for the client, are considered dysfunctional (e.g., “An internship in this company looks like a solution to your problem”).

**Observer Training**

Before coding the analysis sample, three observers received 200 hours of training. All observers coded a training sample, and were only allowed to code the analysis sample when their inter-reliability exceeded $\kappa \geq .70$, that is, after having a substantial agreement (Landis & Koch, 1977). Biweekly and supervised observer meetings gave coders the possibility of addressing questions, deciding on rules, and solving disagreements. All observers kept a diary to note specifics, which were discussed within these regular meetings in order to reach coding accuracy.

**Behavioral Measures for Construct Validity**

**Behavioral measures.** Convergent validity was assessed using the MITI (German version MITI-d; Brueck et al., 2009; Klonek et al., 2015), which is a behavioral coding instrument to assess behavioral skills in the psychological approach of MI (Miller & Rollnick, 2013). Although MI coding instruments are mainly used for MI-based interventions, they include
conceptually overlapping micro-behaviors with ACT (e.g., open/closed questions, summarizing/reflecting are codes that exist both in MITI and ACT).

A new pair of observers used a software-supported MITI-d version (Klonek et al., 2015) to code the seven behavioral categories (open/closed questions, simple/complex reflections, MI adherent/non-adherent behavior). Before coding, they also received a coder training for the MITI-d, that is, they read theoretical literature, studied the coding manual (Brueck, Frick, & Loessl, 2006), and participated in gradual learning tasks (i.e., we used coded transcripts provided in Klonek et al., 2015). None of the training material was used in subsequent analysis. Both observers had the possibility to ask the coding coordinator (T.W.) to discuss coding discrepancies. To assess reliability of the MITI-d, we calculated the intraclass correlations (ICCs, Shrout & Fleiss 1979) for 53 double-coded samples. All seven MITI categories (ICCs > .84) achieved excellent levels of interrater reliability (Cicchetti, 1994).

**Outcome Measures for Criterion Validity**

**Empathy of practitioner.** We used a global rating scale from the MITI-d (i.e., gestalt ratings) to assess counselors’ empathy. This global rating is based on practitioner’s behavior of a 20-minutes sequence, and aims to capture an overall impression (i.e., “gestalt rating”) using a 7-point scale to evaluate the extent of counselor empathy (1 = low empathy to 7 = high empathy). High empathy ratings were given when practitioners showed “an active interest in making sure they [understood] what the client [was] saying,” whereas low empathy ratings were given when there was “little interest in the point of view and in the experiences of the client” (Brueck et al., 2006, p. 2). Interrater reliability for 53 empathy ratings was excellent (ICC = .84). To further reduce the measurement error of our rating measure, we used the average empathy ratings from both observers as an outcome variable in subsequent analyses.
**Working Alliance Inventory.** We used the short WAI-S version from Tracey and Kokotovic (1989). Items for the short version (WAI-S) and its three subscales—bonding, goal orientation, and task orientation—were obtained from a factor analysis of the WAI (cf., Tracey & Kokotovic, 1989). Furthermore, this short version has been used in previous studies (e.g., Gessnitzer & Kauffeld, 2015; Mutlon et al., 2003). The WAI-short scales each comprise four items, and were originally developed as a self-report measure. As we applied an independent observers’ perspective to infer working alliance directly from the videos (cf., Meinecke, Klonek, & Kauffeld, 2016), items were adapted accordingly (e.g., “Counselor/coach and client agree on the things the client will need to do to help improve his/her situation”). Items were scored from both observers using a 7-point Likert-type scale (1 = *strongly disagree*, 7 = *strongly agree*).

Cronbach’s alpha for both observers was $\alpha = .71/.81$, and interrater reliability had an ICC of .83. As both observers rated all 53 sessions, we used the observer-mean WAI ratings as a dependent variable in subsequent analyses to further reduce measurement error. (Note that observers who rated empathy and working alliance were not the same observers who coded the counselors’ micro-behaviors.)

**Overview of Statistical Plan**

**Reliability analysis.** Reliability for behavioral codes was estimated by calculating ICCs for 18 double-coded sessions (Fleiss & Shrout, 1978). The ICC is a statistical index commonly used to estimate reliability because it adjusts for chance agreement and systematic differences between observers (Fleiss & Shrout, 1978); it is also a more conservative estimate than the Pearson correlation. We computed absolute (criterion-referenced) agreement between two observers using the following formula (Bakeman & Quera, 2011), which is stricter than the relative (norm-referenced) agreement (Bakeman & Quera, 2011): $ICC =$
\[ \frac{MS_b - MS_e}{(MS_b + MS_e - (2/n)(MS_o - MS_e))} \], where \( n \) is the number of observations, or summary values, obtained by each observer; \( MS_b \) is the between-measures mean square; \( MS_o \) is the observer, or repeated measures, mean square; and \( MS_e \) is the error mean square. According to Cicchetti (1994), we classified ICCs by cut-off criteria: below .40 = poor; .40–.58 = fair; .60–.74 = good; and .75–1.00 = excellent.

**Validity analysis.** Convergent validity was assessed by comparing 53 sessions that were coded with both ACT and the MITI-d by independent sets of coders for each instrument. First, we selected conceptually overlapping codes. Second, we derived session summary counts for each behavioral code for both instruments for all 53 sessions. Finally, we calculated correlations of behavioral code measures between both instruments.

For convergent validity (H2), we expected high correlations for conceptually similar codes from ACT and MITI, such as ACT-closed questions with MITI-closed questions, ACT-open questions with MITI-open questions, ACT-paraphrase with MITI-simple reflections, ACT-summary with MITI-complex reflections, ACT-interrupt with MITI-MI non-adherent behavior, and ACT-appreciation with MITI-MI adherent behavior. In terms of discriminant validity, we expected close-to-zero and non-significant correlations for codes that were conceptually different.

For criterion validity, we expected that socioemotional skills explain variance in empathy (H3 and H4). Furthermore, we expected procedural behavioral skills to explain variance in working alliance, particularly with regard to goal and task-orientation (H5 and H6). Finally, we expected that ACT counseling skills explain variance above and beyond behavioral MITI measures (H7).
Results

Reliability

Sample size for reliability analysis was chosen a priori following two guidelines:
Bakeman, Deckner, and Quera (2005) recommend a sample of 15–20% from the corpus to check reliability. Furthermore, for the calculation of ICCs, a minimal sample of > 5 is recommend—with ten or more sessions resulting in more robust results (Bakeman & Quera, 2011). To fall above these recommended minimal criteria for reliability analyses, we randomly selected a subsample of N = 18 (i.e., 33%) sessions.

-----Insert Table 1 about here-----

H1 stated that micro-behaviors and counseling skills obtained from a video-based career counseling instrument will provide reliable scores. Table 1 shows reliabilities for all behaviors, as well as for the respective superordinate counseling skills. On average, micro-behavioral codes showed $M_{\text{ICC}} = .77 \ (SD_{\text{ICC}} = .22)$, that is, code-specific reliability mostly showed excellent agreement (Cicchetti, 1994). Measures of functional and dysfunctional socioemotional skills all showed excellent interrater reliability (i.e., ICCs > .74), according to Cicchetti’s criteria (1994)—with the exception of the codes humor (ICC = .73), addressing client’s feelings (ICC = .71), disagreement/ criticism (ICC = .65), and summaries (ICC = .35). Measures of functional and dysfunctional procedural counseling skills showed good to excellent reliabilities (ICCs > .63)—with the exception of clarifying one’s own contributions (ICC = .48), and goal orientation (ICC = .06).

Following procedures from Moyers et al. (2003), we collapsed behavior codes that theoretically belong to a superordinate category (i.e., counseling skill), and calculated interrater

---

3 Information about descriptive statistics for study variables (means, standard deviations, and correlation matrices) are available as supplemental material by contacting one of the authors (F.K., P.I, or T.W.).
reliability estimates for these categories as well. Reliabilities for these four superordinate counselor skill measures were in the good to excellent range (ICCs = .66 - .90; cf., Table 1), thus supporting H1.

**Construct Validity**

For convergent validity (H2), we compared ACT codes with conceptually similar and different codes from the MITI. In H2, we stated that objectively coded micro-behaviors (i.e., close questions, open questions, paraphrasing, summarizing, appreciation, interruption of the client) will be positively associated with conceptually similar categories of the MITI (i.e., closed question, open question, simple reflection, complex reflection, MI adherent behavior, MI non-adherent behavior).

---Insert Table 2 about here---

Table 2 shows correlations for MITI codes with conceptually similar ACT codes. The table also contains a mono-trait hetero-method block from the multi-trait, multi-method (MTMM) matrix (Campbell & Fiske, 1959). Conceptually overlapping codes (from both instruments) are ordered in the same way, such that the diagonal allows estimating convergent validity, whereas the off-diagonal cells indicates discriminant validity results.

The MTMM-diagonal of Table 2 shows significant correlations for the following ACT and MITI variables: ACT-closed questions with MITI-closed questions (r = .55, p < .01), ACT-open questions with MITI-open questions (r = .56, p < .01), ACT-paraphrase with MITI-simple reflection (r = .25, p = .07), ACT-interrupt with MITI-MI non-adherent (r = .54, p < .01), ACT-appreciation with MITI-MI adherent (r = .33, p < .05). Only ACT-summaries showed a non-significant correlation with MITI-complex reflection (r = .18, p = .19), which might be
explained due to low reliabilities for the ACT-code (summaries; reliability determining an upper bound for validity estimates, Bakeman & Quera, 2011).

Overall, these correlations support the convergence of central ACT codes with MITI-codes. Conversely, off-diagonal cells in Table 2 showed non-significant correlations that are, on average, close to zero ($M_r = .02$), thus supporting discriminate validity. By providing empirical results for both convergence and discrimination, these results support H2 (i.e., ACT codes showing convergent and discriminant validity with MITI-codes).

**Criterion Validity**

To test criterion validity, we performed multiple linear regression analyses for two outcome criteria: counselor empathy and working alliance.

**Counselor Empathy**

Predictors for empathy were functional and dysfunctional socioemotional skills ($F = 2.2, R^2 = .08$) In H3, we stated that there will be a positive relationship between functional socioemotional skills and empathy. Results indicated that functional socioemotional skills positively predict counselor’s empathy ($\beta = .30, p < .05$), thus, supporting H3. In H4, we stated that there will be a negative relationship between dysfunctional socioemotional skills and empathy. Contrary to expectations, dysfunctional socioemotional skills were not significantly related to empathy ($\beta = -.17, p > .05$). Thus, H4 was not supported.

**Working Alliance**

Predictor variables for working alliance were functional and dysfunctional procedural skills ($F = 3.5, R^2 = .12$). In H5, we stated that there will be a positive relationship between functional procedural skills and working alliance. Functional procedural skills did not significantly predict working alliance ($\beta = .22, p > .05$), thus not supporting H5. In H6, we stated
there will be a negative relationship between dysfunctional procedural skills and working alliance. Dysfunctional procedural skills negatively predicted working alliance ($\beta = -.37, p < .05$), thus supporting H6.

To test H7, we performed hierarchical multiple regression analyses for the working alliance scales goals and tasks (Table 3). Predictors were entered in three steps: In Model 1, we entered the session duration as a control variable. In Model 2, we entered the behavioral MITI measures. In the Model 3 (which tested H7), we entered procedural counseling skills measured with ACT.

-----Insert Table 3 about here-----

In H7, we stated that counselor skills (measured with ACT) explain working alliance (in terms of goals and tasks) above and beyond client-centered and motivational counselor skills (measured with MITI). With respect to working alliance (i.e., goals), Model 2 and 3 significantly explained variance in this outcome variable (Model 2: $F = 7.98, df = 3, p < .01, R^2 = .33$; Model 3: $F = 6.09, df = 5, p < .01; R^2 = .39$). Further, inclusion of procedural counseling skills (measured with ACT) explained incremental variance in working alliance goals above and beyond the MITI observational measures ($\Delta R^2 = .07, p < .10$), thus supporting H7.

With respect to working alliance (i.e., tasks), Model 2 and 3 significantly explained variance in this outcome variable (Model 2: $F = 10.82, df = 3, p < .01, R^2 = .40$; Model 2: $F = 8.41, df = 5, p < .01; R^2 = .47$). Counseling skills (measured with ACT) explained incremental variance in working alliance (tasks) above and beyond the MITI measures ($\Delta R^2 = .07, p < .01$), and further supported H7.

Discussion
The current research study investigated the psychometric properties of a video-based instrument for career counselors. By using behavioral analyses of counselors in action, our aim was to better understand the micro-behaviors and counseling skills that contribute to empathy and working alliance in career guidance sessions. Overall, two main findings occurred from this study.

First, we provided detailed, category-specific reliability analyses for both micro-behaviors and counseling skills, showing that the video-based instrument can be used to reliably code micro-behaviors of career counselors.

Second, results from construct validity analyses showed that counselor micro-behaviors of the coding instrument not only conceptually, but also empirically, overlap with micro-behaviors that are assessed by an established instrument (i.e., the MITI). In addition to construct validity, our study also supported criterion validity of measures obtained from the instrument: whereas functional socioemotional skills positively predicted counselor’s empathy, dysfunctional procedural skills significantly and negatively predicted working alliance, as rated by independent observers. Moreover, our results suggest that the measures from the instrument explained incremental variance in working alliance above and beyond established MITI measures.

**Theoretical Implications**

Unraveling the career counseling process is a major and current challenge with regard to counseling professionalization (cf. OECD, 2003, Grant et al., 2010). This is where our findings, based on behavioral observations, may make an important contribution. First, our results show that relevant counselor behaviors can be observed systematically and reliably over the course of the interaction with their clients. In line with several authors (e.g., Heppner & Heppner, 2003, Miller et al., 2001, 2004), our study suggests that genuine career guidance process measures are
well suited to capture micro-behaviors of career counselors during a counseling session. Thus, researchers who want to better understand the behavioral ingredients that actively contribute to good career interventions can use the video-instrument for these purposes. For example, we have no systematic knowledge of whether professionals trained in specific theoretical approaches (e.g., systems theory framework, Patton & Mahon, 2006, or theory of career construction, Rehfuss, Del Corso, Galvin & Wykes, 2011) display characteristic behavioral profiles within their sessions. However, this knowledge would allow us to assess training effectiveness in different theoretical schools of thought, and allow us to develop benchmarks or empirical comparisons of how different career theories manifest into specific behavioral counseling profiles. Furthermore, the software integration of the behavioral instrument automatically records on- and offsets for all micro-behaviors. By means of integrating temporal processes into counseling research, researchers can better pinpoint temporal sequences that are crucial for counselor–client interactions.

Second, our study also contributes to a better understanding of behavioral skills that are related to counselor empathy and working alliance. Previous therapy research has suggested that empathy is a crucial ingredient within a counseling relation that positively impacts client health and adherence to treatment (Moyers & Miller, 2013). In line with several authors (e.g., De Haan, 2008, 2009; Rogers, 1961), the current study has shown that functional socioemotional skills in particular, such as using summaries, bonding with the client, or appreciation, positively and significantly predict counselor empathy. Similar to findings from psychotherapy (Sultanoff, 2003), our results suggest that humorous expressions are also related to empathic behavior.

Contrary to expectations, dysfunctional socioemotional skills were not significantly related to empathy. Interrupting the client or disruptive counselor behavior did not contribute to
lower ratings on empathy or working alliance. It is possible that such behaviors—rather than being considered as real disruptions—are considered as supportive or essential for successful processes, if well dosed and applied in a sensitive way. Counselors are responsible for structuring and controlling the process, and therefore may sometimes have to interrupt clients from time to time (Ianiro, Schermuly, & Kauffeld, 2013). However, behaviors, such as disagreement or criticism, were negatively and significantly related to working alliance (in terms of agreeing about tasks and goals). These findings support the assumption that empathy versus working alliance cover different and complementing aspects of the counselor–client relationship.

Another potential explanation for the lacking association between dysfunctional socioemotional skills, empathy, and working alliance may be that socioemotional skills had a very low mean and small standard deviation, whereas working alliance and empathy scores were above average in our sample of counselors. These statistics suggest a restricted variance in socioemotional skills and further, that counselors in our study were generally above average (in terms of working alliance and empathy). A reduced range of a variable can lessen the strength of an association with another variable and this interpretation may offer an alternative explanation why we did not find support for H4 and H5.

Interestingly, findings on dysfunctional procedural behaviors appear to be more insightful compared to functional procedural skills. As expected, closed questions and directive behaviors were negatively related to working alliance. The latter results might indicate that effective counselors need to help their clients in a client-centered fashion, so that clients (rather than the counselors) discover the problems and voice their own solutions (cf. Markland et al., 2005; Van Quaquebeke & Felps, 2016). By contrast, counselors who engage in problem-solving themselves, and suggest solutions, might undermine their clients’ active reflection and
understanding of what drives their career goals, as well as the obstacles that they are facing. Counselors who heavily engage in problem-solving instead of helping their client to work out their problems might miss out on opportunities for their client to have both innovative moments and self-exploration of the pros and cons of a specific career dilemma (Rochart & Rossier, 2016).

Taken together, these findings suggest differential effects for specific counselor behaviors. To evaluate which counselor behavior is beneficial for counseling success, a micro-analysis of specific behaviors appears to be a promising approach. Moreover, the results also suggest that a distinction between functional and dysfunctional skills may be too simplistic. It is conceivable that some of the chosen micro-behaviors are neutral, or only have a negative impact if they are applied too often, or expressed in an extreme way (e.g., interrupting the client). Still, other behaviors may only have a negative impact if missing (e.g., open questions). The impact of some micro-behaviors may only be fully understood when the corresponding nonverbal expressions are considered as well. Overall, we are only beginning to understand the interactional and behavioral dynamics in career interventions. The current study provides a tool that magnifies the behavioral aspect of effective career counseling.

Finally, the current study provides an observational instrument that can inform research about behavioral processes in career guidance interventions. This is an important contribution, especially because scholars have repeatedly criticized the overreliance on self-report survey measures because these can be flawed by social desirability, memory distortions, or limited attention spans (Miller & Mount, 2001). There is strong evidence that self-reported behavior is often little, or not at all, related to how participants behave if we observed them (Miller & Mount, 2001).
Practical Implications

The current study provides important implications for career counseling practice. First, our study informs career practitioners that displaying socioemotional behaviors (e.g., paraphrasing client statements, showing appreciation, and building contact) will help them to convey empathy toward their clients. Furthermore, practitioners learn from our study that interrupting or disagreeing with clients does not necessarily reduce their level of empathy, but that it is more important to engage in positive socioemotional behaviors in this context. By contrast, practitioners who want to focus on building a good working alliance with their clients are advised to refrain from displaying disruptive behaviors or frequently interrupting their clients (i.e., dysfunctional procedural skills). For the working alliance, displaying functional procedural behaviors (e.g., prioritizing or making procedural suggestions) seems to be less important. Taken together, our study implies that practitioners who want to both build a good working relationship and express empathy toward their clients should both display functional socioemotional skills and refrain from displaying dysfunctional procedural skills.

Second, our study directs practitioners to use video-based analyses to better understand their own career counseling approach. By coding their own sessions, they have a tool to systematically understand their micro-behavioral repertoire within a session. Practitioners with interests in co- or supervision can use the instrument as a means of providing high-quality feedback. As coded recordings can be easily played back by means of the software, supervisors can give feedback based on visible behavioral sequences, which facilitates discussions with counselors-in-training. As a result, practitioners have a tool to reflect and improve their skills.

Third, agencies providing career counseling training can use video-based analyses for evaluation purposes. Although counseling approaches in psychotherapy have been using
observational instruments for nearly two decades to evaluate how trainings affect practitioners’ counseling behaviors (De Roten, Zimmermann, Ortega, & Despland, 2013; Miller & Mount, 2001), these assessments are still relatively uncommon in career research (Hirschi, 2017). We presented a tool that training agencies can use to evaluate specific skills in career counseling, and knowledge of how these effect the relationship quality. As self-reported skills do not share much overlap with objectively assessed behavioral skill measures (Miller et al., 2004), the video-based instrument can provide a more objective quality assurance of career counseling skills (McMahon, 2004).

**Limitations and Future Research**

We acknowledge several limitations of this study. First, our study has used a relatively homogenic sample of young counselors with a limited amount of counseling experience. This implies a limited generalizability of our findings, and we can only cautiously infer how far our results also generalize to more age-diverse samples. Therefore, we encourage future studies to use video-based analyses in samples that show more variance in terms of age, gender, and experience. As the profession of career counselors constitutes a large variety of professionals, we also expect that behavioral profiles of career counselors with different education and training should differ. Nonetheless, and to the authors’ knowledge, the current study thus far provides the largest reported study that has used behavioral observation of career counselors—therewith exceeding the study from Multon et al. (2003), who observed behavioral responses of 19 career counselors-in-training.

Second, some behavioral categories showed low reliabilities. Even though our study is not the only study that has reported problems with low reliabilities classifying counselor behaviors (for example, Moyers et al., 2003, reported poor reliabilities for 50% of counselor
behaviors), the aggregate counseling behaviors all provided good to excellent reliability. Furthermore, our study provides a first orientation in category-specific reliability, and therewith indicates which behaviors are potentially difficult to assess for observers. Future studies can use our results to focus on behavioral categories and different nuances of nonverbal behaviors that potentially need stronger emphasis during observer training.

Third, in terms of behavioral focus, our study only investigated behaviors of the career counselor. However, there is gathering evidence that suggests behaviors of the counselor directly interact with clients’ motivational response within the session (cf., Klonk et al., 2016). Future studies should use behavioral analyses of career counselors to investigate the interactional dynamics between counselor and clients by means of advanced statistical methods, such as sequential or analysis of behavioral patterns (Bakeman & Quera, 2011; Meinecke et al., 2016).

Fourth, our study does not provide information on how counselor skills affect client development outcomes (e.g., improved career planning, achievement of career goals, or higher self-efficacy in career decisions). We encourage future studies to use the new tool presented here, and link counselor behaviors to effectiveness measures of career interventions. We would hypothesize that functional counselor skills positively relate to outcomes such as clients’ career planning behaviors. This research could further extend the applicability of this tool, and inform career professionals.

**Conclusion**

The current study introduced a video-based instrument and examined the micro-behaviors and counseling skills of career practitioners. By analyzing recorded career guidance sessions, we have shown that video-based scores are reliable and valid. Counselor procedural skills contributed to empathy and working alliance. A better understanding of micro-behaviors in
career guidance can contribute both to training and practice, as well to our theoretical understanding of effective career guidance.
References


Grant, A. M. (2005). What is evidence-based executive, workplace, and life coaching? In M. Cavanagh, A. M. Grant, & T. Kemp (Eds.), *Evidence-based coaching: Theory, Research*


### Table 1.

**ACT Coding scheme for career counselor skills, underlying micro-behaviors, and reliabilities**

<table>
<thead>
<tr>
<th>SKILLS and micro-behaviors</th>
<th>ICC</th>
<th>Classification</th>
<th>M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SOCIO-EMOTIONAL SKILLS (+) (Mean)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contact</td>
<td>.75</td>
<td>excellent</td>
<td></td>
</tr>
<tr>
<td>Paraphrasing</td>
<td>.93</td>
<td>excellent</td>
<td>0.77 (0.9)</td>
</tr>
<tr>
<td>Summaries of client statements</td>
<td>.35</td>
<td>poor</td>
<td>2.71 (2.51)</td>
</tr>
<tr>
<td>Humor</td>
<td>.73</td>
<td>good</td>
<td>3.83 (4.71)</td>
</tr>
<tr>
<td>Bonding</td>
<td>.81</td>
<td>excellent</td>
<td>6.78 (6.6)</td>
</tr>
<tr>
<td>Addressing client’s feelings</td>
<td>.71</td>
<td>good</td>
<td>2.84 (3.07)</td>
</tr>
<tr>
<td>Appreciation</td>
<td>.77</td>
<td>good</td>
<td>7.99 (6.52)</td>
</tr>
<tr>
<td><strong>SOCIO-EMOTIONAL SKILLS (-) (Mean)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interruption of the client</td>
<td>.94</td>
<td>excellent</td>
<td>3.09 (3.74)</td>
</tr>
<tr>
<td>Disruptive behavior</td>
<td>1.00</td>
<td>excellent</td>
<td>0.09 (0.25)</td>
</tr>
<tr>
<td>Self-promotive behavior (i.e., self-praise)</td>
<td>1.00</td>
<td>excellent</td>
<td>0.13 (0.72)</td>
</tr>
<tr>
<td>Disagreement / criticism</td>
<td>.65</td>
<td>good</td>
<td>0.88 (1.12)</td>
</tr>
<tr>
<td><strong>PROCEDURAL SKILLS (+) (Mean)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open question</td>
<td>.98</td>
<td>excellent</td>
<td>36.2 (16.46)</td>
</tr>
<tr>
<td>Goal orientation</td>
<td>.06</td>
<td>poor</td>
<td>1.35 (1.95)</td>
</tr>
<tr>
<td>Clarifying own contributions</td>
<td>.48</td>
<td>fair</td>
<td>10.55 (10.81)</td>
</tr>
<tr>
<td>Procedural suggestion</td>
<td>.93</td>
<td>excellent</td>
<td>43.62 (19.17)</td>
</tr>
<tr>
<td>Prioritizing</td>
<td>.63</td>
<td>good</td>
<td>0.94 (1.37)</td>
</tr>
<tr>
<td>Time management</td>
<td>.78</td>
<td>excellent</td>
<td>0.87 (1.29)</td>
</tr>
<tr>
<td>Visualization with Media</td>
<td>.76</td>
<td>excellent</td>
<td>8.68 (8.86)</td>
</tr>
<tr>
<td><strong>PROCEDURAL SKILLS (-) (Mean)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Closed question</td>
<td>.94</td>
<td>excellent</td>
<td>35.5 (14.19)</td>
</tr>
<tr>
<td>Pointing out a problem / suggesting or explaining a solution</td>
<td>.85</td>
<td>excellent</td>
<td>2.52 (3.31)</td>
</tr>
<tr>
<td>Losing train of thought and running off topic</td>
<td>.72</td>
<td>good</td>
<td>2.39 (4.9)</td>
</tr>
</tbody>
</table>

*Note.* Code frequencies are standardized to counts per hour. *a* = categorization based on Cicchetti (1994) for ICCs, (+) = functional, (-) = dysfunctional
Table 2.

*Construct (Convergent and Discriminant) Validity for Counselors’ Micro-Behaviors Measured with Two Conceptually Related Coding Instruments*

<table>
<thead>
<tr>
<th>MITI-closed question</th>
<th>MITI-open question</th>
<th>MITI-simple reflection</th>
<th>MITI-complex reflection</th>
<th>MITI-MI non-adherent behavior</th>
<th>MITI-MI adherent behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT-closed question</td>
<td>.55**</td>
<td>.12</td>
<td>.03</td>
<td>.02</td>
<td>.09</td>
</tr>
<tr>
<td>ACT-open question</td>
<td>.03</td>
<td>.56**</td>
<td>-.02</td>
<td>.19</td>
<td>-.19</td>
</tr>
<tr>
<td>ACT-paraphrase</td>
<td>-.07</td>
<td>.13</td>
<td>.25†</td>
<td>.16</td>
<td>.17</td>
</tr>
<tr>
<td>ACT-summaries</td>
<td>-.11</td>
<td>.12</td>
<td>-.03</td>
<td>.18</td>
<td>-.19</td>
</tr>
<tr>
<td>ACT-interrupt</td>
<td>.02</td>
<td>-.10</td>
<td>.24</td>
<td>.06</td>
<td>.54**</td>
</tr>
<tr>
<td>ACT-appreciation</td>
<td>.05</td>
<td>.00</td>
<td>.16</td>
<td>.25</td>
<td>.05</td>
</tr>
</tbody>
</table>

*Note. N = 53. †p < .10, *p < .05, **p < .01 (two-tailed).*
Table 3.

**Multiple Hierarchical Regression Using MITI and ACT Measures for Working Alliance Subscale Goals and Tasks**

<table>
<thead>
<tr>
<th></th>
<th>Working alliance (goals)</th>
<th></th>
<th>Working alliance (tasks)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F$</td>
<td>$R^2$</td>
<td>$\Delta R^2$</td>
<td>$B$</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session duration</td>
<td>.27</td>
<td>.01</td>
<td>.01</td>
<td>-.00</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session duration</td>
<td>7.98*</td>
<td>.33</td>
<td>.32**</td>
<td>-.01</td>
</tr>
<tr>
<td>MITI(+)</td>
<td>.04</td>
<td>.01</td>
<td>.44**</td>
<td></td>
</tr>
<tr>
<td>MITI(-)</td>
<td>-.06</td>
<td>.02</td>
<td>-.41**</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Session duration</td>
<td>6.09*</td>
<td>.39</td>
<td>.07†</td>
<td>-.01</td>
</tr>
<tr>
<td>(covariate)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MITI(+)</td>
<td>.04</td>
<td>.01</td>
<td>.46**</td>
<td></td>
</tr>
<tr>
<td>MITI(-)</td>
<td>-.03</td>
<td>.02</td>
<td>-.23</td>
<td></td>
</tr>
<tr>
<td>ACT procedural(+)</td>
<td>.00</td>
<td>.00</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>ACT procedural(-)</td>
<td>-.01</td>
<td>.01</td>
<td>-.33*</td>
<td></td>
</tr>
</tbody>
</table>

*Note. N = 53. MITI = Motivational Interviewing Treatment Integrity; MITI(+) = MITI functional skills (open question, simple and complex reflection, MI adherent behavior); MITI(-) = MITI dysfunctional skills (closed question, MI non-adherent behavior); ACT procedural(+) = ACT functional procedural skills; ACT procedural(-) = ACT dysfunctional procedural skills; †$p < .10$, $p < .05$, **$p < .01$ (two-tailed).