

Syracuse University

**SURFACE**

---

College Research Center

David B. Falk College of Sport and Human  
Dynamics

---

March 2012

## Alliance and termination status in couple therapy: A comparison of methods for assessing discrepancies

Suzanne Bartle-Haring

*Ohio State University - Main Campus*, [rgangamm@syr.edu](mailto:rgangamm@syr.edu)

Tatiana Glebova

*Alliant International University*

Rashmi Gangamma

*Ohio State University - Main Campus*, [rgangamm@syr.edu](mailto:rgangamm@syr.edu)

Erika Grafsky

*Ohio State University - Main Campus*

Robin Delaney

*Ohio State University - Main Campus*

Follow this and additional works at: <https://surface.syr.edu/researchcenter>

 Part of the [Marriage and Family Therapy and Counseling Commons](#)

---

### Recommended Citation

Bartle-Haring, Suzanne; Glebova, Tatiana; Gangamma, Rashmi; Grafsky, Erika; and Delaney, Robin, "Alliance and termination status in couple therapy: A comparison of methods for assessing discrepancies" (2012). *College Research Center*. 10.

<https://surface.syr.edu/researchcenter/10>

This Article is brought to you for free and open access by the David B. Falk College of Sport and Human Dynamics at SURFACE. It has been accepted for inclusion in College Research Center by an authorized administrator of SURFACE. For more information, please contact [surface@syr.edu](mailto:surface@syr.edu).

This article was downloaded by: [Syracuse University Library]

On: 26 February 2013, At: 10:01

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Psychotherapy Research

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tpsr20>

### Alliance and termination status in couple therapy: A comparison of methods for assessing discrepancies

Suzanne Bartle-Haring<sup>a</sup>, Tatiana Glebova<sup>b</sup>, Rashmi Gangamma<sup>a</sup>, Erika Grafsky<sup>a</sup> & Robin Ostrom Delaney<sup>a</sup>

<sup>a</sup> Human Development and Family Science, The Ohio State University, Columbus, OH, USA

<sup>b</sup> Marriage and Family Therapy, Alliant International University, Sacramento, USA

Version of record first published: 05 Apr 2012.

To cite this article: Suzanne Bartle-Haring, Tatiana Glebova, Rashmi Gangamma, Erika Grafsky & Robin Ostrom Delaney (2012): Alliance and termination status in couple therapy: A comparison of methods for assessing discrepancies, *Psychotherapy Research*, 22:5, 502-514

To link to this article: <http://dx.doi.org/10.1080/10503307.2012.676985>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Alliance and termination status in couple therapy: A comparison of methods for assessing discrepancies

SUZANNE BARTLE-HARING<sup>1\*</sup>, TATIANA GLEBOVA<sup>2</sup>, RASHMI GANGAMMA<sup>1</sup>,  
ERIKA GRAFSKY<sup>1</sup>, & ROBIN OSTROM DELANEY<sup>1</sup>

<sup>1</sup>Human Development and Family Science, The Ohio State University, Columbus, OH, USA & <sup>2</sup>Marriage and Family Therapy, Alliant International University, Sacramento, USA

(Received 6 May 2010; revised 7 March 2012; accepted 14 March 2012)

### Abstract

Much of the empirical data available about therapeutic alliance and its relationship to termination status come from individual psychotherapies. We know less about therapeutic alliance in couple therapy. A unique characteristic of alliance in couple or family therapy is the possibility of discrepancies in alliance between system members. In this study we sought to demonstrate three statistical techniques: standard deviations, the intraclass correlation to assess discrepancies in alliance over time during the initial stage of couple therapy, and the use of these various measures to predict termination status using a sample of 72 couples from a university-based training clinic. Differences in partners' alliances operationalized either as categorical or continuous variables but when analyzed separately at each time point were not predictive of termination status. When multilevel modeling was used, a difference in the way the discrepancies changed over a period of time was related to termination status.

**Keywords:** split alliance; couple therapy; discrepancies in alliance; termination; drop-out

The objective of this study was to demonstrate three methodological approaches to assess discrepancies in alliance in couples, and use these measures of discrepancy in alliance to predict termination status when measuring alliance multiple times during the initial stage of couple therapy. Therapeutic alliance has been identified as a substantial common factor in treatment continuance and success across a variety of modalities and clinical problems (e.g., Horvath, 2001; Kazdin, Marciano, & Whitley, 2005; Knobloch-Fedders, Pinsof, & Mann, 2007; Martin, Garske, & Davis, 2000). The vast majority of research on therapeutic alliance has been based within an individual psychotherapy framework, and there remains a paucity of research on this concept from a Couple and Family Therapy perspective (CFT). Developing alliances in a triad rather than a dyad is a more complex process. The results of the few studies available indicate that alliance in CFT develops in specific ways (Garfield, 2004; Johnson & Greenberg, 1985; Knerr et al., 2009) and presents unique challenges such as discrepancies in alliance (e.g., Muniz de la Pena, Friedlander, & Escudero,

2009; Pinsof & Catherall, 1986; Robbins, Turner, Alexander, & Perez, 2003). A discrepancy in alliance occurs when one family member perceives more or less alliance with the therapist than another family member. Just how large of a discrepancy matters is one of the questions being addressed in this project.

There is emerging evidence of the role of discrepancies in alliance in couple and family therapy. Recent studies of family therapy with drug-using adolescents showed that similarities in alliance across family members were significant predictors of retention in treatment while overall level of family alliance did not always predict retention (Robbins et al., 2003, 2006, 2008). These studies have highlighted a critical role of similarity in alliance for drop-out across racial /ethnic groups (Hispanic and White, non-Hispanic) and across two family therapy models (brief strategic family therapy and functional family therapy). Another study of discrepancies in alliances in a sample of US families ( $n = 29$ ) and Spanish families ( $n = 21$ ) and couples ( $n = 16$ ) found that discrepancies in alliance did not invariably predict premature termination, though families with

Correspondence concerning this article should be addressed to: Suzanne Bartle-Haring, The Ohio State University, Human Development and Family Science, 151B Campbell Hall, 1787 Neil Ave, Columbus, 43210, USA. Email: sbartle@che.osu.edu

relatively more severe differences in alliance dropped out of treatment most often (Muniz de la Pena et al., 2009).

We used three separate analyses using couple as the unit of analysis to investigate whether discrepancies in alliance between partners in couple therapy would be predictive of termination status: standard deviations of the difference, the intraclass correlation, and the intraclass correlation over time. In order to do these analyses one very important decision needed to be made. That was how to define termination status. Termination statuses have been operationalized in the literature in various ways (e.g., Masi, Miller & Olson, 2003), again using individual therapy for the most part. Definitions of premature termination or dropout include but are not limited to failure to attend a certain number of sessions, failure to attend the last scheduled appointment, therapist judgment, or combination of therapist judgment and number of sessions (e.g., Sharf, 2007). Accordingly, definitions of completers vary as well. Because the therapists who participated in this study were in training and used different treatment approaches it was meaningless to use a particular number of sessions as a criterion for completion. Therapist judgment as agreement on termination appeared to be a more relevant criterion for this study. There also appears to be no consensus among researchers on definition or inclusion of “no show” (Masi et al., 2003). It is our contention that there is something different about an individual client who “no shows,” versus couple and/or family clients. In essence, for a couple or family to “no show” all members of the client system have to agree to simply not show up. This seems to us to be very different from an individual who “forgets” an appointment or simply doesn’t show up. Therefore, in the clinic in which these data come, therapists note on the termination form whether the clients ended with agreement with the therapist, ended treatment against therapist recommendation or “no showed.”

Terminated with agreement typically means that the clients and the therapists agreed that the goals of treatment had been met. Termination without agreement typically means that the clients decided to leave treatment against the therapist’s recommendation, but at least the couple let the therapist know that they would not be returning. The no-show status is reserved for couples who do not come to their next scheduled appointment and the therapist is then unable to reach the clients again either because the clients do not return phone calls, or they cannot be reached for other reasons. All three termination statuses have clients who come for one or many more sessions. For the purposes of this project we elected to keep these three categories with the caveat

that those in the “left without agreement” category could have included different types of clients (i.e., those who felt they got what they wanted and those who felt the treatment was no longer useful). Since the therapists in this study were in training and new to therapy in general, they may not be as successful in determining when clients are “finished.” Although there is no research evidence to suggest that new therapists are less able to determine when clients have met their goals, it may be the case that new therapists are more “idealistic” about couple relationships and hope to see interactions between couples that may not be realistic for all couples, thus they keep their couple clients coming beyond a time when the couples themselves would say their goals have been met.

The next question to be addressed for the project is how much of a difference in alliance between couple members makes a difference. When other researchers have investigated discrepancies in alliances they have elected to use the standard deviation of difference scores as a marker of difference (e.g., Heatherington & Friedlander, 1990; Knobloch-Fedders et al., 2007; Muniz de la Pena et al., 2009). Researchers have found that discrepancies in alliances of one standard deviation occurred in 32% to 40% of couple cases, and a more than two standard deviation difference occurred in 6% to 14% of couple cases (Coupland & Serovich, 1999; Heatherington & Friedlander, 1990; Knobloch-Fedders, Pinsof, & Mann, 2007; Mamodhousen, Wright, Tremblay, & Poitras-Wright, 2005). Using difference scores is statistically problematic because they can have low reliability, which can increase both Type I and Type II errors (Edwards, 2001). That is, when subtracting scores on a particular measure, the analyst is subtracting both “true” score and error. Thus, with a raw difference score, it is unclear whether the difference in scores is due to the difference in measurement error or the difference in “true” score. The other issues concern where the critical point of discrepancy is. Generally researchers have arbitrarily selected a one standard deviation of the mean difference, and/or a two standard deviation of the mean difference to imply a critical difference in alliance between partners. Edwards (2001) suggests that arbitrarily truncating continuous data leads to a loss of information and a reduction in explained variance. Peters and Van Voorhis (1940) provided an example using a one-half standard deviation to create three groups and demonstrated that the variance explained by the trichotomized score was approximately 26% less than that explained by the continuous difference score. So, by arbitrarily setting a “critical” difference, and then categorizing dyads into groups who reach this critical

difference and dyads who do not, we lose the variance within the categories, and we are able to explain less variance in an outcome variable, possibly deciding that the effect size is small when in fact it may be larger. Thus, using a continuous score for differences in alliance may provide more predictive power for termination status, than categorizing groups based on standard deviations.

Another problem with difference scores is that they only include the difference in scores and not the level of the variable of interest. In other words, if we only know that there is a discrepancy in alliance but we do not know how strong (or weak) the alliance is, we may not have the complete picture. It may be that the system of the family members and the therapist can tolerate a discrepancy in alliance if the alliance is relatively high, while a discrepancy may not be tolerated (the clients will drop-out) if the alliance is relatively low.

In the emerging field of dyadic data analysis (Kenny, Kashy & Cook, 2006), many of the issues surrounding the use of difference scores have been resolved to some extent by creating dyadic indices of dissimilarity or similarity. Results can vary depending on which index is selected; however, Kenny et al. (2006) recommend selecting an index based on the question of interest. They suggest that if the level of the construct is the most relevant, which is the case in understanding discrepancies in alliance as well as the level of alliance in the therapeutic relationship, then an intraclass correlation (ICC) or some type of dissimilarity measure such as a distance should be used. It should be noted, however, that computing a dyadic index implies an idiographic analysis in that the index is computed for each couple separately. This is in contrast to a nomothetic approach which would provide a degree of similarity or difference in the sample, rather than for each dyad (Kenny et al., 2006). When the ICC is computed idiographically, rather than a “central” tendency in the sample, we get an index of similarity for each couple, which then can be used as an independent variable or as a dependent variable (see Appendix 1 for formula). When we use the typical standard deviation technique, we are comparing the couple with the central tendency of the sample, rather than individually or idiographically. With small samples, and with an eye to clinical work, it seems more important to understand unique qualities of dyads rather than how they compare to some central tendency which may or may not be useful when it comes to an individual couple. In this study we compare the ICC and the more traditional standard deviations of the difference groupings in their ability to predict termination status.

A positive ICC is an assessment of the shared variance between partners, or how much they agree about the construct of interest (Kenny et al., 2006). Lower ICCs reflect less shared variance between partners and thus less agreement or more discrepancy in scores. Theoretically, a negative ICC is not possible, but, unfortunately, negative ICCs do occur. Kenny et al. (2006) suggest that this be interpreted to mean that the members of the dyad have relatively low scores, or are even less alike than a randomly matched pair. Taylor (2010) suggests that a negative ICC indicates that the true ICC is very low.

Another issue when investigating the predictive ability of discrepancies in alliance between couple members is the time of measurement. Typically, alliance has been measured once or twice during the treatment (e.g., Knobloch-Fedders et al., 2007). Some researchers suggest that more longitudinal studies measuring the alliance at different points of time may be required for better understanding of therapeutic change in couple and family therapy (Anker, Owen, Duncan, & Sparks, 2010; Bourgeois, Sabourin, & Wright, 1990; Knobloch-Fedders et al., 2007). This may be especially important for understanding the role of discrepancies in alliance. It has been shown in family therapy that there was an increase in disagreement between adolescent and mother about the alliance with the therapist in those who dropped out compared with those who completed treatment (Robbins et al., 2008). Changes in discrepancies in alliance were detected in studies of couple therapy but the relationship between those changes and therapeutic process were not investigated quantitatively (Knobloch-Fedders et al., 2007; Muniz de la Pena et al., 2009). In Muniz de la Pena et al.’s (2009) study it was found that the majority of “split” alliances occurred early in treatment, during the third session. Symonds and Horvath (2004) suggested that the lack of correlation between alliance and outcome for couples who disagreed about the strength of the alliance early in treatment may be explained by the fact that some couples were able to come to an agreement during therapy and achieved positive outcomes and some couples did not overcome their disagreement about alliance and had poorer outcomes. Thus, it seems important to explore discrepancies in alliance over time. We hypothesize that there is a relationship between termination status and changes in discrepancies in alliance, specifically, the more couple members agree about the alliance with the therapist over time, the more likely they will terminate with agreement.

## Method

The objective of this study was to demonstrate three techniques to assess discrepancies in alliance using the couple as the unit of analysis, and use these measures of discrepancy in alliance to predict termination status when measuring alliance multiple times during the initial stage of couple therapy. First, we used a categorical variable to group those couples in which their difference was 1 or 2 standard deviations from the mean difference between couples. Second, we used the ICC of therapeutic alliance for the couple members. The ICC is usually not correlated with the variables used in its calculation (Maguire, 1999), thus we can use the ICC to assess differences along with levels of the variables of interest (alliance from both partners' perspective). Third, we used alliance scores and ICCs of alliance scores between couple members over time to assess whether increases or decreases in the discrepancy in alliance scores might be predictive of termination status. Our goal was to highlight the differences in assessing discrepancies in alliance that could affect measurement issues and conclusions about discrepancies in alliance depending on the strategies used.

## Participants

The convenience sample for this project comes from a larger study of clients receiving treatment at an on-campus clinic operated by a small Couple and Family Therapy (CFT) PhD program ( $n = 195$ ). We excluded couples who were continuing treatment at the end of the study ( $n = 34$ ), those who dropped out after the first session ( $n = 63$  or 32%), and six same-sex couples. This resulted in the sample of 72 heterosexual couples; the male partners in the relationships were on average 31.6 ( $SD = 8.32$ ) years old, and the female partners were 29.6 ( $SD = 7.71$ ) years old. Even though the study was open to all couples, only six same-sex couples sought treatment during the time of data collection. Same-sex couples are indistinguishable dyads (there is no way to distinguish who is partner 1 or partner 2 in statistical terms, it would be merely an arbitrary designation), while the heterosexual couples are distinguishable by sex. Since dyadic data analysis becomes more complex with indistinguishable dyads (Kenny et al., 2006) we elected to remove the six same-sex couples from the sample.

Using same-sex couples in the analysis would also have confounded the issue of gender when it comes to alliance in psychotherapy. In a recent meta-analysis, Friedlander, Escudero, Heatherington, and Diamond (2011) reported that the male partner's alliance scores were a stronger predictor of

outcome for the couple than the female partner's. Anker et al. (2010) also demonstrated this with a dyadic data analysis technique (an Actor, Partner Interdependence Model), which in essence takes into account the other partner's alliance score when examining the relationship between the actor's alliance and the actor's termination status. Thus, the couple is the unit of analysis, but the partner is distinguished by their sex. In both of these research projects it appears that sex of the partner matters when it comes to alliance and termination status, thus, we decided that using only heterosexual couples was best for these analyses.

Of the heterosexual couples remaining, 55% were married for the first time, 7.5% were remarried and 21% were cohabiting at the time they participated in the study. Seventy-four percent of the sample was Caucasian, with 10.8% African American, 4.1% Latino, and 2.5% Asian. Seventy-seven percent of the sample made less than \$50,000 annually, with 24.1% between \$20,000 and \$29,000 annually. When asked during their initial phone call to schedule an appointment at the clinic, 80% of these clients reported a couple issue as their presenting concern including communication, marital problems, intimacy problems, contemplating divorce, problems with conflict, extramarital affairs, etc. The other categories of presenting concerns included individually focused problems (7%) such as depression, anxiety, substance abuse or parenting issues (10.1%).

Treatment at the clinic during the time of this data collection (2003–2006) was provided by 13 student therapists who were seeking their doctoral degrees. The PhD program in which these therapists were enrolled is small, admitting one to four students each year. Thus, our number of therapists for the study is also small. All therapists had already earned a master's degree in marriage and family therapy or a related field. Ten of the therapists were female. The average age was 28.69 years ( $SD = 6.47$ ) with a range from 25 to 44. Since the study data were collected over a 2½-year time span, therapists had between one and 19 cases within this data set. Of the cases they had, they could have had clients early in their training as well as later in their training depending on when they entered the program while the study was ongoing. Four of the therapists who treated the bulk of the cases in this data set ( $n = 57$ ) were female, with one being African American, and one being an international student from India; the other two were Caucasian.

In this training program students are encouraged to select a theory or integrate a number of theories from the theories of change within CFT. The therapists in training within the group of four who

treated the bulk of the clients had selected variations of Contextual Family Therapy and Bowen Family Systems Therapy. All therapists were supervised during the time that they treated clients in the clinic, with case report, video or live supervision. All couples in this data set were seen conjointly only and no limits were set on the number of sessions that clients could attend.

## Procedures

Clients were introduced to the study at the first session, during the time that they were consenting to treatment. The study was explained to them, including its voluntary nature, by their therapist, including that choosing not to participate would in no way impact the services they received at the clinic. If the clients consented to be in the study, they signed a separate consent form for participating in the study versus the consent for treatment form. Clients were told that the clinic was collecting information in order to improve treatment at the clinic and to get a “snapshot” of them for their therapist to view but that their therapists would not see their alliance ratings. Of the 195 couples asked to participate in the larger study, 26 declined (11.8% refusal rate). Based on an independent samples *t*-test, those who declined the research came for fewer sessions (4.70) than those who consented to the research (7.86;  $t = 3.29$ ;  $p < .01$ ). These are the only data available for those who did not participate in the research.

Clients were asked to complete an intake packet that included items on relationship satisfaction, depressive symptoms, locus of control, differentiation of self, and stress symptoms experienced around the problem that brought them to therapy. Clients who participated in the study also completed after-sessions questionnaires and were given a \$10 reduction in their fee. The clinic operates on a sliding fee scale based on income and number of dependents. The range of fees goes from \$10 to \$65.

Therapists were also briefed about the study, including its procedures, and were asked to voluntarily participate. Their participation implied consent. Therapists were told that the study was about factors that lead to successful termination. The protocol for the study was approved by the social and behavioral science IRB for the university in which the clinic operated. Therapists and clients were each asked to complete a therapeutic alliance scale after each session from session 2 through 6. To ensure anonymity, clients completed alliance rating forms after the therapist had completed the session and left the room and then deposited completed forms in a locked box near the exit. Client alliance ratings were then recorded by a clinic assistant to ensure

that therapists did not see the alliance scores of their clients. Although we did have alliance data for the therapists in this study, we did not use these data for the current study in part because of the small number of therapists, which would have made capturing therapists' effects difficult statistically, and also because therapists had more missing data than clients.

## Measures

Therapeutic alliance was assessed with the Working Alliance Inventory—Shortened Version (WAI-S; Tracey & Kokotovic, 1989), which is based on Bordin's three constructs of alliance—tasks, goals, and bonds (Bordin, 1994). The WAI-S is a 12-item self-report measure that uses a 7-point Likert-type scale (1 = never, 7 = always). The shortened version reveals a factor structure similar to the full-length version, as well as an acceptable internal consistency ranging from .90 to .92 for the client version; .98 for the total score; and test-retest reliability of .83 across a two-week period (Horvath & Greenberg, 1994; Tracey & Kokotovic, 1989). The WAI-S is made up of three sub-scales: development of bonds, agreement on goals, and agreement on tasks. The bond subscale ( $n = 4$ ) assesses the emotional bond of trust and attachment between client and therapist (e.g., “My therapist and I trust one another”). The goals subscale ( $n = 4$ ) assesses the degree of agreement concerning the overall goals of treatment (e.g., “My therapist and I are working toward goals that we both agree on”). The tasks subscale ( $n = 4$ ) assesses the degree of agreement concerning the tasks relevant for achieving these goals (e.g., “My therapist and I agree about the things I will need to do in therapy to help improve my situation”). The subscale scores can range from 4 to 28 and all three subscale scores can be combined for a total mean item score. Therefore, total scores can range from 1 to 7, with higher scores reflecting more positive ratings of working alliance. In the present study, we used the total scores since the subscales only had four items each. The internal consistency reliabilities for this sample for women were .81, .78 and .77 for sessions 2–4 respectively. The reliabilities for the male partners were .80, .81 and .82 for sessions 2–4 respectively.

We selected the WAI-S for several reasons. First, it was brief and we wanted to limit the number of items on the after-session questionnaire. Second, the clinic clientele includes individuals, couples, and families. We wanted to keep the distribution of after-session questionnaires as simple as possible, so we elected to use an “individually” oriented alliance scale so that all clients could complete it, rather than having

different versions for different client systems. This enabled us to compare the development of alliance for both individual clients and couple and family clients, which would not have been possible if we used different versions of the WAI or the better-known but much longer Pinsof and Catherall (1986) alliance scale for couple and family therapy clients. Unfortunately, Symonds and Horvath's WAI for couples was published after the initiation of the study from which these data come.

Clients completed the WAI-S after sessions 2 through 6. The major reason for collecting data after each session was to make the data collection procedure as simple as possible for the therapists. In a previous study at the same clinic when data were to be collected at sessions 3, 6 and 9, therapists invariably forgot to distribute the questionnaires. Also from previous research at the clinic we knew that on average clients came for six sessions, thus the decision was to collect the data over the first six sessions of therapy in order to provide the most data on the largest number of clients.

There were several ways that clients ended treatment. One way was to "no show" at some point after the first session. That is, if couples did not come for a scheduled appointment and were not seen again nor did they contact the therapist, they were considered part of the "no show" group. We included this group in the analysis and labeled them "no show" ( $n = 40$ ). On average this group came for 5.15 (2–20) sessions ( $SD = 3.74$ ). A second group that was included was one in which the clients let their therapist know that they would not be returning to treatment, contrary to the therapist's recommendation ( $n = 11$ ). This group was labeled "no agreement." On average this group came for 8.09 (2–30) sessions ( $SD = 9.62$ ). Finally, we had clients who ended treatment "successfully" through mutual agreement with their therapist that the goals for treatment had been met; this group was labeled "with agreement" ( $n = 21$ ). On average this group came for 8.46 (2–47) sessions ( $SD = 9.64$ ). In an ANOVA comparing number of sessions by termination status, the number of sessions attended among the three termination statuses differed significantly ( $F(2,1) = 3.27$ ;  $p < .05$ ).

## Results

### Missing Data

Missing data were replaced at the item level to create scale scores only when fewer than one-third of the items were missing for the WAI items. When fewer than one-third of the items for a particular instrument were missing (i.e., WAI), they were replaced

with the mean of that item. Cases in which all data were missing were not replaced in the analysis procedures. At session two, 82 wives had complete data and 11 had one to three items missing, 83 husbands had complete data and six had one to three items missing on the 12-item WAI. At session three, 69 wives and 71 husbands had complete data, while seven wives and four husbands had one to three items missing. At session four, 60 wives and 62 husbands had complete data and five wives and two husbands had one to three items missing. The final sample of 72 couples was created based on complete data for each session and termination status data. In the multilevel modeling procedure, the analysis allows for unequal numbers of cases at each time point, and weights the cases differently based on completeness.

### Termination Status using Standard Deviation

We examined differences of more than one and two standard deviations from the sample mean in partners' scores of therapeutic alliance. To create these scores we used the absolute value of the difference between the husband and wife on the WAI. We then used the standard deviation of the mean of that difference to create a group that was within one standard deviation of the mean difference, a group that was at one standard deviation from the mean but not two standard deviations of the mean, and a group who had a difference score that was two or more standard deviations from the mean difference score. We elected to create three groups rather than conducting the analyses several times in order to decrease the number of tests for the analyses, as well as to eliminate overlap among the groups. In previous work, groups were selected who were not one standard deviation from the average difference and then a group who was at least one standard deviation different. This group obviously includes others who are more than one standard deviation from the difference. Since, one of the objectives of this project was to show the "arbitrary" nature of these divisions, we elected to show whether the three groups created had any predictive ability. The numbers of couples having a one standard deviation difference and two standard deviation difference in partners' scores are presented in Table I by termination group.

Multinomial logistic regressions using both partners' therapeutic alliance scores at sessions 2, 3 or 4 and the categorical variable that represented their difference scores (not one, one, or two or more standard deviations) showed significant results for predicting termination status (dependent variable) depending on the session. The means and standard



Table I. Number of couples with differences in partners' WAI scores at one standard deviation but less than two standard deviations and with two standard deviations or more from the sample mean at sessions 2, 3 and 4 by termination status groups

	Agreement ( <i>n</i> = 21)		No agreement ( <i>n</i> = 11)		No show ( <i>n</i> = 40)	
	1 <i>SD</i>	2 <i>SD</i>	1 <i>SD</i>	2 <i>SD</i>	1 <i>SD</i>	2 <i>SD</i>
Session 2	3	0	3	1	4	3
Session 3	0	1	0	1	6	0
Session 4	1	0	3	2	3	0

deviations for the alliance scores for each session in each termination group can be found in Table II. It should be noted that Symonds and Horvath (2004) using the couple version of the WAI reported means of 6.0 and 6.04 for female and male partners at session 3 in their study, in which they eliminated the couples who dropped out of therapy and the therapists in the study were more experienced. The difference in the means was significant only in the "no agreement" group at session 4. There appeared to be no significant differences in the odds of "no show" or "no agreement" in comparison to "with agreement" when using session 2 or 3 data and the standard deviation of the difference group. When using session 4 data, the overall equation was significant ( $\chi^2(8) = 23.44$ ) and the pseudo  $R^2$  using Cox and Snell was .38, suggesting that the independent variables explained about 38% of the variance in termination status. (In multinomial logistic regression "pseudo  $R^2$ " is the proportional reduction in residual variance and can be thought of as an effect size.) The likelihood ratio tests for each of the independent variables were significant for male partner's and female partner's level of alliance, but the standard deviation group was not significant. In comparing the odds of being in a particular category, only the male partner's alliance score decreased the

odds of being in the "no agreement" group in comparison to the "no show" group (about .8 times) and the coefficient was negative. This suggests that the higher the male partner's alliance at session 4, the more likely it was that the couple would leave without agreement at some point in therapy in comparison to simply "no showing."

### Termination Status using the ICC

Table III provides the average ICC for each session for couples who ended with agreement, without agreement, and "no showed." The central tendency in the data would suggest that couples share very little variance in their perspectives of the alliance; however, in a one-way analysis of variance there were differences in mean ICC among the groups at session 4 ( $F(2,45) = 3.58, p < .05$ ), with those who terminated with agreement having a higher ICC than the other two groups. Kenny et al. (2006) suggest that an ICC of .45 or greater suggests consequential nonindependence. That is, when the ICC is .45 or greater the standard errors of statistical tests are biased and thus any significance tests are suspect. The average ICC in this case was quite low considering the number of negative ICC's calculated; however, as can be seen in the ranges of ICC in each group and at session, there were many couple clients who had at least a .45. This suggests that treating the sample of couples as individuals for statistical purposes would have been misguided.

In a series of multinomial logistic regressions using these intraclass correlatons we included each session's therapeutic alliance scores for both partners and the corresponding ICC (independent variables) to predict termination status (dependent variable). The reference group included those clients who left with agreement. None of the independent variables

Table II. Means (with range) and standard deviations of WAI scores for male and female partners by session for termination status categories

	Males		Females	
	Mean	<i>SD</i>	Mean	<i>SD</i>
<b>With agreement</b>				
Session 2	5.57 (3.58–6.83)	.94	5.69 (4.08–6.67)	.66
Session 3	5.45 (3.42–6.75)	1.00	5.54 (4.0–6.75)	.68
Session 4	5.70 (3.43–7.75)	.89	5.58 (4.13–6.9)	.72
<b>No agreement</b>				
Session 2	5.42 (3.75–6.75)	.87	5.64 (3.58–6.75)	.91
Session 3	5.40 (4.08–6.67)	.75	5.97 (5.00–7.00)	.62
Session 4	4.62 (1.33–6.58)	1.42	6.11 (5.33–7.00)	.52
<b>No show</b>				
Session 2	5.72 (3.33–7.00)	.80	5.91 (4.25–7.00)	.72
Session 3	5.85 (4.25–7.00)	.67	6.03 (4.00–7.00)	.844
Session 4	5.97 (4.08–7.00)	.70	6.10 (3.25–7.00)	.84

Table III. Comparison of mean ICCs (with ranges) for the therapeutic alliance scores of both partners at sessions 2, 3 and 4 among the termination status categories

	Mean	SD
With agreement		
ICC for session 2	-.18 (-.79:.50)	.41
ICC for session 3	-.22 (-.95:.95)	.55
ICC for session 4	.25 (-.71:.93)	.48
No agreement		
ICC for session 2	-.44 (-.93:.16)	.34
ICC for session 3	-.07 (-.98:.79)	.56
ICC for session 4	-.11 (-.52:.26)	.27
No show		
ICC for session 2	-.17 (-.98:.88)	.49
ICC for session 3	-.34 (-.95:.68)	.47
ICC for session 4	-.06 (-.71:.70)	.41 <sup>a</sup>

<sup>a</sup>One-way ANOVA  $F(2,45) = 3.58, p < .05$ .

significantly predicted group membership for termination status for session 2 or session 3 alliance scores. The equation was significant when it included session 4 data, and male partner's alliance was predictive of the odds of being in the "no agreement" versus in the "with agreement" group. The overall equation was significant ( $\chi^2(6) = 21.31; p < .01$ ). The pseudo  $R^2$  was .358; so about 36% of the variance in termination status could be explained by the independent variables in the model. In the likelihood ratio tests for the full equation, only male partner alliance scores had a significant effect ( $\chi^2(2) = 12.36; p < .01$ ). The 95% confidence interval of the odds of leaving without agreement ranged from .10 to .95 with the odds ratio estimated at .31. This suggests that when male partner's alliance score is higher taking into account the female partner's score, the odds of leaving without agreement decreases in comparison to the odds of leaving with agreement.

Using these two methods of determining discrepancies did not appear to be predictive of termination status. That is, when examining data from a single session, discrepancies in alliance between the members of the couple were not predictive of termination status, regardless of the way in which discrepancy was assessed (continuous vs. categorical variable). Only the level of alliance for the male partner was predictive of termination status, and only at session 4.

### Termination Status using the ICC Overtime

In the next analyses, we used hierarchical linear modeling to determine whether the intercepts and slopes of the ICC were different among the three termination status groups. In essence, this analysis showed whether changes in the level of agreement between the couple would be related to termination

status. Since we had very few therapists ( $n = 13$ ), testing for therapists' effects was difficult statistically. We created a three-level model with the ICCs for the three sessions at level 1 (cases over time), the case at level 2, and therapist at level 3. Thus we had a model in which ICCs over time were nested within case, which were nested within therapists. We tested an unconditional intercept and slope model (we had no predictors at level 2 or 3) in order to estimate the amount of variance attributable to case effects and therapist effects. In this unconditional model, therapists' effects contributed about 3% of the overall variance in ICC scores over time, while client effects contributed the remainder. Also, the chi-square statistics for the random effects for the intercept and for the slope at the therapist level were not significantly different than zero (intercept  $\chi^2(11) = 14.05; p = .229$ ; and slope  $\chi^2(11) = 10.89; p < .50$ ). Given this, we elected to use a two-level model, with the limitation that therapists' effects were not controlled.

To test this model, we first calculated the "unconditional" model, or the model with an intercept and a slope for the ICC along with time varying covariates for male partner's alliance, and female partner's alliance. The results of these analyses can be seen in Table IV. The intercept and slope were significantly different from zero, while only the intercept had significant random variance. Neither partner's alliance scores were predictive of the ICC at baseline (as Maquire (1999) suggested the ICC is unrelated to the scores from which it is calculated). This suggests that on average the intercept or initial ICC was .25 but varied significantly among the couples. The slope was positive, suggesting that on average the ICC increased over the three sessions. That is, overall the couples tended to agree more about alliance by session 4. We then added dummy coded terms for termination status with ending "with agreement" as the reference category. The coefficient for "no show" was significant for the slope; those in the "no show" group had a lower slope score in comparison to the "with agreement" group. This relationship can be seen in Figure 1. This suggests that couples in the "no show" group reported a lower rate of change in ICC over sessions 2, 3, and 4 in comparison to the "with agreement" group.

It should be noted that the above analysis tests for a relationship between termination status as an "independent" variable and changes in the ICC or similarities in alliance scores between partners as the "dependent" variable. We examined the trajectories of the ICC in different groups, which allowed us to investigate whether there were differences among the termination status group, thus suggesting the

Table IV. Results from hierarchical linear modeling analysis of the ICC across sessions 2 through 4

	Unconditional model		Final model	
	Coefficient	<i>t</i> -value	Coefficient	<i>t</i> -value
Intercept	-.25	-4.54*	-.25	-2.66*
No show			.033	.275
No agreement			-.100	-.619
Male alliance	.001	.621	.002	.459
No show			-.002	-.288
No agreement			.001	.266
Female alliance	.002	.268	-.011	-.481
No show			.016	.614
No agreement			.009	.345
Slope	.11	2.88*	.21	2.85*
No show			-.178	-1.97*
No agreement			-.034	-.340
Random effects				
	Variance	$\chi^2(56 df)$	Variance	$\chi^2(54 df)$
Intercept	.067	76.03*	.079	77.37*
Slope	.002	57.72	.004	54.46
# of parameters	4		4	
Deviance	256.42		287.23	

\**p* < .05

relationship between termination status and differences in the trajectories of ICCs across early sessions of therapy.

To summarize, in comparing these different methods of assessing discrepancies and then using those discrepancies over time, it appears that changes in the ICC over time are related to termination status, but, within this analysis, level of alliance was not related to termination status. So, if we view alliance as developing over time and assess discrepancies in alliance over time, we are better able to determine

termination status. When we assess alliance at a single session and use that single session's discrepancy, it does not predict termination status, but the level of the male partner's alliance does.

## Discussion

In this study we sought to demonstrate three techniques to assess discrepancies in alliance over time during the initial stage of couple therapy, and to use these measures to predict termination status. Our aim was to highlight the differences in the conclusions that can be drawn about discrepancies in alliance depending on the strategies used. Standard deviation groupings at the three sessions were not predictive of termination status. However, when using this strategy, the male partners' alliance score at session 4 was predictive of termination status. The ICC at each of the three sessions was also not predictive of termination. Using this strategy, however, also resulted in male partners' alliance scores at session 4 being predictive of termination status. These results confirm the results of Symonds and Horvath (2004) and others (cf., Anker et al. 2010; Friedlander et al., 2011) in finding that the male partner's alliance score is more predictive of the outcomes of therapy than the female partner's. Finally, when using the ICC scores across time in a multi-level model, the pattern of change in ICCs was related to termination status. As couple members increasingly agreed about their alliance with the therapist over sessions 2 through 4 controlling for

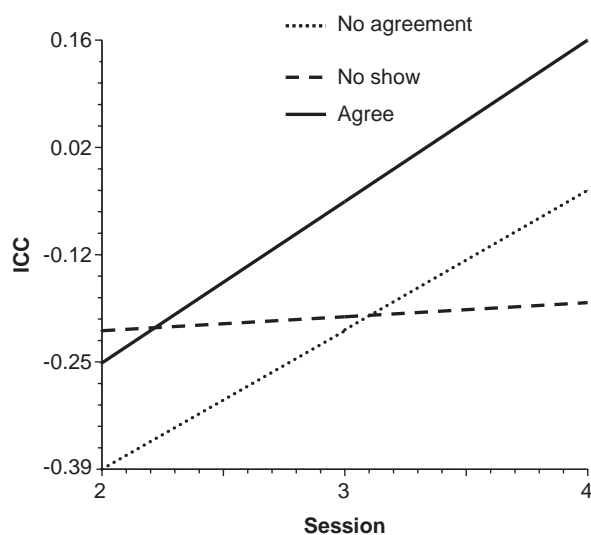


Figure 1. The ICC across sessions for those who ended with agreement, those who "no showed," and those who ended without agreement.

the level of alliance at each session, the more likely it was that they ended treatment with agreement.

From the results of this project it is clear that the method of assessment of difference makes a difference. Differences in partners' alliances operationalized either as categorical (standard deviation groupings) or continuous variables (the ICC) but analyzed separately at each time point were not predictive of termination status. However, analysis of the ICC over time appeared to predict termination status. Those couples who ended their treatment with mutual agreement with their therapist that the goals for treatment had been met increased their agreement about the alliance with the therapist more so than those who dropped out by not showing for a scheduled appointment at some point in the treatment. The results of this project suggest that couples who perceive greater differences in alliance by session 4 are more likely to drop out of treatment. Anker et al. (2010) suggested that very few studies have investigated change in alliance over time because most studies do not assess alliance more than twice. Anker et al. (2010), using three alliance assessments, found that couples whose alliance scores were high early on and increased over time were more likely to be in a group of clients in which both members of the couple experienced clinically significant change.

The findings from our study support literature from individual therapy that suggests that establishing an alliance between session 3 and 5 is essential (Horvath & Symonds, 1991). It also may suggest that it is not just the level of alliance that is important in couple therapy but also the couples' perceptions of similar alliances with the therapist that is also important for ending treatment with agreement. This study finding supports the notion in couple and family therapy literature that discrepancies in alliances change over time (Knobloch-Fedders et al., 2007; Muniz de la Pena et al., 2009), at least during the initial stage of treatment. Moreover, that change in couple agreement on alliance was related to termination status. The results confirm Symonds and Horvath's (2004) suggested explanation of the lack of correlation between alliance and outcome for couples who disagreed about the strength of the alliance early in treatment. This finding is also in accordance with the findings in family therapy that there was a greater increase over time in mother-adolescent discrepancies in alliance in drop-out cases compared with completers (Robbins et al., 2008).

When multilevel modeling was used, we see that there is a difference in the way the discrepancies change over a period of time that could point to a better understanding of process in therapy. Couples who belonged to the "with agreement" group

showed a greater rate of increase in their agreement about the alliance. Evaluating discrepancies in alliance as a continuous, dyadic variable over time seems to shed a different light on its significance to termination status than merely treating it as a categorical variable. This is important given that alliances formed between couples and therapists are known to be more complex (Friedlander et al., 2011; Glebova et al., 2010). Even utilizing an individually oriented instrument for measuring alliance, this study was able to detect dyadic processes in couple therapy by using longitudinal multilevel analysis. It is noteworthy that this study's results are in agreement with the recent findings from the study which used the System for Observing Family Therapy Alliances (SOFTA-o; Friedlander, Escudero, & Heatherington, 2006a; Friedlander et al., 2006b), the instrument developed to assess the systemic nature of family therapy alliance. That study concluded that alliance-related behaviors of individual family members played a greater role early in therapy (session 3), while behaviors indicative of a strong shared sense of purpose (within-system alliance) were more important later on (session 6).

There are limitations in our study that need to be mentioned here. First and foremost, the sample used was one of convenience, including the trainee therapists and the clients, and small especially given the breakdown into termination status groups. Clients were assigned to therapists based on which therapist was "up" next and needed a new client. Thus, replication of this project with more experienced therapists and across different settings is important before trying to place the findings of the current study in context. The findings from the current project cannot be generalized. The instrument used to assess alliance was meant for individual therapy and did not assess the partners' perspectives of the other's alliance with the therapist, or the partners' perception of their sense of being a couple allied with the therapist. These perspectives are addressed in newly developed alliance assessments that are also brief (i.e., SOFTA-s). The data for this project were collected before some of these instruments were published. Evidence suggests that early in treatment the individual alliance with the therapist is important but later, the unit's sense of a common purpose in treatment becomes more important (Friedlander et al., 2011). This may be part of the reason that the ICCs' change was associated with "ending with agreement" in this study. It could be that those couples who begin to agree that *they* can trust the therapist and that *they* and the therapist agree about what needs to be done to reach the ultimate goal of the therapy make better progress

because they begin to see themselves more as a unit than individuals in a couple.

In addition, we were unable to take into account any other factors, such as therapist characteristics, because we had very few therapists, although the three-level model results suggested that therapist effects contributed about 3% to the overall variance. The therapists were trainees with varying levels of experience and were for the most part young and new to providing therapy. This would certainly impact alliance and these results cannot be generalized to more experienced therapists. We also asked clients to complete the same questionnaire after the first six sessions. The questions were repeated each time. Some of the variance in the scores on alliance may be due to repeated exposure to the same instrument. Anker et al. (2010) conducted a similar study using an assessment of alliance after each session attended. They also mentioned that asking clients to complete the same assessment frequently can lead to response bias. In the Anker et al. study, clients completed the alliance measures in the presence of their therapist. In this study, the alliance scale was completed after the therapist left the room, and clients were told that their therapists would not see their answers. Our results are quite similar to those of Anker et al. (2010) so it continues to be important to further research the area of alliance, and how best to assess it and when.

Further, it should be noted that this study did not ask couples to discuss their perspectives with each other. We have no information as to whether or not the members of these couples actually discussed how much they liked or disliked their therapist, or whether they agreed with the therapist about the tasks and goals of their treatment. This may have affected their report of alliance. Another methodological limitation is the approach to the operationalization of termination status. The “no agreement” and “no show” groups might have included different types of clients—those who believed that their goals were met even if their therapist disagreed and those who thought that the provided treatment did not meet their goals at all. We recommend more detailed data collection of termination decision for future research to separate those groups. Also cases of dropouts after the first session were not included in the analyses due to a lack of available data on alliance. Investigation of patterns of change in alliance discrepancies from the very first session could be a focus of future research.

Our goal here was to highlight the differences in assessing discrepancies in alliance that could affect measurement issues in future studies on alliance in couple and family therapy. We could have done this with data created in a Monte Carlo study, just to

show the differences in data analysis techniques. We elected to use an existing data set with several limitations as have been discussed above. It is possible that with data over longer periods of time, the complexity of alliance will emerge more completely. However, despite the limitations, the study contributes to research on the impact of differences in alliance in couple therapy.

Clinically it is important for therapists to assess and monitor therapeutic alliance, and to be able to detect severe discrepancies in alliance between system members. How can we translate these two research methodologies into clinical practice utilization? The standard deviation method of categorization of split alliance is based on the central tendency in the particular sample, and how normally distributed alliance scores actually are. Clinically it may mean that the therapist assesses a couple’s discrepancies in alliance by comparing it with discrepancies of other couples she or he has been working with. This assessment may require a lot of experience and can be problematic. One study (Muniz de la Pena et al., 2009) which investigated congruence between self-reported and observed splits using the standard deviation method of categorization reported modest findings that only 20% and 27% (in the Spanish and the US samples, respectively) of the observed splits were congruent in terms of severity with the self-reported splits. On the other hand, the ICC methodology is grounded in idiographic analysis of the similarity index which is computed for each couple separately. Clinically this corresponds to the therapist’s assessment of couple alliance based on the unique qualities of each case, including alliance dynamics over time. It has been suggested that clients’ and therapists’ perceptions reflect the cumulative process of therapy (Friedlander, Lambert, Escudero, & Cragun, 2008). Thus, the ICC methodology may be more sensitive to that process. However, the ICC methodology does not provide categorization of splits into “severe” or “mild” groups, which can be useful for considering “critical” cases. Therefore, both methodologies have their own merits and limitations. Obviously, more research investigating congruence between clients’ and therapists’ perceptions of severity of discrepancies in alliance using both methodologies is needed. Using the ICC could result in finding “cut-offs” for severe discrepancies vs. not so severe which would be based more on the data, and not on whether or not the distribution of alliance follows a normal curve. As more studies are completed with the same measures of alliance these ranges of ICCs can be used to compare individual cases vs. other cases.

As couple and family therapists, we need to be aware of the multiple relationships we are forming

with couple and family members. It is probably obvious to most CFTs that we often have difficulty “joining” with each person in the therapeutic system in the same way or with the same strength. The results of this study suggest that these differences in alliance matter for treatment outcomes. That is, those who drop out of treatment were more likely to disagree about how allied they were toward their therapist, at least by session 4, in comparison to those who completed treatment. In future work it would be important to include the therapists’ perspective about how much they thought the members of the therapeutic system agreed about their alliance with them. If it is the case that the disagreement or split in alliance is crucial in determining whether clients will remain in treatment or drop out then talking about the “disagreement” between the members about the therapeutic relationship may be a way of intervening to keep the clients in therapy. Anker et al. (2010) and others have recommended that alliance be assessed at each session and used as feedback for the therapist to then discuss with clients.

On the other hand, the differences in alliance may be isomorphic to what is happening with the couple in general. That is, the process underlying the presenting concern which the couple brought to therapy may be about differing perspectives on issues that are crucial to the success of the relationship. Again, discussing this or intervening at this level may be just as useful as discussing their differing views about the therapeutic relationship.

## References

- Anker, M.G., Owen, J., Duncan, B.L., & Sparks, J.A. (2010). The alliance in couple therapy: Partner influence, early change, and alliance patterns in a naturalistic sample. *Journal of Consulting and Clinical Psychology, 78*, 635–645.
- Bordin, E.S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice, 16*, 252–260.
- Bordin, E.S. (1994). Theory and research on the therapeutic working alliance: New directions. In A.O. Horvath & L.S. Greenberg (Eds.), *The working alliance: Theory, research, and practice* (pp. 13–37). New York: Wiley.
- Bourgeois, L., Sabourin, S., & Wright, J. (1990). Predictive validity of therapeutic alliance in group marital therapy. *Journal of Consulting and Clinical Psychology, 58*(5), 608–613.
- Coupland, S. K., & Serovich, J. M. (1999). Effects of couples’ perceptions of genogram construction on therapeutic alliance and session impact: A growth curve analysis. *Contemporary Family Therapy, 21*, 551–572.
- Eames, V., & Roth, A. (2000). Patient attachment orientation and the early working alliance – a study of patient and therapist reports of alliance quality and ruptures. *Psychotherapy Research, 10*(4), 421–434.
- Edwards, J.R. (2001). Ten difference score myths. *Organizational Research Methods, 4*, 265–287.
- Friedlander, M.L., Escudero, V., & Heatherington, L. (2006a). *Therapeutic alliances in couple and family therapy: An empirically informed guide to practice*. Washington DC: American Psychological Association.
- Friedlander, M.L., Escudero, V., Heatherington, L., & Diamond, G.M. (2011). Alliance in couple and family therapy. *Psychotherapy, 48*, 25–33.
- Friedlander, M.L., Escudero, V., Horvath, A.O., Heatherington, L., Cabero, A., & Martens, M.P. (2006b). System for observing family therapy alliances: A tool for research and practice. *Journal of Counseling Psychology, 53*, 214–225.
- Friedlander, M.L., Lambert, J.E., Escudero, V., & Cragun, C. (2008). How do therapists enhance family alliances? Sequential analyses of therapist-client behavior in two contrasting cases. *Psychotherapy: Theory, Research, Practice, Training, 45*, 75–87.
- Garfield, R. (2004). The therapeutic alliance in couples therapy: Clinical considerations. *Family Process, 43*(4), 457–465.
- Glebova, T., Bartle-Haring, S., Gangamma, R., Knerr, M., Delaney, R.O., Meyer, K., Silverthorn, B., McDowell, T., Adkins, K., & Gafsky, E. (2010). Therapeutic alliance and progress in couple therapy: Multiple perspectives. *Journal of Family Therapy*. doi: 10.1111/j.1467-6427.2010.00503.x.
- Heatherington, L., & Friedlander, M.L. (1990). Applying task analysis to structural family therapy. *Journal of Family Psychology, 4*, 36–48.
- Horvath, A.O. (2001). The therapeutic alliance: Concepts, research, and training. *Australian Psychologist, 36*, 170–176.
- Horvath, A.O., & Greenberg, L. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36*, 223–233.
- Horvath, A.O., & Symonds, B.D. (1991). Relation between working alliance and outcome in psychotherapy: A meta-analysis. *Journal of Counseling Psychology, 38*, 139–149.
- Johnson, S., & Greenberg, L. (1985). Differential effects of experiential and problem-solving interventions in resolving marital conflict. *Journal of Consulting and Clinical Psychology, 53*(2), 175–84.
- Kazdin, A.E., Marciano, P.L., & Whitley, M.K. (2005). The therapeutic alliance in cognitive-behavioral treatment of children referred for oppositional, aggressive, and antisocial behavior. *Journal of Consulting and Clinical Psychology, 73*(4), 726–730.
- Kenny, D.A., Kashy, D.A., & Cook, W.C. (2006). *Dyadic data analysis*. New York: Guilford.
- Knerr, M., Bartle-Haring, S., McDowell, T., Adkins, K., Delaney, R.O., Gangamma, R., Glebova, T., Gafsky, E., & Meyer, K. (2009). The impact of initial factors on therapeutic alliance in individual and couples therapy. *Journal of Marital and Family Therapy*. doi: 10.1111/j.1752-0606.2009.00176.x.
- Knobloch-Fedders, L.M., Pinsof, W.M., & Mann, B.J. (2007). Therapeutic alliance and treatment progress in couple psychotherapy. *Journal of Marital and Family Therapy, 33*, 245–257.
- Maguire, M.C. (1999). Treating the dyad as the unit of analysis: A primer on three analytic approaches. *Journal of Marriage and Family, 61*, 213–223.
- Mamodhousen, S., Wright, J., Tremblay, N., & Poitras-Wright, H. (2005). Impact of marital and psychological distress on therapeutic alliance in couples undergoing couple therapy. *Journal of Marital and Family Therapy, 31*, 159–169.
- Martin, D.J., Garske, J.P., & Davis, K.M. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta-analytic review. *Journal of Consulting and Clinical Psychology, 68*, 438–450.
- Masi, M.V., Miller, R.B., & Olson, M.M. (2003). Differences in dropout rates among individual, couple, and family therapy clients. *Contemporary Family Therapy, 25*, 63–75.

- Muniz De La Pena, C., Friedlander, M., & Escudero, V. (2009). Frequency, severity, and evolution of split family alliances: How observable are they? *Psychotherapy Research*, 19, 133–142.
- Peters, C.C., & Van Voorhis, W.R. (1940). *Statistical procedures and their mathematical bases*. New York: McGraw-Hill.
- Pinsof, W.M. (1994). An integrative systems perspective on the therapeutic alliance. In A.O. Horvath & L.S. Greenberg (Eds.), *The working alliance: Theory, research and practice* (pp. 173–198). New York: John Wiley & Sons.
- Pinsof, W.M., & Catherall, D.R. (1986). The integrative psychotherapy alliance: Family, couples, and individual therapy scales. *Journal of Marital and Family Therapy*, 12, 137–151.
- Robbins, M.S., Liddle, H.A., Turner, C.W., Dakof, G.A., Alexander, J.F., & Kogan, S.M. (2006). Adolescent and parent therapeutic alliances as predictors of dropout in multidimensional family therapy. *Journal of Family Psychology*, 20, 108–116.
- Robbins, M.S., Mayorga, C.C., Mitraini, V.B., Szapocznik, J., Turner, C.W., & Alexander, J.F. (2008). Adolescent and parent alliances with therapists in brief strategic family therapy with drug-using Hispanic adolescents. *Journal of Marital and Family Therapy*, 34, 316–328.
- Robbins, M.S., Turner, C.W., Alexander, J.F., & Perez, G.A. (2003). Alliance and dropout in family therapy for adolescents with behavior problems: Individual and systemic effects. *Journal of Family Psychology*, 17, 534–544.
- Sharf, M. (2007). *Psychotherapy dropout: A meta-analytic review of premature termination*. Unpublished doctoral dissertation, Adelphi University.
- Symonds, D., & Horvath, A.O. (2004). Optimizing the alliance in couple therapy. *Family Process*, 43(4), 443–455.
- Taylor, P.J. (2010) An introduction to intraclass correlation that resolves some common confusions. Manuscript viewed on the internet. Programs in Science, Technology & Values, Critical & Creative Thinking, and Public Policy, University of Massachusetts, Boston, MA. peter.taylor@umb.edu.
- Tracey, T.J., & Kokotovic, A.M. (1989). Factor structure of the Working Alliance Inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 1, 207–210.

### Appendix 1

To calculate the ICC at each session for each couple we used the following equations:

1. First we calculated the grand mean for the three subscales of the working alliance scores for each couple for each session (sum of Males' bond, task, goal and females' bond, task and goal divided by 6).
2.  $SS_{\text{between}} = .5 \sum (\text{Male Score} + \text{Female Score} - (2 * \text{Grand Mean}))^2$
3.  $SS_{\text{within}} = .5 \sum (\text{Male Score} - \text{Female Score})^2$
4.  $MS_{\text{between}} = SS_{\text{between}} / (\text{degrees of freedom between, which is the number of scales (6) divided by 2 minus 1})$
5.  $MS_{\text{within}} = SS_{\text{within}} / (\text{degrees of freedom within, which is the number of scales divided by 2})$
6.  $ICC = \frac{MS_{\text{between}} - MS_{\text{within}}}{MS_{\text{between}} + MS_{\text{within}}}$