Shorter communication

Öst’s (2008) methodological comparison of clinical trials of acceptance and commitment therapy versus cognitive behavior therapy: Matching Apples with Oranges?

Brandon A. Gaudiano a,b, *

a Department of Psychiatry & Human Behavior, Alpert Medical School of Brown University, USA
b Psychosocial Research Program, Butler Hospital, 345 Blackstone Boulevard, Providence, RI 02906, USA

A R T I C L E   I N F O

Article history:
Received 17 February 2009
Received in revised form 10 June 2009
Accepted 30 July 2009

Keywords:
Acceptance and commitment therapy
Cognitive behavior therapy
Randomized controlled trials
Research methodology
Empirically supported treatments

A B S T R A C T

Öst (2008) recently compared the methodological rigor of studies of acceptance and commitment therapy (ACT) and traditional cognitive behavior therapy (CBT). He concluded that the ACT studies had more methodological deficiencies, and thus the treatment did not qualify as an “empirically supported treatment.” Although Öst noted several important limitations that should be carefully considered when evaluating early ACT research, his attempt to devise an empirical matching strategy by creating a comparison sample of CBT studies to bolster his conclusions was itself problematic. The samples were clearly mismatched in terms of the populations being treated, leading to differences in study design and methodology. Furthermore, reanalysis showed clear differences in grant support favoring CBT compared with ACT studies that were not reported in the original article. Given the actual mismatch between the samples, Öst’s methodological ratings are difficult to interpret and provide little useful information beyond what could already be gathered by a qualitative review of ACT study limitations. Such limitations are characteristic of the earlier randomized controlled trials of any emerging psychotherapeutic approach.

© 2009 Elsevier Ltd. All rights reserved.

Acceptance and commitment therapy (ACT) is a novel acceptance/mindfulness-based behavioral psychotherapy that has an emerging base of empirical support from clinical trials and supporting studies (Hayes, Luoma, Bond, Masuda, & Lillis, 2006). In recent writings, some critics have argued that ACT offers relatively minor variations compared with traditional CBT and thus may not warrant the widespread clinical and research attention that the treatment has been receiving as of late (Arch & Craske, 2008; Corrigan, 2001; Hofmann, 2008; Hofmann & Asmundson, 2008; Leahy, 2008; Velten, 2007). Öst (2008) recently presented an analysis of ACT and other “third wave” behavior therapies, providing both qualitative and quantitative critiques of their research base of currently published clinical trials. Although an independent methodological critique of ACT research is a welcome addition to the literature, some of the quantitative analyses that Öst conducted to support his conclusions had questionable validity, which are described below.

Öst’s (2008) methods and results

Of the 30 total “third wave” randomized controlled trials (RCTs) identified by Öst (2008), 13 were studies of ACT. He assessed the quality of each study using a methodology rating scale developed from a measure originally designed for posttraumatic stress disorder (PTSD) trials. Öst devised an unorthodox strategy for interpreting the methodological quality ratings for the ACT studies:

In order to assess whether the obtained methodology score is equal to or different from CBT studies, a “comparison” sample of studies was collected. For each of the third wave studies a “twin” CBT study published in the same journal the same, or ±1, year was retrieved. (p. 299)

However, Öst could not find a “twin” CBT study meeting these criteria in all 30 cases. Therefore, to complete the comparison

1 Öst (2008) also reported results for other specific “third wave” approaches, including dialectical behavior therapy (Linehan, 1993), functional analytic psychotherapy (Kohlenberg & Tsai, 1991), cognitive behavioral analysis system of psychotherapy (McCullough, 2000), and integrative behavioral couple therapy (Jacobson & Christensen, 1996). However, only Öst’s findings regarding ACT are discussed here.
sample, he decided to include 11 additional unmatched CBT studies taken from the following top-tier psychology journals instead: *Journal of Consulting and Clinical Psychology, Behavior Therapy*, and *Behaviour Research and Therapy*. In fact, 38% of the ACT sample could not be “matched” based on Öst’s original criterion. Öst reported that ACT studies had significantly lower total scores on the rating scale compared with CBT studies. Furthermore, Öst also analyzed each item of the rating scale separately and reported that ACT had significantly lower scores than CBT on the following items: reliability of the diagnosis, reliability and validity of outcome measures, checks for treatment adherence, control of concomitant treatments, representativeness of the sample, assignment to treatments, number of therapists, and therapist training/experience. Öst reported that ACT studies were not significantly less likely to be grant funded than CBT studies. Thus, he rejected grant funding as a possible explanation for the methodological differences observed.

Finally, Öst conducted a meta-analysis of the 13 ACT clinical trials and found that ACT had significantly greater effect size improvements compared with no treatment (ES = .96), treatment as usual (ES = .79), and active treatments (ES = .53), which were similar to the effect sizes reported in the earlier meta-analysis by Hayes et al. (2006). However, given the methodological limitations described above, Öst concluded that ACT did not meet formal criteria as an “empirically supported treatment” (Chambless & Ollendick, 2001).

### “Matching” Apples with Oranges?

First, it should be acknowledged that Öst (2008) provided a much needed and potentially informative methodological critique of the specific ACT studies published to date. Furthermore, Öst’s critique highlighted important methodological limitations that must be taken into account when attempting to generalize findings from these investigations. Clearly, there is room for improvement and a need for increased independent replication to confirm early positive findings from pilot studies. It should be noted that Hayes and colleagues have described similar methodological weaknesses in their previous reviews of ACT research (Hayes et al., 2006; Hayes, Masuda, Bissett, Luoma, & Guerrero, 2004; Hayes, Strosahl, & Wilson, 1999).²

Perhaps in an attempt to bolster the strength of his own arguments, Öst (2008) chose to devise an empirical strategy for testing whether ACT studies were objectively methodologically deficient compared with similar CBT studies. However, post hoc matching relies primarily on the validity of the choices and assumptions made during the matching process (Rubin, 1973). Unfortunately, Öst’s methods relied on the rather questionable assumption that ACT studies published around the same time period in the same journal as CBT studies should have similar methodologies and could be judged by the exact same standards. This sole criterion ignored the fact that ACT and CBT interventions are at very different stages of clinical trial testing, and thus ACT studies have historically had less funding to support this early research. In contrast, CBT clinical trials are in the most advanced stage of any psychotherapy outcome research to date.³

Table 1 shows the actual mismatch between ACT and CBT trials based on the clinical populations studied in these trials. All of the purportedly “matched” CBT trials were studies of emotional disorders, whereas the majority of the ACT studies were conducted in more difficult-to-treat and often treatment-resistant populations. In fact, a Fisher’s exact test of the percentage of non-depression/anxiety trials in the CBT (0%) versus ACT (62%) clinical trials analyzed was highly significant (p = .002).

It is not surprising that there would be differences between trials of ACT and CBT trials driven by factors such as: (1) the previously published efficacy data at the time the study was conducted; (2) the population being treated; (3) the nature of the intervention and its stage of development at the time; (4) the resources/funds available for conducting the study; (5) the appropriateness of the comparison group in relation to the currently available and empirically-supported treatments for the target population; and (6) the specific research aims of the study. A closer examination of Öst’s (2008) methodological ratings reveals potential bias against any newer psychotherapy in early clinical trial research, such as ACT. Most of the methodological differences reported by Öst could be accounted for by factors such as the amount of grant support/funds available to support the trial, the particular stage of treatment development at the time the study was conducted, and the specific population being studied.

As cases in point, I co-authored one of the studies (Gaudiano & Herbert, 2006) included in the ACT sample and one of the studies (Herbert et al., 2005) included in the CBT sample created by Öst. The Gaudiano and Herbert study tested a brief ACT intervention for inpatients diagnosed with DSM-IV psychotic-spectrum disorders. This trial was only the second study ever attempted using ACT in this population. Our findings were consistent with those obtained from an earlier independent trial (Bach & Hayes, 2002), and extended these initial findings by showing additional benefits in terms of clinically significant symptom reductions. I also was a co-author on a study of cognitive-behavioral group therapy (CBGT) for social phobia in which we examined the incremental efficacy of social skills training when added to the standard treatment protocol (Herbert et al., 2005). The fact that CBGT had been used for decades and was the gold-standard treatment for social phobia at the time of our investigation required that we employ a more sophisticated design to test our hypotheses and contribute to the scientific literature. In contrast, the ACT study was clearly labeled as a “pilot” study in the title, and was considered a preliminary investigation to support the need for future larger scale studies to verify initial positive findings. Based on this early promising research, a larger grant-funded study is now being conducted by a group of researchers in Australia on ACT versus supportive therapy for psychosis.

Öst’s (2008) comparison of CBT and ACT studies attempted to draw sweeping conclusions about study methodology while ignoring context. For example, he also criticized ACT studies for using treatment as usual comparison groups. However, in the case

---

² Also see: [http://www.contextualpsychology.org/state_of_the_act_evidence](http://www.contextualpsychology.org/state_of_the_act_evidence).

³ Some may point out that Hayes has reported that work on ACT began as early as the 1980’s (Hayes, 2008; Zettle, 2005). Furthermore, the first small clinical trial comparing ACT versus cognitive therapy for depression was conducted in 1986 (Zettle & Hayes, 1986). However, one must be careful to distinguish between the theoretical develop of ACT and the history of its RCTs. Most of the RCTs reviewed by Öst were conducted after the treatment protocol was formally available to independent researchers with the publication of the ACT book in 1999 (Hayes et al., 1999). Thus, the theoretical work was started years earlier, and other experimental and basic research studies have been conducted along the way to build the foundation for the clinical application of ACT as tested in formal RCTs, which came later.

<table>
<thead>
<tr>
<th>Clinical population</th>
<th>Acceptance and commitment therapy</th>
<th>Cognitive behavior therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Anxiety or stress</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Chronic medical conditions</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Psychosis</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Pain</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Addiction</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Personality disorder</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
of psychotherapy for psychotic disorders, treatment as usual has been the most frequently used comparison group in past CBT studies (Gaudiano, 2005). This is because psychosocial treatments for this population are universally adjunctive in nature and provided in conjunction with pharmacotherapy. Thus, it was reasonable to use a treatment as usual comparison group in our preliminary ACT study for psychosis in hospitalized patients (Gaudiano & Herbert, 2006). Öst “matched” our ACT study with a CBT study for acute stress disorder (Bryant et al., 2006), which would require a very different type of study design. Another study included in the sample tested ACT for institutionalized patients with treatment-resistant epilepsy (Lundgren, Dahl, Melin, & Kies, 2006). Öst matched the ACT epilepsy trial with a CBT social phobia trial (Clark et al., 2006). Such comparisons make little sense. Treatment as usual would not have been an appropriate comparison group for a social phobia trial, given that this is an established frontline and primary treatment for the disorder. The attempt by Öst to create a matching variable based on year of publication in a particular journal alone missed crucial factors that necessarily informed the various methodological decisions made by investigators. These choices are frequently dictated by factors specific to the investigation and should be debated or defended within that context.

Öst (2008) also reported that ACT studies were less likely to use samples defined by DSM diagnostic criteria. He states: “This is difficult to understand, since there does not seem to be an ideological resistance to diagnosing among ACT researchers” (p. 312). Öst’s statement implies that a study is not valid unless it investigates a diagnostic group included in the current version of the DSM. However, whether or not the sample is defined in terms of the DSM is largely irrelevant to the issue of appropriately describing the sample. Coming from the behavior analytic tradition, ACT does not subscribe to the assumption that the approach is only potentially useful for those who fit into a certain DSM category (Hayes et al., 1999). In other words, the issue is not whether ACT researchers use the DSM to diagnose the sample, as Öst suggests. Instead, the appropriateness of the use of the DSM for defining a sample should be based on the actual aims of the investigation.

Finally, it is unclear why Öst (2008) chose to focus his attention only on “horse race” trial results, in which two treatments are pitted against one another to look for differences principally in outcome. Given that most psychotherapies share components and many have been found to produce similar outcomes, many researchers have begun to advocate that more attention be paid to the actual mechanisms of action in effective treatments (Borkovec & Sibrava, 2005; Herbert & Gaudiano, 2005; Kazdin, 2005; Lohr, DeMaio, & McGlynn, 2003; Rosen & Davison, 2003). Unfortunately, this question has been slow to be investigated in traditional CBT (Gaudiano, 2005, 2008; Longmore & Worrell, 2007). In contrast, even early ACT studies have attempted to provide data on the treatment’s possible mechanisms of action (Hayes et al., 2006). Thus, Öst’s review pointed out weaknesses in some methodological areas, while ignoring the potential strengths in early ACT versus CBT research.

Reanalysis of grant funding

Öst (2008) rejected the idea that differences in grant funding could have accounted for the methodological differences between the CBT and ACT studies after finding no significant differences between the groups on this variable. Upon request, Öst provided his data for reanalysis to explore some of the potential confounds described above. I independently verified grant funding status for the ACT and CBT study samples. Öst reported that 46% (n = 6/13) of the ACT studies were supported by grants. One of these studies was incorrectly coded and not in actuality grant funded (i.e., Bach & Hayes, 2002). In addition, one of the CBT studies was incorrectly listed as being directly supported by a grant (i.e., Herbert et al., 2005), and another was not included as grant funded, but the primary author confirmed that it was (Barlow, Craske, Cerny, & Klosko, 1989) Fig. 1.

Originally, Öst (2008) conducted a statistical analysis of grant funding in the 13 ACT studies compared with the 13 “matched” CBT study subsample only. This use of the subsample of CBT studies led to low statistical power in the analysis, even though the differences appeared meaningful (77% for CBT vs 38% for ACT). Furthermore, this analysis using a subsample of the CBT studies was inconsistent because Öst used the full 30 study CBT sample when comparing ACT studies to CBT studies on methodological quality ratings. In the full CBT sample, 80% (n = 24) of studies were grant funded, compared with only 38% (n = 5) of studies in the ACT sample. A Fisher’s exact test showed that significantly more CBT studies were grant funded compared with ACT studies (p = .013). In addition, there was a moderately strong, positive correlation between grant funding and methodological quality as assessed by the total score in the ACT and CBT studies (r = .52, p = .001).

Öst also failed to assess differences in the actual funding amounts between the CBT and ACT studies. Review of the study funding data showed that only 1 of the 5 grant-supported ACT studies was supported by a large grant from the National Institutes of Health, whereas many of the CBT studies appeared to be supported by large grants (e.g., RO1s). I obtained estimates of the amount of funding for these studies either from publicly available data or directly from the principal investigators. Given the difficulty of collecting this data, only the 13 study CBT subsample was used in this analysis. In the two cases where data were unavailable, a conservative estimate was used based on amounts of similar size grants; in a third case with multiple projects totaling a known amount, a conservative portion for the published study was estimated. A total of $1,448,570 went to the 13 ACT studies or an average of $111,428 each. Only one of the five funded studies exceeded $150,000. A total of $6,438,157 went to the 13 CBT studies or $495,242 each, and 9 of the 10 funded studies exceeded $150,000. A Mann–Whitney U test showed a highly significant difference between the ACT and CBT studies in terms of amount of grant funding (Z = 2.64, p = .008). Although approximately 1.5 million might sound like a considerable amount of money to spend on early ACT trials, the amount of money supporting CBT trials was far greater and likely permitted the use of more advanced methodological procedures when compared with the ACT studies.

![Fig. 1. Grant funding for clinical trials of ACT (n = 13) vs CBT (n = 30).](image-url)
Concluding comments

Öst (2008) highlighted some important limitations in the currently published clinical trials of ACT that must be carefully considered. However, methodological evaluations do not occur in a vacuum as Öst’s analysis assumed, and attempts to create a matched CBT sample for comparison using a summary index of methodological quality should have been based on justifiable criteria. The above discussion demonstrates that Öst’s methodology did not meet these standards. Furthermore, Öst failed to fully investigate several important confounds in his analyses as they pertained to the ACT subsample of studies specifically.

One potentially fairer approach that Öst rejected was to compare early ACT studies with early CBT studies. Of course, this could have caused its own problems as these studies were conducted in different eras. There have been many advances in psychotherapy research over the years, including an increased emphasis on mechanisms of action, investigations of which have been present even in early ACT studies, but largely missing from CBT studies. In the end, ACT studies and CBT studies funded by similar grant mechanisms (e.g., National Institutes of Health R01 grants) are likely to have similar methodological features. However, at the time of the Öst review, there were not sufficient numbers of large, grant-funded ACT studies to make meaningful comparisons with similar CBT studies. Such studies are increasingly being conducted and this issue can be reevaluated when further evidence is available. Early studies of ACT have methodological limitations that will require independent replication with increasingly larger samples to confirm. This process is a natural part of psychotherapy research.

In his review, Öst (2008) concluded that ACT did not meet formal criteria as an empirically-supported treatment (EST) (Chambless & Ollendick, 2001). However, after the publication of Öst’s paper, the American Psychological Association’s Division 12 posted an updated list of ESTs. ACT for depression is now listed as having “moderately strong” empirical support. This is based on the evaluation that ACT meets the criteria for a “probably efficacious” treatment for depression, which is a level of empirical support that is arguably more applicable to treatments supported by early clinical trials.

Also since the publication of Öst’s (2008) review, another independent meta-analysis of ACT has been conducted. Powers, Zum Vorde Sive Vording, and Emmelkamp (2009) meta-analyzed 18 clinical trials of ACT and reported that ACT was significantly better than waiting lists/psychological placebos (ES = .68) and treatment as usual (ES = .42), but not more efficacious than established treatments (ES = .18). However, Levin and Hayes (in press) had the opportunity to reanalyze the Powers et al. data, and after correcting a number of problems (e.g., process or secondary variables mislabeled as primary outcomes) reported that that ACT was significantly more efficacious than established treatment comparisons in these clinical trials (ES = .27...49, depending on the studies included as using “established treatments”).

Perhaps most importantly, ACT proponents and ACT critics should be held to similarly high methodological standards. It is legitimate to criticize the shortcomings of ACT research. However, criticisms must be based on solid evidence and sound underlying reasoning. If we do not hold both camps to these standards, little productive dialogue will be produced. As new, promising treatment approaches are developed and tested, it is crucial that researchers carefully evaluate the outcome studies in their proper context, and apply the empirically supported treatment criteria as they were originally intended, and not in ways that serve to prematurely discount emerging evidence. Critics of ACT should be welcomed and then debated, as they provide opportunities for a spirited discussion of the issues that hopefully will highlight ways of continuing to advance the scientific status of ACT.

Acknowledgments

I would like to thank Lily Brown for her help in data collection. I thank Lars-Goran Öst for making his data available for reanalysis. Steve Hayes also collected some additional information from Ps and I thank him for sharing it with me.

The preparation of this manuscript was supported in part by a grant from the National Institute of Mental Health (MH076937).

References


See: http://www.psychology.sunysb.edu/eklonsky-division12/treatments/depression_acceptance.html.


Rosen, G. M., & Davison, G. C. (2003). Psychology should list empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behavior Modification, 27*, 300–312.


