The empirically supported status of acceptance and commitment therapy: An update

Matthew F. Smout, Louise Hayes, Paul W. B. Atkins, Jessica Klausen and James E. Duguid

Abstract

Acceptance and commitment therapy (ACT) is a transdiagnostic cognitive behavioural therapy that predominantly teaches clients acceptance and mindfulness skills, as well as values clarification and enactment skills. Australian treatment guideline providers have been cautious in recognising ACT as empirically supported. This article reviews evidence from randomised controlled trials published since Öst's review, and examines the extent to which the methodology of ACT research has improved since. Since 2008, ACT research has improved its use of adherence and competence monitoring. Good-quality studies could be considered to offer National Health and Medical Research Council Level II evidence for chronic pain, obsessive–compulsive disorder, and a subset of other anxiety disorders (panic disorder, social phobia, and generalised anxiety disorder). The majority of studies demonstrated that ACT significantly improved primary outcomes but used comparison conditions that did not rule out therapy-unspecific factors, including use of concurrent treatments, as explanations for the improvements. Recommendations for future ACT research are presented.

Key Points

1. On average, acceptance and commitment therapy (ACT) research methodology has improved its use of monitoring treatment adherence and competence since Öst’s (2008) review.
2. Use of treatment as usual unmatched for contact and unmonitored for competence, and unmonitored use of concurrent treatments are the primary factors preventing the attribution of better outcomes for ACT recipients to therapy-specific effects.
3. Good-quality studies have been published since Öst’s (2008) review supporting ACT’s efficacy in chronic pain, obsessive–compulsive disorder, and a subset of other anxiety disorders (panic disorder, social phobia, and generalised anxiety disorder).

Funding: None.
Conflict of interest: All authors are members of the APS ACT Interest Group.
guidelines, an updated review of recent ACT research literature seems warranted.

**Previous Reviews**

There have been three meta-analyses of ACT outcome studies, each with a different collection of studies and study classification system. Each found ACT superior to wait-list control (WTC), placebo, and treatment as usual (TAU) with average post-treatment effect sizes ranging from moderate (Powers, Zum Vording Sive Vording, & Emmelkamp, 2009) to strong (Hayes, Luoma, Bond, Masuda, & Lillis, 2006). Hayes et al. (2006) and Öst (2008) found ACT superior to active interventions with moderate post-treatment effect sizes, but Powers et al. (2009) found ACT no more effective than other empirically supported treatments because of differences in the classification of control conditions (see Levin & Hayes, 2009; Powers & Emmelkamp, 2009).

While Öst’s (2008) meta-analysis found ACT quantitatively superior to active treatments, he argued that none of the seven studies were methodologically sound enough to classify ACT as an empirically supported therapy, based in part on comparisons with CBT studies (see Gaudiano (2009) for criticisms of Öst’s methodology). The seven areas in which Öst rated ACT as consistently inferior to CBT were (1) either not using structured diagnostic interview schedules or not reporting how diagnoses were made; (2) including outcome measures with unknown psychometric properties; (3) providing insufficient description of TAU; (4) having only one clinician administer therapies in both conditions, thereby confounding the therapy and the therapist; (5) having treatments provided by therapists with relatively little experience in providing that treatment, treating that problem, or both; (6) failing to include therapy adherence and competence information or verifying participants’ perceptions of treatment credibility; and (7) failing to provide information about whether participants were receiving any other kinds of additional treatment. RCTs employing credible control groups were criticised for having small sample sizes, and those with credible sample sizes were criticised for high attrition and resulting small cell sizes. Our aim here is to review the methodology of ACT RCTs published since Öst’s review and assess the extent to which these studies have addressed the methodological concerns Öst raised.

**Method**

Inclusion criteria for the present review were studies (a) employing random assignment, (b) to ACT as one treatment arm, with (c) either a wait list or comparative treatment, (d) published in English and published from 2008. The Association for Contextual Behavioral Science, the professional association for ACT, was contacted and a list of recent RCTs was provided. Additionally, electronic databases (Ovid, Medline, EBSCO, PsychINFO, Web of Science, and SCOPUS) were searched using the terms “acceptance” and “randomized (or randomised)” in title. Thirty-three RCTs were retrieved by 31 January 2012, listed in Table 1.

For comparison with the studies reviewed in Öst (2008), we considered the methodological quality of post-2008 studies that were clinic-delivered, did not employ a WLC group, were published in English, and were not described as ACT training workshops (see asterisked studies in Table 1). Studies were then rated using Öst’s (2008) Psychotherapy Outcome Study Methodology Rating Form (POSMRF), which consists of 22 items (see Table 2) rated on a 3-point scale (0–2) providing brief descriptions of the design features characteristic of “poor,” “fair,” and “good” ratings. The acceptable inter-rater reliability reported by Öst was a mean kappa coefficient of 0.75. In this present review, the lead author rated every study, while three of the co-authors rated a third each. The first attempt at independent ratings yielded unacceptable kappas (below 0.5) for 10 of the scale items. The raters then discussed and reached consensus on ratings for two example studies before conducting a second set of independent ratings; however, this still did not achieve acceptable kappa ratings for 10 of the scale items. These difficulties may represent a weakness in the scale, in our use of it, or both, so ratings for these low-agreement items should be treated cautiously. Table 2 reports the results of the first attempt at ratings and the range of kappas for each scale item. Where raters differed in their rating, the mean of the ratings was used. Independent samples t-tests were calculated to compare studies reviewed by Öst (2008) and the studies since, applying Bonferroni correction for significance levels.

**Results**

**Methodology Rating Outcome Using the Öst Scale**

The means of our ratings across the set of ACT studies since 2008 corresponded to “fair” to “good” points of the scale for 12 of the 22 methodology items. The scale items we gave mean ratings well below “fair” were: reliability of diagnosis (most studies did not use interviews with trained interviewers), blind assessors (most studies did not report using or did not use assessors blind to treatment condition), and assessor training. Most studies did not report assessor training or
<table>
<thead>
<tr>
<th>Study</th>
<th>Disorder</th>
<th>Participants</th>
<th>Diagnostic System</th>
<th>Comparison</th>
<th>N at start</th>
<th>Attrition (%)</th>
<th>Cell size</th>
<th>% female</th>
<th>M age (weeks)</th>
<th>Tx sessions</th>
<th>Tx hours</th>
<th>FUp months</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lundgren et al. (2008)</td>
<td>Epilepsy</td>
<td>EEG Yoga</td>
<td></td>
<td>18</td>
<td>0</td>
<td>9</td>
<td>33</td>
<td>23.6</td>
<td>5</td>
<td>4</td>
<td>12</td>
<td>6, 12</td>
<td>I + G</td>
</tr>
<tr>
<td>Roemer, Orsillo, &amp; Salters-Pedneault (2008)</td>
<td>GAD</td>
<td>DSM-IV</td>
<td>Delayed Tx</td>
<td>31</td>
<td>13.3</td>
<td>13</td>
<td>71</td>
<td>33.6</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>3, 9</td>
<td>I</td>
</tr>
<tr>
<td>Varra, Hayes, Roget, &amp; Fisher (2008)</td>
<td>Drug and alcohol counsellors</td>
<td>NI</td>
<td>Education</td>
<td>60</td>
<td>0</td>
<td>29.5</td>
<td>58</td>
<td>53.7</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>G</td>
</tr>
<tr>
<td>Wickel, Ahlqvist, Bring, Melin, &amp; Olsson (2008)</td>
<td>Whiplash disorders</td>
<td>NI</td>
<td>WLC</td>
<td>22</td>
<td>0</td>
<td>10</td>
<td>76</td>
<td>51.5</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>4, 7</td>
<td>I</td>
</tr>
<tr>
<td>Li, Hayes, Bunting, &amp; Masuda (2009)</td>
<td>Attempting to lose weight</td>
<td>NI</td>
<td>WLC</td>
<td>84</td>
<td>0</td>
<td>42</td>
<td>90</td>
<td>50.8</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>G</td>
</tr>
<tr>
<td>Petersen &amp; Zettle (2009)</td>
<td>Depression in alcohol dependence</td>
<td>DSM-IV</td>
<td>TAU</td>
<td>29</td>
<td>0</td>
<td>12</td>
<td>50</td>
<td>37.8</td>
<td>3.2 ACT</td>
<td>5.2 ACT, 3.1 ACT</td>
<td>0</td>
<td>I</td>
<td></td>
</tr>
<tr>
<td>Tapper et al. (2009)</td>
<td>Woman attempting to lose weight</td>
<td>NI</td>
<td>Continue with normal diet</td>
<td>25.8</td>
<td>29.5</td>
<td>100</td>
<td>40.8</td>
<td>15</td>
<td>4</td>
<td>8</td>
<td>2</td>
<td>G</td>
<td></td>
</tr>
<tr>
<td>Wadsworth et al. (2009)</td>
<td>Paediatric pain</td>
<td>NI</td>
<td>Multidisciplinary Tx</td>
<td>32</td>
<td>6.3</td>
<td>14.5</td>
<td>78</td>
<td>14.8</td>
<td>12</td>
<td>12</td>
<td>14</td>
<td>3, 5, 8</td>
<td>I</td>
</tr>
<tr>
<td>Flaxman and Bond (2010a)</td>
<td>Worksite stress</td>
<td>NI</td>
<td>WLC</td>
<td>311</td>
<td>41.2</td>
<td>15.5</td>
<td>40</td>
<td>30.9</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>3</td>
<td>I</td>
</tr>
<tr>
<td>Flaxman and Bond (2010b)</td>
<td>Worksite stress</td>
<td>NI</td>
<td>SIT, WLC</td>
<td>107</td>
<td>13.5</td>
<td>22</td>
<td>70</td>
<td>39</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>G</td>
</tr>
<tr>
<td>Fledderus, Bohlmeijer, Smit, &amp; Wijers (2011)</td>
<td>Psychological distress</td>
<td>NI</td>
<td>WLC</td>
<td>18</td>
<td>14.7</td>
<td>40.5</td>
<td>82</td>
<td>49</td>
<td>NI</td>
<td>8</td>
<td>16</td>
<td>3</td>
<td>G</td>
</tr>
<tr>
<td>Hinton and Geynor (2010)</td>
<td>Psychological distress (uni students)</td>
<td>Questionnaire screens</td>
<td>WLC</td>
<td>22</td>
<td>20</td>
<td>11</td>
<td>73</td>
<td>20.1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>I</td>
</tr>
<tr>
<td>Johnston, Foster, Shannan, Starkey, and Johnson (2010)</td>
<td>Pain</td>
<td>NI</td>
<td>WLC</td>
<td>24</td>
<td>50</td>
<td>7</td>
<td>63</td>
<td>43</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>S</td>
</tr>
<tr>
<td>Smout et al. (2010)</td>
<td>Methamphetamine</td>
<td>DSM-IV</td>
<td>OBT</td>
<td>104</td>
<td>67.4</td>
<td>15.5</td>
<td>40</td>
<td>30.9</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>3</td>
<td>I</td>
</tr>
<tr>
<td>Tawilghe et al. (2010)</td>
<td>OCD</td>
<td>DSM-IV</td>
<td>PMR</td>
<td>79</td>
<td>12.9</td>
<td>34.5</td>
<td>61</td>
<td>37</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>3</td>
<td>I</td>
</tr>
<tr>
<td>Bohlmeijer, Fledderus, Roks, &amp; Peterse (2011)</td>
<td>Depression</td>
<td>NI</td>
<td>WLC</td>
<td>93</td>
<td>10.2</td>
<td>40.5</td>
<td>82</td>
<td>49</td>
<td>8</td>
<td>8</td>
<td>NI</td>
<td>3</td>
<td>G</td>
</tr>
<tr>
<td>Birnberg, Michaela, Hesser, &amp; Berglund (2011)</td>
<td>Stress (social workers)</td>
<td>NI</td>
<td>WLC</td>
<td>18</td>
<td>3.2</td>
<td>47</td>
<td>82</td>
<td>49</td>
<td>8</td>
<td>8</td>
<td>16</td>
<td>0</td>
<td>G</td>
</tr>
<tr>
<td>Brown et al. (2011)</td>
<td>Test anxiety</td>
<td>Screening test</td>
<td>CT</td>
<td>16</td>
<td>0</td>
<td>8</td>
<td>69</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>M = 3.4 weeks</td>
<td>G</td>
</tr>
<tr>
<td>Butryn, Forman, Hoffman, Shaw, and Juanosoto (2011)</td>
<td>Mild to moderate depression</td>
<td>Questionnaire screens</td>
<td>WLC</td>
<td>376</td>
<td>22.4</td>
<td>113.3</td>
<td>70</td>
<td>42.5</td>
<td>9</td>
<td>NA</td>
<td>M = 4</td>
<td>3</td>
<td>S</td>
</tr>
<tr>
<td>Fledderus, Bohlmeijer, Peterse, &amp; Schreurs (2012)</td>
<td>Adolescent depression</td>
<td>DSM-IV</td>
<td>TAU</td>
<td>38</td>
<td>13.6</td>
<td>15</td>
<td>71.1</td>
<td>15</td>
<td>NI</td>
<td>NI</td>
<td>M = 20.8</td>
<td>3</td>
<td>I</td>
</tr>
<tr>
<td>Muto, Hayes, &amp; Jeffcut (2011)</td>
<td>Japanese students</td>
<td>NI</td>
<td>WLC</td>
<td>10</td>
<td>25.7</td>
<td>30.5</td>
<td>63</td>
<td>23.6</td>
<td>8</td>
<td>NA</td>
<td>NI</td>
<td>2</td>
<td>S</td>
</tr>
<tr>
<td>Weiland et al. (2012)</td>
<td>Bariatric surgery patients</td>
<td>NI</td>
<td>TAU</td>
<td>39</td>
<td>6.3</td>
<td>16.5</td>
<td>89.7</td>
<td>43.1</td>
<td>8</td>
<td>8</td>
<td>6</td>
<td>0</td>
<td>I + S</td>
</tr>
<tr>
<td>Wetherall et al. (2011)</td>
<td>Pain</td>
<td>SCID</td>
<td>OBT</td>
<td>114</td>
<td>12.2</td>
<td>57</td>
<td>509</td>
<td>54.9</td>
<td>8</td>
<td>8</td>
<td>12</td>
<td>6</td>
<td>G</td>
</tr>
<tr>
<td>Pearson, Follette, &amp; Hayes (2012)</td>
<td>Body image</td>
<td>NI</td>
<td>WLC</td>
<td>73</td>
<td>30.1</td>
<td>25.5</td>
<td>100</td>
<td>43.4</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>2 weeks</td>
<td>G</td>
</tr>
<tr>
<td>Griffin et al. (2011)</td>
<td>Smoking</td>
<td>Screening test</td>
<td>Bupropion</td>
<td>303</td>
<td>35.8</td>
<td>106</td>
<td>58.7</td>
<td>46</td>
<td>10</td>
<td>20</td>
<td>NI</td>
<td>6, 12</td>
<td>I + G</td>
</tr>
<tr>
<td>Westin et al. (2011)</td>
<td>Tinnitus</td>
<td>Multiple criteria</td>
<td>Tinnitus Retraining Therapy</td>
<td>64</td>
<td>0</td>
<td>19.5</td>
<td>483</td>
<td>50.6</td>
<td>10</td>
<td>8.4</td>
<td>10.25</td>
<td>6, 18</td>
<td>I</td>
</tr>
<tr>
<td>Thorpe et al. (2011)</td>
<td>Pain</td>
<td>NI</td>
<td>WLC</td>
<td>115</td>
<td>36.4</td>
<td>27.5</td>
<td>644</td>
<td>46</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>6, 12</td>
<td>S</td>
</tr>
<tr>
<td>White et al. (2011)</td>
<td>Psychosis</td>
<td>ICD-10</td>
<td>TAU</td>
<td>27</td>
<td>0</td>
<td>12</td>
<td>22.2</td>
<td>34</td>
<td>11.5</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>I</td>
</tr>
<tr>
<td>Shawyer et al. (2012)</td>
<td>Psychosis</td>
<td>SCID</td>
<td>Befriending</td>
<td>44</td>
<td>8.3</td>
<td>19</td>
<td>44</td>
<td>39</td>
<td>15</td>
<td>15</td>
<td>12.5</td>
<td>6</td>
<td>I</td>
</tr>
<tr>
<td>Lozza et al. (2012)</td>
<td>Shame in addicts</td>
<td>NI</td>
<td>TAU</td>
<td>133</td>
<td>17.6</td>
<td>56.5</td>
<td>45.9</td>
<td>33.6</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>G</td>
</tr>
<tr>
<td>Hesser et al. (2012)</td>
<td>Tinnitus</td>
<td>Questionnaire screen</td>
<td>99</td>
<td>3</td>
<td>31.7</td>
<td>43</td>
<td>48.5</td>
<td>8</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>S</td>
<td></td>
</tr>
</tbody>
</table>

AR, applied relaxation; I, individual; G, group; NI, not included; PMR, progressive relaxation training; S, self-help; SIT, stress inoculation training; TAU, treatment as usual; WLC, wait-list control; N/A, not applicable.

* included in Table 2 ratings.
accuracy), power analysis, and control of concomitant treatments. The reference group of CBT studies in Öst (2008), by comparison, had mean ratings corresponding to the “fair” point of the scale or higher for diagnosis reliability and control of concomitant treatments.

We rated ACT studies since 2008 significantly higher than Öst (2008) rated studies before 2008 on inclusion of checks for treatment adherence and competence. Otherwise, ratings were not significantly different. There were trends towards improvement for reliability of diagnosis ($t(28) = 2.53, p = .02$) and number of therapists used ($t(28) = 3.33, p = .003$), two areas that Öst (2008) had previously identified as deficient.

### Qualitative Review

We now present the findings of studies of 17 clinician-delivered ACT for clinical populations involving comparison with TAU, placebo, or active control published from 2008. We evaluate the design features that impact whether the study can contribute to conclusions about the efficacy of ACT. While undergoing publication review, a further 10 RCTs have appeared in press, nine of which were available as online prepublication. Seven of these targeted clinical populations, six of which were clinician delivered. Although we could not include them in the previous ranking, we have attempted to provide the most recent publications in this qualitative review section.

#### Anxiety disorders

Twohig et al. (2010) randomised 79 adults with chronic ($M = 20.5$ years) OCD to ACT or progressive muscle relaxation (PMR). ACT was offered without exposure and response prevention (ERP) even though it would typically be combined. ACT produced greater reductions in observer-rated OCD symptoms than PMR (between-groups post-treatment: $d = 0.84$), maintained at 3-month follow-up, with a greater proportion demonstrating clinically significant change (ACT: 46%; PMR 18%), which is the average rate of recovery for CBT of OCD (Fisher & Wells, 2005). Participants who were at least mildly depressed at baseline showed greater depression improvement in ACT than PMR (between-groups follow-up

---

1Excludes Flaxman and Bond (2010b), Butryn et al. (2011), and Brown et al. (2011), which were included in quantitative methodology ratings but excluded from this clinical discussion.

### Table 2

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Clarity of sample description</td>
<td>1.23 (0.73)</td>
<td>1.26 (0.46)</td>
<td>−0.09 to 0.78</td>
</tr>
<tr>
<td>2. Severity/chronicity of disorder</td>
<td>1.31 (0.86)</td>
<td>1.03 (0.65)</td>
<td>−0.13 to 0.47</td>
</tr>
<tr>
<td>3. Representativeness of sample</td>
<td>1.08 (0.76)</td>
<td>1.47 (0.47)</td>
<td>N/A to −0.27</td>
</tr>
<tr>
<td>4. Reliability of the diagnosis</td>
<td>0.15 (0.38)</td>
<td>0.56 (0.48)</td>
<td>0.23 to 0.25</td>
</tr>
<tr>
<td>5. Specificity of the outcome measures</td>
<td>1.77 (0.60)</td>
<td>1.85 (0.37)</td>
<td>N/A to 0.60</td>
</tr>
<tr>
<td>6. Reliability and validity of outcome measures</td>
<td>1.54 (0.66)</td>
<td>1.76 (0.30)</td>
<td>N/A to 0.60</td>
</tr>
<tr>
<td>7. Use of blind evaluators</td>
<td>0.31 (0.48)</td>
<td>0.56 (0.66)</td>
<td>N/A to 0.55</td>
</tr>
<tr>
<td>8. Assessor training</td>
<td>0.31 (0.63)</td>
<td>0.41 (0.65)</td>
<td>N/A to 0.20</td>
</tr>
<tr>
<td>9. Assignment to treatment</td>
<td>0.85 (0.38)</td>
<td>1.26 (0.30)</td>
<td>−0.22 to 0.25</td>
</tr>
<tr>
<td>10. Design</td>
<td>1.23 (0.73)</td>
<td>1.44 (0.45)</td>
<td>0.25 to 1.0</td>
</tr>
<tr>
<td>11. Power analysis</td>
<td>0.00 (0.00)</td>
<td>0.47 (0.78)</td>
<td>0.47 to 1.0</td>
</tr>
<tr>
<td>12. Assessment points</td>
<td>0.92 (0.64)</td>
<td>0.88 (0.66)</td>
<td>0.47 to 1.0</td>
</tr>
<tr>
<td>13. Manualised, replicable, specific treatment programs</td>
<td>1.54 (0.66)</td>
<td>1.56 (0.59)</td>
<td>0.11 to 0.76</td>
</tr>
<tr>
<td>14. Number of therapists</td>
<td>0.23 (0.44)</td>
<td>0.82 (0.51)</td>
<td>0.24 to 0.73</td>
</tr>
<tr>
<td>15. Therapist training/expertise</td>
<td>0.69 (0.75)</td>
<td>0.91 (0.75)</td>
<td>0.25 to 0.61</td>
</tr>
<tr>
<td>16. Checks for treatment adherence*</td>
<td>0.15 (0.38)</td>
<td>0.94 (0.64)</td>
<td>0.06 to 0.35</td>
</tr>
<tr>
<td>17. Checks for therapist competence*</td>
<td>0.00 (0.00)</td>
<td>1.06 (0.64)</td>
<td>−0.02 to 0.35</td>
</tr>
<tr>
<td>18. Control of concomitant treatments</td>
<td>0.23 (0.60)</td>
<td>0.47 (0.65)</td>
<td>N/A to 0.22</td>
</tr>
<tr>
<td>19. Handling attrition</td>
<td>0.85 (0.80)</td>
<td>1.12 (0.44)</td>
<td>−0.11 to 0.22</td>
</tr>
<tr>
<td>20. Statistical analysis and presentation of results</td>
<td>1.69 (0.63)</td>
<td>1.97 (0.12)</td>
<td>N/A</td>
</tr>
<tr>
<td>21. Clinical significance</td>
<td>0.69 (0.75)</td>
<td>0.97 (0.58)</td>
<td>0.39 to 0.59</td>
</tr>
<tr>
<td>22. Equality of therapy hours</td>
<td>1.55 (0.82)</td>
<td>1.21 (0.88)</td>
<td>N/A to 0.75</td>
</tr>
</tbody>
</table>

*Independent t-test $p < .002$ (Bonferoni correction).

N/A, rating could not be calculated because at least one rater’s ratings did not vary for that item.
achieved with less-educated samples. The results suggest that ACT is a distinct protocol that can produce sustained symptomatic improvement in anxiety disorders among relatively well-educated, low-severity samples.

**Behavioural medicine**

Lundgren, Dahl, Yardi, and Melin (2008) randomised 18 adults living in India to either ACT or yoga. Both groups received two 1.5-hour individual sessions and two 3-hour group sessions over 5 weeks, with 1.5-hour booster sessions 6 and 12 months later. Both ACT ($d = 1.3$) and yoga ($d = 1.4$) reduced seizure index (quantity \times duration). ACT produced greater change in seizure index at 12 months, although this is likely because the yoga group had a significantly lower mean seizure index pretreatment. Quality-of-life results were inconsistent. ACT recipients improved on the World Health Quality of Life Brief Version (Amir et al., 2003, $d = 0.81$) while yoga recipients improved on the Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985, $d = 0.58$).

There was considerable overlap in the ACT and yoga protocols, providing a strong test of ACT, and the measures used were typical of studies of psychological therapy for epilepsy. However, given that both interventions were experimental for this population, a placebo or delayed control condition should have been included to exclude natural remission and regression to the mean as explanations for improvement.

Westin et al. (2011) randomised 64 normal-hearing adults with tinnitus (duration $M = 7.7$ years) without severe psychiatric disorder (77% \geq 1 co-morbid medical problem) to ACT, tinnitus retraining therapy (TRT) or WLC. ACT was delivered in up to 10 weekly 60-minute sessions (except session 2, 75 minutes, a total of 10.25 hours). TRT involved only one 2.5-hour consultation consisting of medical evaluation, sound-generator fitting, psychoeducation, and instruction in sound therapy. Participants were told to aim to wear the device for at least 8 hours per day. ACT recipients reported lower tinnitus handicap ($d = 1.04$), sleep problems ($d = 0.22$), and anxiety ($d = 0.80$) than WLC at post-test. Over the 18-month follow-up period, the ACT group reported lower tinnitus handicap ($d = 0.71$ post-test) and sleep problems than the TRT group, with no differences in quality of life, anxiety, or depression. A greater proportion of ACT recipients achieved reliable improvement (54.5%) than TRT (20%) in tinnitus handicap at 6-month follow-up but no difference in reliable deterioration (4.5% ACT, 10% TRT). High-end functioning in tinnitus handicap was achieved by 36% of the ACT group and by 10% of the TRT.

Although TRT is widely used, a Cochrane review found only one low-quality RCT, albeit demonstrating TRT to be superior to tinnitus masking (Phillips & McFerran, 2010).
While TRT constituted a greater amount of impersonal therapy, ACT provided a greater amount of personal contact, so it is unclear whether non-specific contact could also account fully, or in part, for the superior outcomes in the ACT group.

Rost, Wilson, Buchanan, Hildebrandt, and Mutch (2012) assigned 47 women with stage III or IV ovarian cancer to 12 × 1-hour individual sessions of ACT or a manualised CBT-based TAU over 4 months. ACT recipients reported significantly greater reductions in distress \(d = 0.89\), mental disengagement \(d = 3.49\), emotional control \(d = 6.11\), anxiety \(d = 1.26\), and depression \(d = 1.69\) and improvements in quality of life \(d = 1.35\) and acceptance \(d = 2.02\) than TAU. Seven participants from the ACT condition and five from TAU died after randomisation, before post-assessment. Changes in post-treatment distress were mediated by changes in mental disengagement and active planning after eight sessions. The study employed only a single therapist providing both treatments without measuring adherence, competence, or treatment credibility so the stability and generalisability of the findings is unknown, but the large effect sizes in favour of ACT suggest that the protocol was effective in improving quality of life in this group.

**Borderline personality disorder (BPD) pathology**

Morton, Snowdon, Gopold, and Guymer (2012) randomised 41 (93% female) individuals who met four or more diagnostic criteria for BPD to TAU (at least fortnightly contact for supportive counselling, medication management, and inpatient admissions or crisis contacts if needed) or TAU + ACT (12 × weekly 2-hour groups). ACT recipients reported significantly greater improvement in self-reported BPD symptoms \(d = 0.81\) with 29.4% achieving clinically significant change (compared with 0% in TAU), as well as improvements in anxiety symptoms \(d = 0.83\), hopelessness \(d = 0.91\), psychological flexibility \(d = 0.98\), emotion regulation skills \(d = 0.78\), mindfulness \(d = 0.80\), and fear of emotions \(d = 1.16\) but not depression or stress symptoms. The study had several methodological limitations, most notably low sample size, unmatched contact time between conditions, and no assessment of concomitant treatment. Nevertheless, the findings encourage more rigorous investigation.

**Depression**

Petersen and Zettle (2009) assigned 24 involuntary inpatients of a substance disorder clinic exhibiting depression (83% mood disorder not otherwise specified) to receive individual ACT or TAU sessions (unspecified, but typically 12-step counselling) in addition to other TAU (anti-depressant medication, group therapy, nightly Alcoholics Anonymous meetings, and other health education groups). Both conditions achieved equivalent reductions in observer-rated \(r^2 = 0.86\) and self-reported depression \(r^2 = 0.72\). ACT recipients required shorter lengths of stay than TAU (22.7 rather than 33.3 days, \(d = 0.97\)) and received less individual sessions of therapy. The study suggests that ACT might convey a cost-effectiveness advantage, but a larger sample with a lengthy follow-up, manualised control condition, and monitoring of treatment utilisation would be needed to have confidence in the finding.

Louise Hayes, Boyd, and Sewell (2011) randomised 38 depressed adolescents to outpatient ACT or TAU (manualised CBT-based). Post-treatment data were available for 79% of the sample, but only 32% of the sample provided 3-month follow-up data, with attrition worse in TAU. Self-reported depression improved in ACT but not in TAU, which deteriorated between post-treatment and follow-up (between-groups pretreatment to post-treatment: \(d = 0.38\); reliable improvement: ACT 58%, TAU 36%). Both groups improved during treatment on the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997). ACT but not TAU recipients continued to improve by 3-month follow-up. The study was limited by a small sample size, reliance on self-report measures, and high research attrition over a short follow-up period. Problematically, staff members provided both treatments but were not randomly allocated to condition, and there was no independent verification of treatment adherence and competence to rule out contamination between interventions or ensure the control condition constituted a worthwhile comparison group. However, given the degree of improvement was similar to other established psychotherapies (Watanabe, Hunot, Omori, Churchill, & Furukawa, 2007) and a potential treatment retention advantage, more rigorous replication seems warranted.

Finally, Folke, Parling, and Melin (in press) randomised 35 unemployed individuals on sick leave for depression \(M = 351\) consecutive sick leave days prior to treatment to TAU or TAU + ACT \(1 \times 60–90\) minutes individual + 5 × 120–180 minutes group sessions). TAU + ACT showed significantly greater improvements in self-reported depression \(d = 0.86\), general mental health \(d = 0.52\), and quality of life \(d = 0.71\) from pre-treatment to 18-month follow-up. There were no differences between groups in the proportions at follow-up who remained unemployed or on disability pensions (66.6% ACT, 62.5% TAU). Unfortunately, amount of contact in TAU and use of concomitant treatments were not evaluated, so it is unclear to what extent therapy-unspecific factors contributed to ACT recipients’ better outcomes.
Disordered eating

Juarascio, Forman, and Herbert (2010) analysed a subsample of 55 adults presenting to a university counselling centre who reported sub-threshold eating disorder pathology according to the Mini International Neuropsychiatric Interview (Sheehan et al., 1998) at intake. Participants were randomised to individually delivered ACT or cognitive therapy. Therapy was not manualised and the number of sessions not limited ($M = 12.7$). At post-treatment, ACT recipients evidenced greater reductions in eating pathology ($\eta^2_p = 0.08$). There were no significant group differences in self-reported depression, anxiety, or quality of life. Design strengths included the use of a large number of therapists ($n = 23$) and evidence that random recordings from the pool of cases from which this subsample was drawn were independently rated as competent and faithful to the assigned condition (Forman, Herbert, Moitra, Yeomans, & Geller, 2007). However, as a post hoc study there were numerous methodological shortcomings, notably the use of a single screening questionnaire to evaluate eating disorder pathology. The study suggests that ACT may constitute a distinct intervention, so a more rigorous investigation is warranted.

Up to 30% of bariatric surgery patients regain their weight within 24 months, so Weineland, Arvidsson, Kakoulidis, and Dahl (2012) explored whether ACT could reduce emotionally disordered eating in this population. Patients were assigned to either ACT (two face-to-face sessions plus six weekly Internet-delivered modules with 30-minute telephone support) or TAU (post-surgery appointments with the surgeon, nurse, and dietician, post-surgery guidelines, and follow-up telephone support as needed). The ACT group made greater reductions in self-reported shape concern ($\eta^2_p = 0.12$), weight concern ($\eta^2_p = 0.12$), subjective binge eating ($\eta^2_p = 0.19$), preoccupation with body shape ($\eta^2_p = 0.13$), and improvements in quality of life ($\eta^2_p = 0.13$) and acceptance of weight-related thoughts and feelings ($\eta^2_p = 0.18$). There were no differences in eating concern or restraint. While ACT recipients achieved large effect size improvements on three of the nine measures, contact time was not matched between groups so improvements cannot be conclusively attributed to ACT rather than non-specific contact. The results encourage further investigation with a well-specified active control matched for amount of contact, and follow-up periods incorporating objective measures of weight.

Pain

Wicksell, Melin, Lekander, and Olsson (2009) randomised 32 participants (aged 10–18) to ACT or a multidisciplinary treatment TAU (MDT) including amitriptyline for long-standing (duration $M = 32.4$ months) paediatric pain. Both groups improved over time on all primary measures except MDT did not improve general mental health. ACT produced significantly greater reductions in pain impairment beliefs than MDT considering all follow-up periods ($\eta^2_p = 0.23$) and greater improvements in mental health ($\eta^2_p = 0.15$) and pain interference ($\eta^2_p = 0.16$) considering only pre-test to post-test. ACT produced greater improvements than MDT on secondary outcomes of kinesiophobia ($\eta^2_p = 0.12$), pain intensity ($\eta^2_p = 0.13$), and pain-related discomfort ($\eta^2_p = 0.15$) throughout follow-ups.

The main limitations of the study were a small sample size which may limit the stability of the findings and no formal assessment of therapist competence in either condition, so it is unclear whether ACT or greater therapist ability or both contributed to the better outcomes. The number of sessions offered was flexible in each condition, and as a result, the MDT group received a greater number of sessions ($M = 22.8$) than the ACT group ($M = 13$), suggesting that a formal cost-effectiveness study is warranted. The improvements in this ACT group compare favourably in magnitude and breadth of outcomes to existing psychological treatments for paediatric pain (Palermo, Eccleston, Lewandowski, Williams, & Morley, 2010).

Wetherell et al. (2011) randomised 114 adults with chronic pain (duration $M = 15$ years) to either ACT or CBT. Both groups significantly improved pain interference from baseline to follow-up (ACT: $d = 0.35$, CBT: $d = 0.45$) and depression (ACT: $d = 0.80$; CBT: $d = 1.0$) and pain-related anxiety (ACT: $d = 1.0$; CBT: $d = 0.81$) but not pain severity, activity, or quality of life during treatment. Participants rated CBT as more credible, and ACT recipients reported greater satisfaction with treatment. Both groups made equivalent improvements in pain acceptance and perceived control of pain (but only the latter were correlated with pain interference).

The study’s main strengths were its sample size and inclusiveness (e.g., only 29.8% employed and 53.5% psychiatric co-morbidity) and independent measurement of treatment adherence. The study was limited to self-report measures where objective outcomes (e.g., return to work, number of sick days) would have been informative. Although results are typical for psychological therapies on chronic pain (Eccleston, Williams, & Morley, 2009), ACT produced weaker effects than in previous ACT for pain trials. The authors acknowledge that their protocol may have underemphasised the acceptance component and its group stand-alone delivery may also have reduced its potency.
Mo’temadi, Rezaiemaram, and Tavallaie (2012) randomised 30 Iranian females with chronic headache (M = 3.9 years) to medical TAU or ACT + TAU group (8 x 90-minute weekly sessions). Only ACT recipients improved by post-test on affective (d = 1.35), disability (d = 0.93), and distress (d = 2.54) but not sensory (d = 0.28) domains of pain intensity. There were several methodological limitations, most notably no control of concomitant treatment, no follow-up, and unmatched amount of contact between groups, so it is unclear whether the greater improvements can be primarily attributed to ACT.

Psychosis
White et al. (2011) randomised 27 adults with psychotic disorders (from casenotes: schizophrenia 48%; unspecified non-organic psychosis 26%; schizoaffective 11%; bipolar and psychosis 11%) to ACT + TAU or TAU (pharmacotherapy, including review by psychiatrist, case management and/or psychotherapy, and multidisciplinary input as needed). ACT recipients achieved greater reductions in blinded observer-rated negative symptoms (between group change d = 0.47) but not positive symptoms, and a lower proportion met “caseness” for depression (14% ACT versus 46% TAU), but not anxiety. There were fewer crisis contacts in the ACT group, but otherwise there was no information about how much TAU was received by each group. Given the importance of reducing negative psychotic symptoms, the results encourage the continued development of ACT protocols for this group.

Shawyer et al. (2012) randomised 43 adults with command hallucinations despite antipsychotic medication at therapeutic doses to Treatment of Resistant Command Hallucinations (TORCH; combining ACT, CBT, and motivational interviewing) or Befriending. The study was well designed, featuring blinded assessors, structured psychiatric assessment, independent adherence ratings, and a severe population (72.1% schizophrenia; 21% schizoaffective; 7% mood disorder with psychotic features; duration: M = 14.7 years).

Compliance with harmful command hallucinations proved an univiable primary outcome measure because of low base rates. Both groups equally improved their confidence to cope and resist complying with command hallucinations by post-test, but these changes were not maintained at 6-month follow-up. In post hoc within-group analyses of secondary outcomes, the TORCH but not the Befriending group demonstrated significant improvements on positive and negative symptoms and global assessment of functioning. Befriending but not TORCH reported improvement in distress and disruption from hallucinations at post-test, although TORCH improved disruption by follow-up to catch up. TORCH made significant pretreatment to post-treatment and pretreatment to follow-up improvements on 2/4 quality-of-life measures, whereas Befriending made significant pretreatment to follow-up improvement on one. Consistent with the literature on CBT without ACT components (Jones, Hacker, Cormac, Meaden, & Irving, 2012), TORCH was not clearly superior to a less complex psychosocial therapy. The trial was ultimately underpowered (only 72% planned sample recruited) and the effect sizes were lower than anticipated, so a larger trial is needed to reliably estimate ACT’s efficacy with treatment-resistant psychosis. Nevertheless, it provides preliminary evidence that ACT may be more effective than non-specific therapy in improving psychotic symptoms.

Substance use disorders
Smout et al. (2010) randomised 104 methamphetamine users to ACT or CBT. Both groups reduced methamphetamine measured by hair samples (time effect: d = 0.61) and self-report (d = 2.69), self-reported other drug use (d = 0.81), negative consequences of methamphetamine use (d = 2.53), psychological dependence (d = 3.08), depression (d = 2.25), and general mental health (d = 2.09) from baseline to 24 weeks post-entry. CBT significantly increased the proportion of methamphetamine-free hair samples from baseline to 12 weeks post-entry and reduced negative consequences from 12 to 24 weeks. Serious methodological problems limit its conclusiveness. Attrition was acceptably high (70% at 12 weeks and 86% at 24 weeks post-entry), leaving the study underpowered to detect true differences. Treatment completion in both groups was low (Mdt = 3.0, IQR 5.5), which would likely have reduced their potency.

Gifford et al. (2011) randomised 303 cigarette smokers to bupropion alone or buproprion plus a combined ACT and functional analytic psychotherapy intervention, an ACT-consistent approach focused on the therapeutic relationship. Combined treatment recipients achieved higher quit rates (7-day self-report confirmed with carbon monoxide breath samples), both 10 weeks after quit date (50% versus 27.9% medication alone, d = 0.46) and at 1-year follow-up (31.6% versus 17.5%, d = 0.33).

Ost (2008) highlighted several limitations with a precursor study by this group (Gifford et al., 2004), some of which Gifford et al. (2011) addresses. Attrition was again high in both groups (41% combined, 48% medication); however, a large proportion of this group’s data was available for intent-to-treat modelling. While treatment credibility was not rated, the combined treatment achieved higher client satisfaction ratings than medication alone. Again, no information about whether
additional treatment was sought during the follow-up period was reported, although this is somewhat less critical as Gifford et al. (2011) found group differences post-treatment unlike Gifford et al. (2004). A significant limitation with Gifford et al. (2011) is that the medication-only group did not receive equivalent therapist contact time, so the improvements cannot be conclusively attributed to ACT, although evidence that post-treatment acceptance of smoking cues mediated the effect of treatment condition on follow-up quit rates suggests some therapy-specific effect. The study suggests that ACT may be a valuable adjunct to pharmacotherapy for smoking, but an ACT + bupropion design is needed to confirm this.

Luoma, Kohlenberg, Hayes, and Fletcher (2012) assigned individuals from a 28-day residential substance dependence treatment centre to TAU-only or TAU substituting 6 hours of the first week for three 2-hour ACT group sessions aimed to reduce internalised shame, which is associated with poor treatment engagement and functioning. ACT + TAU showed a smaller reduction in shame by post-test (1 week after entry, \( d = 0.26 \)) than TAU but a larger reduction at follow-up \( (d = 0.66) \). By 4-month follow-up, more ACT recipients reliably reduced shame (30.9% versus 19.7%) and fewer reliably deteriorated (2.9% versus 15.2%). As predicted, ACT recipients used 82% more drug and alcohol treatment services during follow-up than TAU-only. ACT recipients were more likely (OR 2.30, 95% CI (1.14, 4.74)) to have been abstinent from all substances during any follow-up week than TAU-only. There were no group differences in reliable improvement of general mental health, but TAU-only had a higher proportion that reliably deteriorated (32.5% versus 5.4%). Higher post-treatment shame was associated with greater treatment utilisation, which in turn mediated reduced substance use over follow-up. The authors speculated that ACT may have increased acceptance of shame and reduced avoidant or overoptimistic suppression.

There were several design limitations. Attrition was substantial (41%, equivalent between conditions) and the 4-month follow-up modest. Participants were recruited by treatment staff, and it is unclear how systematic this selection was. No formal psychiatric diagnosis or systematic recording of substance use at baseline was undertaken, although social impairment was evident in the sample, with 57% on probation, parole, bail, or awaiting trial and only 15% employed. Finally, while ACT sessions were independently rated as highly competent, no such ratings of TAU were made, so it is unknown whether differences may be due to general therapist competence. Nevertheless, the impact of a 6-hour variation in protocol is remarkable and worth further replication.

Stotts et al. (2012) explored the use of ACT to assist opioid-dependent clients detoxify from methadone maintenance. Participants \( (n = 56) \) were assigned to 24 × 50-minute sessions of either ACT or manualised drug counselling, concurrent with a 5-month linear methadone-reduction protocol. Completion of the programme was significantly higher for ACT (60%) than drug counselling (46.2%) recipients, and a higher proportion receiving ACT successfully detoxified (37% versus 19%), although there were no differences in opioid use during treatment. Study strengths included independent rating of random sessions from both conditions, finding equivalently high adherence and competence in both. The main limitations were: (1) small sample size, so results may not be stable; and (2) longer training time for the ACT condition, which possibly confounds amount of training with intervention. Nevertheless, the effect size compares favourably with existing interventions for opioid detoxification and, given its importance and difficulty to attain, encourages continued investigation of ACT for this purpose.

Discussion

ACT research has continued to proliferate since 2008. The majority of RCTs have been pilot studies, which share many of the shortcomings Öst (2008) identified with previous ACT literature. Significant improvements are evident in monitoring the adherence and competence of ACT. Two of the most common factors that limit the conclusiveness of ACT research remain the widespread use of TAU, which is not matched for amount of contact or monitored for competence, and failure to assess the use of other treatments. Thus, despite the ACT condition achieving superior results to TAU in the majority of studies, explanations such as general competence or increased attention cannot be ruled out as contributing in part or in whole to the outcomes. Future research should include comparison conditions that receive an equivalent level of training, competence monitoring, and amount of contact to the ACT condition.

There are some noteworthy post-2008 exceptions to the above, which we regard as being of sufficient quality to be regarded as “properly conducted” RCTs and contribute to “Level II” evidence (National Health and Medical Research Council, 2000). Wicksell et al.’s (2009) ACT protocol for paediatric pain produced large improvements over a more contact-intensive TAU condition that included pharmacotherapy. Twohig et al.’s (2010) study was methodologically rigorous and demonstrated that ACT even without ERP was efficacious in reducing OCD symptomatology. While longer follow-up periods would be desirable with both populations, the studies...
demonstrated that the benefits persisted at least 3 months beyond the phase of active treatment. Arch et al.’s (2012) study was similarly rigorous and demonstrated that ACT was efficacious in reducing anxiety disorder symptoms, at least in well-educated adults. Finally, while the areas were specialised and the studies small, the improvements seen in the ACT condition of Stotts et al.’s (2012) methadone detoxification trial and Rost et al.’s (2012) end-stage ovarian cancer trial strongly suggest that ACT rather than therapy-unspecific factors was effective with these populations.

In evaluating the empirical status of ACT, it is important not to overlook that one of its authors’ primary intents was to create a transdiagnostic model with broad applicability, including the treatment of problems that do not fit neatly into diagnostic categories. ACT protocol content does not vary much between applications, and this creates potential efficiencies in training and developing competence in settings where it is difficult to constrain the range of presenting complaints. To date, there has been no systematic investigation of the potential cost-effectiveness of ACT. There is some indication from Luoma et al. (2012) and Arch et al. (2012) that ACT is associated with accessing more therapy, which may be advantageous if the alternative is deterioration and re-presentation in more costly treatment settings but disadvantageous if the alternative is developing greater independence and self-management skills in an alternative treatment. These possibilities warrant further investigation.

A method of reliably assessing clinical trial methodology is important because it extends our knowledge beyond the statistical analysis of outcomes. However, the quantitative conclusions in this review are limited because of the unacceptable kappa reliability ratings on 10 of the 22 items on Ost’s (2008) POSMRF. Despite training in the scale and two independent attempts at ratting, the three reviewers were unable to reach kappa ratings above 0.5 for 10 scale items. Therefore, future studies should address the weaknesses in this scale and aim to improve usability.

In conclusion, we would argue that there is sufficient quality evidence to warrant the use of ACT in Australian treatment guidelines in the treatment of OCD, chronic pain, and in well-educated populations, anxiety disorders. For other conditions, the evidence is not yet conclusive enough for treatment guidelines to recommend using ACT. However, the existing evidence suggests ACT is capable of achieving significant improvements in addiction, psychosis, depression, personality disorder, disordered eating, and behavioural medicine populations. We would argue that intelligent evidence-based practice entails considering contextual factors in the application of treatment protocols. These may include experience and competence with a particular approach, lack of response to first-line interventions, and a case formulation that fits with the model guiding the approach, and we would consider the application of ACT in the areas reviewed here under these conditions to be consistent with evidence-based practice.

References


pain (excluding headache) in adults. Cochrane Database of Systematic Reviews, (2)CD007407. doi: 10.1002/14651858.CD007407.pub2.


acceptance and commitment therapy versus education on stigma toward people with psychological disorders. *Behaviour Research and Therapy*, 45(11), 2764–2772.


Powers, M. B., & Emmelkamp, P. M. G. (2009). Response to “is acceptance and commitment therapy superior to established treatment comparisons?” *Psychotherapy and Psychosomatics*, 78, 380–381.


