Shepherding the Crowd Yields Better Work

Steven P. Dow  
HCI Institute  
School of Computer Science  
Carnegie Mellon University

Björn Hartmann  
Anand Kulkarni  
Computer Science Division  
Industrial Engineering & Operations Research  
University of California, Berkeley

Scott R. Klemmer  
Stanford HCI Group  
Computer Science Department  
Stanford University
How can crowdsourcing environments better support learning?
work complexity

skill level

novice

simple, transactional

mobileworks

amazon mechanical turk
Project Oscar

123 Arrowhead Ave.
Westborough, CO 12345

Oscar Smith
Lead Research Manager

Phone: (213)233-1950
Email: (213)574-4213

Please Copy Text from Business Card into Text Fields:

Name
Title
Email
Address:
Address Line 1
add line
City
State
Zip Code

Please Accept HIT First.
work complexity

skill level

simple, transactional

novice

mobileworks beta

amazon mechanical turk
Graphic Design Contractors

43,304 Contractors were found based on your criteria.

Andrey Sazonov
Certified 3D designer and instructor
Skills: 3D Design, Graphics, Photoshop, Recruiting
$16.67/hr - Hours: 1,937 - Ukraine - Last active: Today - Tests: 25 - Portfolio: 8
work complexity

creative, multi-faceted

simple, transactional

novice

expert

skill level

mobileworks βeta

amazon mechanical turk

Chirply

oDesk

99 designs
How to get better work from unskilled crowds

- Quality mechanisms such as parallelism, redundancy and iteration (Little, HCOMP 10)
- Specific workflows for complex work, such as Find-Fix-Verify for copy editing (Bernstein, UIST 10)
- Breaking work into smaller pieces, such as Crowdforge for article production (Kittur, CSCW 11) & Turkomatic for essay writing (Kulkarni, CSCW 12)
Hypothesis: Shepherding the crowd will lead to better work, learning, and perseverance.

Providing meaningful real-time feedback
Benefits of feedback

• Understand key metrics in a task domain supports learning *(Shute, 2008; Sadler, 1989)*

• Awareness of judgment can be motivating *(Annett, 1969)*

• Social interaction can support peripheral learning *(Lave and Wenger, 1991)*
Example task: write a review

Nikon D3100 14.2MP Digital SLR Camera with 18-55mm f/3.5-5.6 AF-S DX VR Nikkor Zoom Lens
by Nikon

Price: $649.00 & this item ships for FREE with Super Saver Shipping. Details

In Stock.
Ships from and sold by Amazon.com. Gift-wrap available.
Want it delivered Tuesday, May 37 Order it in the next 2 hours and 26 minutes, and choose One-Day Shipping at checkout. Details

12 new 15 used from $569.00 3 refurbished from $640.00

587 of 600 people found the following review helpful:

⭐⭐⭐⭐⭐ Nikon D3100 VS. D300 VS. D700, September 23, 2010

By Michael Allebach "Allebach Photography" (Philadelphia, Pa) - See all my reviews

This review is from: Nikon D3100 14.2MP Digital SLR Camera with 18-55mm f/3.5-5.6 AF-S DX VR Nikkor Zoom Lens (Electronics)

For the cost of this camera, I don’t think you can get anything better. The low light performance is off the charts. As a wedding photographer I regularly shoot with Nikon’s high end professional equipment and I was amazed how close this camera is to a pro camera. Now let me get specific. In order to compare I took a look at 100% files out of each camera I own.

Which camera excels Nikon D3100($Cheap) VS. D300($1600) VS. D700 ($2,700):
* Lens = The D3100 is the only camera that comes with a lens at it's normal price
* ISO Performance = Tie between D3100 and D700! (It could be Nikon's new processing but the JPEG looks fantastic I was shooting D3100 on 6400 iso with very little noise at all)
* Low Light Focusing = D700
Shepherd concept

Worker -> Product review -> Assessor

Worker receives Shepherd feedback from Assessor.
Instructions:

1) After you accept the HIT, you will see a list of products or media content. Select six products you would like to review.

2) For each customer review, write original useful comments for other people who might want to buy the product. Do not copy a review you find online. We will not pay for copied text.

Guidelines:

- Provide the exact product name and model (if applicable).
- Do NOT copy text from some other source. Do NOT plagiarize.
- Write an honest and useful customer review.
- Share personal stories about your experience with the product.
- List good and bad aspects of the product.
- Make comparisons between this product and similar products.
- Talk about the product's value given its price.
- Try to mention the product name/model in the review.
- Use correct spelling and grammar.
- Avoid excessive punctuation and capitalization.
- The review should be long enough to provide useful information, but not too long that it causes the reader to lose interest.

Your product reviews can be positive or negative. Just be honest and write a real review
Use these GOOD example reviews as a guideline:

Positive review

"A great everyday camera" 4 out of 5 stars
I purchased this camera after my Nikon Coolpix S3000 lost its shutter button because I dropped it on a marble floor. Got the Canon Powershot SD1300IS and have to say, I love it! I took pictures of my dog running through wildflowers in the wooded area near my home. They look awesome!

Negative review

"A camera with too many limitations" 1 out of 5 stars
Similar to the old Canon model, the Canon Powershot SD1300IS takes good, clean photos, and it's compact. However, instead of having an external menu for choosing modes (such as low-light, macro, or action shot), you need to toggle through multiple submenus on the display. I couldn't find a way to get used to it.
At first, the Nintendo Wii was great fun...
An expert provided the following assessment:

- The worker wrote an original review. The worker did not plagiarize.
- The worker wrote an honest and useful customer review
- The worker included personal stories/anecdotes
- The worker listed good and bad aspects of the product
- The worker compared this product to others
- The worker assessed the product's value given its price
- The worker mentioned the product's name/model
- The worker did not have spelling and grammar mistakes
- The worker used appropriate punctuation and capitalization
- The worker wrote the right amount (1-2 paragraphs)

How effective is this customer review?

- [ ] 9 Excellent
- [ ] 8
- [ ] 7 Very Good
- [ ] 6
- [ ] 5 Acceptable
- [ ] 4
- [ ] 3 Borderline
- [ ] 2
- [ ] 1 Poor

How can the worker improve their work?
great. Can you also share some personal stories?

Submit
Worker A1JNK917EIFZU4 wrote this customer review for a video game console:

Nintendo - Wii
Rating: 2
Nintendo Wii gets old, fast.
At first, the Nintendo Wii was great fun, but the appeal quickly faded into boredom. The controls are gimmicky and tiresome. Compared to other consoles out there, PS3 and XBOX 360, the Nintendo Wii just doesn't have the staying power. I had quite a bit of fun the first month or so of owning the Wii, but lost interest not too long after. I came to realize most of the games were the same concept over and over and over. Very boring. If you're buying this for kids, it may be a good plan, as it is very simple. For adults, I would recommend getting a more complex and versatile system with a better variety of games. I eventually sold my Nintendo Wii with all the games for a measly $75. Don't waste your money like I did.

An expert provided the following assessment:

- The worker wrote an original review. The worker did not plagiarize.
- The worker wrote an honest and useful customer review
- The worker included personal stories/anecdotes

The worker listed good and bad aspects of the product

- The worker assessed the product's value given its price
- The worker mentioned the product's name/model
- The worker did not have spelling and grammar mistakes
- The worker used appropriate punctuation and capitalization
- The worker wrote the right amount (1-2 paragraphs)

How can the worker improve their work?

great. Can you also share some personal stories?
Worker A1JNK917EIFZU4 wrote this customer review for a video game console:

Nintendo - Wii
Rating: 2

Nintendo Wii gets old, fast.
At first, the Nintendo Wii was great fun, but the appeal quickly faded into boredom. The controls are gimmicky and tiresome. Compared to other consoles out there, PS3 and XBOX 360, the Nintendo Wii just doesn't have the staying power. I had quite a bit of fun the first month or so of owning the Wii, but lost interest not too long after. I came to realize most of the games were the same concept over and over and over. Very boring. If you're buying this for kids, it may be a good plan, as it is very simple. For adults, I would recommend getting a more complex and versatile system with a better variety of games. I eventually sold my Nintendo Wii with all the games for a measly $75. Don't waste your money like I did.

An expert provided the following assessment:

- The worker wrote an original review. The worker did not plagiarize.
- The worker wrote an honest and useful customer review
- The worker included personal stories/anecdotes
- The worker listed good and bad aspects of the product
- The worker compared this product to others
- The worker assessed the product's value given its price
- The worker mentioned the product's name/model
- The worker did not have spelling and grammar mistakes
- The worker used appropriate punctuation and capitalization
- The worker wrote the right amount (1-2 paragraphs)

How effective is this customer review?
- 7 Very Good
- 6
- 5 Acceptable
- 4
- 3 Borderline
- 2
- 1 Poor

How can the worker improve their work?
great. Can you also share some personal stories?
Worker A1JNK917E1FZU4 wrote this customer review for a video game console:

**Nintendo - Wii**
**Rating:** 2

**Nintendo Wii gets old, fast.**
At first, the Nintendo Wii was great fun, but the appeal quickly faded into boredom. The controls are gimmicky and tiresome. Compared to other consoles out there, PS3 and XBOX 360, the Nintendo Wii just doesn't have the staying power. I had quite a bit of fun the first month or so of owning the Wii, but lost interest not too long after. I came to realize most of the games were the same concept over and over and over. Very boring. If you're buying this for kids, it may be a good plan, as it is very simple. For adults, I would recommend getting a more complex and versatile system with a better variety of games. I eventually sold my Nintendo Wii with all the games for a measly $75. Don't waste your money like I did.

An expert provided the following assessment:

- The worker wrote an original review. The worker did not plagiarize.
- The worker wrote an honest and useful customer review
- The worker included personal stories/anecdotes
- The worker listed good and bad aspects of the product
- The worker compared this product to others
- The worker assessed the product's value given its price
- The worker mentioned the product's name/model
- The worker did not have spelling and grammar mistakes
- The worker used appropriate punctuation and capitalization

**How effective is this customer review?**

- 9 Excellent
- 8
- 7 Very Good
- 6
- 5 Acceptable
- 4
- 3 Borderline
- 2
- 1 Poor

**How can the worker improve their work?**
great. Can you also share some personal stories?
Comparative study

No assessment

work on task → submit work

External assessment

work on task → review feedback and edit task → submit work

external assessment
Comparative study

No assessment
- work on task
- submit work

Self assessment
- work on task
- assess & edit own work
- submit work

External assessment
- work on task
- review feedback and edit task
- submit work

external assessment
Study hypotheses

- **H1**: Assessment will help crowd workers produce better quality work.

- **H2**: Assessment will help crowd workers learn and improve their work over time.

- **H3**: Assessment will motivate crowd workers to produce more output.

- **H4**: The effect sizes for H1, H2, and H3 will be larger for external assessment than self-assessment.
Study measures

- **Task Performance**: expert judgment on 10 criteria and a 9-point overall scale
- **Learning**: change in performance over time (regression slope coefficient)
- **Perseverance**: amount of revisions to reviews (ratio and string edit distance)
Study procedure

Worker

Task page

Shepherd database
Write your review of a video game console

<table>
<thead>
<tr>
<th>Product brand</th>
<th>Nintendo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product model</td>
<td>Wii</td>
</tr>
<tr>
<td>Title for your review</td>
<td>Nintendo Wii gets old, fast.</td>
</tr>
<tr>
<td>How do you rate this item?</td>
<td>Poor</td>
</tr>
<tr>
<td>Your review:</td>
<td>At first, the Nintendo Wii was great fun...</td>
</tr>
</tbody>
</table>
Study procedure

Worker → Task page → Shepherd database → IM notification → Assessor
Study procedure

8:55 am: Worker 123984 submitted a new review. Please assess it at the following URL: http://...

9:05 am: Worker 340394 submitted a new review. Please assess it at the following URL: http://...

9:12 am: Worker 123984 submitted a new review. Please assess it at the following URL: http://...
<table>
<thead>
<tr>
<th>Worker</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFW82A1CWITW</td>
<td>cell phone</td>
<td>video game</td>
<td>digital camera</td>
<td>tv show</td>
<td>5.8</td>
</tr>
<tr>
<td>2011-05-26 11:14:32</td>
<td>Time: 8m</td>
<td>Time: 8m</td>
<td>Time: 7m</td>
<td>Time: 4m</td>
<td></td>
</tr>
<tr>
<td>A338XRY1M5OV0J</td>
<td>cell phone</td>
<td>digital camera</td>
<td>TV</td>
<td>computer</td>
<td></td>
</tr>
<tr>
<td>2011-05-26 11:11:43</td>
<td>Time: 8m</td>
<td>Time: 8m</td>
<td>Time: 10m</td>
<td>Time: 17m</td>
<td></td>
</tr>
<tr>
<td>A3DIA04SJ1N6KL</td>
<td>cell phone</td>
<td>Score: 5</td>
<td>See FB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011-05-26 11:07:36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1M69H9REHPES</td>
<td>mp3 player</td>
<td>tv show</td>
<td>music album</td>
<td>printer</td>
<td>6.0</td>
</tr>
<tr>
<td>2011-05-26 11:06:18</td>
<td>Time: 7m</td>
<td>Time: 3m</td>
<td>Time: 4m</td>
<td>Time: 3m</td>
<td></td>
</tr>
<tr>
<td>A1VAG01PMX7DS3</td>
<td>mp3 player</td>
<td>cell phone</td>
<td>dvd player</td>
<td>digital camera</td>
<td></td>
</tr>
<tr>
<td>2011-05-26 11:02:09</td>
<td>Time: 5m</td>
<td>Time: 1m</td>
<td>Time: 2m</td>
<td>Time: 1m</td>
<td></td>
</tr>
<tr>
<td>A1EDXCMHYRS9UE</td>
<td>mp3 player</td>
<td></td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>2011-05-26 11:01:15</td>
<td>Time: 4m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A20IW9A1X2TIH5</td>
<td>mp3 player</td>
<td>video game</td>
<td></td>
<td></td>
<td>4.0</td>
</tr>
<tr>
<td>2011-05-26 11:00:44</td>
<td>Time: 8m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A29OFT1AULRFU</td>
<td>tablet computer</td>
<td>Score: 0.0</td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>2011-05-26 10:17:31</td>
<td>Time: 6m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A8AUTPJ0PJJUG</td>
<td>digital camera</td>
<td>Score: 0.0</td>
<td></td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>2011-05-26 10:04:12</td>
<td>Len: 362c</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Worker A1JNK917EIFZU4 wrote this customer review for a video game console:

Nintendo - Wii
Rating: 2
Nintendo Wii gets old, fast.
At first, the Nintendo Wii was great fun, but the appeal quickly faded into boredom. The controls are gimmicky and tiresome. Compared to other consoles out there, PS3 and XBOX 360, the Nintendo Wii just doesn't have the staying power. I had quite a bit of fun the first month or so of owning the Wii, but lost interest not too long after. I came to realize most of the games were the same concept over and over and over. Very boring. If you're buying this for kids, it may be a good plan, as it is very simple. For adults, I would recommend getting a more complex and versatile system with a better variety of games. I eventually sold my Nintendo Wii with all the games for a measly $75. Don't waste your money like I did.

Great. Can you also share some personal stories?
Study procedure

Worker: Task page → Shepherd database → IM notification → Assessment form
Assessor: Assessment form → Shepherd database → IM notification → Task page
Study procedure

Worker

Task page → Shepherd database → Assessment form → Assessor

Assessment review page ← Shepherd database ← IM notification ← Assessor
Worker A1JNK917E1FU4 wrote this customer review for a video game console:

**Nintendo - Wii**
Rating: 2
**Nintendo Wii gets old, fast.**
At first, the Nintendo Wii was great fun, but the appeal quickly faded into boredom. The controls are gimmicky and tiresome. Compared to other consoles out there, PS3 and XBOX 360, the Nintendo Wii just doesn't have the staying power. I had quite a bit of fun the first month or so of owning the Wii, but lost interest not too long after. I came to realize most of the games were the same concept over and over and over. Very boring. If you're buying this for kids, it may be a good plan, as it is very simple. For adults, I would recommend getting a more complex and versatile system with a better variety of games. I eventually sold my Nintendo Wii with all the games for a measly $75. Don't waste your money like I did.

An expert provided the following assessment:

- The worker wrote an original review. The worker did not plagiarize.
- The worker wrote an honest and useful customer review
- The worker included personal stories/anecdotes
- The worker listed good and bad aspects of the product
- The worker compared this product to others
- The worker assessed the product's value given its price
- The worker mentioned the product's name/model
- The worker did not have spelling and grammar mistakes
- The worker used appropriate punctuation and capitalization

**How effective is this customer review?**
- 9 Excellent
- 8
- 7 Very Good
- 6
- 5 Acceptable
- 4
- 3 Borderline
- 2
- 1 Poor

**How can the worker improve their work?**
great. Can you also share some personal stories?
Take a minute to reflect...

A few moments ago you wrote this customer review for a video game console:

Nintendo - Wii
Rating: 2
Nintendo Wii gets old, fast.
At first, the Nintendo Wii was great fun, but the appeal quickly faded into boredom. The controls are gimmicky and tiresome. Compared to other consoles out there, PS3 and XBOX 360, the Nintendo Wii just doesn't have the staying power. I had quite a bit of fun the first month or so of owning the Wii, but lost interest not too long after. I came to realize most of the games were the same concept over and over and over. Very boring. If you’re buying this for kids, it may be a good plan, as it is very simple. For adults, I would recommend getting a more complex and versatile system with a better variety of games. I eventually sold my Nintendo Wii with all the games for a measly $75. Don't waste your money like I did.

Assess your own work:

Check all statements that apply.

☐ I wrote an original review. I did not plagiarize.
☐ I wrote an honest and useful customer review
☐ I included personal stories/anecdotes
☐ I listed good and bad aspects of the product
☐ I compared this product to others
☐ I assessed the product's value given its price
☐ I mentioned the product's name/model
☐ I did not have spelling and grammar mistakes
☐ I used appropriate punctuation and capitalization
☐ I wrote the right amount (1-2 paragraphs)

How effective is my customer review?

☐ 9 Excellent
☐ 8
☐ 7 Very Good
☐ 6
☐ 5 Acceptable
☐ 4
☐ 3 Borderline
☐ 2
☐ 1 Poor

Describe your rating and how you can improve your work:
Study procedure

Worker

Task page

Assessment review page

Shepherd database

Assessment form

IM notification

Assessor

6 reviews per worker
Study details

- Tasks posted to MTurk over 1 week
- Round-robin condition assignment
- 277 workers created 945 product reviews
- 9% plagiarism; 52% dropped early (no diff b/n conditions)
- Analyzed 105 participants (538 reviews)
- Total cost of experiment: ~$200 total ($1.50 for 6 reviews)
Expert and self-assessment both lead to better work.

F(2,102) = 3.02, p = 0.05
Assessment helped workers improve over time

- **External assessment**: 95% HPD interval = [-0.03, 0.23], p=0.08
- **Self assessment**: 95% HPD interval = [0.09, 0.44], p=0.001
- **No assessment**: 95% HPD interval = [-0.13, 0.06], p=0.38
Why does self assessment prove as effective as external assessment?
Self assessment in learning

• Improved essay writing for students in grades 4&5 *(Ross and Starling, 2005)*

• Higher student self-direction for those with low pretests of self-direction *(Henry, 1994)*

• Greater intrinsic motivation *(Schunk, 1996)*

• Helps students align their work with concrete grading criteria *(David, 1995; Taras, 2002)*
Revisions

original product review

real-time assessment

revised product review
Revisions

original product review -> real-time assessment -> revised product review

post assessment
Revised reviews were better

\[ t(188)=2.09, p<0.05 \]
External assessment encourages more revisions

χ²=47.3, p<0.05
External assessment encourages more revisions

χ²=47.3, p<0.05

F(1,81)=3.26, p=0.07
No interaction effect between condition (self vs. external) and version (original vs. revised).
Benefits of shepherding

• Improved overall results

• Increased performance over time (better skilled workers?)

• More work output per dollar (especially after external assessment)
Discussion
Discussion

How much did attrition affect the results?
Discussion

Can crowd peers perform assessments?
Worker distribution over time
Can the crowd produce quality assessments?

- Low agreement between expert and crowd assessments (Cohen’s Kappa=0.20)

- Crowd filtering mechanisms (Ipeirotis et al, 2010; Karger et al., 2011)

- Qualifications
Future Work
real-time remote apprenticeship
Task page

Real-time remote apprenticeship

Shadowing

LEARNER

EXPERT
Task Q&A

Worker-to-worker communication
Recap

- Shepherding the crowd improves work.
- Self-assessment can provide guidance comparable to external assessment.
- External assessment leads to more revisions.
- Future work will focus on other learning methods (peer assessment, apprenticeship, ...).
Thank you.

Steven Dow
spdow@cs.cmu.edu
www.cs.cmu.edu/~spdow

Coauthors
Björn Hartmann, UC Berkeley, CS Division
Anand Kulkarni, UC Berkeley, Industrial Engineering
Scott Klemmer, Stanford HCI Group

Financial support
Hasso Plattner Design Thinking Research Program