Online Diaries for Everyone: Data Quality, Device Usage and Compliance with Personal Expenditure Diaries

Douglas Williams, Westat
Hanyu Sun, Westat
Outline

- Background & Online Diary Improvement Project
- Usability test design and methods
- Findings: compliance
- Findings: diary and device usage
- Findings: data quality
- Summary
• CE Diary one of two components of the CE Survey capturing minor and frequent expenditures.

• Current diary uses PAPI and a household respondent covering two one-week reporting periods.

• Continued review and development of online platform for CE Diary (see Elkin, To, and Williams, 2017 – FedCASIC).
  – Web Diary Feasibility Test – 2013
  – Individual Diaries Feasibility Test (IDFT) – 2014
  – Proof of Concept Test – 2015
  – Electronic Diary Desktop Design Improvements – 2016
Online Diary Improvement Project

- Developed collaboratively with BLS building on prior research and diary data requirements
  - Modernization – flexible instrument that can add and adapt to new technologies
  - Accessible across multiple platforms and mobile optimized
    - Smartphone ownership: 55% in 2013; 77% in 2016
    - Designed for personal diary assignment
- Instrument developed using open-source programming software
  - Adaptive (web) design – optimized for mobile
  - Design elements (look and feel) same across device types used to access
Design & Methods: Online Diary Improvement Project

• Recruited households for usability test
  – Desktop/Laptop and Smartphone ownership
  – Multiple age eligible household members (71% of households)
• Procedure: placement interview; diary reporting; pick-up interview
  – In-person interviews conducted March, 2017 – May, 2017
  – Completed 62 placement interviews; 61 pick-up interviews
  – Incentive: $40 for the main diarist; $10 for other HHM
# Mobile Diary View

## My Expenses

<table>
<thead>
<tr>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wednesday, Jul 5</td>
<td></td>
</tr>
<tr>
<td>Thursday, Jul 6</td>
<td></td>
</tr>
<tr>
<td>Friday, Jul 7</td>
<td></td>
</tr>
<tr>
<td>Saturday, Jul 8</td>
<td></td>
</tr>
<tr>
<td>Sunday, Jul 9</td>
<td></td>
</tr>
<tr>
<td>Monday, Jul 10</td>
<td></td>
</tr>
<tr>
<td>Tuesday, Jul 11</td>
<td></td>
</tr>
</tbody>
</table>

### Add Expenses

- **Date:** Wednesday, Jul 5
- **Type:** Meals, Snacks, and Drinks Away From Home
- **Details:**
  - **Describe meal:** Enter description
  - **Total cost (include tax and/or tip):** $0.00
  - **What meal was this?**
    - Breakfast
    - Lunch
    - Dinner
    - Snack/Other
  - **Select any alcohol included:**
    - None
    - Beer
    - Wine
    - Other
Usability Test Results
Research Questions

• Compliance & Use
  – How well do different household members comply with the diary task?

• Device Usage
  – What are the characteristics diary usage by device?
  – Are there differences in types of expenditures entered by device type?
  – Are there any barriers to accessing the online diary?

• Data Quality
  – Do expenditure descriptions indicate problems or category mismatches?
Compliance

• How well do different household members comply with the diary task?
  – Composition of recruited sample
    • 62 main diarist – responsible for encouraging other HHM participation
    • 72 other household members

• Access & Use
  – Access: log-in and create password
  – Use: enter any expense
**Compliance: Results**

- **Access & Use**
  - 61 of 62 (98%) main diarist accessed the diary
  - 50 of 72 (69%) other household members accessed the diary
  - Similar results for use

- **Comply with directions to access and change password day of visit**

<table>
<thead>
<tr>
<th>Days from visit</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Diarist</td>
<td>51%</td>
<td>16%</td>
<td>8%</td>
<td>7%</td>
<td>5%</td>
<td>5%</td>
<td>2%</td>
<td>7%</td>
</tr>
<tr>
<td>Other</td>
<td>22%</td>
<td>18%</td>
<td>10%</td>
<td>10%</td>
<td>6%</td>
<td>14%</td>
<td>14%</td>
<td>6%</td>
</tr>
</tbody>
</table>
Expense Reporting Behavior

Main Diarist

<table>
<thead>
<tr>
<th>Reporting Day</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.3</td>
<td>6.4</td>
<td>6.9</td>
<td>8.3</td>
<td>6.9</td>
<td>5.5</td>
<td>9.9</td>
<td>11.3</td>
</tr>
</tbody>
</table>

Other Household Members

<table>
<thead>
<tr>
<th>Reporting Day</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4.0</td>
<td>2.4</td>
<td>4.2</td>
<td>8.1</td>
<td>2.8</td>
<td>5.3</td>
<td>6.0</td>
<td>5.9</td>
</tr>
</tbody>
</table>
Expense Entry Timeliness

• Timely entry of expenses
  – Suggest attentiveness to the diary task
  – Better data: less time for recall; less likely to forget expense

<table>
<thead>
<tr>
<th>Difference</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Diarist</td>
<td>31.0</td>
<td>33.2</td>
<td>11.3</td>
<td>11.2</td>
<td>4.6</td>
<td>2.5</td>
<td>6.2</td>
</tr>
<tr>
<td>Other</td>
<td>15.1</td>
<td>36.4</td>
<td>9.0</td>
<td>9.2</td>
<td>9.0</td>
<td>6.7</td>
<td>14.6</td>
</tr>
<tr>
<td>All</td>
<td>27.0</td>
<td>34.0</td>
<td>10.7</td>
<td>10.7</td>
<td>5.7</td>
<td>3.6</td>
<td>8.3</td>
</tr>
</tbody>
</table>
What are the characteristics of diary usage by device?

- Half of main diarist used mobile / one-third only used mobile
- Why not more mobile: smaller screen, lack of tactile keyboard
- Desk/Laptop used more (1,522; 679), mobile entries were more timely

<table>
<thead>
<tr>
<th>Device Type</th>
<th>Desk / Laptop</th>
<th>Mobile</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Diarist (n = 61)</td>
<td>43</td>
<td>31</td>
<td>13</td>
</tr>
<tr>
<td>Other (n = 49)</td>
<td>37</td>
<td>20</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Difference</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desk/Laptop</td>
<td>24%</td>
<td>31%</td>
<td>12%</td>
<td>12%</td>
<td>21%</td>
</tr>
<tr>
<td>Mobile</td>
<td>33%</td>
<td>41%</td>
<td>8%</td>
<td>8%</td>
<td>10%</td>
</tr>
</tbody>
</table>
Device Usage: Expense Types

• Are there differences in types of expenditures entered by device type?
  – ODIP – CE expense categories expanded from 4 to 10

• Food and Drinks Away from Home
• Food and Drinks for Home Consumption
• Clothing, Shoes, Jewelry, and Accessories
• All Other Products and Services

• Meals, Snacks, and Drinks Away From Home
• Food and Drinks for Home Consumption
• Clothing, Shoes, Jewelry, and Accessories
• Medical Expenses and Supplies
• Entertainment and Recreation
• Home Furnishings and Decorative Items
• Education Expenses and Supplies
• Transportation Expenses
• Personal Care or Hygiene Items
• All Other Products, Services, and Expenses
<table>
<thead>
<tr>
<th>Expense Type</th>
<th>Desk/Laptop</th>
<th>Mobile</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Meals, Snacks, and Drinks Away From Home</td>
<td>15.0%</td>
<td>30.2%</td>
<td>19.7%</td>
</tr>
<tr>
<td>2 - Food and Drinks For Home Consumption</td>
<td>52.9</td>
<td>21.5</td>
<td>43.2</td>
</tr>
<tr>
<td>3 - Clothing, Shoes, Jewelry, and Accessories</td>
<td>6.0</td>
<td>6.8</td>
<td>6.2</td>
</tr>
<tr>
<td>4 - Medical Expenses and Supplies</td>
<td>1.9</td>
<td>2.4</td>
<td>2.0</td>
</tr>
<tr>
<td>5 - Entertainment and Recreation</td>
<td>2.1</td>
<td>5.7</td>
<td>3.2</td>
</tr>
<tr>
<td>6 - Home Furnishings and Decorative Items</td>
<td>0.3</td>
<td>2.5</td>
<td>1.0</td>
</tr>
<tr>
<td>7 - Education Expenses and Supplies</td>
<td>0.4</td>
<td>0.7</td>
<td>0.5</td>
</tr>
<tr>
<td>8 - Transportation Expenses</td>
<td>6.6</td>
<td>9.7</td>
<td>7.6</td>
</tr>
<tr>
<td>9 - Personal Care, or Hygiene Items</td>
<td>3.7</td>
<td>8.0</td>
<td>5.0</td>
</tr>
<tr>
<td>10 - All Other Products, Services, and Expenses</td>
<td>11.1</td>
<td>12.5</td>
<td>11.5</td>
</tr>
</tbody>
</table>
Barriers to Access: Failed Login Attempts

• Are there any barriers to accessing the online diary?
  – Setting/remembering password most reported barrier
  – Password
  • Eight characters: number, upper & lower-case letter, special character (!@#$%^&*)

<table>
<thead>
<tr>
<th>Total number of failed attempts</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent</td>
<td>46.9</td>
<td>12.6</td>
<td>8.1</td>
<td>3.6</td>
<td>3.6</td>
<td>3.6</td>
<td>4.5</td>
<td>17.1</td>
</tr>
</tbody>
</table>

• One-half of all failed attempts occurred before setting password
Data Quality

• Do expenditure descriptions indicate problems or category mismatches?
  – Description field: open-ended text field for describing expense
    • Red polo shirt; 2 liter coke; Bananas; Gasoline; Etc…
  – Type of issues identified – independent coding and review
    • Including multiple items in one entry
    • Vague descriptions: item or expense is unclear (e.g., food, clothes, cleaning)
    • Establishment name: (e.g., restaurant name)
    • Incorrect category
    • Unidentifiable: gibberish
### Data Quality: Expense Description

- Overall, poor quality descriptions low (of n = 2,255 total expenses)
- Issues clustered within respondents, or affect specific category
  - 14 respondents account for over half (55%) of coded issues
  - Establishment name generally used for meals away from home (restaurant name)

<table>
<thead>
<tr>
<th>Description Issue Type</th>
<th>Count</th>
<th>Percent/All Expenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple items</td>
<td>52</td>
<td>2.3%</td>
</tr>
<tr>
<td>Vague description</td>
<td>37</td>
<td>1.6%</td>
</tr>
<tr>
<td>Establishment name</td>
<td>59</td>
<td>2.6%</td>
</tr>
<tr>
<td>Incorrect category</td>
<td>41</td>
<td>1.8%</td>
</tr>
<tr>
<td>Unidentifiable</td>
<td>13</td>
<td>0.6%</td>
</tr>
</tbody>
</table>
Summary
Summary

- Personal Diaries
  - Near uniform use of online diary by main diarist (household respondent)
  - Usage lower for other HHM, but encouraging
  - Minimum level? Estimate the proportion of expenses lost for household diaries versus loss due to nonparticipation in personal diaries.
  - Not ideal to rely on the main diarist to encourage and motivate other HHM to participate
  - Indications that other HHM start later and abandon sooner
Summary

• Online Diaries
  – Offer dimensions not available with paper: use of multiple devices, mobility
  – Mobile devices were frequently used, but not as expected
    • Most respondents used at home – still convenient
    • Receipt complexity influencing device type used?
  – Expenses entered via mobile device were more timely – better data?
  – Password the biggest barrier to access – and mobility
  – Data quality – low incidence of issues, but still problematic
    • E.g., 52 instances of multiple entries can equal 100-150 missed expenses
Thank you!
DouglasWilliams@Westat.com