ALL POSSIBLE WORLDS: ADVANCING PASSIVE ELECTRONIC MEASUREMENT OF READERSHIP

Dr. Pasquale Pellegrini and Adam Gluck, Arbitron Inc

1. Introduction

The past decade has witnessed the development of new passive or near passive, electronic measurement for radio, TV, outdoor, out of home, and internet consumption using devices like the Arbitron Portable People Meter (PPM™) for radio and TV, amongst others (e.g., Eurisko's Media Monitor, Ipsos' cell meter, GfK's mediawatch). Most recently, census-based measurement using digital set-top boxes has emerged as another example of using digital clickstream data for passive audience measurement with similar logic being applied to other digital media (Pellegrini, 2006).

PPM technology has provided currency audience estimates in Canada (since 2003), Norway (since 2006) and now Kazakhstan, with Iceland and Denmark launching panels next year. In the US, Philadelphia and Houston are PPM radio markets with New York now being recruited – all part of the Arbitron plan to electronically measure the top 50 US markets. PPM is an electronic measurement system for TV and radio that detects an inaudible, identifying audio code in a broadcaster's output, along with a time stamp, that permits proper crediting and the calculation of ratings estimates (see Patchen and Webb, 2002; Pellegrini and Purdye, 2004; 2005). Perhaps the most attractive feature of the PPM is its ability to alleviate respondent burden via the passive ability to collect media consumption. In addition, it provides more accurate audience measurement because it precludes respondent recall and bias that is often laden with errors. The case in favor of passive electronic audience measurement of personal media consumption is discussed in more detail elsewhere (see Pellegrini, 2005); the extension of PPM measurement into non-audio environments such as readership – permitting true, passive, multi-media measurement – is at the cornerstone of this paper.

The limitations of today's recall-based readership measurement, and the potential to elevate readership measurement to the electronic domain, necessitate the development and testing of electronic readership measurement. Current recall, paper and pencil techniques appear antiquated in comparison to state-of-the-art electronic measurement of TV and radio. In addition, readership needs to be measured and reported in a shorter timeframe than presently exists. Moreover, there is no way to know when readership or ad exposure actually occurs. This limitation makes it a challenging exercise to infer causal relationships between magazine ad exposure and purchasing. Taken together, these limitations put print at a comparative disadvantage to electronically measured media. This measurement gap will only widen if print measurement is not added to the electronic measurement roster.

This paper aims to advance and fine-tune the technology and methodology of passive electronic measurement of readership, and builds directly on the encouraging 'proof of concept' results reported by McConochie, Bailey and Gluck (2006). We present results from both laboratory and field based tests of prototype RFID tags placed into select consumer magazines detected by well-tested PPM technology capable of picking up signals from the tags. More specifically, this paper reports on the testing of a comprehensive set of magazine types (binding types, thickness), comparisons of recall (diary) versus passive electronic measurement, and the results of laboratory tests designed to examine range of detection, incidence of false positive detections (spillover) and multi-media detection. This last test sets the stage for ongoing development of the PPM technology as a multi-media passive measurement device. A discussion of possible data edit rules to preclude false positive readership detection is included. As a summary, the advantages and challenges of passive, electronic readership measurement are outlined.

2. Passive Electronic Audience Measurement and RFID

The key to passive measurement of readership is the link between the PPM technology and RFID. RFID, simply stated, is a technology that consists of two parts – a tag and a reader. The tag part is attached to the item you want to track, identify or measure. Examples include clothing, boxes of goods (e.g., for inventory control or security), animals (e.g., for lost pets) and even people (e.g., a nightclub in Barcelona will tag VIP guests). Each tag has a microscopic chip that encodes information about what is being tracked or measured. So, for instance, E-Z pass type applications at toll booths would contain information about the vehicle passing the toll – this is detected by the reader, the second part of RFID technology. The reader is a device that detects nearby chips and will read their embedded information – the range is variable and can literally be from only a few inches to 50 or 60 feet. For our purposes, it is very easy to calibrate the PPM to be an RFID reader.

The motivation for an initial surge of companies to enter the RFID market was the technology's positioning as the replacement for the ubiquitous barcode that emerged in the 1970s (but initially invented in the 1940s). Despite their apparent utility, barcodes are limited by their need to be read by a laser and the barcode and laser must be lined up for proper detection, as in the check out of a grocery store. RFID tagged items, in contrast, could be detected without any of the items being removed from the grocery cart. Although it is typically unnoticed by most consumers, RFID technology is amongst them everyday; security cards in office and residential buildings, E-Z Pass toll applications, new credit cards that offer express pay services or new passports with embedded RFID. Companies like Wal-Mart track their suppliers' deliveries with RFID as does the US military.

It is easy to imagine dozens of possibilities using RFID technology, yet the general use RFID technology has slowed in its deployment for a number of reasons. In early applications, a closed system was used in the sense that the tag and reader were from the same company with communications between the two kept simple (e.g., E-Z Pass type applications). The expansion of RFID into applications with products from myriad companies, and the need to have access to backend systems, complicates matters. The complications stem, partly, from the existence of multiple RFID standards, such ISO and EPC, while other standards exist in some individual countries, like Japan and potentially China. An additional challenge is the debate about frequencies since RFID can operate at low, high or ultra-high frequency; again, the standards vary from country to country. A final concern is patent control by a few large corporations which restricts the entry of smaller RFID companies due to large royalty fees.

As with many other companies who realize the potential benefits that RFID technology can bring, Arbitron continues to pay close attention to developments in this technology. It is beyond the scope of the present paper to provide a detailed status report on the RFID industry and tag technology, but some key points are worth noting. First, it is important to consider that RFID technology can be divided into passive and active; the former describes systems where the tags are 'dumb' and the reader essentially does all the work, while the latter system are those which contain a power source in the tag itself. In fact, the birth of the RFID industry was a result of the expectation of a low cost methodology associated with passive tags being the new technology for inexpensive tracking of tagged items (less than $.05/tag) and the removal of a line of site restriction. More recently, active tag adoptions are on the rise as their inherent advantages of considerable read range, long tag life and compatibility with GPRS, Wi-Fi and other technologies are exploited. Arbitron has explored, reviewed and considered a wide range of RFID technology, both passive and active, and continues to keep abreast of the latest developments – a formidable task with almost daily news of new advancements (Octave Technology, 2007).

RFID technology has a broad reach with capabilities unmatched by other technologies. The readership audience measurement possibilities of RFID technology are numerous: audience accumulation, issue-specific readership, removal of recall error and bias, social desirability bias disappears, model bias is confronted, multi-media measurement is attained and, potentially, page and ad-specific views are delivered - all on a timely basis. The potential is clear; the challenges await us. The research reported here presents, for the first time, documented test results of this important technology in a passive print measurement application.

3. Research Design

In the past three years, a series of tests were performed in order to examine the efficacy of using RFID technology for passive measurement of readership. Arbitron research staff explored the sensitivity of RFID signals across various surfaces (metal, glass, etc…) and the PPM has been tested as the "reader" of RFID codes across a series of other applications. In this paper, we focus on the most recent set of laboratory tests that permit a view of the most pressing issues of magazine readership detection and exposure, cross-media code detection and false positive detection.

In an early test of RFID technology – the so-called "Friends and Family test" - respondents were asked to keep a paper diary of start and stop times for their reading of a set of RFID-equipped magazines while also carrying PPMs. The PPMs are the reader in this scenario, while still maintaining their original functionality with respect to media code detection and motion compliance data gathering. The test duration was 1 week and the RFID data were analyzed to compare PPM detected exposures with those reported in respondent diaries. A variety of magazines were RFID enabled representing a variety of sizes and binding types. In an earlier paper, McConochie, Bailey and Gluck (2006) reported on the encouraging results related to the match between diary recall and PPM detection of readership. We expand that analysis here.

A second, more recent, series of tests were pure laboratory, structured protocol in design. These tests involved volunteers situated in a room, at various distances apart and tasked with reading RFID tagged magazines for periods of fifteen minutes while wearing their PPMs. Once again, a variety of magazines were tested. In addition, for some tests, encoded audio was broadcast in the background. The tests were therefore able to measure, under rather rigorous conditions, range of detection, incidence of false positive detection, and the capability to detect other media codes as well as RFID signals using the same PPM. In each of the trials, control PPMs were placed around the room as well as control RFID-tagged magazines, for comparison and validation purposes.

4. Results

We present the results in two parts. First, we revisit the earlier Friends and Family test described above to look more closely at false positive detections and the refinement in measurement from the use of a simple edit rule. Second, we devote considerable attention to the results of our most recent structured protocol laboratory tests designed to measure the readership detection and exposures with distance, background audio and the presence of other magazine readers.

Friends and Family: Diary recall versus PPM detections of RFID

Overall, the results from our past diary/meter RFID detection test appeared sensible. While the diary was not an exact measure of "truth" for tracking when respondents actually read their magazines, there was a great deal of convergence between the reading sessions listed in panelist's diaries and those detected by their meters. In fact, almost 80% of all panelists' diary reading sessions had at least one correct corresponding RFID detection during the same time frame. Furthermore, the frequency of detections during a diary documented reading session occurs with almost twice the frequency of detections outside of diary documented reading sessions which attests to the granularity that passive detection brings.

However, the fact that the meter was picking up detections outside of the diary reading sessions posed an interesting problem. Comparing the diaries and meters, it became clear that many of detections noted by panelists' meters did not correspond to the panelists' diary entries. We concluded that these detections were in all likelihood caused by one of three sources:

1. The panelist who detected the codes was actually reading, but neglected to record the reading in their diary (i.e., recall error).
2. The panelist was not reading, but the RFID tag in the magazine became activated, say, by rigorous movement of the magazine (i.e., false positive from tag sensitivity).
3. Another member of the household was reading the magazine in close proximity to the panelist (i.e., false positive from spillover).

Looking at the meter logs and the diary entries, it is not possible to determine when the first or second cases occurred and from this point forward we will refer to these as "unclassified detections"; however it is possible to determine when the third case occurred. Comparing the meter detections and diary logs of all members of a household, we determined that 25% of all undocumented detections were caused by another person reading in close proximity.

RFID code edit rule test: Off by One

Another area of investigation concerns the issue that the RFID codes detected did not always perfectly match up with the list of expected codes. As the tags transmit their code only for a few seconds, it is possible that the entire code may not be read perfectly by the meter every time. In this paper, we will call these instances Off by One codes for clarity. Throughout the course of our early Friends and Family field test, a small number of invalid codes met this description.

As a result, we decided to analyze what would happen if we could recover these ambiguous codes and therefore count the Off by One codes as valid detections. In order to do this, we first looked at the number of diary-noted reading sessions for which we detected a valid or an Off by One code. We then calculated the average number of these detections across all diary listed reading sessions, as well as the average number of detections across all diary reading sessions that had at least one detection. We then excluded the Off by One codes and repeated these calculations, and the difference in the average number of detections can be attributed to the Off by One codes. The results of this exercise are shown below in Table 1. Note that the average number of detections per diary event increases slightly, but the increase is not overwhelming.

Table 1: Comparison of valid detections and valid detections plus Off by One codes.

	Valid Detections Plus Off by One Detections (Denominator Sample Size)	Valid Detections Only (Denominator Sample Size)
Detections / Total Events in Diary	19.4 (n = 204)	18.6 (n = 199)
Detections / Total Events in Diary with at least one Detection	24.0 (n = 252)	23.6 (n = 252)

We then looked at the number of events that had a valid detection again, this time, examining the average number of code detections per 5 minutes during these reading sessions. Again, the impact of including the Off by One detections was minimal, and the results are shown in Table 2 below.

Table 2: Comparison of diary documented valid detections per 5 minutes with and without Off by One codes.

	Valid Detections Plus Off by One Detections (Denominator Sample Size)	Valid Detections Only (Denominator Sample Size)
Detections /5 Minute Period*	5.0 (n = 868)	4.9 (n = 854)

* These five minute periods correspond to diary reading sessions that have at least one RFID detection.

We then examined the impact of the Off by One Detections on unclassified detections. These detections are likely caused by one of the first two sources cited above (e.g. neglecting to record the reading in the diary or moving the magazine around). For this analysis, we again looked at the frequency of detections per 5 minutes, with the results shown in Table 3. The impact of Off by One on these unclassified detections is very small, though it is interesting to note that unlike the earlier analyses, including the Off by One detections actually decreases the average number of detections per 5 minutes. This is a result of the fact that many of the Off by One detections included here occur in 5 minute periods that were not originally counted.

Table 3: Comparison of unclassified detections per 5 minutes with and without Off by One codes.

	Valid Detections Plus Off by One Detections (Denominator Sample Size)	Valid Detections Only (Denominator Sample Size)
Detections /5 Minute Period**	2.31 (n = 2506)	2.35 (n = 1660)

** These five minute periods correspond to all time periods outside of diary reading sessions that have at least one RFID detection.

Lastly, we looked at the impact of the Off by One detections and the unclassified detections that are traced back to someone else's reading. These detections can be considered as false positives from spillover. For this analysis, we looked at both the average number of detections per reading event and the average number of detections per 5 minutes. Note that the number of diary reading sessions that corresponded to another panelist's unclassified reading sessions is relatively low. As such, the overall number of mistakenly classified detections per event during these times may seem high at first glance, but it is driven by the overall low number of false positives from spillover. Looking at the Off by One detections, their impact is again rather negligible.

Table 4: Comparison of false positive detections with and without Off by One codes.

	Valid Detections Plus Off by One Detections (Denominator Sample Size)	Valid Detections Only (Denominator Sample Size)
Detections / (Total Events in Diary with at least one Detection)	25.3 (n = 47)	23.9 (n = 46)
Detections /5 Minute Period*	3.8 (n = 316)	3.6 (n = 306)

* These five minute periods correspond to diary reading sessions that have at least one RFID detection.

Taken together, the ratio of valid detections per five minutes coinciding with diary noted reading sessions, compared to unclassified detections or false positives per five minutes changed marginally when including an Off by One edit. Said another way, the total number of detections requiring this edit in order to be credited as valid detections were marginal and equally distributed between known false positives and unclassified detections. In order to better evaluate false positive detections, the unclassified detections from the diary need to be more clearly defined in a controlled laboratory environment permitting a more rigorous analysis. We performed the laboratory tests in the summer of 2007 and the results are described below.

Structured Protocol Laboratory Tests

The laboratory tests provided a rich and detailed database from which to examine the detection capabilities of the PPM and RFID tagged magazines. At the most granular level, the individual PPMs can report media codes and RFID detections at the 15 second level. Thus, there are some 5,163 units of analysis across all the tests and test subjects. In Table 5, a number of results are presented showing the possible detection categories, instances, instances as a percent of the total in a column, and whether other test subjects were in the same room at the time of detection (and if so, how many people). The other detection categories include those where there are known false positives from other magazines in the same room, unused magazines that were in the same room or a magazine that was being read in another room. It is important to note that the number of Off by One codes (recall these data from the earlier, diary-based Friends and Family tests) have all but disappeared for the more recent laboratory tests. This reduction and virtual elimination of Off by One codes is the result of an improved design for the tag technology used in this test. Operationally, reading is defined as the opening of the cover of a magazine as the RFID tag, in this application, is designed to transmit codes upon the magazine cover being opened.

Table 5: All possible detections and outcomes, across all tests, by number of 'other' test subjects present in same room.

Other Persons Present	0		1		2	
Result	**Instance**	**%Total**	**Instance**	**%Total**	**Instance**	**%Total**
No Code Detected	**1,028**	**93.5%**	**1,950**	**88.3%**	**1,539**	**83.0%**
Correct Code Detected	**60**	**5.5%**	**171**	**7.7%**	**148**	**8.0%**
Inexact Code	0	0.0%	0	0.0%	8	0.4%
FP from Other in Same Room	0	0.0%	85	3.8%	152	8.2%
FP from Unused Magazine Detected	1	0.1%	0	0.0%	8	0.4%
FP from Other in Another Room	11	1.0%	2	0.1%	0	0.0%

Note: Other test subjects were 5 to 15 feet apart, while persons outside of the room were at least 50 feet away from test subjects.

The most apparent result from Table 5 is that the vast majority of the 15 second blocks (the raw data units of analysis) have a "no code" status. The RFID tag does not transmit constantly so this result is consistent with our expectations. This is unlike media codes which are constantly being detected by the PPMs. The table also shows that correct code detections – detections for the correct magazine that the test subject was looking at during the specific test – are the next most populated variable. The numbers appear small because the number of possible detections is quite low given the fact that the tagged magazine would not send a constant data stream for detections. In addition, a pattern of increased false positive (denoted FP) detections[1] increases from the left to the right of the table, with zero other panelists being an obvious result all the way to 2 others being present having almost the same number of overall detections as correct code detections. However, this aggregate table hides some important details such as the distances between panelists and the presence/absence of other media codes – both of which could impact the detection levels.

In Table 6, we remove the row designated for the no code detected category to focus on the other categories more closely and recalculated the column percentages. More importantly, the detections are now divided into two groups: detections with encoded audio in the background during the reading test and those without. The outcomes are virtually identical whether background audio was present or not, and the distributions matched the overall test distribution. This is a significant finding because the PPM has been designed with an eye towards the measurement of cross-media usage in a single-source, passive sample design, and this test illustrates that the PPM can provide accurate code detection of media whilst simultaneously gathering RFID print detections. This result is not unexpected because the PPM technology was designed to detect media codes on another 'layer' rather than compete for media detections with the RFID codes, so no difference should be attained in detections. In terms of its potential for multi-media measurement, this result is extremely encouraging and significant. There is a small difference in false positive detections from other test subjects, but this is more than likely a random result.

Table 6: Detections by category with and without background encoded audio content.

	No Background		Yes Background		
Result	Instance	% Total	Instance	% Total	% Overall
Correct Code Detected	186	60%	193	59%	59%
FP from Other in Another Room	12	4%	1	0%	2%
FP from Other in Same Room	107	35%	130	40%	37%
FP from Unused Magazine Detected	4	1%	5	2%	1%

Another question arises from the test of cross-media code detections: how well did the PPM continue to detect media codes for encoded audio (both radio and TV) while being used as a RFID reader for passive print detections? Here, the big news is that there is really that there is no news to report: all media codes tested during the RFID laboratory tests were detected as they would be under normal circumstances. In all cases, the proper media outlet would have been credited for the correct quarter hour.

Turning to the effects of distance between test subjects, Table 7 presents the same data as above but this time broken out by the distance range between panelists. It should be noted that the RFID tags that were used in this test were not 'tuned' down (or up) in terms of their range, to any extent, and so we expect detections to occur across subjects and across some considerable distances. It is possible to tune the tags to restrict detections but for this laboratory test we preferred to be conservative and test the tags in a raw form. Again, no code detections are the dominant outcome but are omitted from the table for clarity. Comparing the correct code detections with false positive code detections, we can see that the incidence of false positive codes from other magazines in the same room increases as distance between panelists decreases. The 50+ feet distance refers to test subjects reading in rooms some considerable distance from the main testing area. Once the distance between the tagged magazines reaches a large enough threshold, the odds of any false positive detection are reduced considerably. Now the question is whether the false positive codes from other magazines are related more to the distance range of the tag code transmission or to the number of persons in the room reading, or both equally. To answer this, we can look more closely at the various interactions involved in Tables 8, 9 and 10, below.

[1] We use the term 'detections' as opposed to 'exposures' as we have not yet defined a proper reading exposure at this point, and the data remain unedited. We return to this question later in the paper.

Table 7: Detections by category and various distance thresholds.

Distance thresholds (feet)	5		10		15		50+	
Result	**Instance**	**%Total**	**Instance**	**%Total**	**Instance**	**%Total**	**Instance**	**%Total**
Correct Code Detected	**155**	**49.7%**	**62**	**54.4%**	**102**	**68.9%**	**60**	**83.3%**
Inexact Code	6	1.9%	0	0.0%	2	1.4%	0	0.0%
FP from Other in Same Room	147	47.1%	50	43.9%	40	27.0%	0	0.0%
FP from Unused Magazine Detected	3	1.0%	2	1.8%	3	2.0%	1	1.4%
FP from Other in Another Room	1	0.3%	0	0.0%	1	0.7%	11	15.3%

In Table 8, we isolate the detections where no other test subjects were present in the same room and so the distance threshold is essentially 50+ feet. These numbers are identical to those in table 7 but are presented again for clarity; where there are no others in the room there remain false positive detections because some of the tagged magazines had obviously large transmission ranges. Table 9 isolates the detections observed with 1 other test subject reading in the same room and so there are multiple cases of distance thresholds ranging from 5 to 15 feet away from the test subject. The expected pattern of increased correct code detections and decreased false positive detections from other test subjects in the same room is evident here. At 5 feet apart, test subjects reported some 58% correct code detections versus 41% of false positives from a test subject reading nearby.

Clearly, the distance between the test subjects is important in terms of correct code detections, but Table 10 makes plain the fact that the number of other test subjects in the room makes the larger impact on correct code detections overall. The pattern of correct code detections increases only slightly with increasing distances between test subjects and is clearly impacted by the presence of 2 other tagged magazine readers in the same room. It is important to note, however, that the number of false positives shown on the chart is an aggregation from the other 2 test subjects in the room. The interaction between distance and the number of competing readers will diminish in importance as tags are tuned appropriately in future tests.

Table 8: Detections by category, with zero other persons in room and a 50+ feet distance threshold.

0 persons	50+ feet	
Result	**Instance**	**%Total**
Correct Code Detected	**60**	**83.3%**
FP from Unused Magazine Detected	1	1.4%
FP from Other in Another Room	11	15.3%

Table 9: Detections by category, with one other person in room and three distance thresholds.

1 person	5		10		15	
Result	**Instance**	**%Total**	**Instance**	**%Total**	**Instance**	**%Total**
Correct Code Detected	**59**	**57.8%**	**38**	**62.3%**	**74**	**77.9%**
FP from Other in Same Room	42	41.2%	23	37.7%	20	21.1%
FP from Other in Another Room	1	1.0%	0	0.0%	1	1.1%

Table 10: Detections by category, with two other persons in room and three distance thresholds.

2 persons	5		10		15	
Result	**Instance**	**%Total**	**Instance**	**%Total**	**Instance**	**%Total**
Correct Code Detected	**96**	**45.7%**	**24**	**45.3%**	**28**	**52.8%**
Inexact Code	6	2.9%	0	0.0%	2	3.8%
FP from Other in Same Room	105	50.0%	27	50.9%	20	37.7%
FP from Unused Magazine Detected	3	1.4%	2	3.8%	3	5.7%

A final variable of interest concerns the binding type of the magazines tested. As we mentioned earlier, a variety of magazines were selected for study – a representative sample of varying thicknesses and corresponding binding styles (simple saddle stitch and perfect binding). When the data were analyzed by binding style, we see that the correct code was detected at the same level across both binding styles – at the same level as the total percentage of correct code detections (Table 11). All of the other possible results were virtually identical between the two binding styles; there is no impact of binding style on detection.

Table 11: Detections by category across binding style.

Result	% of Perfect	% of Saddle	% of Total
Correct Code Detected	60%	58%	59%
FP from Other in Another Room	1%	3%	2%
FP from Other in Same Room	37%	38%	37%
FP from Unused Magazine Detected	2%	1%	1%

From 'detections' to 'exposures'

The test results illustrate that detections of RFID enabled magazines are consistent and logical. Overall, our laboratory tests showed that when we detected an exact code, the code came from the expected magazine about 59% of the time. The remainder of the time the code detected was from another magazine. Clearly, these raw data at 15 second units of analysis are not ideal for assigning even a basic metric of magazine exposure (i.e., defining exposure of less than 60% of the time would not be acceptable as a measure of readership). In addition, there would be plenty of bias associated with this metric. However, a single obvious choice for an electronic, passive readership metric is not clear and is the subject of considerable debate in the print industry. From page and advertisement views, to average issue audience, there are a number of useful metrics required by the industry. We certainly do not assume that we are prepared to answer this question, but we can say that providing an electronic and more granular measurement than what is available today using passive electronic recall methodology is a significant step forward.

From our Friends and Family test, we know that the average number of detections per 5 minutes for a "correct" code is higher than the average number of detections per 5 minutes for unclassified or false positive codes. Using this information, we examined the total detections that were sensed by a test subject's PPM that were attributed to the various magazines during the laboratory tests. We then developed a relatively simple algorithm that assigned a person to each magazine that was read during each 15 minute time period. Hence, we were able to rather quickly generate a quarter hour reading metric.

While the laboratory tests provided a limited amount of data, and the algorithm has not been refined fully, we were able to correctly assign the magazine to the actual panellist who was reading it 88% of the time. With more tuning of the strength of the tags and a refinement of the matching algorithm for editing, it should be rather straight-forward to increase the accuracy with which we are able to detect reading sessions at a highly granular level. While it would not be surprising to have correct match rates for quarter hours in the 90% to 95% range with continued development, it is important to note that the highly granular 15 second measurement scale used in the laboratory tests should be redefined in follow up tests to better match the eventual metric required by the industry.

Our operational definition of a reading exposure is associated with the activity of opening the magazine cover; different tag designs can be used for variations on this definition. Here, important ethnographic research started by McConochie and Bailey (2004) designed to guide passive measurement of print, provides the most useful advice on the operational definition of reading. The PPM cannot judge actual reading (not unlike peoplemeters in TV measurement panels) so some proxy definition of reading must be used. In our case, we felt it appropriate to use "opening the cover" as a definition for this initial proof of concept research. The ethnographic research shows that magazine reading is highly dynamic across individuals, but that "page flipping" is the most common description of reading amongst both light and heavy readers. In the end, we need to determine a behaviorally based metric that is superior to the recall and opinion based metrics of current practice.

6. Discussion

Two years ago at the Worldwide Readership Research Symposium, Mattlin and McDonald (2005) discussed, in considerable detail, the promise and perils of RFID technology. It is worth revisiting some key elements of their discussion here. The implementation of RFID will most certainly require the use of panels rather than large 'sweep' or 'one shot' samples as in today's print measurement. Evaluating whether equivalent sample sizes can be achieved is an important issue and one that we feel might have been oversimplified and certainly presents the worst case scenario.

Our extensive experience with building and running PPM panels shows us that there is planned and unplanned turnover, plus daily editing-out of members. In practice, the number of different people appearing on the panel will be larger than the average panel size. In fact, the total sample size over a month is 50 percent bigger than the average panel size over the same month. Over 26 weeks (the number of observations for bi-weeklies), the increase will be even larger. Using this information to develop new estimates of 'effective sample size' will produce numbers considerably larger than those discussed by Mattlin and McDonald (2005). Suffice to say, it if far too early to suggest that panels are not an option for readership research.

Similarly, we believe panels can be used for the estimation of average issue audiences. Because of panel churn, we can not estimate average issue audience directly as this is a cume figure (i.e., how many people looked into this issue over the course of, say, a month). While in the past, Arbitron and other measurement companies have used models for cume measurements, panels afford us the opportunity to use objective, empirical data to estimate cume audiences over a week, month, etc…(Gluck and Stinnett, 2007). We could argue that this is an improvement over the existing one shot sample 'recent reading' approach to estimation and the inherent biases associated with it. The remaining research question is: How do the limitations of 'recent reading' compare with the supposed limitations of RFID and passive electronic measurement of print?

The development of passive, electronic readership measurement provides a viable and, arguably, improved, alternative to today's readership currency measurement. Future research in laboratory settings with different tag technology, and in the field with human subjects, will offer even greater insights into the potential for passive print measurement. In addition, continued ethnographic research is a necessity to ensure that the metrics and measurement standards used provide improved and reliable audience information to the print industry. A fruitful avenue for future testing of passive, electronic readership measurement falls within Project Apollo™ - the single-source, multi-media and consumer purchasing behavior service jointly run by Arbitron and Nielsen. The inclusion of electronic print measurement into Project Apollo is an exciting possibility.

References

Gluck, A. and Stinnett, S. (2007) "What happens after seven days?" Proceedings of the 2nd Annual ARF Audience Measurement Symposium, NY, NY

Mattlin, J.A. and McDonald, S.C. (2005) "Passive measurement of print: The promise and perils of RFID technology." Worldwide Readership Research Symposium, Prague.

McConochie, R. and Bailey, J. (2004) "Progress toward passive measurement of print: How consumers "really" read magazines" Proceedings of the ARF-Esomar WAM, Geneva.

McConochie, R., Bailey, J. and Gluck, A. (2006) "Toward silo-shattering measurement of readership" Proceedings of the 1st Annual ARF Audience Measurement Symposium, NY, NY.

Octave Technology (2007) "RFID Industry Status Report" Prepared for Arbitron., Ellicott City, MD.

Patchen, R. H. and Webb, B.M. (2002) "The future is now: The very latest findings from the US market launchof the Portable People Meter in Philadelphia." Proceedings of the ARF-Esomar WAM, Cannes.

Pellegrini, P.A. (2006) "The mainstream and the clickstream." Proceedings of the 1st Annual ARF Audience Measurement Symposium, NY, NY.

Pellegrini, P.A. (2005) "Listen without prejudice: Passive measurement for radio, TV and beyond." VUE (the monthly magazine of the Marketing Research and Intelligence Association, Canada), Toronto.

Pellegrini, P.A. and Purdye, K. (2005) "Station to station: Measuring radio audiences with a PPM panel in Quebec." Proceedings of the ARF-Esomar WAM, Montreal.

Pellegrini, P.A. and Purdye, K. (2004) "Passive versus button-pushing: A comprehensive comparison from parallel TV panels in Quebec." Proceedings of the ARF-Esomar WAM, Geneva.

Acknowledgements

The authors thank Ken Purdye, Scott Stinnett and Erwin Ephron for helpful comments. In addition, we are grateful to Brett Hoffman and numerous test subjects volunteers for their help in completing this research.