electronic International Journal of Time Use Research

volume 10
number 1
november 2013
ISSN 1860–9937
dx.doi.org/10.13085/eIJTUR.10.1

contents

electronic ***International Journal of Time Use Research***
2013, Vol. 10, No. 1, 1-8 ***dx.doi.org/10.13085/eIJTUR.10.1.1-8***

Student time allocation and self-rated performance – Evidence from a sample survey in Sicily (Italy)

Massimo Mucciardi

Assistant Professor Massimo Mucciardi
Department of Economics, Statistics, Mathematics and Sociology
University of Messina
Messina, Italy
e-mail: mucciard@unime.it

Abstract

The paper collects the results of a survey performed in 2010 aimed at analysing how high school students in the province of Messina in Sicily (Italy) spend their time. Principally, is analyzed the interaction between use of time, scholastic performance and time dedicated to study. So, we propose an estimation model for the daily study-time of students. From a methodological point of view, using a two stage regression procedure to estimate self-rate performance (S_{rpe}) and time devoted to study (T_{stu}) allows to correct the estimates by simultaneity effects between these variables. In the first stage, the self-rate performance at school is estimated in a reduced form and is used as a proxy of scholastic performance in the second step. Next, we run an ordinal regression in order to estimate the hours dedicated to study declared by the student. The results obtained show that students with a high expected value of study-time come from lyceum, they are mostly females, and tend to read more. Furthermore, they have satisfactory scholastic performance, are helped by their mothers when they do their homework, have a lower-than-average age difference with their mothers, but a higher-than-average age difference with their fathers.

JEL-Codes: C13; C30; C36

Keywords: Study-time; Use of time; Self-rate performance; Two stage regression procedure; High school

The first results of this study come from the pilot study were presented at the Population Study Days, Padua (Italy) 16 to 18 February 2005.

1 Introduction

Most of the previous studies on the effect of performance and time devoted to study have been very focused. Schmidt (1983) analyzes a survey sample of 216 students at the University of Wisconsin-Madison and estimates a Cobb-Douglas educational production function using both Ordinary Least Square and Full Information Maximum Likelihood methods. He finds an elasticity of performance with respect to hours of class attendance of 0.215 and of study hours of 0.017. Romer (1993) considers the class attendance as an endogenous factor and tries to correct the endogeneity effect by introducing some proxies for motivation in the estimates of the performance function. Bratti and Staffolani (2002) consider the students' performance as a direct consequence of the allocation between time devoted to study and leisure time. Dolton, Marcenaro and Navarro (2003) find that the lectures are four times more productive than self-study. As pointed out by Olivares (2002), the study time-grade association literature has provided inconsistent findings: some researchers have found a positive association, others a negative association, and yet others no association between study time and grades. Unlike the academia, high school students are obliged to attend the lessons. This implies that lessons attendance can be unable to explain student's performance. On the other hand, several factors can influence the high school students' performance; in general, the level of the grades may depend on the time devoted to study and vice-versa, even if this relation may be influenced by other factors such as the different courses of study, the efficiency of the teachers, and environmental and motivational factors.

In light of these considerations, this paper collects the results of a survey performed in 2010, aimed at analysing how high school students in the province of Messina in Sicily (Italy) spend their time. The main purpose of research is to understand the relationship between use of time, self-rated performance (proxy of scholastic performance) and time devoted to study (study-time). So we propose an estimation model for the daily study-time of students. From a methodological point of view, using a two stage regression procedure to estimate self-rated performance (S_{rpe}) and time devoted to study (T_{stu}) allows to correct the estimates by simultaneity effects between these previous variables. The paper outline is as follows: in the next section we present the research model and the data utilized; in the third, we deal with the problem of the estimation model used for the study-time, and in the last section we discuss the main results obtained and on the possible ways the research may be developed.

2 The research model and data

The research model used is based on a study performed by Sabbadini and Palomba (1994) on the use of time by men and women. We divided the students' time into 4 categories: "physiological activities", "activities for the family", "study" and "leisure time". We interviewed about

1800 students from various types of public educational institutions[1] (Lyceum, Technical institute and Vocational school), using a specially drawn-up questionnaire (see table 1 for sample features).

Table 1
Sample features

Variable	Students
Gender	
Male	865 (47.5%)
Female	955 (52.5.%)
Age	
Mean (SD)	16.87 (1.55)
Ethnicity	
Italian	1827 (100%)
Type of school	
Lyceum	798 (43.7%)
Technical institute	751 (41.1%)
Vocational school	278 (15.2%)
Year attended	
1st	474 (25.9%)
2nd	341 (18.7%)
3th	368 (20.1%)
4st	400 (21.9%)
5st	245 (13.4%)

n=1827, Source: Own performed survey 2010, own calculations.

The sampling plan takes into account the percentage and geographical distribution of educational facilities in the province of Messina in Sicily (Italy). With regard to the questionnaire used for the research, students first completed the section on personal and family details, and, later, the specific sections on the "use of time". The analysis was performed using the "overall average duration" of each activity during a standard day. To this purpose, we also considered an indicator of frequency (every day/3-5 times a week/1-2 times a week/1-2 times a month/never) for all the activities except physiological activities and study (see table 2). The questionnaire was self-compiled, but under the supervision of expert testers. Furthermore, detailed information on family composition, educational level and working activity of parents are

[1] It is important to highlight that education in Italy is compulsory from 6 to 16 years of age, and is divided into five stages: Kindergarten, Primary School, Lower Secondary School, Upper Secondary School (which corresponds to the High-School level) and University. In particularly, the Upper Secondary School situation varies, since there are several types of schools differentiated by subjects and activities. The main division is between the "Lyceum", the "Technical Institute" and the "Vocational School". Any kind of Upper Secondary School that lasts 5 years (age 14 to 18) grants access to the final exam. This exam grants access to University.

considered. Sadly, for confidentiality reasons, it was not possible to detect the students'grade point average.

Table 2
Average distribution of time during the typical day of a student

Activity	Time		Frequency of activity (%)				
			Every day	3-5 times/ week	1-2 times/ week	1-2 times/ month	Never
Sleep (night and/or afternoon)	8 h	7 min	*100.00*	-	-	-	-
Personal hygiene and bodily care	1 h	20 min	*100.00*	-	-	-	-
Eating breakfast	0 h	13 min	*70.63*	-	-	-	*29.37*
Eating lunch	0 h	40 min	*99.86*	-	-	-	*0.14*
Eating the evening meal	0 h	41 min	*99.54*	-	-	-	*0.46*
Housework	1 h	5 min	*29.74*	*16.04*	*29.19*	*6.66*	*18.38*
Shopping	0 h	31 min	*3.34*	*10.28*	*32.99*	*16.20*	*37.20*
Looking after younger brothers/sisters	0 h	22 min	*10.70*	*2.93*	*3.93*	*1.89*	*80.56*
Work outside the family	0 h	39 min	*5.49*	*3.04*	*5.58*	*3.25*	*82.64*
Other family activities	0 h	10 min	*2.21*	*0.84*	*0.65*	*0.15*	*96.17*
Time spent travelling to and from school	0 h	51 min	*100.00*	-	-	-	-
Time spent at school	5 h	44 min	*100.00*	-	-	-	-
Study at home	2 h	33 min	*100.00*	-	-	-	-
Sport and gym	1 h	13 min	*10.42*	*28.60*	*19.86*	*2.45*	*38.67*
Watching television	2 h	9 min	*87.50*	*7.28*	*2.19*	*0.06*	*2.97*
Listening to music	1 h	44 min	*64.87*	*19.04*	*11.41*	*1.09*	*3.59*
Use of computer and the internet	1 h	5 min	*23.97*	*19.32*	*20.91*	*5.35*	*30.45*
Going out with friends	2 h	53 min	*31.80*	*24.96*	*34.54*	*2.43*	*6.26*
Reading (not school books)	0 h	31 min	*6.27*	*10.51*	*15.35*	*19.77*	*48.10*
Reading newspapers or comics	0 h	24 min	*11.52*	*11.34*	*23.84*	*10.59*	*42.70*
Going to the cinema or theatre	1 h	20 min	*0.41*	*1.03*	*10.54*	*52.12*	*35.90*
Going dancing (dancing school, disco)	1 h	25 min	*1.48*	*3.04*	*15.53*	*24.74*	*55.21*
Games arcades, playing cards, playstation	0 h	48 min	*12.74*	*15.07*	*17.01*	*7.84*	*47.34*
Painting, playing musical instruments, crafts	0 h	24 min	*4.70*	*3.90*	*6.98*	*7.84*	*76.59*
Voluntary work and parish activities	0 h	22 min	*1.26*	*1.84*	*10.84*	*4.54*	*81.52*
Other 1	0 h	6 min	*0.97*	*0.98*	*1.10*	*0.52*	*96.43*
Other 2	0 h	1 min	*0.17*	*0.15*	*0.14*	*0.20*	*99.34*

Source: Own performed survey 2010, own calculations.

Thus the problem of the measurement of scholastic performance was solved by asking the students a synthetic judgment on their subjective satisfaction about their own scholastic out-

comes[2]. This implies that the performance of each student is classified as a dichotomous variable (satisfactory/ unsatisfactory). We called this proxy variable self-rated performance (S_{rpe}).

3 Estimation procedure[3]

As noticed in the introduction, we used a well-known technique (Green, 2000) to estimate self-rated performance (S_{rpe}) and the study-time (T_{stu}) so as to correct the estimates by simultaneity effects between these variables. So the model used to estimate the time devoted to study involves a two-stage regression procedure[4]. In the first stage, the student's self-rated performance (P_{stu}) at school (satisfactory/unsatisfactory) is estimated in a reduced form, and is used as a proxy of scholastic performance in the second step.

Then in the first stage we applied a logistic regression:

$$(1) \qquad S_{rpe} = f\left(F_{spo}, P_{ide}, M_{eds}, F_{prs}, T_{pla}, T_{dan}, T_{out}\right)$$

where F_{spo} = frequency with which the student practises sport or goes to the gym (every day/3-5 times a week/1-2 times a week/1-2 times a month/never); P_{ide}= proxy of the type of student (highly idealistic/idealistic/concrete); Meds= mother's educational level (high/medium/low); F_{prs} = father's professional status (high/medium/low); T_{pla}=time devoted to playing at a games arcade; T_{dan}=time devoted to going dancing and Tout=time spent outside with friends. As a research hypothesis we assumed that these variables were exogenous. The student's self-rated performance explained by the regression (1) was used as an instrumental variable in the second stage. So we run an ordinal regression model (McCullagh, 1980) in order to estimate the hours devoted to study (T_{stu}) declared by the student (over 4 hours/between 3 and 4 hours/between 2 and 3 hours/between 1 and 2 hours/less than one hour):

$$(2) \qquad T_{stu} = f\left(D_{fag}, D_{mag}, S^{*}_{rpe}, T_{rea}, T_{sch}, G_{stu}, Y_{att}, M_{hel}\right)$$

where D_{fag} = difference between the student's and his/her father's age, D_{mag} = difference between the student's and his/her mother's age, S^{*}_{rpe}= theoretical values of the student's self-rated performance, explained by (1), T_{rea} = time devoted to reading non-scholastic books; T_{sch} = type of school (Lyceum/Technical institute/Vocational school), G_{stu} = student's gender (male/female), Y_{att} = year attended (1^{st}-5^{st}) and M_{hel} = mother's help with study (yes/no).

In an ordinal regression model, various "link" functions may be used. In this case, the logit function ensured the best fit. The output of an ordinal regression gives the probability that a generic unity falls between the categories of the response variable: in this way we obtained the expected value of study-time for each unit (student).

---

2 It seems useful to point out that the question was referred to the whole of the teachings studied and was not related to the overcoming of the school year.

3 All analyses were performed using SPSS 16.0 and Excel 2003.

4 Among the various models examined the regressions (1) and (2) have provided the best performance according to the theory of "two-stage regression procedure". (For more details see Green and Wooldridge).

4 Main results

An average, high school student in the province of Messina (Italy) spends his/her standard working day as shown in table 2. The most interesting outcome seems to be that students spend around 84% of their time performing daily activities, while the remaining 16% is devoted to leisure time or (rather irregular) family activities. The proportion spent studying is 11% if referred to the entire day, or 13% if referred only to daily activities.

Let us now examine the results of equation 1 and 2 (table 3-4).

Table 3
Results of the logistic regression (first stage)

Variables	Coeff.	Sd.Err.	P-value
Constant	.685	.233	.003
F_{spo}			.006
3-5 times at week	.325	.204	.111
1-2 times at week	.461	.151	.002
1-2 times at month	.527	.168	.002
Never	.760	.503	.130
Every day (ref.) every day			
P_{ide}			.000
Highly idealistic	1.086	.163	.000
Idealistic	1.065	.164	.000
Concrete (ref)			
M_{eds}			.008
Low	-.235	.181	.195
Medium	-.465	.154	.003
High (ref.)			
F_{prs}			.000
Low	-.669	.222	.003
Medium	-.503	.131	.000
High (ref.)			
T_{pla}	-.169	.058	.004
T_{dan}	-.098	.033	.003
T_{out}	-.121	.037	.001

Note: S_{rpe} =1 if the personal assessment of the performance is satisfactory,
S_{rpe} =0 if the personal assessment of the performance is unsatisfactory
Sample size = 1439; Nagelkerke R Square = 0.15; goodness of fit Chi^2 (13) = 149
p-value = 0.000, Hosmer-Lemershow test = 8.494 p-value = 0.387
Ref. =reference category
Source: Own performed survey 2010, own calculations.

From the first stage estimate (logistic regression) it emerges that students who display satisfactory performance (S_{rpe}) have mothers with a high educational level and fathers with a high professional status.

Table 4
Results of the ordinal regression (second stage)

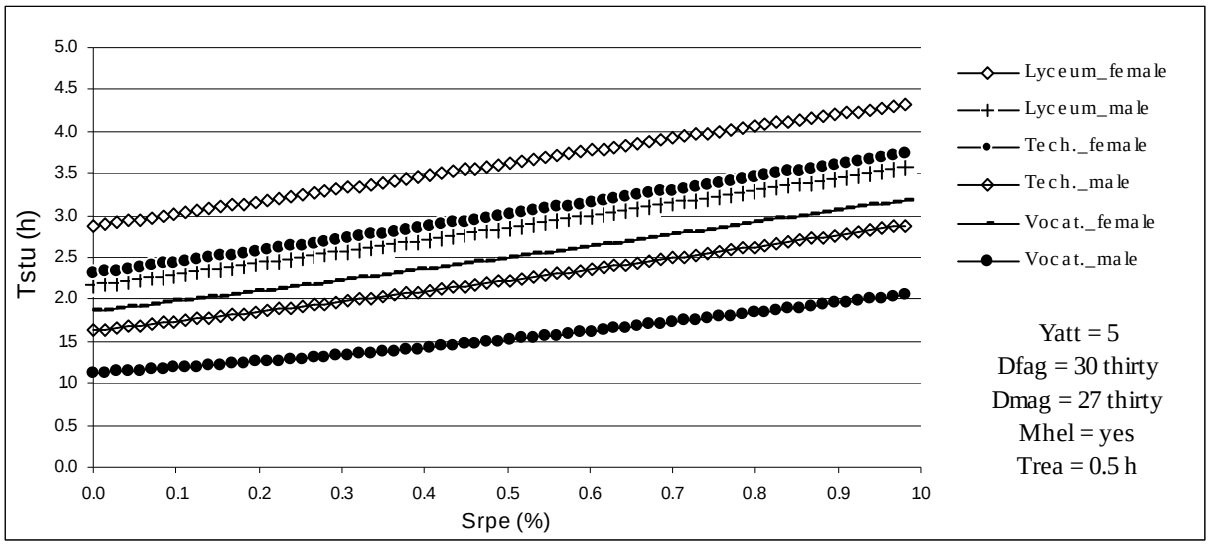
Variables	Coeff.	Sd.Err.	P-value
$1<T_{stu}<=2$	.864	.409	.035
$2<T_{stu}<=3$	2.608	.415	.000
$3<T_{stu}<=4$	4.030	.422	.000
$T_{stu}>4$	5.215	.430	.000
S^*_{rpe}	2.152	.372	.000
D_{fag}	.049	.015	.001
D_{mag}	-.037	.016	.018
T_{rea}	.572	.072	.000
Ytt (1^{st})	-.381	.175	.030
Yatt (2^{nd})	-.894	.185	.000
Yatt (3^{th})	-.556	.181	.002
Yatt (4^{st})	-1.162	.183	.000
Yatt (5^{st}) (ref)	0	.	.
Gender (male)	-1.232	.110	.000
Gender (female) (ref.)	0	.	.
T_{sch} (lyceum)	1.978	.187	.000
T_{sch} (technical institute)	1.071	.176	.000
T_{sch} (vocational school) (ref.)	0	.	.
M_{hel} (Yes)	.366	.104	.000
M_{hel} (No) (ref.)	0	.	.

Note: Dependent variable =1 if $T_{stu}>4$; 2 if $3<T_{stu}<=4$, 3 if 2<Tstu<=3;
4 if $1<T_{stu}<=2$ and 5 if $T_{stu}<=1$,
Sample size = 1440; Nagelkerke R Square = 0.37; goodness of fit Chi^2 (12) = 579 ,
p-value = 0.000, Ref. =reference category,
Source: Own performed survey 2010, own calculations.

Moreover, they are highly idealistic, play little sport, on average, tend to spend little time in discos and games arcades, and go out with their friends only rarely (see first stage, table 3). Ordinal regression (second stage, table 4) shows that students with a high expected value of study-time (T_{stu}) come from lyceums, they are mostly females, and tend to read more. Furthermore, they have satisfactory scholastic performance, are helped by their mothers when they do their homework, have a lower-than-average age difference with their mothers, but a higher-than-average age difference with their fathers. The year attended plays an interesting role. In fact, students attending the 1^{st}, 3^{nd} and 5^{th} year have an expected value of study-time higher than students in their 2^{rd} and 4^{th} years: this tends to confirm the importance of 'transition years'. Having obtained the estimates it is possible to suitably modify the variables of interest and ex-

trapolate various profiles of study-time (T^{stu}) by simulating hypotheses such as the provenance from different schools, the self-rate performance and the year attended (see figure 1).

Figure 1
Study-time (T_{stu}) vs self-rate performance (S_{rpe}) analyzed for different schools



Source: Own performed survey 2010, own illustration.

As one can see, the gender and type of school strongly influence the relationship between student's self-rated performance and study-time.

5 Conclusion

The present research sought to better understand the nature of the self-rated performance and study-time by examining the effects of the student time allocation, individual characteristics of the students and some socio-demographic characteristics of the parents. In order to do so, we have introduced a two-stage regression procedure for the student's self-rated performance and student time allocation. Although most of the work reported in the literature concerns the context academic (Olivares, 2002), the results obtained in this study seem to confirm that the satisfaction in school performance is a good predictor of the study time. Respect to previous studies, we identify new predictors such as gender of the student, type of school attended (proxy of the course difficulty) and time devoted to reading non-scholastic books. However, we believe that the model proposed and the results obtained should also be evaluated in relation to the empirical nature of the study and the geographical context. In fact there is a high risk that inaccuracies may occur in this type of sample survey especially in terms of the exact measurement of the timing of the daily, weekly and monthly activities. As mentioned above, the survey is missing some important variables such as the characteristics of the teachers (teacher effectiveness) and student's grades. We consider this work as a pilot study therefore, we aim at replicating this

survey in other scholastic contexts in order to validate the results obtained. A study of this kind is currently in progress by the author.

References

Bratti, M. and S. Staffolani (2002), Student time allocation and educational production function, Working Paper No. 170, Department of Economics, University of Ancona (Italy).

Dolton, P., Marcenaro, O. and L. Navarro (2003), The effective use of student time – A stochastic frontier production function case study, in: *Economics of Education Review*, Vol. 22, No. 6, 547-560.

Greene, W. H. (2000), Econometric Analysis, Prentice-Hall.

McCullagh, P. (1980), Regression models for ordinal data, in: *Journal of the Royal Statistical Society, Series B*, Vol. 42, No. 2, 109-142.

Olivares, O. J. (2002), An analysis of the study time-grade association, Radical Pedagogy, CAAP.

Romer, D. (1993), Do students go to class?, in: *The Journal of Economic Perspectives*, Vol. 7, No. 3, 167-174.

Sabbadini, L. and R. Palomba (1994), Tempi diversi – L'uso del tempo di uomini e donne nell'Italia di oggi, Italian Institute of Statistics (ISTAT).

Schmidt, R. M. (1983), Who maximizes what? – A study in student allocation, in: *American Economic Review*, Vol. 73, No. 2, 23-28.

Wooldridge, J.M. (2002), Econometric Analysis of Cross Section and Panel Data, MIT Press.

electronic ***International Journal of Time Use Research***
2013, Vol. 10, No. 1, 9-37 ***dx.doi.org/10.13085/eIJTUR.10.1.9-37***

A measure of concentration of the use of time, with an application to the pattern of daily leisure activities

Jorge González Chapela

Dr. Jorge González Chapela
University of Girona and Centro Universitario de la Defensa de Zaragoza
Academia General Militar
Ctra. de Huesca s/n
E-50090, Zaragoza, Spain
e-mail: jorgegc@unizar.es

Abstract

Drawing an analogy with industry concentration, a well-grounded measure of individual concentration (or specialization) of the use of time is presented. Equipped with this measure, we explain and provide evidence of a "division of leisure" effect on the organization of daily leisure activities. A demand model featuring subsistence daily leisure shows that the concentration of leisure can vary with the quantity of leisure available. Sequential moment conditions and the exogenous possibility of more leisure brought about by the weekend unveil an asymmetrically U-shaped response in a sample of employed German men.

JEL-Codes: J22

Keywords: Time allocation; Herfindahl-Hirschman index; IV estimation

This paper originated from a conversation with Ricard Torres and was written during my stay at the University of Girona, whose hospitality I gratefully appreciate. I am also indebted to the anonymous referee, Gerhard Bosch, Dan Hamermesh, Joachim Merz, Xavier Sala-i-Martin, Sabine Sonnentag, and seminar participants at the Sophia Antipolis campus of the SKEMA Business School for helpful comments, and to Kristina Kott, of the German Federal Statistical Office, for able assistance with the data. Financial support from the Spanish Ministry of Education (ECO2011-29751/ECON) is gratefully acknowledged.

1 Introduction

The objective of this paper is twofold. The first objective is to investigate a measure of concentration of an individual's use of time. For good reasons, the statistical analysis of time-use observations has become widespread in economics and other social sciences. A measure of individual concentration of the use of time seems necessary for any quantitative study attempting to explain the causes or consequences of variation in the degree of concentration (or specialization) of the use of time. The second objective is to examine theoretically and empirically the pattern of daily leisure activities in terms of this measure[1]. While on working days many people relax just watching the TV, on non-working days many people watch the TV but also go out. Can variations in the concentration of leisure activities be explained on the basis of the quantity of leisure available? In other words, is there a "division of leisure" effect on the organization of daily leisure activities? From a theoretical viewpoint, the answer to this question informs about the structure of consumer preferences driven the demand for daily leisure. From a more empirical perspective, it contributes to understanding the determinants of the demand for variety (Gronau and Hamermesh, 2008), which, in turn, may prove useful for organizing the diversity of the supply of recreational activities.

Drawing an analogy with the concentration of firms within an industry, Section 2 derives a measure of individual concentration of the use of time that possesses properties considered desirable for measures of industry concentration. Section 3 develops a simple theoretical model for the concentration of daily leisure activities, whose main purpose is to show that the sign of the reaction of concentration to variations in the quantity of leisure available is theoretically ambiguous. Section 4 examines empirically the pattern of concentration of daily leisure activities in a sample of male workers extracted from the German Time Budget Survey (ZBE) 2001/2002. Although both the quantity of leisure and its degree of concentration are certainly chosen by the individual, we argue that the ZBE panel structure as well as the exogenous reduction in market work brought about by the weekend for many workers, offer an avenue for identifying the causal effect of the former on the latter. Conclusions and some directions for future research are provided in Section 5.

[1] Other aspects of leisure such as quantity, quality, inequality, timing, togetherness, or recovery, have been investigated by a large socio-economic and psychological literature interested on behavioral and welfare comparisons. See, among many others, Owen (1971), Juster and Stafford (1985), Kooreman and Kapteyn (1987), Robinson and Godbey (1999), Bittman and Wajcman (2000), Hamermesh (2002), Mattingly and Bianchi (2003), Bittman (2005), Jenkins and Osberg (2005), Kahneman and Krueger (2006), Aguiar and Hurst (2007), Sonnentag et al. (2009), and Sevilla et al. (2012).

2 A measure of concentration of the use of time

The kind of concentration that we aim to measure is that implicit in a vector with the times spent on a series of activities, i.e. in an activity profile. Existing measures of time-use concentration respond to alternative aims, such as measuring intra-household specialization (Bonke et al., 2008) or assessing variability in a sample of activity profiles (González Chapela, 2006), and require at least two profiles to be computed. Closely connected with the concept of concentration are the notions of diversity, variety, and specialization. A time-use profile heavily concentrated on a few activities would be typically classified as little diversified and varied, or as very specialized. Hufnagel (2008) deals with the evaluation of time-use diversity across consecutive days, what requires at least two activity profiles. Variety, understood as the number of different activities undertaken (Gronau and Hamermesh, 2008; Ray, 1979; Sonnentag, 2001), ignores how time is distributed across activities engaged in. Baumgardner's (1988) measure of physicians' degree of specialization is constructed from quantities of outputs, but there may be cases where the times spent producing the outputs offer a more accurate way of measuring specialization, or are, indeed, the only available information.

In the 1960s several articles began to appear that examined the concentration measures employed in empirical analyses of industrial structure. Perhaps the best summary of those articles is Hall and Tideman (1967), who developed a set of desirable properties for measures of concentration in an industry. By equating activity profiles with industries, activities with manufacturing plants, and time with firm size, advantage of those efforts is taken here for defining a well-grounded measure of concentration of the use of time.

Let P_m denote the relative time share spent on activity m, $m = 1,\ldots,M$. Following Hall and Tideman (1967), a measure of concentration of the use of time ought to be: (1) One dimensional, i.e. unambiguous. (2) Independent of the total time analyzed, but a function of all the P_m's. (3) Affected by a change in any P_m, with concentration increasing (respectively, decreasing) if there is a shift from activity m to n and $P_m < P_n$ ($P_m > P_n$). (4) Reduced by one-Kth if each activity is divided into K more specific activities of equal duration. (5) A decreasing function of M when time is spent on M activities of equal duration. (6) Between 0 and 1. Although these properties cannot determine the best measure of concentration to use, they serve to discard measures that are undesirable for theoretical reasons.

A well-known measure of industry concentration that possesses all of the properties set forth by Hall and Tideman is the Herfindahl-Hirschman index (HHI)[2]. Hence, the measure

$$HHI = \sum_{m=1}^{M} P_m^2 \tag{1}$$

which is the sum of the squares of all M relative time shares, immediately suggests itself as a measure of concentration of the use of time: It is one-dimensional and utilizes all the P_m's. Its maximum value is 1, which corresponds to the case of complete concentration: $P_m = 1$ for some

[2] For an explanation of the origins of this index, see Theil (1967, p. 316).

m, and $P_n = 0 \ \forall n \neq m$. For given M, the minimum value is $1/M$ which is attained when all relative time shares are equal. This minimum approaches zero as M increases. Properties 3 and 4 are also satisfied because a (small) shift from m to n changes expression (1) by $2(P_n - P_m)$, and because $\sum_{m=1}^{M} K(P_m/K)^2 = 1/K \sum_{m=1}^{M} P_m^2$. The denomination *HHI* is kept to refer to expression (1) hereafter.

In spite of the theoretical appeal of the *HHI*, the concentration ratio, i.e. the fraction of an industry size held typically by its 4, 8, or 20 largest firms, has been frequently employed in empirical studies of industrial structure. Besides being highly correlated with the *HHI* (e.g., see Bailey and Boyle, 1971), the concentration ratio is more operational, for its calculation does not require knowing the size of every firm in the industry. But when it comes to measuring concentration of the use of time, and the time-use information has been collected by the time diary methodology, that shortcoming of the *HHI* is less marked: One of the main reasons behind the current popularity of the time diary is that it permits distinguishing a large number of activities. For example, if (as is typical of European time-use surveys) diarists record activities in 10 minute slots and the activity coding list distinguishes more than 144 activities, a researcher could discern up to 144 main activities on the diary day. (As the referee pointed out, in practice this number is much lower, as people cannot survive on 10 minutes of sleep and 10 minutes of eating to then accomplish 142 other activities.) Indeed, this wealth of detail raises a problem since, for analysis purposes, researchers end up classifying the recorded activities into a few time-use aggregates. As properties 4 and 5 suggest, this practice may alter the degree of concentration observed in the data, and since there are many ways one might classify activities, it calls for assessing the robustness of the findings to different activity aggregations.

The entropy measure of information theory has been also used as an index of industrial concentration, e.g. see Theil, 1967, and Horowitz and Horowitz, 1968. In the interpretation of the latter, the entropy quantifies the degree of uncertainty as to which of the firms in the industry will secure the custom of a buyer chosen at random. Analogously, the entropy measure

$$H = -\sum_{m=1}^{M} P_m \ln P_m \tag{2}$$

evaluates the degree of uncertainty implicit in an activity profile: The greater the entropy the greater the uncertainty as to which activity the individual is carrying out on a minute chosen at random. Measure (2), however, fails to satisfy properties 4 (the entropy is reduced by $\ln K$ when each activity is divided into K more specific activities of equal duration) and 6 (as the entropy ranges between 0 and infinity). According to Hall and Tideman (1967), property 6 is not strictly necessary (it simply makes the measure easier to use), but property 4 is very necessary if we are to have confidence in the measure's cardinal properties.

Besides appraising existing measures of industry concentration, Hall and Tideman (1967) proposed a new measure of concentration that satisfied all of their properties. The main difference between the *HHI* and the so-called *TH* index is that while the former weights each firm by its relative share, the latter weights by the firm rank—the mth largest firm receives weight m, whereby the number of firms in the industry becomes emphasized. As the number of activities

engaged in may indicate in part the presence of constraints on the use of time (more activities suggest fewer constraints), one could argue that the number of activities should be stressed in a measure of time-use concentration. This goal is accomplished by the *THI* equivalent

$$THI = \frac{1}{\left(2\sum_{m=1}^{M} mP_m\right)-1} \tag{3}$$

where (abusing somewhat the notation) the mth longest activity receives weight m.

3 The concentration of daily leisure activities – A simple theoretical model

In this section, a simple demand model for daily leisure activities is developed that allows analyzing the effect of the quantity of leisure on its degree of concentration. This model can be viewed as a particular case of Becker's (1965) general theory of choice, where, for analytical convenience, market goods are abstracted. Hence, too, it can be seen as the obverse of classical demand models. Although the definition of leisure is not completely specified until Section 4, our concept of leisure tallies with Ås (1978) notion of free time, i.e. time that is left after satisfying basic physiological needs, working for pay, and doing things we are committed to.

On a certain day, an individual is faced with the choice of dividing a certain amount of leisure (L) into two activities[3]:

$$L = L_1 + L_2 \tag{4}$$

where L_m denotes time devoted to activity m, $m = 1, 2$. Preferences over activities are represented by the utility function

$$U\left(L_1, L_2\right) = \left(L_1 - \gamma_1\right)^{\alpha_1}\left(L_2 - \gamma_2\right)^{\alpha_2} \tag{5}$$

(Geary, 1950-51, Stone, 1954; see also Prowse, 2009), which possesses some desirable properties. In expression (5), $\alpha_m > 0$ and (without loss of generality) $\alpha_1 + \alpha_2 = 1$. Although there is no requirement that any γ_m be positive, the sum $\gamma_1 + \gamma_2$ is interpreted here as the minimum daily leisure needed by the individual to live his/her life. In the terminology of Goodin et al. (2005, 2008), $\gamma_1 + \gamma_2$ would be defined as necessary time in leisure, although these authors do not consider that everyone needs to devote some time to leisure on a daily basis. Nevertheless, the existence of a subsistence quantity of daily leisure may not be an unreasonable assumption.[4] Theory and evidence in the field of work and organizational psychology indicate that leisure allows recovering our physical and mental capabilities from effort expended at work (see e.g. the

[3] This decision can be viewed as the second stage of a 2-stage budgeting where leisure is weakly separable from goods and the price of leisure is normalized to 1.

[4] The model in Leuthold (1968) allows for "minimum required" hours of leisure over the year.

volume edited by Sonnentag et al., 2009; this role of leisure was also suggested by Becker, 1965, p. 498, and Stafford and Cohen, 1974, and is explicit in Schwartz and McCarthy, 2007). Also, people may tend to engage frequently in leisure activities in order to gratify psychological needs such as affiliation, self-expression, or status (Tinsley and Eldredge, 1995). In our sample of employed German men, for example, 98.3 percent report some leisure in each of the three diary days, whereas in Spain and in the US the percentage of the population aged 15+ who reports some leisure on a typical day is, respectively, 97.6 and 95.9.[5] The portion of minimum daily leisure spent on activity m is denoted γ_m. Influencing the distribution of that minimum between activities may be factors such as activity set-up costs (created for example by the need to travel and coordinate with others), the price of the goods consumed in the course of the activities, the degree of recovery obtained from each activity, or the particular psychological needs gratified. Intuitively, set-up costs and goods prices should be inversely related to γ_m, whereas recovery and preferences for the needs gratified should be directly related. Some evidence on this issue is provided in Section 4.

Maximizing (5) subject to the adding-up constraint (4) yields the following system of demand functions:

$$L_m = \gamma_m + \alpha_m \left(L - (\gamma_1 + \gamma_2)\right), \quad m = 1, 2 \tag{6}$$

which expressed in relative shares form produces

$$P_m \equiv \frac{L_m}{L} = \left(\frac{\gamma_m}{\gamma_1 + \gamma_2}\right)\left(\frac{\gamma_1 + \gamma_2}{L}\right) + \alpha_m \left(1 - \frac{\gamma_1 + \gamma_2}{L}\right), \quad m = 1, 2 \tag{7}$$

Analogously to expression (4.7) in Deaton and Muellbauer (1980, p. 145), P_m is a weighted average of demand patterns pertaining to days where L is so small that $P_m = \gamma_m/(\gamma_1 + \gamma_2)$ and days where L is so large that P_m approaches α_m. P_m is increasing and concave in L if and only if $\alpha_m > \gamma_m/(\gamma_1 + \gamma_2)$, and decreasing and convex if the inequality is reversed.[6] Of course, in this two-activity model if P_m is, say, increasing in L, P_n has to be decreasing. The composition of leisure would be independent of L if and only if $\alpha_m = \gamma_m/(\gamma_1 + \gamma_2)$ or $\gamma_m = 0$, $m = 1, 2$, whereby, in the light of this model, the finding of an empirical response of concentration to L would add to the support of the minimum daily leisure assumption.

The direction of the effect of variations in L on the concentration of leisure cannot, in general, be determined *a priori*.[7] Suppose, to be specific, that $\gamma_1 > \gamma_2$, implying that $P_1 > P_2$ on days where

---

5 Author's calculations with data from the German Time Budget Survey 2001/2002, the Spanish Time Use Survey 2002-2003, and the American Time Use Survey (ATUS) 2003. In the first two surveys the definition of leisure is that of Leisure 1 (see Section 4); in the ATUS, leisure is defined as time devoted to ATUS major categories 12 and 13 plus associated travel.

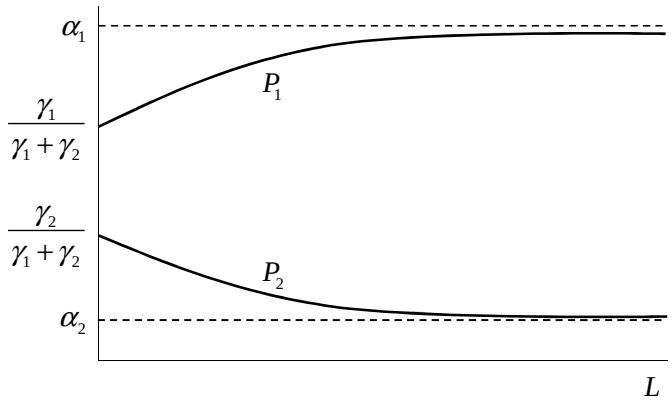
6 The total leisure elasticity ($e_m = d \log L_m / d \log L$) is determined analogously. If $\alpha_m > \gamma_m/(\gamma_1 + \gamma_2)$, then $e_m > 1$ and activity m would be considered a luxury; if $\alpha_m < \gamma_m/(\gamma_1 + \gamma_2)$, then $e_m < 1$ and activity m would be a necessity. Inferiority cannot occur with $\alpha_m > 0$.

7 In the specific case that $\gamma_1 = \gamma_2$ and $\alpha_1 \neq \alpha_2$, concentration would always increase with L.

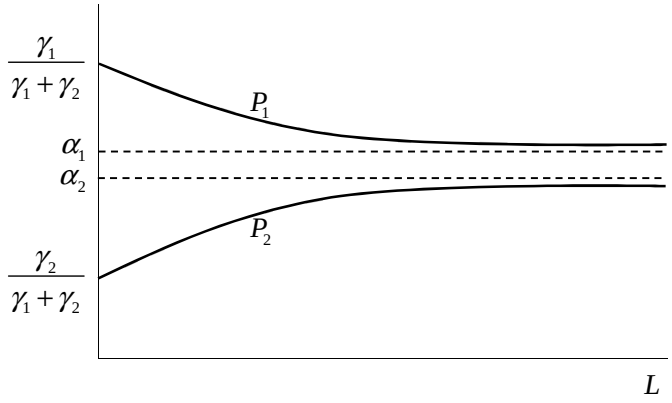
$L = \gamma_1 + \gamma_2$. Then, three cases can be distinguished, illustrated respectively in panels (a)-(c) of Figure 1.

Figure 1
The effect of variations in L on the concentration of leisure

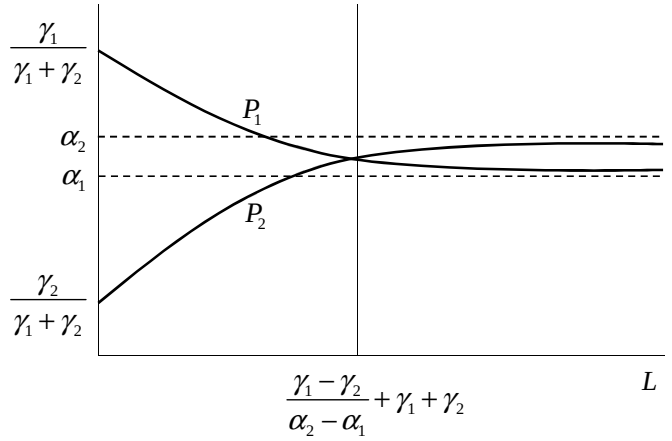
(a)



(b)



(c)



Source: Own illustration.

In the first case $\alpha_1 > \gamma_1/(\gamma_1+\gamma_2)$, so that P_1 would increase with L. Then, by the third property established in Section 2, the concentration of leisure would unambiguously increase on days where the individual had more leisure available. The second case assumes $\alpha_1 < \gamma_1/(\gamma_1+\gamma_2)$ and $\alpha_1 > \alpha_2$. Hence, P_1 would decrease with L, but even if L were very large, P_1 would be greater than P_2. Thus, the concentration of leisure would decrease on days where the individual had more leisure available. The third case shares features of the previous two situations. Suppose again that $\alpha_1 < \gamma_1/(\gamma_1+\gamma_2)$, but now $\alpha_1 < \alpha_2$. Then, P_1 would decrease with L as in the second case, but now P_1 would be above P_2 only in the region where $L < (\gamma_1 - \gamma_2/\alpha_2 - \alpha_1) + \gamma_1 + \gamma_2$, being below P_2 when $L > (\gamma_1 - \gamma_2/\alpha_2 - \alpha_1) + \gamma_1 + \gamma_2$. As a result, the relationship between the quantity of leisure and its degree of concentration would be U-shaped, with the peak of the U located asymmetrically if $(\gamma_1 - \gamma_2/\alpha_2 - \alpha_1) + \gamma_1 + \gamma_2$ deviated from the mean of L.

4 The concentration of daily leisure activities – Evidence from time diary data

4.1 Data, Measures, and Correlations

The German Time Budget Survey (ZBE) 2001/2002, a nationally representative quota sample of private households, is particularly unique as a rich source of information on time use and labor force characteristics. In the households interviewed, all individuals of 10 years and older were requested to complete three time diaries based on 10-minute intervals: two weekdays and a Saturday or Sunday, all pertaining to his/her reference week. If the completion of some diary was to be postponed, it was so for a complete week, so that the effective diary day corresponds to the same day of the week that the designated day. Socio-demographic and socio-economic characteristics of household members, including a sequence of questions about working on weekends, were collected by means of additional questionnaires.[8]

Although Ås (1978) notion of free time is a helpful classificatory principle of activities, the classification of some activities may be disputed. Hence, I will compute more than one measure of leisure, as in for example Aguiar and Hurst (2007) and Sevilla et al. (2012). Leisure 1 gathers time spent on social life and entertainment, sports and outdoor activities, hobbies and games, and mass media, which are activities that we cannot pay somebody else to do for us and that are not biological needs. To these, Leisure 2 adds child care (specifically, reading, playing,

[8] The Institute for the Study of Labor offers metadata for this survey at http://idsc.iza.org/metadata/. See also Statistisches Bundesamt (2005). To avoid seasonal distortion in the use of time, the survey was conducted over the course of one year, distributing the whole survey size evenly between April 2001 and March 2002. The tasks reported in the diaries were coded on the basis of an activity list encompassing some 230 activities capable of aggregation into the standardized Eurostat codes (see Eurostat, 2004). The high average number of activity episodes per day (ranging from the first to the third diary day: 25.3, 24.8, and 23.4, respectively), the very low prevalence of diaries with fewer than 7 episodes (0.1, 0.1, and 0.3 percent, respectively), and the very low presence of diaries missing two or more basic activities (0.4, 0.4, and 0.7 percent, respectively) indicate diary data of good quality (Juster, 1985; Robinson, 1985; Fisher et al. 2012).

and talking with child), gardening and pet care (specifically, the latter refers to outdoor activities with pets), and volunteer work and meetings. Finally, Leisure 3 will sum together Leisure 2 and time spent sleeping, an activity pursued largely for restorative purposes (e.g., see Biddle and Hamermesh, 1990).[9] As usual, the travel time associated to each activity is embedded in the total time spent on it.

As to the number of activities distinguished (M), I will work at a rather aggregated level to avoid unbalances in the measured detail of the different leisure domains. Thus, in the case of Leisure 1, the activities distinguished will be those listed in the previous paragraph, so that $M = 4$. The additional activities included in Leisure 2 will serve to assess the impact on the results of the number of activities distinguished: M will be alternatively set at 5 and 7 for Leisure 2, depending on whether child care, gardening and pet care, and volunteer work and meetings are aggregated together. For the same reason, M will be alternatively set at 6 and 8 for Leisure 3. Previewing the results, our main empirical conclusions will be unaffected by those changes in M.

The study sample is restricted to employed men aged 23-59 to exploit the market work reduction brought about by the weekend for many workers as an exogenous source of leisure. The sex and age selection criteria are intended to reduce sample selection issues. I also discarded persons who completed less than three diaries or who, in some diary, provided unspecified uses of time, presented fewer than 7 activity episodes, missed two or more of the four basic activities defined in Fisher et al. (2012), or reported no leisure (as measured with Leisure 1). The last requirement is a consequence of the *HHI* being only calculated when leisure is greater than zero, and excludes 1.7 percent of the observations that satisfy the other criteria for inclusion in the sample. This leaves us with 2,266 men, contributing a total of 6,798 diary days. Table 1 presents some characteristics of these persons. The sample will be further restricted for some specifications to workers who did not postpone the completion of the first diary, which yields a sample size of 1,431 men. Demographic differences between both samples are statistically insignificant, although the subsample presents, on average, 15 minutes more Leisure 1 (and 13 minutes less market work) per day, and concentration is about 3 percent smaller.

Table 2 presents sample descriptive statistics on the quantity, cross-activity distribution, and degree of concentration of leisure, organized by day of the week. The last row of the table lists the number of diaries used in the calculations. Leisure patterns are pretty stable from Monday to Thursday, irrespective of the leisure measure considered. In terms of our narrowest measure, leisure activities take up (on average) almost 4 hours each of those days, with approximately 61 percent of this time being devoted to mass media and 28 percent to social life and entertainment. Fridays bring about an extra hour of leisure and a change in its distribution, which becomes less tilted towards mass media (sleep, in the case of Leisure 3) and more inclined towards social life and entertainment.

---

[9] Leisure 1 includes the activities classified into 1-digit codes 5-8 of Eurostat (2004, Annex VI). Leisure 2 includes additionally the 1-digit code 4 and 3-digit codes 341, 344, and 383. To these, Leisure 3 adds the 2-digit code 01.

Table 1
Descriptive statistics

Variable	Mean	Std. dev.	Minimum	Maximum
Age	43.2	8.3	23	59
Leisure 1	5.0	2.9	.2	19.7
Leisure 2	5.7	3.1	.2	21.3
Leisure 3	13.5	3.8	1.2	23.8
HHI Leisure 1	.66	.21	.25	1
HHI Leisure 2[a]	.59	.21	.21	1
HHI Leisure 2[b]	.59	.21	.19	1
HHI Leisure 3[a]	.48	.13	.22	.98
HHI Leisure 3[b]	.48	.13	.22	.98
Variable (%)	Mean	Variable (%)		Mean
Married	79.5	Very good or good health		76.9
College graduate	34.1	Works every Sat[c]		15.0
Non-German	1.2	Works every Sun[d]		7.5

Notes: Data are of 2,266 employed men. Leisure is expressed in daily hours. [a]: Child care, gardening and pet care, and volunteer work and meetings are aggregated together; [b]: Those three activities are kept disaggregated. [c]: Percentage of those completing a diary for a Saturday. [d]: Percentage of those completing a diary for a Sunday.
Source: German Time Budget Survey (ZBE) 2001/2002, own calculations.

This one-hour increase in leisure has little impact on its concentration (when averaged across individuals) except for Leisure 3, whose concentration decreases by some 8 percent. On Saturdays, leisure increases substantially, social life and entertainment reaches the weekly maximum and mass media the weekly minimum. Coinciding with these changes, the concentration of leisure decreases noticeably with respect to Fridays (from around 4 percent in the case of Leisure 1 to about 9 percent in the case of Leisure 3). Hours of leisure reach the weekly maximum on Sundays and concentration the weekly minimum. With reference to Saturdays, the modest increase in Leisure 1 and Leisure 2 observed on Sundays is accompanied by a substantial reduction in concentration (11 and 9 percent, respectively), due to the larger importance of sports and outdoor activities. The evolution of concentration over the week as measured by the *THI* is essentially the same. The outstanding preponderance of mass media on those days where leisure is smaller suggests that that activity's minimum daily time might be much larger than that of other activities. As shown in Table 3, a reason for this could be mass media's lower related travel and necessity of coordination with others, which reduce set-up costs. Yet, the fact that hobbies and games present similar figures but a much smaller importance on the time budget indicates that alternative reasons are involved.

At the diary level, the Ordinary Least Squares (OLS) estimates presented in the first two rows of Table 4 suggest a U-shaped relationship between daily leisure and its degree of concentration that is robust to some individual characteristics and the day of the week. The estimated coeffi-

cient associated to the quantity of leisure is negative, and that associated to the square of this positive, both being statistically different from zero at the 0.01 level.[10]

Table 2
Average leisure (hours per day), leisure distribution and leisure concentration, by day of the week – Employed prime-age men

Variable	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Leisure 1	3.8	3.8	3.9	3.9	4.8	6.7	7.3
Social life and entertainment	.27	.27	.29	.28	.34	.38	.32
Sports and outdoor activities	.06	.07	.08	.07	.07	.08	.12
Hobbies and games	.04	.05	.04	.05	.05	.05	.05
Mass media	.63	.61	.59	.60	.54	.49	.50
HHI Leisure 1	.69	.70	.67	.67	.68	.65	.58
Leisure 2	4.3	4.4	4.3	4.4	5.5	7.7	8.1
Social life and entertainment	.23	.23	.26	.25	.30	.34	.29
Sports and outdoor activities	.06	.07	.07	.06	.06	.07	.11
Hobbies and games	.04	.04	.04	.05	.05	.05	.05
Mass media	.56	.55	.54	.54	.48	.42	.46
Child care	.02	.02	.01	.01	.02	.02	.02
Gardening and pet care	.04	.03	.03	.04	.04	.04	.02
Volunteer work and meetings	.05	.06	.05	.05	.05	.06	.05
HHI Leisure 2[a]	.62	.63	.62	.60	.61	.57	.52
HHI Leisure 2[b]	.62	.63	.61	.60	.61	.56	.52
Leisure 3	11.8	11.8	11.8	11.8	12.5	15.8	17.7
Social life and entertainment	.08	.08	.09	.09	.14	.17	.14
Sports and outdoor activities	.02	.03	.03	.03	.03	.04	.05
Hobbies and games	.01	.02	.01	.02	.02	.02	.02
Mass media	.19	.18	.18	.18	.18	.18	.20
Child care	.01	.01	.01	.00	.01	.01	.01
Gardening and pet care	.02	.01	.01	.01	.02	.02	.01
Volunteer work and meetings	.02	.02	.02	.02	.02	.03	.02
Sleep	.65	.65	.65	.65	.58	.53	.55
HHI Leisure 3[a]	.52	.52	.52	.51	.48	.43	.43
HHI Leisure 3[b]	.52	.52	.52	.51	.47	.43	.42
Diaries	904	926	888	944	870	1,180	1,086

Notes: Relative shares showing the distribution of leisure across activities are in italics. [a]: Child care, gardening and pet care, and volunteer work and meetings are aggregated together; [b]: Those three activities are kept disaggregated. Source: German Time Budget Survey (ZBE) 2001/2002, own calculations.

The U-shape, however, is not symmetrical: Ranging from our narrowest definition of leisure to the broadest, the minimum of the U is reached at 9.2, 9.6, and 20.2 hours, respectively, i.e. close to the 90th percentile of the corresponding sampling distribution of leisure (located,

[10] The standard errors listed in Table 4 are robust to heteroskedasticity and clustered at the individual level.

respectively, at 9.3, 10.2, and 19.3 hours). Hence, the partial effect of the quantity of leisure on its degree of concentration is negative for most of the leisure range. Computed, for example, at average leisure values, an extra hour of leisure reduces concentration by approximately -0.027, -0.025, and -0.023,[11] implying a 4 to 5 percent reduction in each case. These results change very little when concentration is assessed with the *THI*: The minimum of the U is reached at 9.3, 9.6, and 20.1 hours, and the reduction in concentration induced by an extra hour of leisure is -0.026, -0.025, and -0.022 when computed at average leisure values.

Table 3
Percentage of waking leisure spent on related travel and not alone, by leisure activity – Employed prime-age men

Leisure activity	Related travel	Not alone
Social life and entertainment	11.6	79.0
Sports and outdoor activities	9.4	63.8
Hobbies and games	1.7	46.8
Mass media	-	46.8
Child care	0.0	97.6
Gardening and pet care	49.3	57.4
Volunteer work and meetings	17.9	64.7

Notes: Mass media has not related travel in the Eurostat activity coding list. In the ZBE 2001/2002, no respondent reports travel related to child care as main activity. Source: German Time Budget Survey (ZBE) 2001/2002, own calculations.

Regarding the other effects presented in Table 4, having a college degree or being in good health is negatively associated with concentration, particularly when sleeping is not included in leisure. Being married is essentially unrelated to concentration when child care is excluded from leisure, but becomes a strong predictor for concentration otherwise: ceteris paribus, married men's Leisure 2 concentration is, on average, 4 to 5 percent smaller, the larger reduction observed when child care is kept disaggregated from other activities. No statistically significant differences in concentration are observed from Monday to Thursday, but concentration (as measured from Leisure 1 and Leisure 2) is greater on Fridays and Saturdays, and smaller on Sundays. These end-of-the-week differences in concentration might be the result of social norms regulating the type of leisure activities allowed on certain days, and/or of the availability of more leisure companions within and outside the household (Bittman, 2005; Jenkins and Osberg, 2005).

The existence of a significant and generally negative partial correlation between the quantity of leisure and its degree of concentration does not demonstrate a causal relationship. At the very least, we are faced with the prospect of omitted variable bias. It is conceivable, for example, that persons who like practicing some sport present less concentrated leisure profiles and demand more leisure.

[11] This partial effect is computed by subtracting the value of the regression function for concentration evaluated at mean leisure from this function's value evaluated at that mean plus 1, holding other regressors fixed.

Table 4
Leisure concentration regressions – Employed prime-age men

	Dependent variable: *HHI*, computed from									
	Leisure 1		Leisure 2[a]		Leisure 2[b]		Leisure 3[a]		Leisure 3[b]	
Independent variables	(1)		(2)		(3)		(4)		(5)	
Leisure	-.067	(.003)**	-.072	(.003)**	-.073	(.003)**	-.075	(.004)**	-.075	(.004)**
Leisure squared	.0037	(.0002)**	.0038	(.0002)**	.0038	(.0002)**	.0019	(.0001)**	.0019	(.0001)**
Age	-6×10^{-4}	(.003)	-.002	(.003)	-.0023	(.0031)	-.0026	(.0016)	-.0027	(.0016)
Age squared	1×10^{-5}	(3×10^{-5})	4×10^{-5}	(3×10^{-5})	5×10^{-5}	(4×10^{-5})	3×10^{-5}	(2×10^{-5})	3×10^{-5}	(2×10^{-5})
Married	.007	(.008)	-.023	(.008)**	-.028	(.008)**	-.004	(.004)	-.005	(.004)
College	-.014	(.006)*	-.012	(.006)*	-.013	(.006)*	.001	(.003)	.001	(.003)
Non-German	.017	(.031)	.036	(.032)	.039	(.032)	-.007	(.015)	-.007	(.015)
Very good or good health	-.017	(.007)**	-.014	(.007)*	-.013	(.007)*	-.006	(.004)	-.006	(.004)
Tue	.011	(.009)	.013	(.009)	.012	(.009)	-.001	(.004)	-.001	(.004)
Wed	-.008	(.009)	-3×10^{-6}	(.009)	.001	(.009)	-.002	(.005)	-.002	(.005)
Thu	-.010	(.009)	-.011	(.009)	-.010	(.009)	-.007	(.005)	-.007	(.005)
Fri	.028	(.009)**	.026	(.009)**	.026	(.009)**	-.027	(.005)**	-.027	(.005)**
Sat	.027	(.009)**	.023	(.009)*	.023	(.009)*	.001	(.005)	.001	(.005)
Sun	-.026	(.009)**	-.012	(.009)	-.009	(.009)	.023	(.005)**	.023	(.005)**
Intercept	.885	(.061)**	.878	(.063)**	.882	(.064)**	1.20	(.043)**	1.20	(.044)**
R^2	.136		.148		.149		.410		.409	

Notes: Data are of 6,798 diaries pertaining to 2,266 individuals. The estimation method is OLS in all columns. The independent variable Leisure is measured in hours and its definition is consistent with that of the dependent variable. Heteroskedasticity robust standard errors clustered at the individual level are in parentheses. [a]: Child care, gardening and pet care, and volunteer work and meetings are aggregated together; [b]: Those three activities are kept disaggregated. *: Significant at 5 percent. **: Significant at 1 percent.
Source: German Time Budget Survey (ZBE) 2001/2002, own calculations.

Or that good weather conditions during the survey reference week encouraged both the range of leisure activities undertaken and the demand for leisure. Therefore, the estimated partial correlation might be influenced by unobserved individual and/or week effects. It is also possible that unobserved diary day effects are biasing the estimates. This would be the case if, for example, a friend's visit on the diary day promoted both the demand for leisure and the range of leisure activities undertaken. For all these reasons, the main estimates presented in Table 4 are to be re-examined.

4.2 Estimation Method

Assume that individual i's leisure concentration and leisure quantity on day d (denoted, respectively, HHI_{id}^{L} and L_{id}) are related according to

$$HHI_{id}^{L} = \beta_0 + \beta_1 L_{id} + \beta_2 L_{id}^2 + \beta_3 I_{id}^{Fri} + \beta_4 I_{id}^{Sat} + \beta_5 I_{id}^{Sun} + \mu_i + u_{id}, \quad d = 1,2,3 \tag{8}$$

where the β's are unknown parameters to be estimated. The three diary days available in the ZBE 2001/2002 are sorted out from Monday to Sunday, so that $d = 1$ and $d = 2$ indicate weekdays and $d = 3$ indicates the Saturday/Sunday. Thus, diaries are arranged chronologically except for individuals who postponed the completion of a weekday diary. The possible convexity of the concentration profile is captured by the terms L_{id} and L_{id}^2, whose associated coefficients, β_1 and β_2, would be respectively negative and positive. I_{id}^{Fri}, I_{id}^{Sat} and I_{id}^{Sun} are indicator variables taking on value one if the diary pertains to the day indicated in the superscript and value zero otherwise. The mean-zero unobserved variable μ_i, which represents individual-level features and circumstances influencing the concentration of leisure that were invariant during the survey week, is allowed to be arbitrarily correlated with the observed explanatory variables. Included in μ_i would be, for example, the total weekly hours of work and the prices of goods consumed in the course of leisure activities if those prices were invariant during the survey week. The mean-zero variable u_{id} stands for unobserved factors altering the concentration of leisure on day d. It is assumed to be weakly exogenous:

$$E\left(u_{id} \middle| \mathbf{x}_{id-1}, \mathbf{x}_{id-2}, \mu_i\right) = 0, \quad d = 1,2,3 \tag{9}$$

where $\mathbf{x}_{id} \equiv (1, L_{id}, L_{id}^2, I_{id}^{Fri}, I_{id}^{Sat}, I_{id}^{Sun})$ and d assumes a chronological ordering (i.e., (9) does not hold if i postponed the completion of a weekday diary). Moment conditions similar to (9) are typical of intertemporal decision making models under uncertainty (e.g., see Hall, 1978, and Altonji, 1986), where a rational expectations assumption makes u_{id} to be uncorrelated with explanatory variables dated at $d-1$ or earlier. In this study, u_{id} is allowed to be correlated with L_{id} and L_{id}^2 because the quantity of leisure is under the individual's control. In this context, it is well-known that the pooled OLS estimator of (8) is biased and inconsistent[12]. To get rid of μ_i ,define $\Delta HHI_{i3}^{L} = HHI_{i3}^{L} - HHI_{i2}^{L}$, $\Delta L_{i3} = L_{i3} - L_{i2}$, $\Delta L_{i3}^2 = L_{i3}^2 - L_{i2}^2$ and $\Delta u_{i3} = u_{i3} - u_{i2}$. Then,

[12] See for example Wooldridge (2002, Ch. 10). The asymptotic analysis is as the number of sample individuals tends to infinity.

$$\Delta HHI_{i3}^{L} = \beta_4 + \beta_1 \Delta L_{i3} + \beta_2 \Delta L_{i3}^2 + \gamma_1 I_{i3}^{Sat-Fri} + \gamma_2 I_{i3}^{Sun-MTW} + \gamma_3 I_{i3}^{Sun-Fri} + \Delta u_{i3} \quad (10)$$

In this expression, each I is an indicator variable taking on value one if the difference was taken as indicated in the superscript and value zero otherwise. For example, $I^{Sun-MTW}$ equals one if the second diary day is Monday, Tuesday, Wednesday, or Thursday, and the third diary pertains to a Sunday. The unknown parameters γ are such that $\beta_3 = -\gamma_1$ and $\beta_5 = \beta_4 + \gamma_2$. It can also be shown that

$$\gamma_3 = \gamma_1 + \gamma_2 \quad (11)$$

a result that will be tested in the data.

While expression (10) is a standard cross section equation that can be estimated by OLS, the key conditions for OLS to consistently estimate β_1 and β_2,

$$E\left(\left(L_{i3} - L_{i2}\right)\left(u_{i3} - u_{i2}\right)\right) = 0 \quad (12a)$$

$$E\left(\left(L_{i3}^2 - L_{i2}^2\right)\left(u_{i3} - u_{i2}\right)\right) = 0 \quad (12b)$$

will not hold if L_{id} or L_{id}^2 are correlated with u_{id}. I use responses to the questions "Does it happen that you work on weekends? If Yes, how often?", which are asked of all workers by the ZBE 2001/2002, as well as L_{i1} and L_{i1}^2, to instrument ΔL_{i3} and ΔL_{i3}^2. Working on weekends is likely to have a substantial negative impact on the quantity of leisure (e.g., Bittman, 2005, has found a big fall in leisure activities associated to Sunday employment in Australia), and is therefore expected to be negatively correlated with ΔL_{i3} and ΔL_{i3}^2. The validity of this information as an instrument relies upon being uncorrelated with preferences for the concentration of leisure on the Saturday/Sunday of the reference week. This assumption would be questioned if, for example, those who work on weekends got more tired and the degree of tiredness influenced the organization of leisure activities. To check for that possibility, I estimated (10) by OLS with the working on weekends instrument (as specified in the following paragraph) included among the explanatory variables. When sleep is not counted as leisure, the coefficient on the instrument is positive but statistically not different from zero (the *p*-values range from 0.10 to 0.28). When sleep is counted as leisure, the coefficient on the instrument is negative and statistically different from zero (*p*-value 0.01; in the subsample, *p*-values are at or around 0.10). These results cast some doubts on the validity of this instrument when sleep is considered leisure. The validity of L_{i1} and L_{i1}^2 rests on a different rationale: Under the weak exogeneity assumption stated in (9), L_{i1} and L_{i1}^2 are uncorrelated with Δu_{i3}.[13] Since this assumption will not hold in the case of individuals who postponed the completion of the first diary, these persons will be excluded

[13] The validity of L_{i1} (respectively, L_{i1}^2) as an instrument for ΔL_{i3} (ΔL_{i3}^2) does not rule out serial correlation in u_{id}: Correlation between L_{id} (or L_{id}^2) and u_{id} can be prompted by a white noise term, whereas correlation in u_{id} can be induced by a serially correlated preference shifter (Arellano and Honoré, 2001, p. 3238). The validity of L_{i1} and L_{i1}^2 does require, however, that L_{id} and L_{id}^2 exert only a contemporaneous effect on HHI_{id}^{L}.

from some estimations. Correlation between L_{i1} (respectively, L_{i1}^2) and ΔL_{i3} (ΔL_{i3}^2) can be induced by a time-varying serially correlated preference for leisure.

Five mutually exclusive answers are possible to the above-mentioned questions on working on weekends: "Never", "Every week", "Every two weeks", "Every three-four weeks", and "More rarely", which are provided separately for Saturdays and Sundays. From this information I constructed a series of indicator variables corresponding to the five possible responses. The indicator for "Every week", for example, takes on value 1 if the worker completed a diary for a Saturday (respectively, a Sunday) and reports working every Saturday (Sunday), and value 0 otherwise. As shown in Table 1, 15.0 percent of those whose third diary day is a Saturday work every Saturday, whereas the corresponding figure for Sundays is 7.5 percent.[14] Reduced form regressions for ΔL_{i3} and ΔL_{i3}^2 on all exogenous variables, including L_{i1}, L_{i1}^2 and the indicators for "Every week", "Every two weeks", "Every three-four weeks", and "More rarely", revealed that the last three indicators are individually insignificant in each regression. As weak instruments can harm the finite-sample properties of instrumental variables (IV) estimators even in large samples (see e.g. the survey article by Murray, 2006), only the indicator for working every weekend will be included in the instrument set.

4.3 Results

Tables 5 and 6 present the estimates of the differenced equation (10). In Table 5, OLS coefficients, which do not control for the endogeneity of ΔL_{i3} and ΔL_{i3}^2 are presented. In Table 6, Generalized Method of Moments (GMM) estimates calculated with optimal weighting matrix are shown. (Auxiliary IV output, including the first-stage regressions for the endogenous variables, is presented in the Appendix.) In both tables, the upper panel lists results for the full sample, whereas results for the subsample of workers who did not postpone the completion of the first diary are shown in the lower panel. Heteroskedasticity robust standard errors appear in parentheses, and probability values in brackets.

When ΔL_{i3} and ΔL_{i3}^2 are treated as exogenous, a U-shaped relationship between daily leisure and its degree of concentration similar to that presented in Table 4 is estimated. Although differencing has reduced the number of observations to 2,266, the relationship is still precisely measured and attains statistical significance at the 0.01 level. According to the estimates for the full sample, and ranging from our narrowest definition of leisure to the broadest, the U function minimum is located at 9.9, 10.5, and 20.8 hours, respectively, whereas an extra hour of leisure is estimated to reduce concentration by approximately -0.028, -0.027, and -0.022 when the effect is computed at average leisure values. Aggregating child care, gardening and pet care, and volunteer work and meetings into one activity leaves the results essentially unaffected, as well as estimating (10) on the subsample.

[14] Although Saturdays are considered working days in the German working time law, most people do not work on Saturdays, and, for those who work, special bonuses are agreed upon in most collective agreements. Working on Sundays is prohibited, but exceptions can be approved by the authorities. See Bosch (2009) for more information on working time regulations in Germany.

Table 5
Linear models for the concentration of leisure – OLS differences estimates

	Full sample: 2,266 employed prime-age men									
	Dependent variable: ΔHHI^{L}_{i3}, computed from:									
	Leisure 1		Leisure 2[a]		Leisure 2[b]		Leisure 3[a]		Leisure 3[b]	
Independent variables	(1)		(2)		(3)		(4)		(5)	
ΔL_{i3}	-.064	(.005)**	-.065	(.005)**	-.066	(.005)**	-.068	(.004)**	-.068	(.004)**
ΔL^{2}_{i3}	.0032	(.0004)**	.0031	(.0003)**	.0032	(.0003)**	.0016	(.0001)**	.0016	(.0001)**
$I^{Sat-Fri}$	-.010	(.016)	-.009	(.015)	-.012	(.015)	.027	(.007)**	.027	(.007)**
$I^{Sun-MTW}$	-.048	(.014)**	-.040	(.014)**	-.037	(.014)**	.029	(.007)**	.030	(.007)**
$I^{Sun-Fri}$	-.082	(.018)**	-.063	(.017)**	-.061	(.017)**	.057	(.008)**	.058	(.008)**
Intercept	.027	(.012)*	.026	(.012)*	.027	(.012)*	-.006	(.006)	-.006	(.006)
R^2	.110		.122		.125		.319		.321	
Test for endogeneity of ΔL_{i3} and ΔL^{2}_{i3} (robust Wald statistic)	1.22	[.54]	1.16	[.56]	1.62	[.44]				
Wald test: $\gamma_3 = \gamma_1 + \gamma_2$	1.06	[.30]	.44	[.51]	.25	[.62]	.01	[.92]	.03	[.87]
Ramsey's (1969) RESET	2.79	[.43]	1.43	[.70]	1.57	[.67]	15.81	[.00]	15.69	[.00]

Table 5 (Cont.)

	Sub-sample: 1,431 individuals who did not postpone the completion of the first diary									
	Dependent variable: ΔHHI^{L}_{i3}, computed from:									
	Leisure 1		Leisure 2[a]		Leisure 2[b]		Leisure 3[a]		Leisure 3[b]	
Independent variables	(6)		(7)		(8)		(9)		(10)	
ΔL_{i3}	-.058	(.006)**	-.060	(.006)**	-.063	(.006)**	-.061	(.006)**	-.062	(.006)**
ΔL^2_{i3}	.0028	(.0004)**	.0028	(.0004)**	.0030	(.0004)**	.0014	(.0002)**	.0014	(.0002)**
$I^{Sat-Fri}$	-.027	(.019)	-.024	(.017)	-.028	(.017)	.036	(.008)**	.035	(.008)**
$I^{Sun-MTW}$	-.055	(.019)**	-.052	(.018)**	-.047	(.018)**	.022	(.009)*	.023	(.009)*
$I^{Sun-Fri}$	-.083	(.021)**	-.071	(.020)**	-.068	(.020)**	.053	(.010)**	.054	(.010)**
Intercept	.029	(.014)*	.027	(.014)*	.029	(.014)*	-.013	(.007)	-.012	(.007)
R^2	.101		.113		.117		.308		.311	
Test for endogeneity of ΔL_{i3} and ΔL^2_{i3} (robust Wald statistic)	7.77	[.02]	5.48	[.06]	5.58	[.06]	9.43	[.01]	9.12	[.01]
Wald test: $\gamma_3 = \gamma_1 + \gamma_2$	.00	[.98]	.02	[.88]	.08	[.78]	.15	[.70]	.11	[.74]
Ramsey's (1969) RE-SET	5.36	[.15]	.29	[.96]	.18	[.98]	15.56	[.00]	15.35	[.00]

Notes: Heteroskedasticity robust standard errors are in parentheses and probability values appear in brackets. The activities included in ΔL_{i3} are consistent with those in the dependent variable. [a]: Child care, gardening and pet care, and volunteer work and meetings are aggregated together; [b]: Those three activities are kept disaggregated. *: Significant at 5 percent. **: Significant at 1 percent.
Source: German Time Budget Survey (ZBE) 2001/2002, own calculations.

Table 6
Linear models for the concentration of leisure – GMM differences estimates

	Full sample: 2,266 employed prime-age men				
	Dependent variable: ΔHHI^{L}_{i3}, computed from:				
	Leisure 1	Leisure 2[a]	Leisure 2[b]	Leisure 3[a]	Leisure 3[b]
Independent variables	(1)	(2)	(3)	(4)	(5)
ΔL_{i3}	-.138 (.069)*	-.090 (.033)**	-.096 (.034)**	-.077 (.018)**	-.079 (.018)**
ΔL^{2}_{i3}	.0067 (.0033)*	.0039 (.0016)*	.0040 (.0017)*	.002 (.0005)**	.0021 (.0006)**
$I^{Sat-Fri}$	-.040 (.031)	-.025 (.021)	-.030 (.021)	.028 (.008)**	.027 (.008)**
$I^{Sun-MTW}$	-.025 (.028)	-.032 (.017)	-.028 (.017)	.025 (.010)*	.026 (.011)*
$I^{Sun-Fri}$	-.104 (.028)**	-.078 (.022)**	-.078 (.023)**	.055 (.009)**	.056 (.009)**
Intercept	.116 (.086)	.075 (.047)	.084 (.048)	-.012 (.017)	-.010 (.017)
Hansen J test of over-identifying restrictions (No. OR: 1)	1.83 [.18]	.29 [.59]	.37 [.54]	9.79 [.00]	9.43 [.00]

Table 6 (Cont.)

	Sub-sample: 1,431 individuals who did not postpone the completion of the first diary				
	Dependent variable: ΔHHI^{L}_{i3}, computed from:				
	Leisure 1	Leisure 2[a]	Leisure 2[b]	Leisure 3[a]	Leisure 3[b]
Independent variables	(6)	(7)	(8)	(9)	(10)
ΔL_{i3}	-.072 (.025)**	-.071 (.021)**	-.074 (.021)**	-.115 (.020)**	-.115 (.020)**
ΔL^2_{i3}	.0010 (.0019)	.0023 (.0016)	.0024 (.0017)	.0034 (.001)**	.0034 (.001)**
$I^{Sat-Fri}$	-.052 (.023)*	-.043 (.020)*	-.048 (.020)*	.032 (.009)**	.031 (.009)**
$I^{Sun-MTW}$	-.023 (.025)	-.042 (.019)*	-.037 (.019)	-.001 (.015)	.001 (.015)
$I^{Sun-Fri}$	-.090 (.025)**	-.089 (.022)**	-.086 (.023)**	.036 (.013)**	.037 (.013)**
Intercept	.129 (.043)**	.089 (.031)**	.092 (.031)**	-.023 (.012)	-.021 (.012)
Hansen J test of over-identifying restrictions (No. OR: 1)	.53 [.47]	.27 [.61]	.56 [.45]	.08 [.78]	.05 [.82]

Notes: Heteroskedasticity robust standard errors are in parentheses and probability values appear in brackets. The activities included in ΔL_{i3} are consistent with those in the dependent variable. In all columns, ΔL_{i3} and ΔL^2_{i3} are instrumented with L_{i1}, L^2_{i1} and the indicator for working every weekend. [a]: Child care, gardening and pet care, and volunteer work and meetings are aggregated together; [b]: Those three activities are kept disaggregated. *: Significant at 5 percent. **: Significant at 1 percent.

Source: German Time Budget Survey (ZBE) 2001/2002, own calculations.

Instrumenting for ΔL_{i3} and ΔL_{i3}^2 tends to increase (in absolute value) the estimated β_1 and β_2, although the implied relationship between daily leisure and its degree of concentration continues being U-shaped. Estimates are more imprecise, but attain statistical significance at the 0.05 level. According to the estimates for the full sample, and presenting again the results from our narrowest notion of leisure to the broadest, the U function minimum is located at 10.3, 11.7, and 19.2 hours, and an extra hour of leisure reduces concentration by approximately -0.064, -0.044, and -0.021 at average leisure values. In the subsample, β_1 is negative and β_2 is zero (at standard significance levels) in the case of Leisure 1 and Leisure 2, thereby implying an inverse linear relationship between leisure and its concentration. When sleep is counted at leisure, however, the implied relationship is again U-shaped, the peak is reached at 17.0 hours, and an extra hour of leisure reduces concentration by approximately -0.020 at average leisure values.

Since the number of excluded instruments (three) exceeds the number of endogenous variables (two), it is possible to test the overidentifying restrictions on the excluded instruments. The test statistic (Hansen, 1982, J-statistic) is the minimized value of the GMM objective function, and is asymptotically distributed as χ^2 with degrees of freedom equal to the number of overidentifying restrictions (one in this case). The main output of the overidentifying restrictions test is presented separately for each leisure definition at the bottom of each panel in Table 6. When sleep is excluded from leisure, the p-value for this test is above standard significance levels, so that the validity of the instruments is not questioned. Yet, when time spent sleeping is counted as leisure, the validity of the instruments is clearly rejected in the full sample (p-value 0.00). In the subsample, the validity of the instruments is well within confidence bounds irrespective of the leisure definition.

The fact that IV estimates do not expose substantial biases in OLS results suggests that ΔL_{i3} and ΔL_{i3}^2 could not be endogenous. To test for endogeneity, the residuals from regressing ΔL_{i3} and ΔL_{i3}^2 on all the exogenous variables were added to each of the regressions presented in Table 5 (except those in columns (4) and (5), where the instruments revealed as invalid). Then, the joint statistical significance of the residual terms in each regression was tested using a robust Wald test (Wooldridge, 2002, p. 121). Listed in the third from last row of each panel in Table 5 are the results of this test. In the full sample, the claim of exogeneity is well within confidence bounds. In the subsample, the test results suggest that ΔL_{i3} and ΔL_{i3}^2 are endogenous, particularly in the cases of Leisure 1 and Leisure 3.

Additional specification checks can be carried out by testing the restriction on the coefficients in (11) and by testing the statistical significance of powers of the fitted values in the regression for ΔHHI_{i3}^L. Under the assumption that model (8) is correct, (11) is obviously true in the population, but estimation biases could impede its verification in the data. Under the same assumption, powers of the fitted values added to (10) must be jointly insignificant (Ramsey, 1969). Results of robust Wald tests for the hypothesis in (11) and for testing the joint significance of $\widehat{\Delta HHI_{i3}^L}^2$,$\widehat{\Delta HHI_{i3}^L}^3$ and $\widehat{\Delta HHI_{i3}^L}^4$ in (10) are presented at the bottom of each panel in Table 5 separately for each regression. The null in (11) is safely within confidence bounds in all cases. When sleep is not counted as leisure, the claim of no functional form mis-specification is not rejected. Therefore,

and in agreement with panel (c) of Figure 1, a quadratic function seems sufficient to represent the leisure concentration profile in that case. Yet, when sleep is counted as leisure the claim of no functional form mis-specification is clearly rejected.

Overall, the preceding specification checks tend to favor the estimates in columns (1)-(3) of Table 5 and (9)-(10) of Table 6, which tell a rather consistent story: The concentration of daily leisure activities decreases with the hours of leisure available until hours are so large (around the 90th percentile of its empirical distribution in the case of Leisure 1 and Leisure 2; around the 75th percentile in the case of Leisure 3) that concentration reverses its trend. Thus, when the quantity of leisure is small individuals concentrate on a few leisure activities, whose relative importance in the time budget diminishes as more leisure becomes available. Interpreted in terms of our theoretical model, this empirical pattern is in agreement with the case shown in panel (c) of Figure 1, if the equivalent of $(\gamma_1-\gamma_2)/(\alpha_2-\alpha_1)+\gamma_1+\gamma_2$ were located well above the mean of L. It likewise rejects the claim that daily leisure is not required for subsistence, i.e. that $\gamma_1=\gamma_2=0$. Regarding the size of the effect, at average leisure values the concentration of Leisure 1, Leisure 2, and Leisure 3 would decrease around 4 percent with an extra hour of leisure, but the reduction would be much stronger at for example the 25th percentile of the leisure empirical distribution: 8, 10, and 10 percent, respectively.

The estimation results also suggest the existence of day-of-the-week effects on leisure concentration. There is some evidence of a Friday effect (given by minus the coefficient associated to $I^{Sat-Fri}$) when sleeping in included in leisure: Keeping constant the quantity of leisure, leisure activities become, on average, less concentrated on Fridays than in the period Monday-Thursday. The Saturday effect (estimated by the intercept) is positive in the case of Leisure 1 and Leisure 2, and suggests that, at average leisure concentration values, the concentration of leisure is about 4 percent larger on Saturdays than in the period Monday-Thursday. Concentration on Sundays (obtained by adding the coefficients associated to the intercept and to $I^{Sun-MTW}$) is smaller to that observed in the period Monday-Thursday in the case of Leisure 1.

I re-estimated the model in (10) by the methods explained above but replacing ΔHHI_{i3}^{L} with ΔTHI_{i3}^{L}. The different weighting pattern implicit in the *THI* revealed empirically insignificant. The most reliable estimates suggest, again, a U-shaped relationship between the quantity of leisure and its degree of concentration. The peak of the U is located at 10.0, 10.5, and 17.5 hours (ranging from our narrowest definition of leisure to the broadest), and an extra hour of leisure is estimated to reduce concentration by approximately 4 percent when the effect is computed at average leisure values. I also re-estimated the model excluding the travel time associated to each activity, finding that the main findings reported here were preserved.

5 Conclusions and directions for future research

We have presented the Herfindahl-Hirschman index (*HHI*) as a well-grounded measure of concentration of an individual's activity profile. The operationality of the *HHI* as a measure of

time-use concentration is highest when information on the allocation of time is collected by the time diary, as this methodology achieves the highest validity and reliability in the measurement of the use of time. The set of weights with which relative time shares are combined in the *HHI* revealed empirically insignificant in the application contained in this study. Similarly, the main empirical conclusions remained unaltered when the number of activities distinguished in the activity profile was expanded.

A daily leisure demand model predicted a linear or convex profile for the concentration of leisure activities in response to variations in the quantity of leisure available. The observed response in a sample of prime-age German men was indeed convex, with the peak of the function located well to the right of average leisure. To identify this behavior we relied on sequential moment conditions for the concentration of leisure and on weekend working arrangements, which revealed as valid and relevant instrumental variables in many of the specifications considered.

The observed leisure concentration profile is consistent with the existence of a minimum quantity of daily leisure postulated in the theory. It likewise suggests that individuals having less leisure opt for a more concentrated (and perhaps less varied in the sense of Gronau and Hamermesh, 2008) pattern of daily leisure activities, whereby recreation sector firms should probably differentiate their products the most on non-working days. The behavior of women as well as of younger and older men will permit judging the generality of this pattern. Controlling additionally for possible self-selection into the labor force, the estimation of our empirical model could be extended to working women. For students, the exogenous reduction in classes and lectures brought about by the weekend could play the role of the weekend working arrangements in the instrument set.

As market work crowds out leisure (e.g., see Hamermesh, 2006, and Donald and Hamermesh, 2009), another implication of our findings is that market work is constraining the pattern of daily leisure activities. Evaluating the effect that this constraint exerts on individual well-being should be also the goal of future research (the evidence on the effect of the breadth of leisure activities undertaken on individual well-being is rather limited and mixed; see e.g. Ray, 1979, and Sonnentag, 2001), as well as estimating the amount of money required to offset that constraint, which seems relevant for designing effective hourly rate and overtime compensations.

Appendix

Table 7 presents OLS regressions for the potentially endogenous ΔL_{i3} and ΔL_{i3}^2 separately for each leisure definition. The upper panel of the table presents results for the full sample, whereas results for the subsample of workers who did not postpone the completion of the first diary are shown in the lower panel. Standard errors (shown in parentheses) are robust to heteroskedasticity.

Table 7
First-stage regression for ΔL_{i3} and ΔL_{i3}^2 – OLS estimates

	Full sample: 2,266 employed prime-age men					
	Dependent variables (the definition of leisure is indicated by the number after the comma):					
	ΔL_{i3},1	ΔL_{i3}^2,1	ΔL_{i3},2	ΔL_{i3}^2,2	ΔL_{i3},3	ΔL_{i3}^2,3
Independent variables	(1)		(2)		(3)	
Works every Sat/Sun	-.747 (.210)**	-12.34 (2.77)**	-.975 (.233)**	-16.05 (3.34)**	-1.66 (.27)**	-52.88 (7.97)**
L_{i1}	-.269 (.104)**	-4.31 (1.71)**	-.460 (.107)**	-7.27 (1.86)**	-1.02 (.19)**	-30.08 (5.68)**
L_{i1}^2	.030 (.009)**	.553 (.159)**	.038 (.009)**	.694 (.157)**	.035 (.006)**	1.06 (.20)**
$I^{Sat-Fri}$	-.648 (.194)**	-4.91 (2.73)	-.897 (.211)**	-9.23 (3.30)**	-.523 (.222)*	-10.94 (6.68)
$I^{Sun-MTW}$	.530 (.174)**	4.52 (2.44)	.333 (.189)	1.03 (2.90)	1.98 (.22)**	62.06 (6.50)**
$I^{Sun-Fri}$	-.460 (.229)*	-3.19 (3.40)	-1.03 (.24)**	-13.51 (3.91)**	.826 (.263)**	31.81 (8.18)**
Intercept	3.28 (.29)**	39.00 (4.18)**	4.52 (.33)**	60.57 (5.16)**	11.22 (1.31)**	318.9 (39.0)**
R^2	.045	.054	.051	.053	.107	.113
Kleibergen-Paap rank test	6.61 [.037]		22.11 [.000]		57.52 [.000]	
Cragg-Donald statistic	2.70		9.47		26.23	
Kleibergen-Paap *F*-statistic	2.20		7.35		19.11	

Table 7 (Cont.)

	Sub-sample: 1,431 individuals who did not postpone the completion of the first diary					
	Dependent variables (the definition of leisure is indicated by the number after the comma):					
	ΔL_{i3},1	ΔL^2_{i3},1	ΔL_{i3},2	ΔL^2_{i3},2	ΔL_{i3},3	ΔL^2_{i3},3
	(4)		(5)		(6)	
Works every Sat/Sun	-.768 (.243)**	-12.56 (3.35)**	-.932 (.278)**	-15.32 (4.18)**	-1.57 (.34)**	-49.75 (10.02)**
L_{i1}	.021 (.108)	4.66 (1.65)**	-.048 (.121)	4.53 (2.01)*	.395 (.212)	21.91 (6.66)**
L^2_{i1}	-.019 (.009)*	-.525 (.155)**	-.023 (.010)*	-.656 (.180)**	-.028 (.008)**	-1.15 (.25)**
$I^{Sat-Fri}$	-.820 (.224)**	-7.19 (3.22)*	-1.11 (.25)**	-12.63 (3.88)**	-.465 (.254)	-10.48 (7.68)
$I^{Sun-MTW}$	.880 (.224)**	10.27 (3.33)**	.546 (.241)*	5.22 (3.93)	2.21 (.26)**	72.25 (8.00)**
$I^{Sun-Fri}$	-.334 (.289)	-1.83 (4.38)	-.953 (.304)**	-13.58 (4.98)**	.937 (.324)**	34.20 (10.11)**
Intercept	3.21 (.30)**	27.93 (4.27)**	4.31 (.37)**	44.92 (5.74)**	3.67 (1.40)**	30.63 (43.2)
R^2	.061	.038	.088	.055	.159	.149
Kleibergen-Paap rank test	26.11 [.000]		34.79 [.000]		55.94 [.000]	
Cragg-Donald statistic	9.68		14.45		26.47	
Kleibergen-Paap *F*-statistic	8.66		11.54		18.56	

Notes: Heteroskedasticity robust standard errors are in parentheses and probability values in brackets. L_{i1} is measured in hours and its definition is consistent with that of the dependent variable. The Cragg-Donald statistic is the minimum eigenvalue of the *F*-statistic matrix analog for testing the joint significance of the excluded instruments on the first-stage regressions. The Kleibergen-Paap *F* statistic equals to a quadratic form of an orthogonal transformation of the smallest singular value of the F-statistic matrix analog. The Kleibergen-Paap *F* statistic reduces to the Cragg-Donald statistic when the reduced-form errors are i.i.d. *: Significant at 5 percent. **: Significant at 1 percent.,
Source: German Time Budget Survey (ZBE) 2001/2002, own calculations.

Given the definition of ΔL_{i3}, by which leisure hours on a weekday are subtracted from leisure hours on a Saturday or Sunday, it is not surprising to estimate a positive and large coefficient associated to the intercept: Ranging from our narrowest definition of leisure to the broadest, the estimates are 3.3, 4.5, and 11.2 hours, respectively. The leisure gain brought about by the weekend is smaller for individuals working every weekend, whose weekend leisure forgone increases as the definition of leisure broadens: 0.7, 1.0, and 1.7 hours less, respectively. This effect is precisely measured and attains statistical significance at the 0.01 level. Irrespective of the leisure definition, the partial effect of L_{i1} on ΔL_{i3} or ΔL_{i3}^2 is negative for most of the leisure range, a result that seems partly driven by the positive correlation between L_{i1} and L_{i2}. (In the case of Leisure 1, for instance, this correlation is 0.26, whereas that between L_{i1} and L_{i3} is -0.01.) Although all excluded instruments are statistically significant at the 0.01 level in the full sample, and most of them achieve standard significance levels in the subsample, with two endogenous regressors the statistical significance of the excluded instruments is not sufficient in general to identify the β's, as identification requires that the matrix with the reduced-form coefficients associated to the excluded instruments have full rank (Wooldridge, 2002, p. 214). We have tested the null hypothesis that this matrix does not have full rank using the Kleibergen and Paap (2006) rank test. The p-values of this test, listed in Table 7, indicate that our instruments are adequate to identify the β's.

As is well-known, when the vector of instruments is weakly correlated with the endogenous regressors, standard IV coefficient estimates tend to be biased toward $plim(\hat{\beta}^{OLS})$ even in very large samples (e.g., see Staiger and Stock, 1997, and Stock and Yogo, 2005). Since weak instruments can also distort the significance levels for tests based upon standard IV, we shall test for weak instruments using the Stock and Yogo (2005) size-based test.[15] Its null hypothesis is that conventional 5%-level Wald tests for the β's based on IV statistics have an actual size that exceeds a certain threshold, for example 10%. The test statistic with two endogenous regressors is the Cragg and Donald (1993) statistic, whose value and definition are provided in Table 7. Table 7 also presents the value and definition of the *F*-statistic form of the Kleibergen and Paap (2006) statistic, which can be interpreted as a generalization of the Cragg-Donald statistic to the case with non-i.i.d. errors in the reduced-forms for the endogenous regressors. Critical values are taken from Stock and Yogo (2005, Table 5.2). To assure, for example, that the actual size of 5%-level tests for the β's is no greater than 10% (respectively, 15% and 25%), the test statistic must be greater than 13.43 (8.18 and 5.45) with three excluded instruments. In this study, the value of both statistics is generally above the 15% threshold critical value, the main exception being the regressions for Leisure 1 on the full sample. Hence, the estimates presented in column (1) of Table 6 may be biased toward $plim(\hat{\beta}^{OLS})$ because the instruments appear as weak.

[15] The alternative Stock and Yogo (2005) bias-based test requires at least four excluded instruments when there are two endogenous regressors.

References

Aguiar, M. and E. Hurst (2007), Measuring trends in leisure – The allocation of time over five decades, in: *Quarterly Journal of Economics*, Vol. 122, No. 3, 969-1006.

Altonji, J. G. (1986), Intertemporal substitution in labor supply – Evidence from micro data, in: *Journal of Political Economy*, Vol. 94, No. 3, 176-215.

Arellano, M. and B. Honoré (2001), Panel data models – Some recent developments, in: Heckman, J. J. and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, 3229-3296, Elsevier Science B.V..

Ås, D. (1978), Studies of time-use – Problems and prospects, in: *Acta Sociologica*, Vol. 21, No. 2, 125-141.

Bailey, D. and S. E. Boyle (1971), The optimal measure of concentration, in: *Journal of the American Statistical Association*, Vol. 66, No. 336, 702-706.

Baumgardner, J. R. (1988), Physicians' services and the division of labor across local markets, in: *Journal of Political Economy*, Vol. 96, No. 5, 948-982.

Becker, G. S. (1965), A theory of the allocation of time, in: The Economic Journal, Vol. 75, No. 299, 493-517.

Biddle, J. and D. Hamermesh (1990), Sleep and the allocation of time, in: *Journal of Political Economy*, Vol. 98, No. 5, 922-943.

Bittman, M. (2005), Sunday working and family time, in: *Labor and Industry*, Vol. 16, No. 1, 59-83.

Bittman, M. and J. Wajcman (2000) The rush hour – The character of leisure time and gender equity, in: *Social Forces*, Vol. 79, No. 1, 165-189.

Bonke, J., Deding, M., Lausten, M. and L. S. Stratton (2008), Intra-household specialization in housework in the United States and Denmark, in: *Social Science Quarterly*, Vol. 89, No. 4, 1023-1043.

Bosch, G. (2009), Working time and working time policy in Germany, in: Working time – In search of new research territories beyond flexibility debates, The Japan Institute for Labor Policy and Training Report no. 7.

Cragg, J. G., and S. G. Donald (1993), Testing identifiability and specification in instrumental variable models, in: *Econometric Theory*, Vol. 9, No. 2, 222-240.

Deaton, A. and J. Muellbauer (1980), Economics and consumer behavior, Cambridge University Press.

Donald, S. G. and D. S. Hamermesh (2009), A structural model of the fixed time costs of market work, in: *Economics Letters*, Vol. 104, No. 3, 125-128.

Eurostat (2004), Guidelines on harmonized European time use surveys, Office for Official Publications of the European Communities, Luxembourg.

Fisher, K., Gershuny, J., Altintas, E. and A. H. Gauthier (2012), Multinational time use study – User's guide and documentation, Version 5, University of Oxford.

Geary, R. C. (1950-1951), A note on "A constant-utility index of the cost of living", in: *The Review of Economic Studies*, Vol. 18, No. 1, 65-66.

González Chapela, J. (2006), On measuring heterogeneity in the use of time, *Electronic International Journal of Time Use Research*, Vol. 3, No. 1, 110-119.

Goodin, R. E., Rice, J. M., Bittman, M. and P. Saunders (2005), The time-pressure illusion – Discretionary time vs. free time, in: *Social Indicators Research*, Vol. 73, No. 1, 43-70.

Goodin, R. E., Rice, J. M., Parpo, A. and L. Eriksson (2008), Discretionary time – A new measure of freedom, Cambridge University Press, New York.

Gronau, R. and D. S. Hamermesh (2008), The demand for variety – A household production perspective, in: *The Review of Economics and Statistics*, Vol. 90, No. 3, 562-572.

Hall, M. and N. Tideman (1967), Measures of concentration, in: *Journal of the American Statistical Association*, Vol. 62, No. 317, 162-168.

Hall, R. E. (1978), Stochastic implications of the life cycle-permanent income hypothesis – Theory and evidence, in: *Journal of Political Economy*, Vol. 86, No. 6, 971-987.

Hamermesh, D. S. (2002), Timing, togetherness and time windfalls, in: *Journal of Population Economics*, Vol. 15, No. 4, 601-623.

Hamermesh, D. S. (2006), The time and timing costs of market work and their implications for retirement, IZA Discussion Paper No. 2030.

Hansen, L. P. (1982), Large sample properties of generalized method of moments estimators, in: *Econometrica*, Vol. 50, No. 4, 1029-1054.

Horowitz, A. and I. Horowitz (1968), Entropy, Markov processes and competition in the brewing industry, in: *The Journal of Industrial Economics*, Vol. 16, No. 3, 196-211.

Hufnagel, R. (2008), Entropy and stability in time use – An empirical investigation based on the German Time Use Survey, in: *Electronic International Journal of Time Use Research*, Vol 5, No. 1, 26-42.

Jenkins, S. P. and L. Osberg (2005), Nobody to play with? – The implications of leisure coordination, in: Hammermesh, D. and G. Pfann (Eds.), *The economics of time use*, 113-145, Elsevier, Amsterdam.

Juster, F. T. (1985), The validity and quality of time use estimates obtained from recall diaries, in: Juster, F. T. and F. P. Stafford (Eds.), *Time, goods, and well-being*, 63-92, Institute for Social Research, University of Michigan.

Juster, F. T. and F. P. Stafford (1985), Time, goods, and well-being, Institute for Social Research, University of Michigan.

Kahneman, D. and A. B. Krueger (2006). Developments in the measurement of subjective well-being, in: *Journal of Economic Perspectives*, Vol. 20, No. 1, 3-24.

Kleibergen, F. and R. Paap (2006), Generalized reduced rank tests using the singular value decomposition, in: *Journal of Econometrics*, Vol. 133, No. 1, 97-126.

Kooreman, P. and Arie Kapteyn (1987), A disaggregated analysis of the allocation of time within the household, *The Journal of Political Economy*, Vol. 95, No. 2, 223-249.

Leuthold, J. H. (1968), An empirical study of formula income transfers and the work decision of the poor, in: *The Journal of Human Resources*, Vol. 3, No. 3, 312-323.

Mattingly, M. J. and S. M. Bianchi (2003), Gender differences in the quantity and quality of free time – The U.S. experience, in: *Social Forces*, Vol.81, No. 3, 999-1030.

Murray, M. P. (2006), Avoiding invalid instruments and coping with weak instruments, in: *The Journal of Economic Perspectives*, Vol. 20, No. 4, 111-132.

Owen, J. D. (1971), The demand for leisure, in: *Journal of Political Economy*, Vol. 79, No. 1, 56-76.

Prowse, V. (2009), Estimating labour supply elasticities under rationing – A structural model of time allocation behavior, in: *Canadian Journal of Economics*, Vol. 42, No. 1, 90-112.

Ramsey, J. B. (1969), Tests for specification errors in classical linear least-squares regression analysis, in: *Journal of the Royal Statistical Society*, Vol. 31, No. 2, 350-371.

Ray, R. O. (1979), Life satisfaction and activity involvement – Implications for leisure service, in: *Journal of Leisure Research*, Vol. 11, No. 2, 112-119.

Robinson, J. P. (1985), The validity and reliability of diaries versus alternative time use measures, in: Juster, F. T. and F. P. Stafford (Eds.), *Time, goods, and well-being*, 33-62, Institute for Social Research, University of Michigan.

Robinson, J. P. and G. Godbey (1999), Time for life, Pennsylvania State University Press.

Schwartz, T. and C. McCarthy (2007), Manage your energy, not your time, in: *Harvard Business Review October 2007*.

Sevilla, A., Gimenez-Nadal, J. I. and J. Gershuny (2012), Leisure inequality in the United States – 1965-2003, in: Demography, Vol. 49, No. 3, 939-964.

Sonnentag, S. (2001), Work, recovery activities, and individual well-being – A diary study, in: *Journal of Occupational Health Psychology*, Vol. 6, No. 3, 196-210.

Sonnentag, S., Perrewé , P. R. and D. C. Ganster (2009), Current perspectives on job-stress recovery – Research in occupational stress and well being Volume 7, Emerald Group Publishing Limited.

Stafford, F. P. and M. S. Cohen (1974), A model of work effort and productive consumption, in: *Journal of Economic Theory*, Vol. 7, No. 3, 333-347.

Staiger, D. and J. H. Stock (1997), Instrumental variables regression with weak instruments, in: *Econometrica*, Vol. 65, No. 3, 557-586.

Statistisches Bundesamt (2005), Technical quality report – The 2001/2002 Time Budget Survey, Statistisches Bundesamt, Wiesbaden.

Stock, J. H. and M. Yogo (2005), Testing for weak instruments in linear IV regression, in: In D. W. K. Andrews and J. H. Stock (Eds.), *Identification and inference for econometric models – A festschrift in honor of Thomas Rothenberg*, 80-108, Cambridge University Press.

Stone, R. (1954), Linear expenditure systems and demand analysis – An application to the pattern of British demand, in: *The Economic Journal*, Vol. 64, No. 255, 511-527.

Theil, H. (1967), Economics and information theory, Amsterdam, North-Holland.

Tinsley, H. E. A. and B. D. Eldredge (1995), Psychological benefits of leisure participation – A taxonomy of leisure activities based on their need-gratifying properties, in: *Journal of Counseling Psychology*, Vol. 42, No. 2, 123-132.

Wooldridge, J. M. (2002), Econometric analysis of cross section and panel data, The MIT Press, Cambridge and London.

electronic ***International Journal of Time Use Research***
2013, Vol. 10, No. 1, 38-54 dx.doi.org/10.13085/eIJTUR.10.1.38-54

Paradata correlates of data quality in an SMS time use study – Evidence from a validation study

Philip S. Brenner and John D. DeLamater

Philip S. Brenner
Department of Sociology
University of Massachusetts Boston
100 Morrissey Blvd.
Boston, MA 02125
E-Mail: philip.brenner@umb.edu

John D. DeLamater
Department of Sociology
University of Wisconsin-Madison
8128 Sewell Social Sciences Bldg.
1180 Observatory Drive
Madison, WI 53706
E-Mail: delamate@ssc.wisc.edu

Abstract

Abstract: Short Message Service (SMS) text messaging is a ubiquitous technology available on the vast majority of cellphones in use in 2013. It provides a common technological denominator between mobile devices of nearly every make and model, supplying researchers an avenue to collect data without the expense and difficulty of designing specific applications for every cellphone or device on the market. SMS/text messaging was used as a method of data collection using a sample of students from a large, Midwestern university. The procedure adapted conventional time use measurement procedures to fit the device, the sample, and the behavior of interest. After answering questions on a brief Web survey, respondents were asked to text researchers for five days, updating major changes in their activities. Following data collection, data from the text condition was compared to that from a conventional (Web) survey and data from a reverse record check from campus recreation facilities to validate reports of the behavior of interest – physical exercise and activity. Findings suggest that respondents provided consistently high quality data on self-reports of the behaviors of interest. Moreover, paradata measures of text data quality (e.g., number of text messages sent, number of days with messages) predict data quality on the behavior of interest.

JEL-Codes: C830

Keywords: time use, paradata, text messaging, cellphones, response validity

1 Introduction

Short Message Service (SMS) text messaging is widely and frequently used by young adults. In a recent study by researchers at Ball State University, 99 percent of students reported having a cellphone, and virtually all of these students (97 percent) reported sending and receiving text messages (Ball State University, 2009). Many of these young adults text prodigiously. A recent study by the Pew Research Center estimated that young adults send an average of 109.5 texts a day.[1] Moreover, the heaviest users of texting prefer text to talk. Over half (55 percent) of adults sending or receiving more than fifty text messages a day prefer a text message to a phone call.[2]

More than just the ubiquity and utilization of the technology makes it of interest to social scientists in search of data collection opportunities (Schober et al., 2013). Perhaps even more important is the manner of its use. In conjunction with other more recent, web-based social networking technologies and applications (e.g., Instant Messaging [IM], Facebook, Twitter, Foursquare, Google+), texting is used to report current activities and locations to others. All of these technological tools provide researchers with new opportunities for data collection, as well as data mining, to address a wide variety of social science concerns.

However, SMS provides researchers a data collection opportunity not shared by its more recent competitors. SMS is a ubiquitous technology available on nearly all cellphones in use today. It does not require state-of-the-art technology or cutting edge consumer electronics (e.g., a smartphone running the latest version of Google's Android or Apple's iOS) nor does it require additional software development or any intermediary Web-based application (e.g., a Twitter client, a Facebook app, Whatsapp or an IM client, or a custom software application) for data collection (Raento, Oulasvirta, & Eagle, 2009). Rather, SMS/text messaging provides a common technological denominator between smartphones and basic cellphones of nearly every make and model, supplying researchers an avenue for data collection without the expense and difficulty of designing specific applications for every cellphone on the market.

Beyond its ubiquity, three other reasons underline the benefits of using SMS, rather than a downloadable software application. First, requiring survey respondents to download an application to participate generates respondent apprehension and can negatively affect response rates (Walton, Buskirk, & Wells 2013). Second, texting provides a perception of privacy and confidentiality unavailable (or not easily available) with Web 2.0-based social networking

[1] The same study put the median number of texts per day at about 50. The difference between the mean and median shows that the distribution is highly positively skewed, suggesting the presence of some very extraordinary texting outliers.

[2] Notably, this survey was conducted, via voice, to landline and cell numbers. The response rate for the cell sample was 11.5 percent, two points less than that for the landline sample.

applications.[3] Other current messaging applications (e.g., Twitter, Facebook, Foursquare) are a one-to-many communication technology by their very nature. Typical use of these services involves sending a report on one's current activity, location, or state of mind for multiple (or all) other users of the service to see. This default, rightfully (and hopefully) leads to a selection bias of what is shared and what is not (see Patchin & Hinduja, 2010). Thirdly, and relatedly, are additional privacy concerns involving ownership of the data and the potential for inadvertently sharing information with a larger audience of users than intended. Texting, however, is inherently a one-to-one communication channel, lending itself more naturally to a data collection procedure in which a confidentiality assurance can be implemented.

However, for all of its benefits, text messaging as a data collection mode has drawbacks. In the US, cellphone and smartphone users pay either a monthly fee or per text for incoming and outgoing messages. These fees are in addition to monthly charges for voice and data, and may potentially lower participation. SMS also presents an obstacle to the conventional standardized question and response scale paradigm (see Fowler & Mangione, 1990). Certainly, standardized questions and response options could be (and have been) sent by way of SMS to the respondent with instructions for the respondent to select an answer to and respond with the numerical code reflecting their answer back to the researcher. However, texting is, by the nature of the medium, idiomatic. Unlike a Web survey with checkboxes or radio buttons, consistency checks and forced response, there is nothing to prevent the SMS/text respondent from answering how s/he sees fit, regardless of the standardized options. While respondents could potentially be trained to respond with a number associated with a response option, changing the expressive nature of the text message to force it into the standardized questionnaire paradigm fails to capture the strength of the method.[4]

Many SMS-based data collection procedures previously used, even those labeled "diaries," have been somewhat more akin to the Experience Sampling Method (ESM) or conventional survey data collection. For example, in an "SMS Pain Diary" Alfvén (2010) asked respondents to reply to six messages a day using a prearranged coding scheme to report intensity, duration, and results of pain. Similarly, Anhøj and Møldrup (2004) used SMS to send a series of yes or no questions measuring the occurrence of asthma symptoms and use of medication to respondents at preselected times during the day.

In these examples and other extant work, researchers fail to leverage the strengths of using SMS for diary data collection. The idiomatic nature of SMS is a strength of the time diary method of data collection. The strength of chronologically based data collection procedures, like time diaries, is in their ability to avoid the measurement bias that plagues direct survey

---

[3] This is not to say that the transmission of text messages is perfectly confidential. However, texters view their phones as private devices and believe that there is a "widely accepted, unwritten rule" about the confidentiality of text messages (Häkkilä & Chatfield, 2005).

[4] Verbatim responses are not without their own problems, of course. Each message requires coding; an expensive and time-consuming proposition. Moreover, the nature of text messaging is miserly with time and effort, with a focus on abbreviation. Many of the abbreviations used in texting are now well known and do not necessarily present coding problems, although idiosyncratic abbreviations or acronyms may.

questions on normative behaviors (Bolger, Davis, & Rafaeli, 2003). Like other normative behaviors, physical exercise is widely understood to be overreported in surveys using conventional direct questions (Ainsworth, Jacobs, & Leon, 1992; Chase & Godbey, 1983; Klesges, 1990). Verbatim responses to open-ended questions (i.e., "What did you do next?") allow researchers to avoid direct questions about specific behaviors of interest (i.e., "Did you go to the gym?") (Robinson, 1985, 1999; Stinson, 1999), thereby avoiding prompting self-reflection on the part of the respondent, and yielding less biased and higher quality data on many normative behaviors (Bolger et al. 2003; Niemi 1993; Zuzanek & Smale, 1999.)

Like all data collection procedures, chronologically based data collection procedures also have weaknesses, two of which are pertinent to this conversation. First, respondents may fail to report activities of very brief duration that happen frequently during the day. For example, trips down the hall to use the restroom or to the water fountain are likely to be omitted as respondents tend to focus on longer activities (e.g., those that last for hours rather than minutes and the sorts of activities around which the day is planned. Therefore, the focal activities of such a data collection procedure should be these sorts of major activities.[5]

A second main weakness of chronologically based data collection procedures is primarily related to the heavy burden they place on respondents. This burden can result in high rates of nonresponse — either through refusals to participate that yield increased unit nonresponse or incomplete participation as respondents quit the study or choose to participate intermittently, resulting in partial interviews and item nonresponse.[6] In order to reduce the burden of the data collection process, diaries can be, either by the researcher's design or by the unilateral decision of the respondent, filled out at the end of the day or at the end of the reference period. However, shifting the timing of diary completion away from the time of occurrence of individual activities can result in poorer data quality as respondents may introduce errors into the data collection procedure, like forgetting to include events or attributing them to incorrect times.

Notably, the SMS procedure may not relieve respondent burden; rather, it may lead to increased time spent on the data collection task, although this time may be more equally distributed throughout the diary day. However, the SMS procedure does offer some promise as it incorporates features that address these weaknesses and may lead to higher quality data. First, respondents can be asked to report on attitudes or behaviors in situ and as they occur. This application of a real-time data collection procedure may help to reduce forgetting and other memory problems. Second, the procedure may overcome another problem with retrospective reporting – editing and judging. Without the time to reflect and put activities and feelings in context, an SMS-based reporting procedure may be able to avoid much of the social desirability effect and other sources of bias inherent to standardized survey questions. While perhaps

---

5 Notably, ESM would also likely fail to adequately measure activities that are very brief in duration.

6 The AAPOR Standard Definitions and other nonresponse terminology, while still very useful, fit somewhat awkwardly in the case of time use data collection. For example, there are not "items," per se, to be skipped, although certainly skipping parts of the data collection process yields a similar outcome.

not true for all behaviors and activities, especially contranormative, illegal, or embarrassing activities (e.g., illicit drug use or sexual activity), or those of high frequency and brief duration (e.g., using the restroom or getting a drink of water), this procedure should allow a more accurate measurement of the normative activities that are often overreported and that could be considered major activities in a day's schedule (e.g., going to religious services, volunteering, or exercising).

Third, using a technology that some hard-to-survey populations (e.g., young adults) find relevant to their daily lives may yield a more representative achieved sample. Commonly used sampling designs, like random digit dialing, typically produce sampling frames that yield undercoverage of the young adult population (Blumberg & Luke 2007; Currivan, Roe, & Stockdale, 2008). Making matters even worse, conventional survey modes commonly result in high rates of nonresponse amongst sampled individuals in this age group (Groves & Couper, 1998). Combined with an appropriate sampling design, this adoption and adaptation of a technology used frequently by young adults may provide an additional level of interest to leverage their participation (Groves, Singer, & Corning, 2000). In sum, using texting in a manner similar to other diary-like Web-based applications (i.e., Facebook and Twitter) may encourage the participation of young adults, garnering higher rates of cooperation than more conventional data collection methods.

While not a panacea, an SMS-based chronological data collection procedure does offer some promise in reducing these forms of error. However, the promise of this procedure strongly depends on three important considerations. First, the target population must be one that fits well with the method (e.g., a population with a high rate of ownership and use of cell- or smartphones, preferably with unlimited texting plans, like young adults).[7] Second, the sampling frame should either contain cellphone numbers or be readily switchable between a recruitment mode (e.g., Web/email, landline phone, mail) to a cellphone number for data collection. Given the requirements of the first point, Web/email would be the obvious choice. Third, the research problem or question must be one that fits the method well (e.g., an interest in major activities, rather than very frequent but short-duration activities).

The current project matches these requirements well. This technology was used to obtain reports from a sample of university undergraduates regarding their daily activities. The research was focused specifically on the validity of measurement of physical exercise although this emphasis was not disclosed to respondents. Since this is one of the first attempts to implement this method in a rigorous research project, it is useful to examine these data to determine how well the method worked, the quality of the data it produced, and what can be done to improve each. To this end, a series of paradata indictors will be used to predict the observed criterion validity of the focal behavior, physical exercise at university recreational sports facilities.

---

[7] If used in a more general population, adequate funding must be available to purchase text-enabled cellphones for respondents, and/or reimburse them for the cost of the text messages, and provide training for their use.

2 Data and methods

A random sample of 325 undergraduates, stratified by gender and year in school, from a large, Midwestern university were sent an email invitation to participate in the "[University Initials] Student Daily Life Survey" in March and April 2011. The invitation was sent to the student's university email address and included a link to a Web survey. An email reminder to complete the survey was sent three days after the initial invitation, and a final reminder was sent five days after the first reminder email.

The Web survey was comprised of approximately twenty questions about usage of university facilities. While the true purpose of the study was to measure use of university recreation facilities, questions about type and frequency of use of campus libraries, the student union, and other facilities were also asked to mask the focus of the study. Respondents were asked about their "typical" use of recreational facilities on campus and their "usual" activities at these facilities (e.g., weightlifting, swimming, aerobics, and cross-training). Respondents received ten dollars upon completion of the Web survey. 124 respondents completed the Web questionnaire yielding a response rate of 38 percent.[8]

The final question of the Web survey was a request to participate in the SMS data collection procedure. Respondents were told that participation in this part of the project entailed sending text messages to the research team reporting all changes in their major activities for a period of five days. In acknowledgment of their participation, respondents were told that they would receive an additional thirty dollars at the conclusion of their participation. If the respondent was amenable to participating, s/he was asked to enter his or her cellphone number. 87 percent (108 of 124) of the respondents who completed the Web survey agreed to continue into the text component of the study.

Respondents were emailed a two-page participant guide, detailing how and what to report. The first page described the purpose of the study, the tasks required of the respondent, an example of a full day of nine text messages, and instructions on how to text updates to the research staff. Respondents were asked to report all changes in their major daily activities and where they were taking place. The second page of the document was a FAQ list, including instructions on how to report late activities and whom to call or email with questions or concerns.

Respondents were assigned to one of five five-day field periods. Cohorts of text respondents were distributed over a two-week period to ensure coverage of both weekday and weekend days. Respondents were reminded multiple times each day to send messages updating their activities. These reminders were more frequent on the first day of their participation (four times, at 10:00 AM, 1:00 PM, 5:00 PM, and 8:00 PM) and less frequent on the final days of participation (two reminders, at 10:00 AM and 8:00 PM). 81 percent (87 of 108) of the re-

[8] All response rates are computed as AAPOR RR 5, as there are no ineligible cases or cases of unknown eligibility.

spondents who agreed to participate in the text condition sent at least one text during the field period.

At the completion of the texting component of the study, each respondent was asked for his or her student identification number so that study staff could request records on each respondent's use of campus recreation facilities. These records are the product of the scanning of students' identification cards upon admission to the facilities. This process records the student's identification number and the time and day of admittance to the facility. 77 percent (67 of 87) of the respondents who completed the text condition permitted access to their record data, yielding final effective response rates of 27 percent for all texters and 20 percent for respondents allowing access to verification data.

2.1 Measures

Six paradata measures of data quality were observed and will be used as independent variables in the following analyses: (1) the total number of text messages sent, (2) the number of days the respondent sent messages, (3) the percent of messages sent late, (4) the number of days the respondent skipped, (5) the percent of reports that are temporally proximal to a reminder text, and (6) the number of messages that are repeats of prior messages. Two outcome variables will be used in the following analyses: (1) the validity (whether overreported or underreported) of the respondent's claim of the number of days s/he exercised at University facilities, and (2) an indicator of respondent compliance with the record check procedure. Each of these will be described in greater detail.

Total number of messages. The number of messages sent is clearly an indicator of data quality. The fewer messages sent by the respondent, the more likely activity has not been reported and the more poorly the corpus of the respondent's messages will represent his or her activity.[9] For example, unless the entire day was spent ill in bed, it is unlikely that one message could capture a respondent's daily activity. Respondents sent a total of 1904 messages, ranging from 2 to 59 messages per respondent (omitting the respondents who agreed to participate in the texting component of the study but did not send a text). Respondents averaged 22 messages (s.d. = 10.8) during their assigned field period of five days.[10]

Number of messaging days. Respondents were assigned to one of five five-day reporting periods to distribute reporting across the seven days of the week. On average, respondents submitted messages for 5.1 days (s.d. = 1.1), ranging from 2 to 8 days. Most respondents (81 percent) reported activities for at least five days. As this suggests, a number of respondents (31, or 36 percent) reported activities for more than the requested 5 days, while 19 respondents (22 percent) reported on fewer than five days. Failing to send updates for a given

[9] Clearly, this will vary by day of the week. Weekdays tended to have more activity than weekend days, especially Sunday, which elicited the fewest number of messages.

[10] In determining date received, messages received after midnight that reported an activity at the end of the day, typically "going to bed," were coded as received the previous day.

day yields missing data, which may contribute error to estimates. Therefore, sending messages on fewer than the five days assigned likely negatively affects data quality.

Percentage of late messages. Sending late messages may be caused by (or at least associated with) a respondent's lack of task conscientiousness. As such, having many late messages could be an indicator of missing data or other data quality issues. Late messages were those flagged by the respondent as reporting on activities occurring prior to the sending of the message. Knowing that respondents would likely forget to report some changes at the time they occurred, respondents were advised that, if necessary, they could report activities late by including a flag (the word "TIME" in all capital letters) and the time of the activity in the report. Respondents averaged five late reports during the field period, ranging from zero to 80 percent of their messages. Approximately 21 percent of reports were sent late (s.d = .23), and more than half (57 percent) of the respondents reported late one or more times. As can be seen from the range, some participants provided many late reports. Thirteen participants, 15 percent, texted more than half of their reports after-the-fact.

Number of skipped days. The integrity and validity of these data depends on every participant reporting each day during their assigned field period. Therefore, skipping days may yield missing data and contribute to poor data quality. Skipped days are not just a mathematical function of the number of messaging days and the number of days in the reference period. Some respondents who skipped a day in the middle of their assigned reference period continued to report after their assigned field period had ended, perhaps in an attempt to make up for the missed day. About 30 percent (26 respondents) skipped one or more days. The average number of skipped days was greater than a third of a day (0.40) per respondent, ranging from zero to three skipped days. Sundays were especially likely to be skipped; almost two-thirds of the respondents with skipped days (16 respondents) resulted from a failure to report activities for an assigned Sunday. Since the sample was drawn from an undergraduate student population and many of the provided examples were student-related activities, respondents may have felt it was unnecessary to report Sunday leisure activities.

Percentage of messages proximate to reminders. Another way to measure data quality is to evaluate responses by their proximity to reminder messages. There is no reason to believe that students would be engaging in new activities in any kind of systematic way at 10am, 1pm, 5pm and 8pm, and only at these times. Therefore, a high rate of messages proximate to these reminder messages suggests that the respondent may only be reporting activity changes in reaction to these prompts, therefore resulting in unreported events that occur at other times of the day. This would yield missing data and possibly result in poor data quality. The average rate of messages sent proximate to a reminder was about 19 percent (s.d. = .15), where "proximate" is defined as within thirty minutes following a reminder message. The observed range of proximity is very large, with minimum and maximum values matching theoretical limits: some respondents sent all of their messages just after a reminder, whereas other respondents did not send any messages proximate to a reminder.

Number of repeated messages. A careful reading of the corpus of messages indicated a small number of cases where the same message was sent twice within a few minutes. A message was considered repeated if two texts reporting the same activity were sent on the same day within 10 minutes of each other. Typically, the activity was reported twice (e.g., "Going to Target.") As a potential result of respondent carelessness, we include it here as a potential indicator of poor data quality. In two cases, the second message expanded the information contained in the first (e.g., "Going to the grocery store." "The one in [building name]."). These later cases were edited into a single message in the analytical dataset.

Validity of the report of exercise. The first outcome measure, the validity of the reporting on exercise activity, was computed as the difference between the reverse record check and the self-report from the respondent. Reported changes in respondents' major activities were coded for exercise activities and, more specifically, for those that occurred at campus recreational sports facilities. Each day with a report of exercise at a campus recreational sports facility was coded as 1, 0 otherwise. This variable was then summed over the days of the reference period.

Each day during the reference period with record of admittance to a campus recreational sports facility was coded as 1,0 otherwise. This procedure yielded a series of variables, one for each day, each coded for the presence or absence of an admittance. These were summed to reflect the number of days during the reference period that the respondent used campus recreational sports facilities.

The difference between the self-report and the record variable provided an estimate of the validity of the self-report of exercise. This procedure resulted in a three-category nominal variable: (0) valid reporters, (+1) overreporters, and (-1) underreporters. Due to small cell sizes, the latter two categories are collapsed in some analyses creating a dichotomous variable for comparison of accurate and inaccurate reports. Notably, these data appear to be of very high quality. About 80 percent of respondents reported accurately, their claims verified by the reverse record check. The remainder of cases was equally split between over- and underreporting suggesting that measurement error was random rather than systematic.

Compliance with record access. Finally, comparisons will be made with respondents for whom these validity data are available and those for whom these data are not available (i.e., those respondents who did not allow access to the record data). It is possible that respondents who disallowed access to their gym facility use records differ in a systematic way in their data quality from compliant respondents who allow access to these records. This analysis addresses this possibility.

2.2 Analysis plan

Two methods were used to examine the quality of these data and the value of the paradata indicators as predictors of the criterion validity of the measure of the focal behavior – physical exercise. The first method applied a cluster analysis to the full dataset (all text respondents, with or without validation data) to generate a typology of respondents in terms of the

paradata indicators of data quality. A k-means cluster analysis was estimated using a set of paradata variables from the text respondents, including number of days the respondent sent text messages, the total number of messages sent, the number of days the respondent skipped, and the number of late messages sent. These clusters were then compared using two outcomes computed from the validation procedure: (1) rate of inaccurate reporting, and (2) and rates of compliance for the reverse record check. Comparisons use Fisher's exact text and Cohen's d to assess statistical and substantive significance of the predictive value of the paradata indicators as a whole on data quality.

The second analysis uses logistic regression to predict the propensity of respondents to over- or underreport, given these paradata indicators of data quality. This analysis expands on the previous comparisons in two ways. First, it assesses individual paradata indicators of data quality given the criterion measure, discerning those that have predictive validity from those that do not. Second, this analysis permits separate prediction of both overreporting and underreporting, allowing for a more nuanced understanding of the nature of the error in the self-report of exercise and the effect of the paradata determinants of data quality in the assessment of validity.

3 Results

Subjective assessment of the results of the cluster analysis suggests that the most parsimonious model allows four clusters of respondents to emerge (see Table 1). For purposes of presentation, these clusters have been given descriptive names: (1) Prodigious texters, (2) Frequent texters, (3) Occasional texters, and (4) Infrequent texters.

The Prodigious texters of the first cluster comprised less than ten percent of the achieved sample (8 of 87 respondents). Respondents in this cluster sent an average of 44 messages during the reference period, yielding over eight messages a day on average, with no skipped days. About thirty percent of their messages were late and about 14 percent of their messages were sent shortly after reminder texts.

The second cluster, Frequent texters, comprised over a third of the achieved sample (31 of 87 respondents). The main difference between the Prodigious and Frequent texters was the number of messages sent: Frequent texters sent about a third fewer messages than the Prodigious texters. The respondents in this cluster sent about 28 messages during the reference period, yielding over five messages a day on average, skipping very few days. Very similar to the Prodigious texters, Frequent texters' messages were late about a quarter of the time and they sent about 15 percent of their messages shortly after reminder texts.

Occasional texters comprised the largest cluster of respondents at nearly 40 percent of the achieved sample (34 of 87 respondents). Occasional texters sent almost half the number of messages than the Frequent texters (approximately 17 messages during the reference period.) The Occasional texters also skipped about a third of a day on average, yielding fewer than 3.5

messages a day. About 22 percent of their messages were sent late and nearly 20 percent were sent shortly after reminder texts.

Table 1
Mean numbers, rates of key independent variables, by cluster

	Number of		Mean number of			Mean percentage of messages %	
Clusters	Respondents	Messages	Days	Skips	Repeats	Late	After reminder
Prodigious	8	44	6.3	0	0.75	31	14
Frequent	31	28	5.5	0.13	0.19	25	15
Occasional	34	17	5.1	0.38	0.09	22	19
Infrequent	14	7	3.7	1.29	0	4	32

Source: Student Daily Life Survey 2011, own calculations.

The final cluster, Infrequent texters, comprised about a sixth of the achieved sample (14 of 87 respondents). The respondents in this cluster sent only about seven messages during their entire reference period, averaging just over one message a day. These respondents shortened the intended reference period by over a day, skipping 1.3 days on average. Infrequent texters, however, sent very few late messages (about three percent). This rate of timeliness is not surprising given how few messages Infrequent texters sent. Moreover, of those messages, almost a third were sent within thirty minutes of a reminder text.

How do these clusters of respondents, generated using the indicators of data quality from the texting paradata, compare given the outcome of the validation procedure? First, consider the distribution of the 20 respondents who did not allow access to their recreational sport facilities admission records. These respondents were evenly distributed across categories: ten were in the top two categories of better respondents and the other ten were in the bottom two categories of poorer respondents. Thus, the respondent's decision to grant access to their record data is not associated with the quality of the respondent's texting performance.

The more important question is whether these clusters based on paradata have predictive validity. Table 2 compares respondents in these clusters by the outcome of the validation procedure. While cell sizes are small, there appear to be a number of important differences emerging. First, the rates of invalid responses (i.e., under- and overreports) appear to be higher for the Occasional and Infrequent texters. Ten percent of respondents in the Prodigious and Frequent clusters inaccurately report their exercise, but nearly 30 percent of the Occasional and Infrequent texters inaccurately report. While the raw between-group difference is quite large (Δ=20 percentage points; Cohen's d = 0.47), the small effective sample size (N=67) leads it to be just outside of conventional levels of statistical significance using either Fisher's exact test ($p = 0.058$) or Chi-square ($p = 0.064$).

Table 2
Validation of exercise reports, by cluster

	Result of the validation procedure				
	Valid		Not valid		Total
Clusters	N	%	N	%	N
Prodigious	5	83	1	17	6
Frequent	21	91	2	9	23
Top two	26	90	3	10	29
Occasional	19	70	8	30	27
Infrequent	8	73	3	27	11
Bottom two	27	71	11	29	38

Source: Student Daily Life Survey 2011, own calculations.

More direct tests of these potential indicators of data quality can be undertaken to predict the validity of the exercise measure. These tests will allow us to see which of these paradata indicators of data quality have the most purchase in explaining the quality of the exercise data. In addition, these tests will allow the error in the exercise measure to be separated into its two components: overreporting and underreporting (Table 3).

Table 3
Bivariate logistic regression coefficients from models predicting underreporting and overreporting

	Underreporting			Overreporting		
	Coeff.	s.e.	p	Coeff.	s.e.	p
Number of messages	-0.080	0.047	+	-0.007	0.037	
Number of days	-0.716	0.369	+	-0.038	0.387	
Number of message per day	-0.459	0.258	+			
Number of skips	0.622	0.469		0.123	0.548	
Percentage late	-2.940	2.420		-2.193	2.231	
Number of repeats	-0.387	1.002		-0.387	1.002	
Percentage reminder	2.914	2.005		1.494	2.115	

Note: +p < .10; N = 67
Source: Student Daily Life Survey 2011, own calculations.

Logistic regression models were estimated predicting overreporting and underreporting using each of the indicators of data quality: number of messages, number of days with messages, number of skipped days, number of repeated messages, and percentages of late messages and messages sent following a reminder. Results show an important difference between the two forms of error. While none of these indicators predict overreporting in bivariate models, two bivariate models approach conventional levels of statistical significance when predicting underreporting. Both the number of messages sent (β = -0.08; p = 0.09) and the number of days with messages (β = -0.72; p = 0.052) predict underreporting, although the p-value of these tests is just outside conventional levels of statistical significance. As would be expected, these

relationships are negative; each additional message sent yields a reduction in the odds of underreporting by eight percent. Moreover, each additional day of messaging leads to a reduction in the odds of underreporting by 50 percent.

Since these two indictors of data quality are highly correlated ($r = 0.7$), including them both in a multivariate model results in multicollinearity. Therefore, a new variable, average number of messages per day, was computed as the dividend of these two indicators. A similar finding emerged when underreporting was regressed on this new variable. Every unit increase in the average rate of messages per day reduces the odds of underreporting by about a third ($\beta = -.46$; $p = 0.07$). This finding, like those from previous models, is of marginal statistical significance, but suggests that the number of messages and the number of messaging days may be predictive of the validity of key measures.

4 Discussion

Clearly, the most important paradata indicators in predicting data quality are (1) the number of messages and (2) the number of days with messages. These two indicators vary a great deal from the best cluster of respondents (44 messages over all five field days) to the worst cluster of respondents (7 messages with 1.3 field days missed). The distinction between the best and the worst clusters of respondents is stark – a 20 percentage point difference in the validity of their responses. Moreover logistic regression modeling supports this finding. Both the number of messages and the number of messaging days predict data quality – the more of each, the less likely the respondent is to underreport their exercise.

The strength of the paradata indicators of data quality in predicting only one of the two forms of error in the exercise measure may be explained by understanding the nature of these two forms. Overreporting is an error of commission; the respondent has made a claim that cannot be verified. The inability of the indicators of data quality to predict overreporting is understandable as the method is more prone to errors of omission than commission. In contrast, underreporting is an error of omission. The most likely cause of this error is missing data generated by nonresponse (i.e., failure to send updates). This could take a couple of guises, like forgetting or intentionally failing to report on an activity, choosing to end participation in the study early, or skipping days in the middle of the reference period.

Surprisingly, this last type of nonresponse – skipped reporting days – does not increase one's likelihood to underreport. This may be due to a problem with nonresponse, typified by many students' Sunday reports. A number of respondents reported very few Sunday activities, texting only a message like "staying in today" or "at home studying." It is possible that other respondents with a similar level and type of activity failed to report days in which they did not venture out from home. If this is the nature of a skipped reporting day, it is clear why this indicator of data quality would not predict underreporting of exercise at a campus recreation facility. In future applications of this method, researchers must more clearly and carefully

specify which types of activities should be reported and emphasize reporting on each day included in the reporting period.

Surprising, at least initially, is the comparability of the rate of late messaging in the two clusters of more conscientious respondents (Prodigious and Frequent texters) with the somewhat less conscientious respondents in the Occasional texter cluster. This, in combination with its weakness as a predictor in the logistic regression models, suggests that lateness, in and of itself, is likely not a good indicator of data quality. Rather, it may be an inevitable result of this sort of in situ data collection procedure. The rate of lateness, and the lack of an effect of lateness on data quality, suggests that respondents should be told that, while not ideal, sending late messages is understandable and a process should be created to allow respondents to send researchers late reports of their activities, such as that which was provided.

But can lateness be combated with well-timed reminder messages to prompt respondents to update researchers on their recent activities? Compare the findings on lateness with those on the percentage of messages sent after a reminder. In the two clusters of more conscientious respondents (Prodigious and Frequent texters), this rate is between 14 and 15 percent. This rate increases to 19 percent for the Occasional texters, and to 32 percent for the Infrequent texters. This suggests that poorer respondents are either less likely to remember the task of reporting or more likely to wait for a reminder whereas better respondents are more proactive in reporting their activities (Brenner & DeLamater, 2013). Nevertheless, the difference between the top three categories is not large. Further research on the role of reminders may help to clarify their role in data quality; that is, do reminders prompt otherwise good respondents to improve the quality of their data, or do they spur poor respondents to give only a barebones effort?

Perhaps the largest single problem with this particular study is the low response rate. In order to meet the requirements of the human subjects review board, the design of the study required multiple requests for participation from respondents, creating multiple opportunities for respondents to decide to discontinue their participation. These include (1) the initial request for participation, (2) the request for the respondent's cellphone number, (3) the instruction to begin the text component of the study, and (4) the request for the respondent's student identification number for the collection of validation data. With each subsequent request, some sample members inevitably failed to continue participation. In spite of the low response rate, additional analyses do not suggest that unit nonresponse has biased estimates (results not shown). For example, the rate of compliance for the reverse record check does not differ between clusters; 79 percent of the Occasional and Infrequent respondents allowed access to their records compared to 75 percent of the Frequent and Prodigious texters. Future research should attempt to combine these requests or better link each step to the payment of incentives.

Relatedly, the second most important problem is the relatively small sample size, exacerbated by the low response rate, which limits the analyses that can be pursued and leads to a lack of statistical power. In some analyses, over- and underreporting were pooled into a single cate-

gory for comparison with accurate reports. Still, the findings are suggestive and are meant to spur further research.

The sample used here was of undergraduate students at an elite public Midwestern university. As such, findings are hardly generalizable to a national population or even to the larger population of young adult Americans. Yet, this sample was good for testing the main hypothesis (see Brenner & DeLamater, 2013) and ideal for avoiding the self-selection bias that is inherent in similar studies. All students in the sampling frame automatically have access to the campus recreation facilities without making the effort to join (and pay) for membership. Unlike a sampling frame from a similar organization comprised of members of the general population, (e.g., membership rolls at a YMCA or a for-profit fitness center), the sampling frame from the university registrar or bursar allows a frame of gym members (i.e., all students) without a self-selection bias. Nevertheless, future research should attempt to use a sampling frame from a more varied target population.

The method itself has weaknesses that must be weighed, along with its benefits, before being employed. While this method may be well paired with some populations and research topics, like this one, there are other populations (e.g., older adults, employees whose workplaces disallow cell phone use) and research topics (e.g., contranormative behavior, very brief focal activities) with which this method may not match. Researchers should carefully consider the fit of this method, as they would any other method, with the details of a particular sampling design and research topic.

Moreover, SMS may increase respondent burden compared to other chronological measurement methods like a (once-a-day) time diary or ESM. While the in situ data collection of the SMS procedure has positive measurement properties, it requires a great deal of the respondent's time and effort. If matched with an appropriate population and research question, the texting procedure may make the data collection procedure more interesting and relevant for respondents, providing leverage to increase participation and decrease unit and item response (Groves, Singer, & Corning, 2000). However, if the method is poorly matched with the survey population and topic, data quality could be harmed.

5 Conclusion

A time use study was undertaken, adapting conventional time diary procedures to fit with the mode of data collection – SMS/text messaging. Data collected using this novel mode were compared to that from a reverse record check from campus recreational sports facilities to validate the behavior of interest – physical exercise and activity. These comparisons suggested that these data were of high quality overall, with 80 percent of cases generating valid data on the variable of interest and the remaining cases equally distributed amongst over- and underreporting, leaving the population estimate unbiased.

A cluster analysis using a set of six paradata indicators predicted nearly 80 percent of the cases with misreported exercise. Moreover, testing the predictive validity of these paradata indicators in a logistic regression model suggested that only two – the number of messages sent and the number of days the respondent sent text messages – are important for distinguishing between cases with valid and invalid data. This finding suggests that improvement to the measurement procedure (i.e., increasing the number of messages sent and ensuring that respondents report on activities during all the days of the field period) may even further improve data quality.

The high quality of these data did not come at a steep price. Costs were limited to incentive payments – forty dollars per completed case. Notably, some of the suggestions made here to further increase data quality (e.g., using an HTTP-to-SMS service that allows automated reminders; increasing incentives to improve the response rate) would increase costs. Moreover, shifting to a general population may increase costs as respondents may either need to be furnished with text-capable cellphones or reimbursed for their text messaging costs. Yet, even with these additional budget lines, this method could still be cost-effective compared to face-to-face or telephone diary interviews.

While not a tool for every population and research question, this method is clearly viable under the right conditions. For an appropriate target population (e.g., one with near saturation of text-capable cellphones, like a college-age sample, young professionals, or teens, among others), and with a suitable sampling frame that accommodates such a procedure, this method provides another tool in the survey researcher's data collection kit. It allows researchers to use cellphones for data collection without the trouble and expense of providing equipment or specially designed applications to respondents. Moreover, the high rate of cellphone adoption in developing countries (in lieu of landlines) makes this method a possibility for data collection in areas where time use studies would otherwise necessitate personal interviews.

References

Ainsworth, B. E., Jacobs, D. R. and A. S. Leon (1992), Validity and reliability of self-reported physical activity status – The Lipid Research Clinics Questionnaire, in: *Medicine and Science in Sports and Exercise*, Vol. 25, No. 1, 92-98.

Alfvén, G. (2010), SMS pain diary – A method for real-time data capture of recurrent pain in childhood, in: *Acta Paediatrica*, Vol. 99, No. 7, 1047-1053.

Anhøj, J. and C. Møldrup (2004), Feasability of collecting diary data from asthma patients through mobile phones and SMS (short message service) – Response rate analysis and focus group evaluation from a pilot study, in: Journal of Medical Internet Research, Vol. 6, No. 4, 1-42.

Brenner, P. S. and J. D. DeLamater (2013), Social disirability bias in self-reports of physical activity – Is an exercise identitiy the culprit?, *Social Indicators Research*, Advance online publication, DOI: 10.1007/s11205-013-0359-y.

Ball State University (2009), Survey finds smart phones transforming mobile lifestyles of college students, Press Release retrieved from http://www.bsu.edu/news/article/0,1370,--61565,00.html.

Blumberg, S. J. and J. V. Luke (2007), Coverage bias in traditional telephone surveys of low-income and young adults, in: *Public Opinion Quarterly*, Vol. 71, No. 5, 734-49.

Bolger, N., Davis, A. and E. Rafaeli (2003), Diary methods – Capturing life as it is lived, in: *Annual Review of Psychology*, Vol. 54, No. 1, 579-616.

Chase, D. R. and G.C. Godbey (1983), The accuracy of self-reported participation rates, in: *Leisure Studies*, Vol. 2, No. 2, 231-35.

Currivan, D., Roe, D. and J. Stockdale (2008), The impact of landline and cell phone usage patterns among young adults on RDD survey outcomes, Presented at the 63rd Annual Conference of the American Association for Public Opinion Research, New Orleans, May 16.

Fowler, F. J., Jr. and T. Mangione (1990), Standardized survey interviewing – Minimizing interviewer-related error, Sage Publications, Newbury Park.

Groves, R. M., Singer, E. and A. Corning (2000), Leverage-saliency theory of survey participation, in: *Public Opinion Quarterly*, Vol. 64, No. 3, 299-308.

Häkkilä, J. and C. Chatfield (2005), It's like if you opened someone else's letter – User perceived privacy and social practices with SMS communication, MobileHCI '05, Proceedings of the 7th International Conference on human computer interaction with mobile devices and services, 219-222.

Klesges, R C., Eck, L. H., Mellon, M. W., Fulliton, W. Somes, G. W. and C.L. Hanson (1990), The accuracy of self-reports of physical activity, in: *Medicine and Science in Sports and Exercise*, Vol. 22, No. 5, 690-697.

Niemi, I. (1993), Systematic error in behavioral measurement – Comparing results from interview and time budget studies, in: *Social Indicators Research*, Vol. 30, No. 5, 229-244.

Patchin, J. W. and S. Hinduja (2010), Changes in adolescent online social networking behaviors from 2006 to 2009, in: *Computers in Human Behavior*, Vol. 26, No. 6, 1818-1821.

Pew Research Center (2011), Americans and text messaging, Retrieved from http://pewinternet.org/Reports/2011/Cell-Phone-Texting-2011.aspx.

Raento, M., Oulasvirta, A. and N. Eagle (2009), Smartphones – An emerging tool for social scientists, in: Sociological Methods and Research, Vol. 37, No. 3, 426-454.

Robinson, J. P. (1999), The time-diary method – Structure and uses, in: W.E. Pentland, A.S. Harvey, M.P. Lawton and M.A. McColl, (eds.), *Time use research in the social sciences*, Klewer/Plenum, New York, 47-89.

Robinson, J. P. (1985), The validity and reliability of diaries versus alternative time use measures in: Juster, F. T. and F. P. Stafford (eds.), *Time, Goods, and Well-Being*, Ann Arbor, 33-62.

Schober, M. F., Conrad, F. G., Antoun, C., Bowers, A. W., Hupp, A. L., and H. Yan (2013), Conversational interaction and survey data quality in SMS text interviews. Presented at the 68th Annual Conference of the American Association for public opinion research, Boston, May 19.

Stinson, L. L. (1999), Measuring how people spend their time – A time use survey design, in: *Monthly Labor Review*, Vol. 122, No. 8, 12-19.

Walton, L. A., Buskirk, T. D. and T. Wells (2013), Smarter online panels for smartphone users – Exploring factors associated with mobile panel participation, Presented at the 68th Annual Conference of the American Association for Public Opinion Research, Boston, May 16.

Zuzanek, J. and B. J. A. Smale (1999), Life-cycle and across-the-week allocation of time to daily activities, in: Pentland, W. E., Harvey, A. S., Lawton, M. P. and M. A. McColl (eds.), *Time use research in the social sciences*, Klewer/Plenum, New York, 127-154.

electronic ***International Journal of Time Use Research***
2013, Vol. 10, No. 1, 55-75 dx.doi.org/10.13085/eIJTUR.10.1.55-75

Interviewer and respondent interactions and quality assessments in a time diary study

Vicki A. Freedman, Jessica Broome, Frederick Conrad and Jennifer C. Cornman

Vicki A. Freedman, Ph.D.
Institute for Social Research
University of Michigan
Ann Arbor, MI 48104, United States
e-mail: vfreedma@umich.edu

Frederick Conrad, Ph.D.
Institute for Social Research
University of Michigan
Ann Arbor, MI 48104, United States
e-mail: fconrad@umich.edu

Jessica Broome, Ph.D.
Jessica Broome Research
Detroit, MI, United States
e-mail: jessica@jessicabroomeresearch.com

Jennifer C. Cornman, Ph.D.
Jennifer Cornman Consulting
Granville, OH, United States
e-mail: jencornman@gmail.com

Abstract

Systematic investigations of the cognitive challenges in completing time diaries and measures of quality for such interviews have been lacking. To fill this gap, we analyze respondent and interviewer behaviors and interviewer-provided observations about diary quality for a computer-assisted telephone-administered time diary supplement to the U.S. Panel Study of Income Dynamics. We find that 93%-96% of sequences result in a codable answer and interviewers rarely assist respondents with comprehension. Questions about what the respondent did next and for how long appear more challenging than follow-up descriptors. Long sequences do not necessarily signal comprehension problems, but often involve interviewer utterances designed to promote conversational flow. A 6-item diary quality scale appropriately reflects respondents' difficulties and interviewers' assistance with comprehension, but is not correlated with conversational flow. Discussion focuses on practical recommendations for time diary studies and future research.

JEL-Codes: C80, C42

Keywords: Time use, survey methods, data quality

Acknowledgements: This research was funded by grants P30-AG024928 and P01-AG029409 from the US National Institute on Aging. The views expressed are those of the authors alone and do not reflect those of the University of Michigan or the funding agency. During the preparation of this research, Dr. Broome was a graduate student at the University of Michigan.

1 Introduction

Time use studies have become a fixture in the statistical data infrastructure of many countries, including the United States, Canada, Australia, and much of Europe. Responses from such collections, like all surveys, are subject to measurement error – a discrepancy between respondents' answers and the true value of the attribute in question (Tourangeau et al. 2000; Sudman et al. 1996). Answering survey questions about time use requires respondents to interpret the questions, retrieve information from memory for the appropriate reference period (whether yesterday, last week, or last month), format their response to fit given alternatives, potentially self-edit if they feel a particular answer is or is not socially desirable, and communicate their answer to the researcher.

When an interviewer is involved, as is generally the case for telephone-based and face-to-face time use collections, further complications may arise during the interaction (Houtkoop-Steenstra, 2000; Maynard, Houtkoop-Steenstra, Schaeffer, & Van der Zouwen, 2002; Suchman & Jordan, 1990). For example, in highly structured interviews, a common technique designed to minimize interviewer variation, conversational flexibility is limited so interviewers typically may not assist respondents in tasks such as interpreting questions or formatting answers (Suchman and Jordan, 1990).

Methodological studies carried out in the 1970s and 1980s helped establish the 24-hour diary, in which retrospective reports of the previous day are collected and systematically coded, as the optimal method for characterizing time use (Juster and Stafford, 1991). In particular, the method of recalling yesterday has been viewed as less prone than "stylized" reports about last week or month to common measurement errors. For instance, stylized reports are considered more cognitively demanding (requiring recall over a longer term period and potentially arithmetic) and may be subject to social desirability for some activities (e.g., religious participation, physical activity).

Although originally administered by paper and pencil, interviewer-administered diaries are increasingly common around the world, as are computer-assisted interviews (CAI). For example, the U.S. Bureau of Labor Statistics' American Time Use Study (ATUS) is conducted over the telephone by an interviewer (see Phipps and Vernon 2008). To avoid the potential pitfalls of highly standardized interviewing, the diary portion of the ATUS is conducted using "conversational" interviewing layered over a standardized instrument. This technique trains interviewers to guide respondents through memory lapses, to probe in a non-leading way for the level of detail required to code activities, and to redirect respondents who are providing unnecessary information (Bureau of Labor Statistics, 2012). Embedded in this approach is the assumption that relative to inflexible standardized interviews, giving interviewers discretion of what to ask and when to ask it can lead to improved data quality.

Indeed, there have been several studies suggesting that conversational interviewing can lead to better comprehension and hence higher quality responses than standardized interviewing, particularly when respondents' circumstances are ambiguous (Conrad and Schober, 2000; Schober and Conrad 1997), as is likely to be the case in a time diary context. In these studies, conversational interviewers were able to clarify survey concepts, i.e., provide definitions, whether respondents explicitly requested help or the interviewers judged that respondents needed it. They could provide definitions verbatim or could paraphrase them. This practice was not strictly standardized in the sense that different respondents could receive different wording because the clarification dialog was not scripted. Otherwise, the wording was typically very similar between respondents.

The way respondents comprehend their task, recall events, and report about time use when completing 24-hour diaries is not well understood. However, it seems likely interviewers can help each of these processes, if they are not constrained by the need to standardize wording across respondents. Moreover, research questions squarely focused on respondent and interviewer interaction and the role of conversational techniques during the 24-hour diary collection remain largely unexplored. Consequently, questions remain about the extent of cognitive difficulty experienced by respondents and the role interviewers play in shaping the 24-hour diary.

Measures of diary quality have also been lacking, typically focusing on the number of activities reported as a measure of quality (where diaries with fewer than five activities are equated with poor quality; Alwin, 2009). A recent study of time diary quality proposed a new scale based on interviewer perceptions of respondent comprehension, engagement, and uncertainty (Freedman et al. 2012), but how these measures might be related to respondent and interviewer behaviors remains unexplored.

In this paper we analyze recordings of a random sample of 24-hour time diary interviews conducted with a subsample of the U.S. national Panel Study of Income Dynamics (PSID). Our aim is twofold: (1) to describe respondent and interviewer behaviors and interactions during a 24-hour recall diary, paying particular attention to behaviors that may indicate difficulty with the interview and therefore likely to be related to response quality; and (2) to determine which of such behaviors are detected by interviewers in their observations used to assess diary quality.

2 Materials and methods

2.1 The diary interview

The Disability and Use of Time (DUST) supplement to the 2009 PSID collected time diary and supplemental information from couples in which at least one spouse was age 60 or older. Both spouses participated in two (same-day) computer-assisted telephone interviews. Response rates were 73%. For details see Freedman and Cornman (2012).

The DUST time diaries built directly upon the ATUS interview design, but replaced several of the non-standardized conversational techniques in ATUS with tailored, yet scripted, content that gave the interviews a conversational tone (Freedman et al. 2013). Respondents were asked to reconstruct the day prior to the interview, beginning at 4:00 AM. For each activity, the respondent was asked what he/she did [next] and how long the activity took, followed by a series of tailored follow-up questions, including where they were, who was with them, and how they felt (see Table 1). The interviewer entered the activity (or activities) on separate lines in open text fields, which were later coded to a detailed 3-digit coding scheme (Freedman and Cornman 2012). If more than one activity was named, the respondent was taken through a series of scripted questions to identify whether the activities were simultaneous or sequential, and if the former, the main activity.[1]

Interviewers then asked how long the (main) activity took. After keying in the type of response, duration in hours and minutes (e.g.1 hour and 15 minutes) or an exact end time for the activity (e.g. 4:00 PM), interviewers were directed to enter values accordingly.

Once a given diary entry (activity and time) was complete, the interviewer read to the respondent a semi-scripted confirmation of the activity, "So you (were) [main activity] from about [start time] to [end time], is that correct?" The respondent, in turn, could either confirm or have the interviewer correct the information.

After the correct main activity and times were entered, the interviewer then selected one of nine categories for the main activity, which determined appropriate follow-up questions (e.g. where the respondent did the activity, who did the activity with the respondent, who else was there, who they did the activity for, how they felt while doing the activity). Some follow-up questions were limited to specific types of activities. For instance, if the first activity was sleeping, respondents were asked several follow-up questions about the quality of that night's sleep. For some, but not all, of the follow-up questions interviewers were instructed (on the computer screen) that they could "Ask or Confirm" (see column 3 of Table 1). Follow-up questions about where the activity occurred and who participated allowed the interviewer to capture other responses in an open text field.

2.2 Sample and unit of analysis

In total, 394 couples participated. Each member of the sample was asked to complete two diaries (one weekday and one weekend day). 33 spouses were not able to participate because of a permanent health condition and a handful of respondents completed only one interview yielding in all 1,506 completed diaries obtained by 25 interviewers. The mean age of respondents was 69.

[1] Interviewers also were given the option of using two scripted probes (available to the interviewers on laminated cards for ease of use) to guide respondents in a non-leading way for the level of detail required to code activities: "Let's break that down" if not detailed enough (such as I worked or I cleaned up) and "To do what?" if too detailed (such as I got up, I went upstairs).

Table 1
Number of utterances by question, type and actor

Question	Variable name	Interviewer allowed to "ask or confirm"	Utterances		
			Number by Respondent	Number by Interviewer	Total
Most activities					
Yesterday at 4:00am, what were you doing? OR At [time] what did you do next? OR What is the next thing that you can remember doing?	ACTIVITY	N	1746	2241	3987
Until what time did you do that OR How long did that take or how long did you do that?	DURATION	N	1529	2256	3785
So you (were) [activity] from about [start time] to [end time], is that correct?	CONFIRM	N	1116	1622	2738
Where were you while you were doing that? Or Where did you (pick up / drop off) your [passenger]?	WHERE	Y[1]	565	896	1461
How did you get there?	HOW	Y[1]	99	179	278
Who did that with you? OR Who went with you? OR Who were you talking to? OR Who did you pick up/drop off?	WHO ACTIVE	Y[2]	758	1100	1858
Who else was [at home / outdoors at home/yard / at work / there] with you? OR Who else went with you?	WHO PASSIVE	Y[3]	582	885	1467
(If household or care activities:) Who did you do that for?	WHO FOR	N	273	420	693
How did you feel while you (were) [DESCRIPTION]? [(If you had more than one feeling, please tell me about the strongest one. Did you feel mostly unpleasant, mostly pleasant, or neither?	HOW FEEL	N	1038	1963	3001
If more than one activity named:					
Just to be clear, were you doing [both / all] of these activities at [time]?	SAME TIME	N	172	222	394
If doing simultaneous activity: If you had to choose, which of these would you say was the main activity? (*By main activity, we mean the one that you were focused on most*)	MAIN	N	148	211	359

Table 1 (Cont.)

Question	Variable name	Interviewer allowed to "ask or confirm"	Utterances		
			Number by Respondent	Number by Interviewer	Total
If first activity & sleep					
We'd like to know a little more about how you slept [DAY BEFORE YESTERDAY] night. About what time did you go to sleep for the night on [DAY BEFORE YESTERDAY]	TIME BED	N	165	223	388
Did it take you more than half an hour to fall asleep?	FALL ASLEEP	N	119	138	257
Did you wake up during the night, that is between the time you fell asleep and [time woke up]?	WAKE DURING	N	109	137	246
Did you have trouble falling back to sleep?	BACK SLEEP	N	83	120	203
How would you rate your sleep? Would you say it was excellent, very good, good, fair, or poor?	RATE SLEEP	N	111	179	290
Other select follow-up questions					
(If gap between activities:) What time did you start doing that?	START TIME	N	14	19	23
(If traveling:) Were you the driver or the passenger?	DRIVER	Y	74	116	190
(If talking to someone else:) Was this on the phone or in person?	PHONE	Y	22	35	57
Total			8723	12962	21685

[1] The interviewer could ask or confirm for all activities except travel to pick up/drop off.
[2] The interviewer could ask or confirm for all activities except travel to pick up/drop off and talking to someone else. [3] The interviewer could ask or confirm for all activities except work and socializing.
Source: Disability and Use of Time (DUST) 2009, own calculations.

Four interviews conducted by each interviewer, for a total of 100 interviews, were randomly selected. Of these, five were excluded because four were inaudible (all of those sampled for one interviewer) and one interview did not have diary quality data, leaving a total of 95 diaries for the analyses reported here.

For these 95 interviews, approximately one-third of each was recorded, on average the first 9 out of 26 activities. Transcripts of the interviews yielded 21,685 “utterances” (132-440 per diary, or 228 on average), defined as one speaker’s turn in the conversation about a given diary question, and 6015 “sequences” (42-78 per diary, or 63 on average), defined as the set of utterances produced by interviewer and respondent about a question. To illustrate, the sequence below has 5 utterances:

Interviewer: So then how long did it take you to have breakfast?

Respondent: Oh, maybe 20 minutes, half an hour.

Interviewer: Which would be closer, 20 minutes or--

Respondent: Half an hour.

Interviewer: Uhhuh.

For each given activity (e.g. ate breakfast) there are at least four sequences (e.g. the activity, duration, confirmation, and any tailored follow-up questions).

2.3 Interaction coding

A coding scheme was developed by the investigators to identify respondent and interviewer verbal behaviors likely, on theoretical grounds, to be related to quality. In doing so we drew upon Ongena and Dijkstra's (2007) model of interviewer-respondent interaction. The model is structured into several distinct stages of question answering, borrowed from Tourangeau et al. (2000): question formulation, interpretation, retrieval, judgment, response formatting, and finalizing the response. For each stage and each actor in the interview (respondent, interviewer), the model highlights behaviors that may be related to quality.

Table 2 shows the mutually exclusive utterance types and non-mutually exclusive behaviors, for both interviewers and respondents, by stage of interviewer-respondent interaction. Because there is some ambiguity as to whether particular interactions reflect interpretation, retrieval, or judgment, we combine them into a single category, which we refer to as "comprehension." An interviewer utterance reflecting potential problems with *question formulation,* for instance, involves departing from reading verbatim the wording on the screen. *Comprehension*-related behaviors by the interviewer include: offering an explanation, use of probes (What is the next thing you remember doing? Let's break that down), reminders about earlier information provided. *Comprehension*-related behaviors by the respondent include: providing an uncodable answer (including "other, specify" answers not on the coding frame), requests for clarification, offering an explanation, thinking aloud as a response (Umm… or Let me think…), mid-utterance pauses, fillers (e.g., um, uh), hedges (e.g., about 3 o'clock), relying on routines rather than memory of events, self correcting (no, I went to get the mail next), or reconstructing events out loud (it must have been 6 o'clock because I was watching the news). We treated interviewers' offering response categories as evidence of a problem with *response formatting*. Finally, we added an additional set of behaviors reflecting the interviewer's attempt to regulate the *conversational flow*, e.g. interviewers filling silence (while typing answers) with repetition, offering "backchannels" that include neutral phrases (mhm hmm, I see) or gratitude (thank you), and answers to such utterances by the respondent.

Coding was carried out by two trained staff members (a graduate student in survey methodology and the transcriber, an undergraduate student) using Sequence Viewer (http://www.sequenceviewer.nl/) software, which is designed specifically for investigating sequential activities, such as patterns of conversational turns. Initially, both coders were assigned the same

small set of diary interviews to code. Discrepancies were discussed and reconciled before coders continued with the remaining diaries. A detailed coding sheet was developed to guide consistent decision making.

Table 2
Utterance types and behaviors by actor and stage of respondent-interviewer interaction

	Coded Utterances/Behaviors by Actor	
	Interviewer	Respondent
Question formulation	Read not verbatim (u)	
Comprehension: (Interpretation; Retrieval and Judgment)	Explanation (u)	Uncodable answer (u)
	Remind R of earlier response (u)	Request for clarification (u)
	Probe (u)	Explanation (u)
		Thinking aloud (u)
		Pauses (b)
		Fillers (b)
		Hedges (b)
		Relying on routine (b)
		Self correction (b)
		Reconstruction (b)
Response Formatting / Finalizing Response	Offer response options (u)	
Conversation Flow	Fill while logging (u)	Response to repetition (u)
	Repeat response (u)	
	Back channel/gratitude (u)	

u=mutually exclusive utterances; b=non-mutually exclusive behaviors
Source: Own definitions.

2.4 Diary quality measures

A measure of perceived diary quality was constructed based on interviewer's subjective assessments of respondent comprehension, engagement, and uncertainty in completing the diary. Such information was obtained through a set of interviewer observations collected after the interview was completed. Interviewers were asked to assess "none," "some," and "a lot" for how much difficulty the respondent had understanding the questions and how much probing was needed for the respondent to complete the diary. Interviewers also assessed how hard the respondent tried to provide correct answers to the diary (tried to answer all, most, some, or few/no questions correctly); how confident they seemed about the answers to the diary (very, mostly, somewhat, little or not at all); how often the respondent seemed to guess at what he/she did next (all, most, some, few, activities, or never guessed); and how often he/she guessed at how long an activity took (all, most, some, few, activities, or never guessed). We reverse coded the indicators as needed so that higher numbers reflected better quality and (following Freed-

man et al., 2012) summed them to form an overall score (Cronbach's alpha=.80). The diary quality measure ranged from 9 to 24 with a mean of 20.

2.5 Analytic approach

We first tallied the number of utterances by question type and actor (respondent, interviewer). We then tabulated for interviewers and then respondents the percentage of (mutually exclusive) utterance types by question and for respondents the prevalence of various other (non-mutually exclusive) behaviors of interest mentioned above. We also characterized the sequence by calculating its complete length and whether it was a long sequence (with five or more utterances). Because these sequence-level measures include conversational flow in addition to utterances designed to elicit answers from respondents, we also calculated for each sequence the number of utterances it took for a codable answer to first be given and whether a codable answer was given anywhere in the sequence. We also identified the typical (most common) patterns of interviewer-respondent interactions by sequence length. We expect to see patterns by type of question that highlight the more challenging nature of recalling activities and times relative to recalling other details about an activity.

Finally, we examined the relationship between respondent-interviewer interactions and diary quality. To do so we first summarized the utterances and behavior data to the diary level, calculating the percentage of actor utterances in a given diary for each mutually exclusive utterance type and for each (non-mutually exclusive) behavior type. We also calculated the mean sequence length per diary, the mean utterance by which a codable answer was obtained, and the percentage of sequences in each diary with no codable answer. We then examined correlations between each of these measures and each diary quality component as well as the overall diary quality scale. We anticipated that behaviors indicative of problems with question comprehension would be reflected in interviewers' perceptions about diary quality. In contrast, we hypothesized that interviewers would not reflect in diary quality measures their own behaviors in formulating questions or response categories or behaviors related to conversational flow.

3 Results

3.1 Interviewer utterances

Across all types of questions, the majority of interviewer utterances involved question formulation (42% verbatim utterances where the interviewer read exactly what was on the screen and 7% departures from verbatim), and 85% of all question formulations were verbatim (42%/49%). Also common were utterances related to conversational flow (27% backchannels or expressions of gratitude, 10% repeating responses aloud, and 2% fills while logging answers). Far fewer utterances involved assistance with comprehension (1% offers of explanation; 2% probes) and answer formulation (6% offers of categories). Differences in interviewer

utterances by question type are highlighted in Figure 1. Four points are noteworthy. First, departures from reading the question verbatim (shown in red in Figure 1) were most apparent for the questions where interviews were allowed to either ask or confirm (where, how, who was actively engaged in the activity with the respondent, and who else was there). Interviewers also departed from verbatim when they asked about activities that occurred at the "same time" and at the confirmation screen, possibly indicating respondents did not always find the repetition necessary.

Second, interviewer behaviors that indicated assistance with comprehension were rare (<2%) across all question types, with only a few exceptions: probes constituted 6% of utterances about the length of an activity, 4% about the activity[2], and 4% about which was the main activity. These finding suggest these three questions may be somewhat more cognitively challenging—at least for some respondents—than the rest of the items in the interview.

Third, with respect to response formatting, interviewers offered response options most often for the item on how the respondent felt (27% of utterances). We attribute this finding to the break between question and closed response categories (How did you feel while you were <doing activity>? Did you feel mostly unpleasant, mostly pleasant, or neither?), which allowed respondents to interject the answer "fine" in between. Interviewers also offered response categories in nearly 10% of utterances about where they were and 7% of utterances about how they got there, both of which had relatively long lists of potential choices that were not intended to be read.

Fourth, although backchanneling and gratitude constituted a high proportion of utterances across all questions (ranging from 17%-36%), repetition of answers was most common for questions about activities and duration-related questions (including the time the respondent went to bed the night before). It may be that the complexity of these questions led interviewers to repeat information; the activity questions involved recording open text while the latter involved multiple screens to record time (first whether exact time or duration, and then hours and minutes).

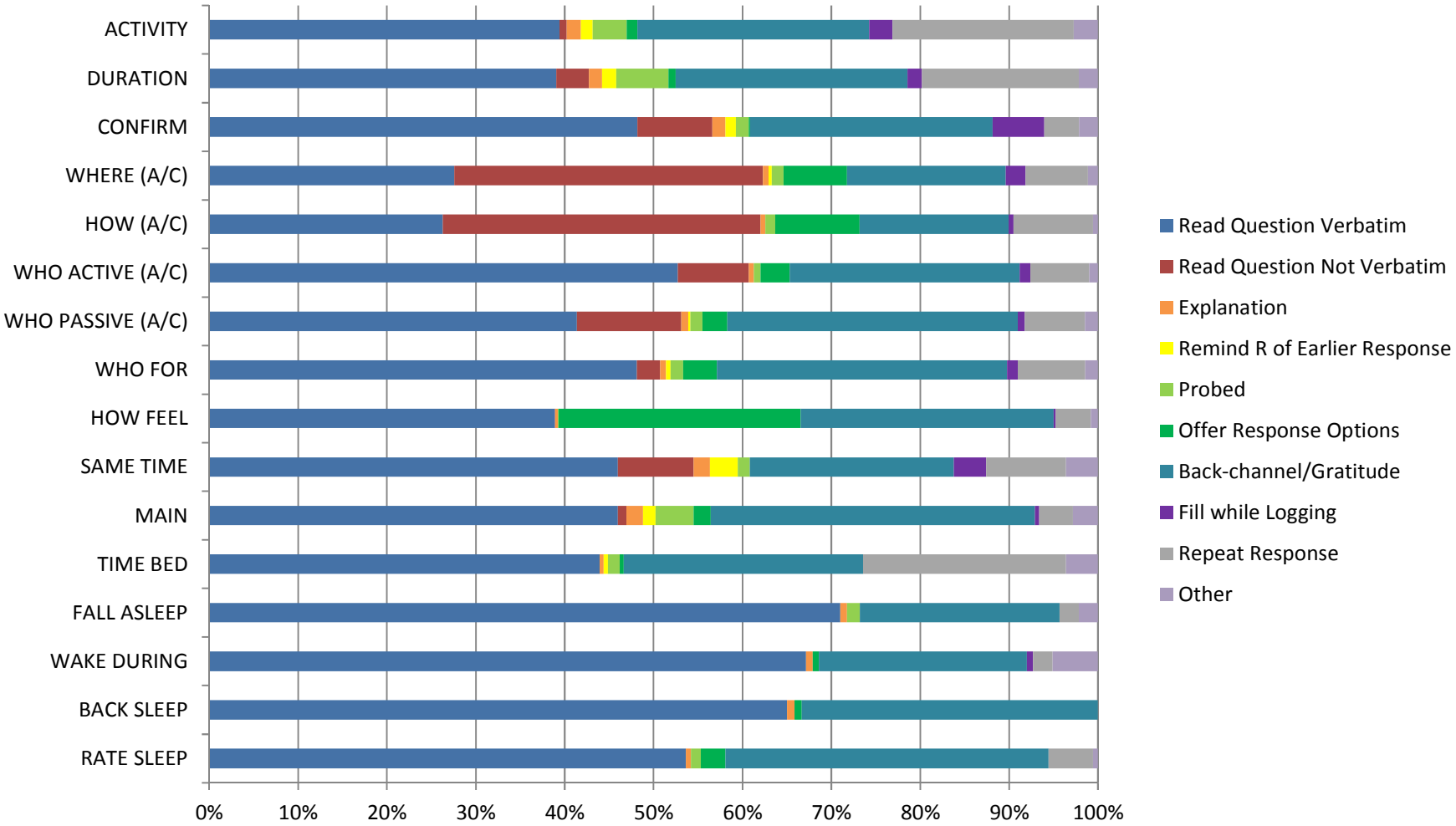
3.2 Respondent utterances and behaviors

Across all types of questions, the majority of respondent utterances (68%) involved codable answers. Far fewer utterances involved utterances related to interpretation difficulties or retrieval and judgment: 10% of utterances were (initially) uncodable answers, 3% involved requests for clarification, and less than 2% thinking. Another 6% of utterances involved conversational flow (response to an interviewer's repetition).

Differences in respondent utterances by question type are highlighted in Figure 2.

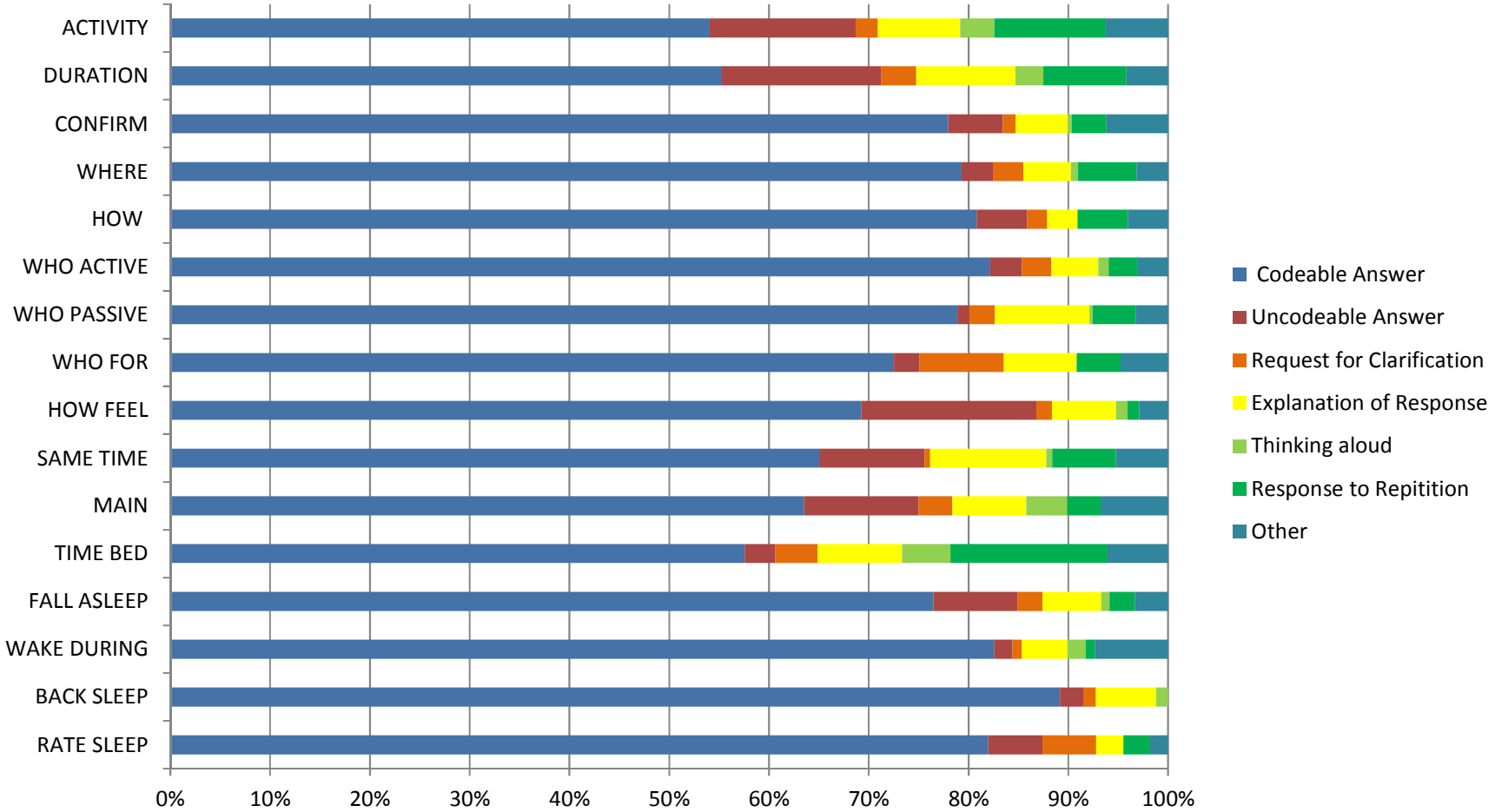
[2] When probing about activities, interviewers used the scripted probes 62% of the time and their own probes 38% of the time.

Figure 1
Interviewer utterances by question type



Source: Disability and Use of Time (DUST) 2009, own illustration.

Figure 2
Respondent utterances by question type



Source: Disability and Use of Time (DUST) 2009, own illustration.

As anticipated, comprehension-related utterances (uncodable answer, request for clarification, and explanation, shown in red, orange, and yellow) were most evident for activity (25% of utterances) and duration questions (30% of utterances). Respondents also appeared to have difficulty with questions about whether activities were done at the same time (23% of utterances) and selecting the main activity (22% of utterances), and as previously mentioned they often offer uncodable answers to the close-ended question asking how they felt (18% of utterances).

Table 3 shows additional respondent behaviors indicative of comprehension challenges by question type. Overall, fillers (14%) and hedges (15%) were most prevalent, followed by pauses (7%). In contrast, reliance on routine (2%), self-correction (2%), and reconstructing events out loud (1%) were rarely heard. Pauses and fillers were most common for questions about what was done next (activity), for the time they went to bed, selection of main activity, and duration of activity. Hedges were most common for duration of activity (46%) and time went to bed (42%). These finding suggest that rather than relying on routine, respondents in this corpus attempted to retrieve information from memory, although the high frequency of hedging about duration suggests times being reported may be better interpreted as approximate rather than exact.

Table 3
Additional respondent behaviors by question type (%)

Question Type	Pauses	Fillers	Hedges	Rely on Routine	Self Correct	Reconstructs Events	N
ACTIVITY	12.4	25.1	11.6	2.4	2.3	2.3	1746
DURATION	8.7	18.7	46.4	4.6	3.3	2.1	1529
CONFIRM	2.0	2.9	13.3	1.0	4.7	0.2	1116
WHERE	1.8	5.7	0.7	0.2	0.5	0.0	565
HOW	0.0	6.1	0.0	1.0	1.0	0.0	99
WHO ACTIVE	3.3	10.8	2.9	1.3	0.7	0.1	758
WHO PASSIVE	1.9	9.6	3.1	0.5	1.5	0.3	582
WHO FOR	4.0	12.1	3.7	1.5	0.0	0.0	273
HOW FEEL	6.1	11.4	8.8	1.6	0.7	0.0	1038
SAME TIME	5.2	12.2	7.6	2.9	2.3	0.0	172
MAIN	11.5	20.3	16.2	0.0	1.4	0.7	148
TIME BED	12.7	24.2	42.4	4.2	3.0	3.0	165
FALL ASLEEP	5.0	12.6	14.3	6.7	0.0	0.0	119
WAKE DURING	4.6	11.9	9.2	3.7	0.9	0.0	109
BACK SLEEP	9.6	10.8	1.2	2.4	0.0	0.0	83
RATE SLEEP	6.3	11.7	18.0	0.9	0.9	0.0	111
ALL	6.5	14.1	15.6	2.1	2.1	1.0	8723

Source: Disability and Use of Time (DUST) 2009, own calculations.

3.3 Patterns within sequences

Across all 6,015 sequences, the average sequence length was 3.4 utterances, 20% of sequences consisted of 5 or more utterances, and 93% of sequences had at least one utterance that was a codable answer, obtained on average after 2.4 utterances.

As shown in Table 4, sequences were longer on average for questions about the activity (4.6), its duration (4.3), time went to bed (4.3), how the respondent felt (4.1), the main activity (3.9) and whether activities that were reported occurred at the same time (3.8). The percentage of sequences with five or more utterances was highest for questions about activity and duration (36% and 32%, respectively).

The average number of utterances to obtain a codable answer ranged from 2.1 to 2.8, with longer than average sequences for questions about the activity, its duration, how the respondent felt, and the main activity. The percentage of sequences with no codable answer was highest for where and how, both of which allowed interviewers to capture "other, specify" (considered for this exercise as not codable).

Table 4
Mean number of utterances per sequence by question type

Question Type	Number of Sequences	Mean number of utterances per sequence	With 5+ utterances in %	Mean utterances until codable answer	No codable answer in %
ACTIVITY	868	4.6	36.3	2.4	3.9
DURATION	876	4.3	32.3	2.7	9.8
CONFIRM	862	3.2	13.1	2.1	4.2
WHERE	542	2.7	10.5	2.3	19.4
HOW	111	2.5	10.8	2.3	28.8
WHO ACTIVE	634	2.9	10.7	2.2	4.4
WHO PASSIVE	440	3.3	15.2	2.2	2.3
WHO FOR	197	3.5	18.8	2.3	5.1
HOW FEEL	740	4.1	25.3	2.8	4.6
SAME TIME	105	3.8	21.0	2.4	4.8
MAIN	92	3.9	22.8	2.7	3.3
TIME BED	91	4.3	26.4	2.3	2.2
FALL ASLEEP	91	2.8	12.1	2.2	2.2
WAKE DURING	91	2.7	4.4	2.1	2.2
BACK SLEEP	74	2.7	4.1	2.1	0.0
RATE SLEEP	89	3.3	13.5	2.2	1.1
ALL	6015	3.4	19.6	2.4	6.7

Source: Disability and Use of Time (DUST) 2009, own calculations.

Common sequence structures by number of utterances are illustrated in Table 5. Regardless of the question type, exchanges between interviewer and respondent in sequences made up of four

or fewer utterances largely followed the same structure. For three utterance sequences, for example, an interviewer's question was typically followed by a codable answer from the respondent, which was then followed by an interviewer backchannel or expression of gratitude. In sequences with four utterances, the typical pattern involved asking the question, providing a codable answer, followed by conversational exchanges such as repeating the respondent's answer, backchannel or expression of gratitude, or a respondent's reply to the interviewer's repetition.

Table 5
Common interviewer-respondent interactions by sequence length

Utterance Number	2	3	4	5 or more – 2nd utterance codable	5 or more – 2nd utterance other than codable
1	Question Asked	Question Asked	Question Asked	Question Asked	Question Asked
2	Codable Answer	Codable Answer	Codable Answer	Codable Answer	Uncodable Answer
3		Conversation	Conversation	Conversation	Attempt to Elicit Answer
4			Conversation	Conversation	Codable Answer
5+				Conversation	Conversation
% of sequences	31%	27%	10%	12%	5%

Disability and Use of Time (DUST) 2009, own calculations.

Among longer sequences (containing five or more utterances; approximately 20% of sequences), two dominant patterns emerged. In one pattern, the interviewer asked a question, the respondent gave a codable answer, and remaining utterances involved interviewer's repetition or gratitude and the respondent's reply to these conversational elements. In the second pattern, the interviewer asked the question, the respondent's utterance reflected difficulty with interpretation or retrieval/judgment (e.g., uncodable answer, request for clarification, explanation, or thinking aloud) and the interviewer attempted to elicit a correct response (e.g. by probing, explaining, repeating the question). After a codable answer was obtained, more conversation typically ensued with the interviewer repeating or expressing gratitude and the respondent sometimes replying to these utterances.

3.4 Relationship to perceived diary quality

Select behaviors reflecting comprehension difficulties were correlated with perceived overall diary quality scores (Table 6). In particular, diaries with a greater percentage of uncodable answers, explanations, and hedging by respondents had lower overall quality scores.

Table 6
Bivariate correlations between diary quality measures and respondent and interviewer behaviors (n=95)

	Stage	How much difficulty under-standing? (1-3)	Amount of probing needed (1-3)	How hard R tried (1-4)	How confident R was with answers (1-4)	How often R guessed at next activity (1-5)	How often R guessed at duration (1-5)	Summary Score (9-24)
Respondent behaviors								
Uncodable answer	Comprehension	-0.24*	-0.24*	-0.25*	-0.30**	-0.32**	-0.37**	-0.41**
Request for clarification	Comprehension	-0.30**	-0.13	-0.13	0.05	-0.07	-0.09	-0.14
Rely on routine	Comprehension	-0.11	-0.19	-0.04	-0.10	-0.06	-0.19	-0.16
Explanation of response	Comprehension	-0.25*	-0.20*	-0.13	-0.21*	-0.17	-0.25*	-0.28**
Thinking aloud	Comprehension	-0.17	0.12	0.12	-0.07	-0.19	-0.17	-0.10
Pauses	Comprehension	-0.09	0.00	-0.06	-0.14	-0.16	-0.14	-0.15
Fillers	Comprehension	0.14	-0.02	0.07	0.05	0.05	0.07	0.08
Hedges	Comprehension	-0.06	-0.04	-0.07	-0.33**	-0.40**	-0.37**	-0.33**
Self correct	Comprehension	-0.20	-0.09	0.08	-0.12	-0.20*	-0.25*	-0.19
Reconstruct events	Comprehension	-0.05	-0.06	0.03	-0.18	-0.15	-0.25*	-0.17
Response to Repetition	Conv. Flow	0.07	-0.12	0.08	0.02	-0.01	0.08	0.03
Interviewer behaviors								
Read not Verbatim	Question Form.	0.09	-0.09	-0.10	-0.09	-0.07	-0.09	-0.09
Explanation	Comprehension	-0.16	-0.20*	0.10	0.11	0.03	-0.11	-0.04
Remind R of Earlier Response	Comprehension	-0.22*	-0.16	-0.11	-0.10	-0.14	-0.21*	-0.21*
Probed	Comprehension	-0.29**	-0.22*	-0.19	-0.16	-0.08	-0.14	-0.23*

Table 6 (Cont.)

	Stage	How much difficulty under-standing? (1-3)	Amount of probing needed (1-3)	How hard R tried (1-4)	How confident R was with answers (1-4)	How often R guessed at next activity (1-5)	How often R guessed at duration (1-5)	Summary Score (9-24)
Interviewer behaviors								
Offer response options	Response	0.06	0.10	-0.01	0.07	0.02	-0.01	0.05
Fill while logging	Conv. Flow	-0.06	0.12	0.07	0.19	0.17	0.04	0.14
Repeat Response	Conv. Flow	0.11	-0.07	0.01	0.02	-0.04	0.04	0.01
Back-channel/Gratitude	Conv. Flow	-0.25**	-0.05	-0.03	-0.07	-0.13	-0.24**	-0.18
Interactions								
Mean number of utter. per se-quence		-0.38**	-0.21*	-0.04	-0.05	-0.17	-0.31**	-0.26**
% of sequences with >5 utterances		-0.39**	-0.23*	-0.12	-0.15	-0.25*	-0.34**	-0.34**
Mean utterances until codable answer		-0.13	-0.21*	-0.15	-0.18	-0.27**	-0.27**	-0.29**
Mean seq. with no codable answer		-0.05	-0.17	-0.13	-0.17	-0.25*	-0.23*	-0.25*
Mean score:		2.8	2.2	3.7	3.2	4.2	4.0	20.1

Note: *p<0.05, **p<0.01
Source: Disability and Use of Time (DUST) 2009, own calculations.

Diaries with higher percentages of reminders to respondents of earlier responses and probing by interviewers also had lower overall quality scores.

Most behaviors reflecting conversation flow alone were not picked up in perceived diary quality evaluations, with one exception. Higher rates of backchanneling and expressions of gratitude by interviewers were associated with lower ratings of respondent understanding and more guessing at activity durations. However, these associations were not strong enough to be reflected in final overall score.

All four indicators of longer sequences were associated with the overall diary quality scores. However, the indicator of sequences with 5+ utterances had the strongest correlation with overall score, and was significantly correlated with four of the six components: having difficulty, probing, guessing at activity and guessing at duration.

4 Discussion

This analysis is the first we know of to systematically describe interviewer-respondent interactions in the context of a time diary and relate them to a new measure of perceived time diary quality. Several findings emerged.

First, evaluation of utterance types and sequences suggests that most time diary questions are answerable by respondents. 93% of all sequences successfully elicited a codable answer and the figure is closer to 96% if "other, specify" responses are considered codable. Only 3% of interviewer utterances and about 15% of respondent utterances signaled potential issues with comprehension (i.e. interpretation, retrieval, or judgment).

Second, consistent with our expectations, questions about what the respondent did next and how long the activity took appeared to be most cognitively challenging for respondents. Respondents signaled uncertainty (Clark and Fox Tree 2002, Schober and Bloom 2004) in responses about what they did next with fillers (um, uh) and about how long it took with hedges (about...), but they did not frequently rely on routine or self-correction, nor did they reconstruct activities aloud. These findings suggest respondents generally try to recall details from the last 24 hours.

Third, time diary questions elicit conversation, even when questions are largely scripted, the purpose of which appears to be to promote the flow of the interview. In our analysis of diary interactions, 40% of interviewer utterances involved backchannels, expressions of gratitude, repeating responses aloud, and filling silence while logging answers and 6% of respondent utterances involved responses to interviewers' repetition. Consequently, unlike more highly scripted interviews, longer than average sequences did not necessarily indicate respondent difficulty with diary questions.

Finally, we provided evidence that a set of six interviewer-provided observations about diary quality appear to appropriately reflect respondents' difficulties with and interviewers' assistance with comprehension. Furthermore, these judgments are not correlated with utterances that simply reflect conversation flow, a finding that further buttresses the validity of the proposed scale.

This study has several important limitations. The DUST diary application is unique in that it purposefully attempted to script, in a flexible way, portions of the questionnaire that in other studies have been left to interviewers to sort out. For instance, unlike ATUS, the DUST diary application has screens that help determine whether activities are sequential or simultaneous. The DUST diary is also purposefully conversational in tone, offering interviewers flexible phrases like "So you (were) [activity] from about [start time] to [end time], is that correct?" It may be that these phrases encourage more conversation than other applications. Notwithstanding these unique features, in other ways, DUST mimics ATUS and other diary applications much more closely; for example, questions about activity, duration, and where/how are standard features of most time diary studies.

An additional limitation is that only a portion of the diary interview was recorded and transcribed. In all cases the first third or so of the interview was recorded – approximately 9 activities out of 26 on average. It may be that respondents learn as they cycle through the interview and that subsequent parts of the interview are less challenging than earlier parts. Future research on this topic would benefit from recording the entire interview and examining utterances by activity number.

Moreover, the DUST sample is limited to older adults, whose mean age was nearly 70, and thus generalizability to all adults is limited. It is not obvious how this limitation influences findings. Given that older adults are likely to have more memory problems than younger adults, this sample may over-represent difficulties with daily diaries. At the same, time, older adults may have fewer time commitments than younger individuals and therefore may be more prone to engage in conversation than their younger counterparts. Future research on time diaries would benefit from widening the age range for evaluations of respondent-interviewer interactions.

Despite these limitations, our analysis suggests several key lessons relevant for future applications and research. One practical finding is that the new measures of diary quality included in DUST appear to capture behaviors and interactions that reflect real problems with diary administration. Since these items are easy to obtain, it may be worthwhile to replicate on other time diary studies in the US and around the world. If such relationships are replicated in other countries, comparisons of quality could be made for the first time using a metric other than number of activities.

Our study also raises potentially important questions relevant to theoretical research on interviewer-respondent interactions. The model advanced by Ongena and Dijkstra's (2007) highlights 5 distinct stages of interaction (question formulation, interpretation, retrieval and judgment, response formatting, and finalizing the response), but we found that, in the case of time

diaries, a sixth category indicating behaviors related to conversational flow may be useful. Such behaviors include repeating information out loud, filling while logging, and backchanneling or offering gratitude.

Why the time diary elicited from interviewers relatively high levels of utterances designed to foster conversation flow (40% of interviewer utterances) is not clear. It may be that the complexity of particular questions led interviewers to repeat information aloud; such a hypothesis would be useful to investigate in future studies. On a more practical level, whether these utterances should be discouraged or encouraged is also not yet clear. We found that such behaviors are not significantly associated with the diary quality measures proposed here. However, we cannot rule out that such behaviors may contribute positively to interview quality in other ways (e.g. by building rapport, filling what would otherwise be awkward silence, or providing the respondent with an opportunity to correct information). Whether such behaviors simply lengthen the interview or provide additional benefit is an important next question.

References

Alwin, D.F. (2009), Assessing the quality of timeline and event history data in calendar and time diary methods, in: Belli, R. F., Stafford, F. P. and D. F. Alwin (eds.), *Life course research*, Thousand Oaks, CA: Sage Publications, 277-301.

Bureau of Labor Statistics. (2012), American time use survey's user guide, Accessed October 27, 2012 http://www.bls.gov/tus/atususersguide.pdf.

Conrad, F. G. and M. F. Schober (2000), Clarifying question meaning in a household telephone survey, in: *Public Opinion Quarterly*, Vol. 64, No. 1, 1-28.

Dijkstra, W. (2002), Transcribing, coding, and analyzing verbal interactions in survey interviews, in: Maynard, D. W., Houtkoop-Steenstra, H., Schaeffer, N. C. and J. van der Zouwen (eds.), S*tandardization and tacit knowledge – Interaction and practice in the survey interview*, New York, Wiley, 401–425.

Freedman, V., Stafford, F. P., Conrad, F. G., Schwarz, N. and J. Cornman (2012), Time together – An assessment of diary quality for older couples, in: *Annals of Economics and Statistics*, Vol. 105-106, No. 1, 271-289.

Freedman, V., Stafford F. P., Schwarz, N and F. G. Conrad (2013), Measuring time use of older couples – Lessons from the panel study of income dynamic, in: *Field Methods*, Vol. 25, No. 2, 1-18, doi:10.1177/1525 822X12467142.

Freedman, V. and J. C. Cornman (2012), Panel study of income dynamics disability and use of time supplement – User's guide, The University of Michigan, Institute for Social Research.

Houtkoup-Steenstra, H. (2000), Interaction and the standardized interview – The living questionnaire, Cambridge, Cambridge University Press.

Juster, T. and F. P. Stafford (1991), The allocation of time – empirical findings, behavioral models, and problems of measurement, in: *Journal of Economic Literature*, Vol. 29, No. 1, 471-522.

Maynard, D. W., Houtkoup-Steenstra, H., Schaeffer, N. C. and J. van der Zouwen (2002), Standardization and tacit knowledge, New York, John Wiley and Sons.

Ongena, Y. P. and W. Dijkstra (2006), Methods of behavior coding of survey interviews, in: *Journal of Official Statistics*, Vol. 22, No. 1, 419–451.

Ongena, Y. P. and W. Dijkstra (2007), A model of cognitive processes and conversational principles in survey interview interaction, in: *Applied Cognitive Psychology*, Vol. 21, No. 1, 145–163.

Phipps, P. and M. Vernon (2008), 24 Hours – An overview of the recall diary method and data quality in the american time use survey, in: Belli, R. F., Stafford, F. P. and D. F. Alwin (eds.), *Calendar and time diary methods in life course research*, Thousands Oaks, CA: Sage, 108-128.

Schober, M. F. and F. G. Conrad (1997), Does conversational interviewing reduce survey measurement error?, in: *Public Opinion Quarterly*, Vol. 61, No. 4, 576–602.

Suchman, L. and B. Jordan (1990), Interactional troubles in face-to-face survey interviews, *Journal of the American Statistical Association*, Vol. 85, No. 1, 232-241.

Sudman, S., Bradburn, N. and N. Schwarz (1996), Thinking about answers – The application of cognitive processes to survey methodology, San Francisco, Jossey-Bass Publishers.

Touranageau, R., Rips, L. and K. Rasinski (2000), The psychology of survey response, Cambridge, Cambridge University Press.

electronic ***International Journal of Time Use Research***
2013, Vol. 10, No. 1, 76-90 dx.doi.org/10.13085/eIJTUR.10.1.76-90

Visualizing multinational daily life via multidimensional scaling (MDS)

John P. Robinson and Jonathan Gershuny

John P. Robinson,
University of Maryland
College Park
Maryland 20742, United States
e-mail: jpr@umd.edu

Jonathan Gershuny
Department of Sociology and
University of Oxford
Wellington Square
Oxford OX1 2JD, United Kingdom
e-mail: jonathan.gershuny@st-hughs.ox.ac.uk

Abstract

One of the notable innovations in social-science methodology developed during the 1960s was Multi-Dimensional Scaling (MDS). MDS made it possible for social scientists to discover, uncover or model the underlying spatial structure of relations between various social collectives (like countries or communities), social objects (like music or artifacts) or social attitudes. One early application of MDS described the dimensional contours of Americans' views of other countries in terms of "perceptual maps of the world". More recently, it has been used to map country differences in the World Values Survey. Spurred by its initial successful applications, MDS was extended to time-diary data collected in the pioneering 1965 Multinational Time-Budget Study, in which it again provided insightful portrayals of daily activity across the 15 national settings in that study. This present article updates and extends these results by applying MDS methods to the most recent diary collection in the Oxford University MTUS data archive – covering more than 20 (mainly European) countries. Once again, the result was plausible (but somewhat different) configurations again emerged from MDS visualizations. Moreover, these mappings were compatible with conclusions from the 1965 mapping and with earlier more conventional analyses.

JEL-Codes: B16, C15, C21, C39, N30, O50

Keywords: Time diaries, Multidimensional Scaling, Multinational comparisions, Social change

Revision of presentation to the 32nd convention of IATUR in Paris, France July 2010.

1 Introduction

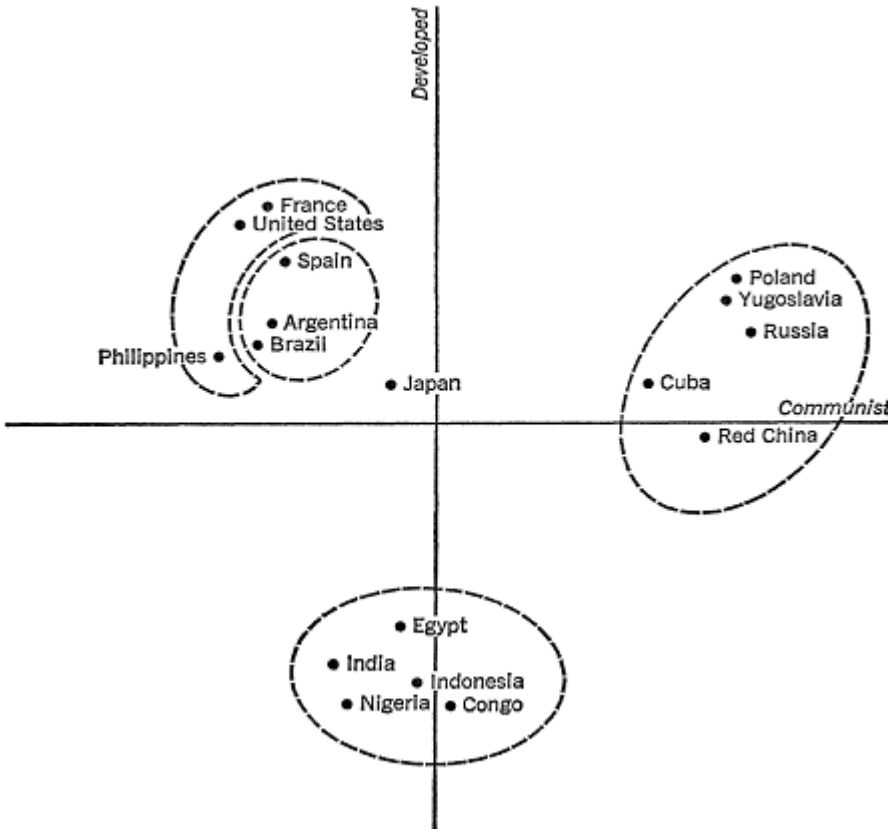
The 1960s marked a decade of great societal experimentation in politics, culture and science. One of the more notable methodological innovations in the social sciences during this decade was a technique called "Smallest Space Analysis" or SSA (Guttman 1968; Kruskal 1964). It later went under the name of Multi-Dimensional Scaling (MDS), and it has become one of the standard analytic tools available in SPSS. Based on calculations and procedures in mathematical topology (or "rubber-sheet geometry", in which the simple order of distances in a space was employed as the central metric, rather than the magnitude of original distances themselves – as in city subway maps), MDS made it possible for social analysts to discover (or uncover) the underlying spatial structure of relations between various groups of people, social collectives (like countries or communities), social objects (like music or artifacts), and social attitudes and values.

Bloombaum (1970) described SSA thusly: Smallest space analysis (SSA) is one among the new methods of nonmetric analysis ....methods recommended for those jobs where the investigator desires a rigorous multivariate analysis under the constraints of no special assumptions. A pleasing related feature of the techniques discussed here is that the results achieved are directly and intuitively interpretable by relatively untutored persons, as well as by the scientist who takes responsibility for his project in its entirety.

One initial application of MDS described the dimensional contours of American perceptions of the countries of the world, or "perceptual maps of the world" (Robinson and Hefner 1968). In this case, a random sample of Detroit respondents and a sample of academic "experts" were given the names of one country (like Argentina or Poland) and asked to which of 16 other countries it was most similar, the term "similar" purposely left undefined in order to allow smallest-space analysis to discover its underlying perceptual structure. Based on these perceptual responses, MDS generated the map in Figures 1 (for the public) and 2 (for the experts), which made it possible to visualize these similarity ratings as reducible to three dimensions, which are highlighted with the dotted circular lines.

In Figure 1, the political (horizontal) perceptual dimension separated mainly Eastern "Iron Curtain" communist countries (like Russia and Poland, but also Cuba and China) on the right from mainly Western capitalist countries, like the US and France, on the left. The second vertical dimension then separated more economically prosperous countries (again like the US and France) at the top from "third world" countries, like India and Nigeria at the bottom. The third cultural dimension (shown by the dotted lines in Figure 1) then separated those countries that had Spanish (or Portugese) roots or lineage, from those that had other cultural connections.

Figure 1
Country positions (for the first two dimensions)
determined by smallest space analysis



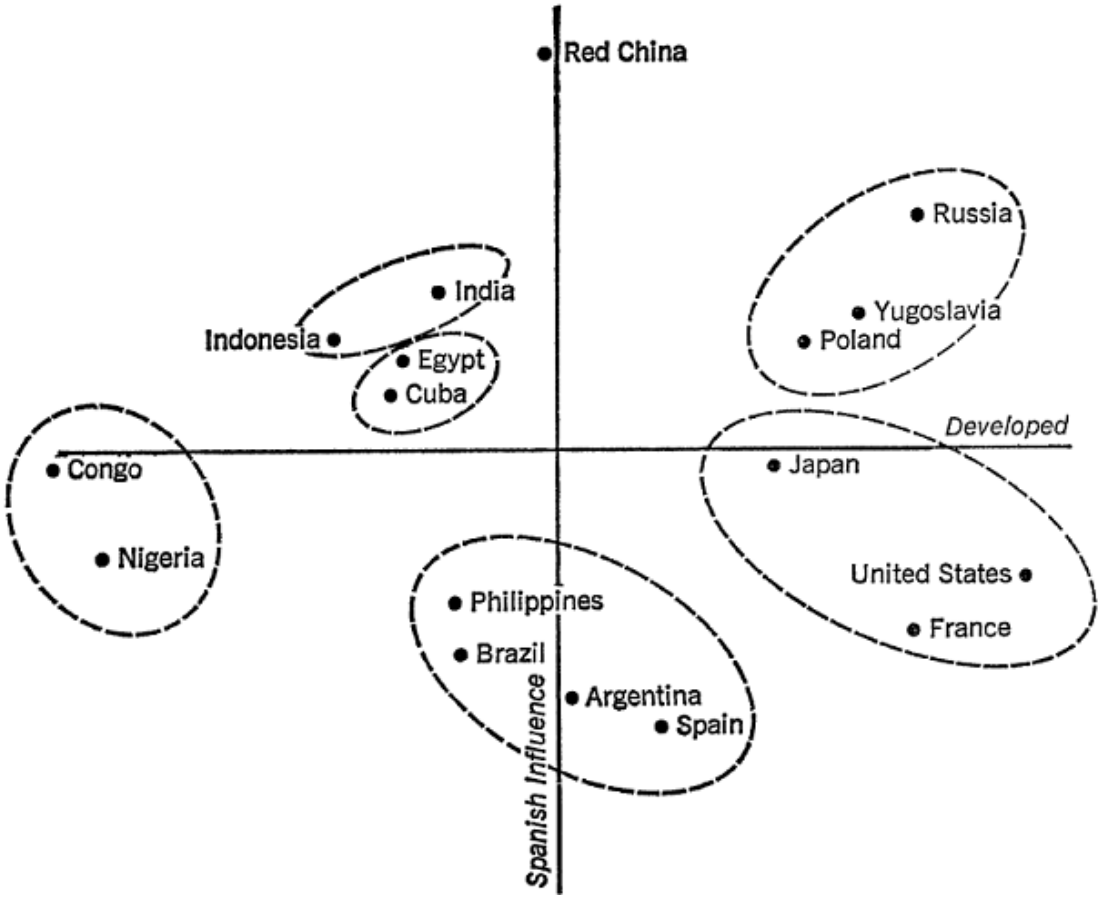
Note: Dashed lines indicating groupings suggested by three dimensional solution.
Source: 1963 Detroit public sample, as reported in Robinson and Hefner(1968),
Own illustration.

Three parallel dimensions were also found in similarity ratings made by a separate sample of academic experts in the Detroit area, but as shown in Figure 2, they differed in the salience or ordering of these three dimensions. The academic sample perceived the economic dimension as of paramount importance, as shown by the horizontal distinction between US, France and Russia on the right and Congo and Nigeria on the left. Their second vertical dimension then emphasized the "Spanish influence" countries (including the Philippines) from the rest, especially China. Their third dimension then separated the politically different communist from capitalist countries, although they saw China as much more distant from this bloc than the public in Figure 1. Indeed, one can see that the countries in Figure 2 are generally more scattered or less clustered than in the public's Figure 1, indicating more indicating more differentiated or nuanced judgments than the public in Figure 2. Here, then, in the two samples, the academics stressed economic factors vs. the public's more political factors.

These mappings, moreover, predicted differences in attitudes toward several foreign policy issues, like the Vietnam War, foreign aid and general isolationism. Members of the Detroit public

who saw more difference economically than politically in their mappings tended to share the academics' greater opposition to that war and support of aid to less developed countries.

Figure 2
Country positions from smallest-space analysis



Note: Dashed lines indicate grouping suggested by three-dimensional solution.
Source: 1964 Academic sample, as reported in Robinson and Hefner (1968),
Own illustration.

Objective Measures: These discoveries then led to the question of how well these MDS perceptual mappings reflected "real world" differences between countries. Here MDS was used to uncover similar dimensions based on "harder" or more accepted measures of national differences, such as a country's GNP, literacy level or type of political representation. Here, two separate dimensions emerged from the available indicators at the time, one economic (mainly based on UNESCO data sources) and one political (based on a set of ratings of political structure types in countries) developed by a Yale University panel of political scientists (Banks and Texter 1963).

The technique has more recently been applied to summarize subjective data collected from the World Values Survey. Based on the public's acceptance of various value statements in different countries, Inglehart and his colleagues (2011) have generated a map that reduced the complex responses of people in these countries to a large battery of value statements to a simple two-

dimensional space. That map can be viewed directly at www.worldvaluessurvey.org, again with the clusterings being of main interest.

Among the wide variety of other social objects and concepts in several academic disciplines analyzed by MDS or SAA are occupations (Laumann and Guttman (1966), occupational interests (Meir 2010), work values (Elizur 1984), workplace values (Singh et al, 2011), leadership styles (Shapira 1976), ,personality beliefs (Kumar, Ryan and Wagner (2012), career adaptability (Johnston et al. 2006), gender differences (Elizur 1994), sex-role attitudes (Ruch 1984), forgiveness likelihood, (Kumar et al. 2009), child intelligence (Fiorello 2006), anthropology of migration (Lalouel and Langaney 1980) and national socio-political characteristics (Bloombaum 1970). Again most of these analyses focus on the clusterings rather than the dimensions that may define them.

2 Data and Methods

The Multinational Time Use Study (MTUS: as described in Fisher and Robinson 2011) is a retrospectively (post-fieldwork) harmonized archive of nationally representative time- diary studies. It currently includes some 60 surveys from 25 countries, the earliest currently dating from 1961 (www.timeuse.org). The statistical approach adopted in the remainder of this article uses a purely inductive method for the investigation of the cross-national record of time use. The authors of this paper intend simply to update the conclusions of Converse (1972) described below. What emerges nevertheless also corresponds to a remarkable degree to the "life-balance triangle" framework discussed in Gershuny (2009).

We employ the same multidimensional scaling technique of Smallest Space Analysis as did Converse. The technique involves, first, constructing **difference half-matrices** by calculating the mean squared differences for each pair of data points, For a pair of data points i and j (representing two countries) and a set of k activities the (generalised Euclidean) distance measure is the square root of the sum of the squared differences in the time devoted to each activity in the pair of countries:

$$D^{ij} = \sqrt{\left(\left(a_1^i - a_1^j\right)^2 ... \left(a_k^i - a_k^j\right)^2\right)}. \quad (1)$$

These 20-country data points yield a total of 380 (20x19) pairs to be arranged in the form of a half-matrix of distances between each pair of points. The straightforward intuitive explanation of SSA technique, is to imagine just such a half matrix but representing distances between cities as in a road atlas, and the SSA program as generating a 2-dimensional mapping of the relative positions of these cities in geographical space. A half matrix of distances among any real set of cities will (disregarding the curvature of the earth) indeed be capable of reconstruction into a map in the two geographical dimensions using a standard SSA programme. Any randomly generated half matrix of distances among n points will be certainly be interpretable as representing

a space in n-1 dimensions, and with increasing degrees of stress in n-2 dimensions, n-3 dimensions, and so on.

It is important to recognize certain limitations in this SSA application, which is intended mainly to illustrate its power to reduce complex time-diary data to provide simple two-dimensional mappings at a single points in time for two data sets (here separated by 40 years in time) examining different countries, and using (somewhat) different diary methods and coding. It is not possible then to reach any conclusions about increasing temporal convergences or divergences across countries or daily activity. We simply present two maps, one for 1965 and one for 1998-2005, that employed different methods and examined different countries, but with the simple conclusion that in both studies, the conclusion about the geo-cultural dominance in country time-use similarity. We are unable to tell whether this convergence is greater or lesser across time.

2.1 SSA/MDS maps of 1965 multinational time-use data

When the multinational time-diary data from Szalai's (1972) pioneering 1965 time-diary study became available soon after the SSA or MDS technique was developed, interest was naturally aroused about how well the method might capture the similarity in daily-life patterns across various countries. MDS techniques here were simply and directly applied to the daily hours/minutes people in each country spent their time – how much time they worked, slept or used the mass media.

Converse (1972) published these MDS results that generated the dimensional visualizations in Figure 3 that provided immediate and plausible insights into how similar life was in the different national settings involved in the study. (It was most helpful in this analysis that Szalai had established a common set of sampling, field and coding procedures that were strictly followed to ensure data comparability across countries.)

Converse succinctly described the resulting MDS diagram in his Figure 3 as follows:

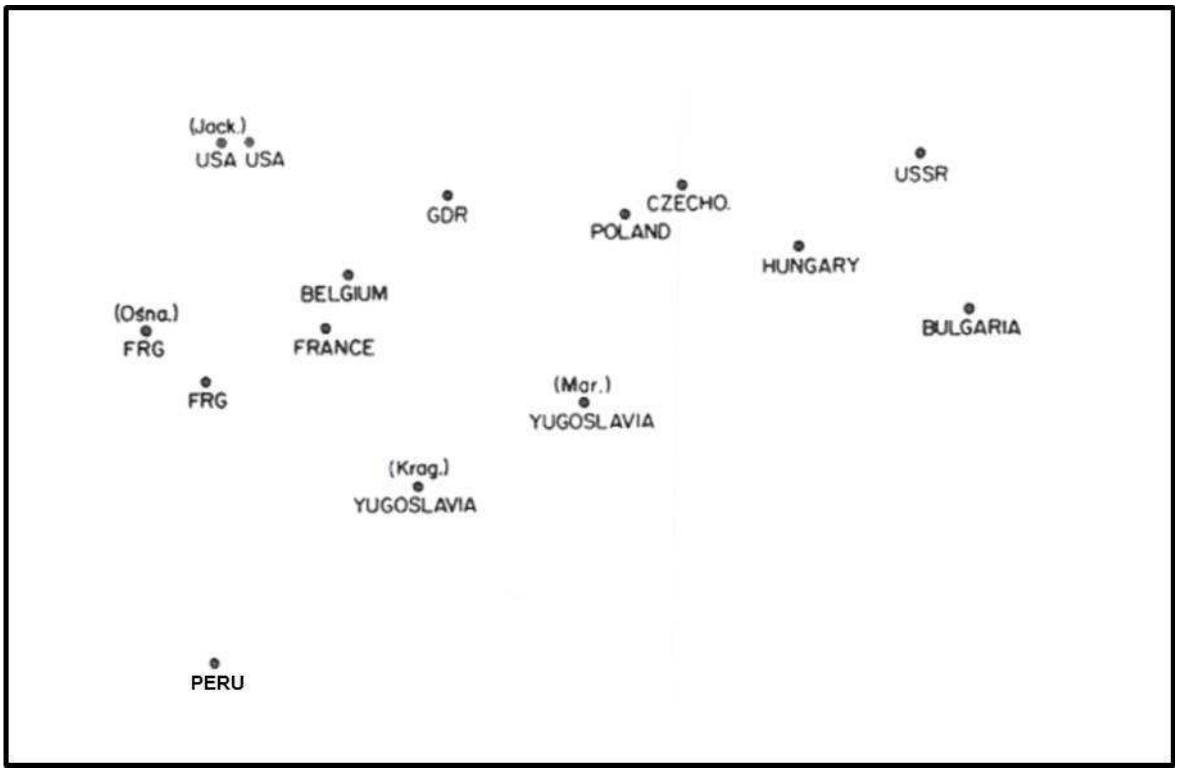
In Figure 3 we have plotted the 'locations' of all our 15 sites with respect to the two major dimensions that arise from such an analysis. We discover to our considerable interest that we have retrieved from these time use profiles a 'picture' that bears a substantial resemblance to a map of the western world, especially if the Atlantic Ocean is removed as though continental drift had not occurred. Peru is off to the 'southwest', both Jackson and the U.S.A. samples are close together to the 'northwest', while Pskov (USSR) and Kazanlik, (Bulgaria), lie fairly near to one another far to the 'eastern' edge of our field of view. The rest of the European sites are filled in along lines, that do only modest violence to a simple geographic representation. (p150)

However, Converse immediately cautioned against this simple explanation on the basis of geographical proximity:

Clearly, the solution is not pure physical geography. The position for the

Kragujevac (Yugoslavia) point is far to the 'West' of its physical location. The Osnabruck (F.R.G.) pair of observations is interchanged with the France-Belgium pair of positions, and so on. However, if we may paraphrase George Bernard Shaw, the marvel is less that our Figure 3 reproduces physical geography poorly, that that it should reproduce it at all. After all, we have not fed the slighted shred of geographical information into the computer, and even if country names rather than code characters had slipped into the machine, the computer would have lacked the wit to impose any kind of geographical ordering whatever onto the results.

Figure 3
Two-dimensional solution for time-use map of 1965



Source: Multinational data from Szalai (As reported in Converse 1972), own illustration.

All that entered the computer were 455 proportions indicating how people at 15 anonymous sites distributed their 24-hour day across 37 disparate and unidentified activity categories. It is remarkable that statistical compression of these raw data yields anything a physical map.

Anticipating the type of analysis to be undertaken next with subsequent diary data collections below, Converse speculated: "Finally, it is natural to wonder how solutions of this sort might look if it were possible to carry them out on data collected at different points in time".

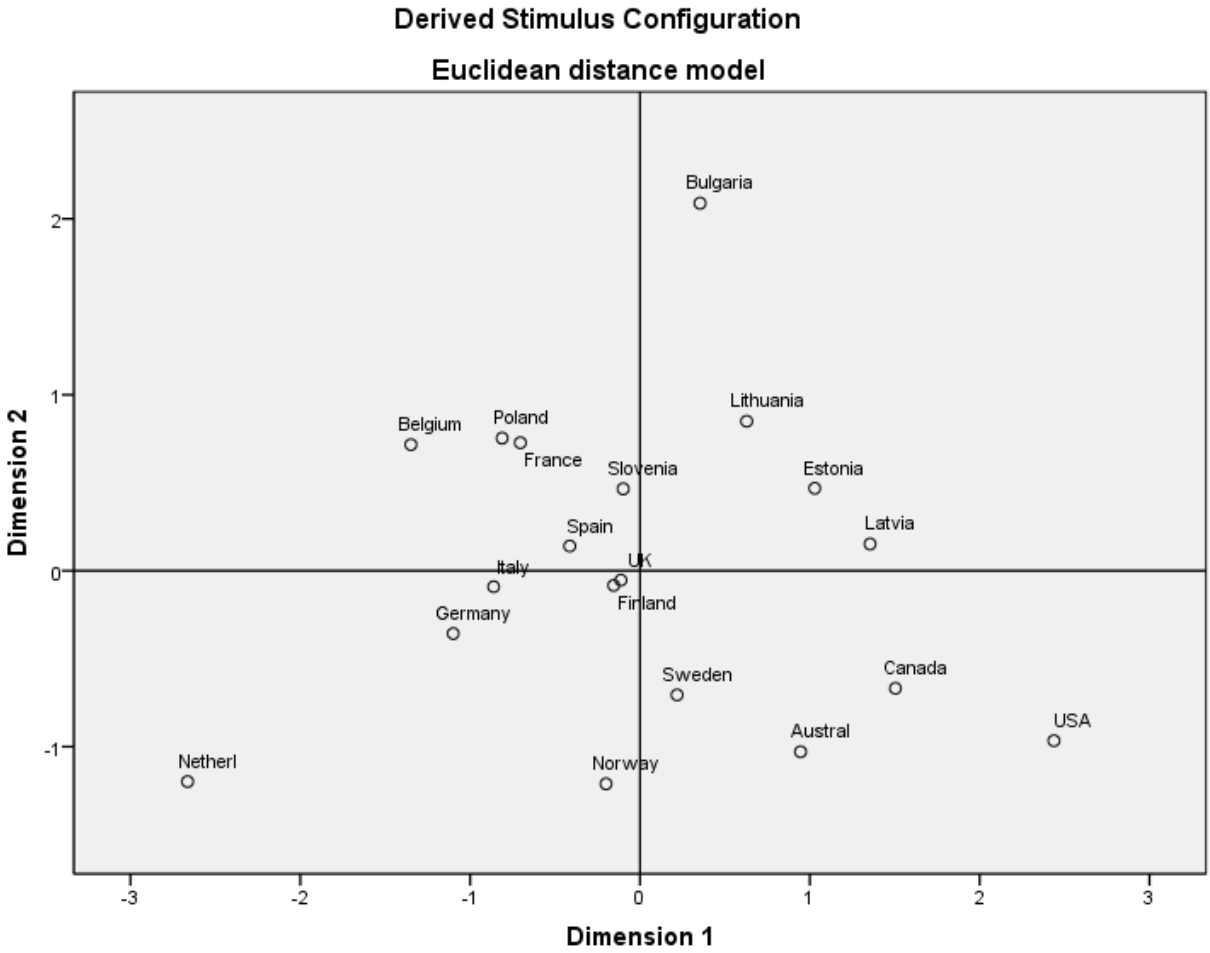
2.2 Updated 1998-2005 MTUS mappings

The recent availability of parallel "harmonized" diary data from the MTUS data archive project initiated and housed at Oxford University – involving more than 25 (mainly European) countries – allows the possibility of replicating, updating and extending these 1965 results to con-

temporary life. Appendix B shows the daily activity differences across these countries by rough geographic categories, as reported in Fisher and Robinson (2011) from the MTUS cross-country files covering 30 daily activities between 1998 and 2005. Here, there is more cross-national variation in diary methods and field procedures than in the Szalai study, although most of the MTUS countries paid very close attention to ensuring multinational and cross-time comparability using agreed-upon statistical guidelines.

Here again, MDS generated maps that represented the major differences between countries in mainly geographic terms, as shown in Figure 4.

Figure 4
MDS plot of multinational positions based on 1995-2005 MTUS diary data



Source: MTUS 1995-2005 (Aggregate data shown in Table 1-Table 4), own illustration.

Using the same basic procedures as Converse employed, the Euclidian distances between countries were calculated from the raw data in Appendix B before entering them into the MDS program in SPSS. Figure 4 reflects different configurations in these MTUS data than in 1965, but then again, there are far more counties available in the MTUS archive (along with different ways of spending time within these countries). Only five of these countries were common to

those in 1965 (France, Germany, Poland Bulgaria and the US), but several other countries had begun collecting national diary data in the 1970s and 80s to track cross-decade trends.

Figure 4 also clearly shows the influence of geography, but often more along language/culture lines than pure physical proximity. For example, the first horizontal dimension contrasts the US and Canada with the Netherlands (and less so Belgium, Germany and Italy), reflecting the sort of continental separation absent from Figure 3. While continental differences are not reflected in the proximity of Australia to the US and Canada, they are for several other countries on the right side of Figure 4 including the three Baltic states, which have less in common with these three Anglophone countries. However, both Baltic and Anglophone counties have more in common, than either does in their difference from Netherlands. In Table 1-Table 4, it can be seen that the Dutch can be seen to be relatively unique in their lower paid work hours, combined with higher socializing and much lower TV hours during free time. These seem to underlie and define most of the difference along the horizontal dimension in Figure 4

Similarly, the second (vertical) dimension mainly serves to contrast Bulgaria at the top from Nordic countries of Norway and Holland at the bottom. While turning Figure 4 upside down does better preserve a north-south dimension, the inclusion of the US and Australia in the "north", and Poland with Lithuania in the "south" does not fit this interpretation particularly well; nor does the placement of Italy and Spain, in the middle of this dimension, make the north-south interpretation any more plausible. What does define Bulgaria's isolation at the top of the vertical dimension are its greater hours on housework, sleep and home meals, combined with lower hours on educational activity, shopping, grooming and various forms of leisure.

Along with the proximity of the three Baltic states (Estonia, Latvia and Lithuania), a number of blocs or groupings in Figure 4 also reflect geographic location: 1) the three Nordic states of Sweden, Norway and Finland, and the pairings of Belgium with France and Italy with Spain. Nonetheless, there are too many "strays" in Figure 4 to consider it a simple replication of the 1965 map in Figure 3.

At the same time, however, these MTUS mappings are consistent with previous analyses of broad trends and shifts in time use using the more conventional procedures reported in Gershuny (2009).

Converse (1972) thus appears to have been too cautious in concluding that:

…..Certainly the reader has reflected on that fact that the strong gradients associated with home use of television are almost certainly transient, being mere functions of the specific period (1965-1966) during which the field works took place. In the United States at one extreme, television use had certainly approached saturation by that period; and in due course of time, it might be expected that its use will have approached saturation as well at the other extreme of our field of view. If this occurs, one of the mainstays of our geographic patterning will have disappeared. (p. 176)

Table 1
Multinational differences weekly hours spent on 31 activities –
South and North America/English speaking

Total hours and minutes per week – Whole population aged 18 to 64	Brazil 2001	Australia 2006	Canada 2005	USA 2003
Paid work/related activity (away from home)	25.8	26.1	28.7	28.6
Paid work at home	2.6	2	NA	1.6
Study & job or skill training	2.1	0.7	1.2	1.1
Homework	1.2	0.9	1.1	0.9
Commuting, job & study-related travel	5.8	3.2	3.0	2.5
Cooking & food related housework	5	6.2	4.8	3.5
All other housework and repairs, gardening	6.2	7.2	8.2	7.8
Shopping, services, other domestic work	3.2	4.6	4.3	3.7
Housework & personal care travel	1.4	2.7	2.7	4.3
Physical/medical child care	1.4	2.2	1.9	2
Interactive & other child care	0.7	3.2	1.0	2
Child care-related travel	0.5	0.8	0.6	0.6
Pet care (excluding walking dogs)	0.1	0.6	0.5	0.4
Sleep & naps	56.4	58.7	58.7	58.6
Wash, dress, & other personal care	7.2	6.2	4.5	5.6
Meals (at home & packed luches)	7.1	6.7	6.2	5.8
Walking (including walking dogs)	0.7	0.7	0.7	0.5
Sport & other exercise	0.9	1.9	2.5	1.6
Organizational & voluntary	3.2	1.3	3.7	3.6
Restaurant, bar, pub, café	2.6	1.3	2.6	1.8
Party, visits & socialise away from home	3.4	2.2	1.5	0.7
Party, visits & socialise at home	2.9	0.4	4.3	6.1
Leisure away from home	0.6	2.5	2.5	1.1
Other travel	3	2	2.0	2.2
Relax, do nothing	1.6	1.5	2.9	1.9
Computing & internet (including games)	0.5	0.4	1.7	1.2
Television	13.3	12.3	13.5	15.6
Radio, Ipod, other audio	0.8	2.3	0.2	0.4
Read	0.7	2.1	2.0	1.9
Other leisure and hobbies	1.3	4.6	0.3	0.3
Unrecorded time (average day)	5.8	0.5	0.3	0.1
Total	168	168	168.0	168

Note: Activities from Fisher and Robinson 2010,
Source: MTUS 1995-2005, own calculations.

Table 2
Multinational differences weekly hours spent on 31 activities – Central European

Total hours and minutes per week – Whole population aged 18 to 64	United Kingdom 2000-01	Belgium 2005-06	France 1998-99	Germany 2001-02	Nether-lands 2000
Paid work/related activity (away from home)	23	18.8	22.1	20.4	18.7
Paid work at home	2.1	1.1	1.3	1.2	1.1
Study & job or skill training	0.9	2	1.9	1.6	1.6
Homework	0.4	1.2	1.2	0.7	0.9
Commuting, job & study-related travel	3.2	3.2	2.8	3	2.8
Cooking & food related housework	6	5.8	6	4.9	6.4
All other housework and repairs, gardening	6.9	8.8	7.9	8.4	7.1
Shopping, services, other domestic work	4.8	4.2	4.7	4.8	4.3
Housework & personal care travel	2.2	1.9	0.1	2.5	2.1
Physical/medical child care	2.3	1.3	1.9	1.3	2.1
Interactive & other child care	1.4	0.7	0.9	0.9	1.9
Child care-related travel	0.9	0.4	0.5	0.4	0.7
Pet care (excluding walking dogs)	0.4	0.4	0.6	0.4	1.2
Sleep & naps	58.8	58.3	61.1	57.3	59.5
Wash, dress, & other personal care	5.4	5.1	5	6.1	6.1
Meals (at home & packed luches)	8.8	11	12.4	10.9	9
Walking (including walking dogs)	1.9	1.8	1.9	1.9	NA
Sport & other exercise	1.3	1.3	1.1	1.6	1.8
Organizational & voluntary	1.5	0.8	1.3	2.2	3.2
Restaurant, bar, pub, café	1.1	1.5	3.2	0.8	1.9
Party, visits & socialise away from home	5.3	4.4	3.2	4.6	8.2
Party, visits & socialise at home	1.9	2.5	1.8	3.3	2.9
Leisure away from home	0.9	1.3	1.4	1.6	1.1
Other travel	3.3	5	3.6	4.2	3
Relax, do nothing	2.2	3	0.7	1.8	1.4
Computing & internet (including games)	1.2	2.6	0.6	2	1.8
Television	15.6	15.4	13.2	12.1	8.1
Radio, Ipod, other audio	0.7	0.5	0.4	0.6	4
Read	2.5	2.5	2.2	3.9	3.7
Other leisure and hobbies	0.7	1.1	3	2.2	1.4
Unrecorded time (average day)	0.4	0.1	NA	0.4	0
Total	168	168	168	168	168

Note: Activities from Fisher and Robinson 2010,
Source: MTUS 1995-2005, own calculations.

Table 3
Multinational differences weekly hours spent on 31 activities – Northern European/Nordic/Baltic

Total hours and minutes per day – Whole population aged 18 to 64	Norway 2000-01	Sweden 2000-01	Finland 1999-2000	Estonia 1999-2000	Latvia 2003	Lithuania 2003
Paid work/related activity (away from home)	24.5	26.7	22.2	27.1	29.3	24.9
Paid work at home	1.2	1.2	2.1	1.5	2.6	5.6
Study & job or skill training	1.4	1.4	1.8	1.1	1.9	2
Homework	0.8	0.7	0.7	0.5	0.7	0.7
Commuting, job & study-related travel	3.2	2.9	2.5	3.3	4.3	3.4
Cooking & food related housework	5.6	5.8	5.1	7.4	5.7	7
All other housework and repairs, gardening	6.3	6.8	7.7	9.5	7.7	9.6
Shopping, services, other domestic work	4.1	4.2	3.9	3.5	2.6	2
Housework & personal care travel	1.6	2.2	1.8	2.1	2.5	2.2
Physical/medical child care	2.3	2	1.9	0	1.1	1.4
Interactive & other child care	0.8	0.9	0.8	0.9	0.6	0.8
Child care-related travel	0.4	0.6	0.2	0.2	0.2	0.1
Pet care (excluding walking dogs)	0.1	0.2	0.4	0.2	0.1	0.1
Sleep & naps	56.2	56.4	59	59.5	59.9	58.9
Wash, dress, & other personal care	5.5	5.3	4.9	6.2	4.7	6.4
Meals (at home & packed luches)	8.5	10.3	8.4	8.4	9.8	10
Walking (including walking dogs)	1.8	2	2	1.6	1.9	1.2
Sport & other exercise	2.1	2	2.3	1.1	1.5	1.1
Organizational & voluntary	1.5	1.6	2	1.8	1.4	1.9
Restaurant, bar, pub, café	0.9	0.4	0.7	0	0.5	0.1
Party, visits & socialise away from home	5.6	4.1	3.7	2.3	2.7	2.5
Party, visits & socialise at home	6.5	3.2	2.6	1.4	1.4	1.5
Leisure away from home	0.9	0.7	0.7	0.6	0.7	0.2
Other travel	4.1	4.6	4.2	2.6	3.1	2.8
Relax, do nothing	1.3	2.6	2.1	1.8	2.1	1.2
Computing & internet (including games)	1.3	1.4	0.9	0.4	0.5	0.8
Television	12.6	11.9	14.7	15.4	13.8	15.3
Radio, Ipod, other audio	0.7	0.5	0.9	0.7	0.5	0.6
Read	3.7	3.3	4.9	4.1	2.8	2.5
Other leisure and hobbies	2.1	1.6	1.8	1.3	1	0.9
Unrecorded time (average day)	0.4	0.5	1.1	1.5	0.4	0.3
Total	168	168	168	168	168	168

Note: Activities from Fisher and Robinson 2010,
Source: MTUS 1995-2005, own calculations.

Table 4
Multinational differences weekly hours spent on 31 activities – Eastern/Southern Mediterrean Europe

Total hours and minutes per day - whole population aged 18 to 64	Poland 2003-04	Slovenia 2000-01	Bulgaria 2001-02	Turkey 2006	Italy 2002-03	Spain 2002-03
Paid work/related activity (away from home)	20.1	23.6	23.7	20.8	23.6	24.6
Paid work at home	3.5	1.1	0.2	NA	0.5	0.7
Study & job or skill training	2	1.5	0.6	2.8	1.1	2
Homework	1.3	1.6	0.5	NA	1.4	1.2
Commuting, job & study-related travel	2.9	2.9	2.8	NA	3.5	3.6
Cooking & food related housework	8.2	7.2	8.6	8.9	7.1	7.1
All other housework and repairs, gardening	8.1	11.9	11.6	7.5	8.9	6.7
Shopping, services, other domestic work	2.9	2.5	1.8	1.6	3.6	4.3
Housework & personal care travel	2.1	1.9	1.9	NA	1.8	1.3
Physical/medical child care	1.8	1.4	1.1	3.4	1.5	2.1
Interactive & other child care	1.6	1.1	1.1	NA	1.2	0.6
Child care-related travel	0.2	0.2	0.1	NA	0.5	0.6
Pet care (excluding walking dogs)	0.2	0.2	0.1	NA	0.1	0.1
Sleep & naps	58.7	58.1	62.4	59.3	57.3	59
Wash, dress, & other personal care	6.1	4.7	4.4	18.8	7.1	5.6
Meals (at home & packed luches)	10.4	9.6	12.6	NA	11.7	11.3
Walking (including walking dogs)	2.1	2.5	2.1	NA	2.3	3.9
Sport & other exercise	1.1	1.6	0.9	0.8	1.3	1.3
Organizational & voluntary	2.9	1.4	1.1	4.4	1.8	1.4
Restaurant, bar, pub, café	0.2	0.6	1.8	NA	1.5	0.9
Party, visits & socialise away from home	3.4	4.1	2.6	0.4	4.6	5.1
Party, visits & socialise at home	2.8	2.9	1.9	8.3	1.9	1.4
Leisure away from home	0.4	0.6	0.1	NA	0.7	0.8
Other travel	3.1	3.2	2.4	9.3	4.7	3.3
Relax, do nothing	1.3	3.4	0.9	4	3.3	2.7
Computing & internet (including games)	1.1	0.7	0.1	NA	0.7	1.1
Television	15.3	13.2	16.6	13.8	10.6	12
Radio, Ipod, other audio	0.9	0.6	0.5	0.5	0.4	0.4
Read	2.6	2.5	2	1.3	2	1.6
Other leisure and hobbies	0.5	1	1.3	2.1	1	1.1
Unrecorded time (average day)	0.2	0.2	0.2	NA	0.3	0.2
Total	168	168	168	168	168	168

Note: Activities from Fisher and Robinson 2010,
Source: MTUS 1995-2005, own calculations.

Indeed, as can be seen in the substantial and leisure-dominating TV figures for all countries (except the Netherlands, one of the main activities isolating the Netherlands in Figure 4) , all countries have come close to TV saturation in the early 21_{st} century, but with viewing hours that are closer to 40% of free time (in the 12-15 weekly hour range). This, in contrast to the 25% of free time among TV set owners across countries in the 1965 Szalai study, where viewing hours were less than 10 hours per week.

3 Summary and conclusions

MDS has again generated useful visualizations that summarize differences between countries over the last half century, using its two-dimensional plot from these differences in time use across countries. The present article updates and extends Converse's (1972) conclusion about applying MDS methods to the more recent time-diary collection in the Oxford University MTUS data archive – covering more than 20 (mainly European) countries. Again, plausible and insightful (but somewhat different from 1965) configurations emerged from MDS visualizations, even though there were only five of the 1965 countries for which updated diary data were available.

Even though it is not possible to quantify whether this represents any increasing convergence in time-use across countries, the MDS-generated country groupings from the 1998-2005 multinational diary data in Table 1 - Table 4 were again largely based on geographical or cultural proximity, much as Converse concluded four decades earlier. Moreover, these updated mappings were compatible with conclusions from earlier more conventional analyses of these recent data described in Gershuny (2009).

Figure 4 makes it possible to confirm that differences in methods across MTUS countries did not obscure the fundamental uniqueness of life in each country. These results extend Converse's geographic interpretation, but not in all respects:

…..There is, however, a difference between the transient weight of specific activities on these patterns, and the persistence of the patterns themselves. If we had completed our field work 25 years earlier, mass television use would have exerted no influence whatever on the outcome, but it is very likely that radio and movie gradients, working in an opposite sense from those we have seen here, would have sustained these geographic patterns with much the same strength (p180).

At least over the last half century, television may have diminished in its ability to differentiate daily life in different countries, but it has been replaced by paid work, family care and other activities that reflect strong geographic/cultural connections (as shown in Table 1 - Table 4 and as described further in Robinson and Martin 2010).

References

Banks, A. S. and R. B. Textor (1964), World handbook of social and political indicators, Cambridge MA, MIT Press.

Bloombaum, M. (1970), Doing smallest space analysis, in: *Journal of Conflict Resolution*, Vol. 14, No. 3, 409-416.

Converse, P. (1972), Country differences in time use, in: Szalai, A. (ed.), *The use of time*, The Hague, Mouton.

Elizur, D. and A. Sagie (1999), Facets of personal values – A structural analysis of life and work values, in: *Journal of Applied Psychology*, Vol. 48, No. 1, 73-87.

Elizur, D. (1984), Facets of work values – A structural analysis of work outcomes, in: *Journal of Applied Psychology*, Vol. 69, No. 3, 379-389.

Elizur, D. (1994), Gender and work values – A comparative analysis, in: *The Journal of Social Psychology*, Vol. 134, No. 2, 201-212.

Fiorello, C. (2006), The cylindrical structure of the Wechsler Intelligence Scale for children –IV – A retest of the Guttman Model of Intelligence, New York, Elsevier.

Foster, J., Barkus, E. and C. Yavorsky (2006), Understanding and using advanced statistics, Thousand Oaks, Ca, Sage.

Gershuny, J. (2009), Veblen in reverse, in: *Social Indicators Research*, Vol. 93, No. 1, 37-45.

Gershuny, J. (2000), Changing times Oxford, Oxford University Press.

Guttman, L. (1968), A general non-metric technique for finding the smallest coordinate space for a configuration of points, in: *Psychometrika*, Vol. 33, No. 4, 469-506.

Inglehart, R. (1997), The Inglehart Values Map in modernization and postmodernization, Princeton University Press. The map can be viewed directly at worldvaluessurvey.org.

Johnston, C., Broonen, J., Stauffer, S., Hamtiaux, A., Pouyaud, J., Zecca, G., Houssemand, C. and J. Rossier (2013), Validation of an adapted French form of the Career Adapt-Abilities Scale in four francophone countries, in: *Journal of Vocational Behavior*, Vol. 83, No. 1, 1-10.

Kruskal, J. (1964), Non-metric multidimensional scaling – A new method, in: *Psychometrika*, Vol. 21, No. 2, 118-129.

Kumar, V., Ryan, R. and K. Wagner (2012), Structure of the Personality Beliefs Questionnaire-Short-Form – Smallest space analysis, in: *Current Psychology*, Vol. 31, No. 4, 386-392.

Kumar, V. K., and R. B. Ryan (2009), On measuring forgiveness – Implications from smallest space analysis of the Forgiveness Likelihood Scale, in: *Current Psychology*, Vol. 28, No. 1, 32-44.

Lalouel, J. and A. Langaney (1980), Bedik and Niokholonko of Senegal – Inter-village relationship inferred from migration data, in: American Journal of Physical Anthropology, Vol. 45, No. 3, 453-466.

Laumann, E. O. and L. Guttman (1966), The relative associational contiguity of occupations in an urban setting, in: *American Sociological Review*, Vol. 31, No. 2, 169-178.

Meir, E. (2010), The structure of occupations by interests – A smallest space analysis, in: *Journal of Vocational Behavior*, Vol. 76, No. 1, 16-24.

Robinson, J. and R. Hefner (1968), Perceptual maps of the world, in: *The Public Opinion Quarterly*, Vol. 32, No. 2, 273-280.

Robinson, J. and S. Martin (2010), IT use and declining social capital – More cold water from the GSS and ATUS, in: *Social Science Computer Review*, Vol. 29, No.1, 45-63.

Ruch, L. (1984), Dimensionality of the bem sex role inventory – A multidimensional analysis, in: *Sex Roles*, Vol. 10, No. 1-2, 99-117.

Shapira, Z. (1976), A facet analysis of leadership styles, in: *Journal of Applied Psychology*, Vol. 61, No. 2, 136-139.

Singh, P., Bhandarker, A., Rai, S. and A. Jain (2011), Relationship between values and workplace – An exploratory analysis, in: *Facilities*, Vol. 29, No. 11/12, 499-520.

Szalai, .A. (1972), The use of time, The Hague, Mouton.

electronic ***International Journal of Time Use Research***
2013, Vol. 10, No. 1, 91-111 dx.doi.org/10.13085/eIJTUR.10.1.91-111

time-pieces

news on time use research in the
electronic **International Journal of Time Use Research**

New developments in time technology – projects, data, computing and services

COMING FULL CIRCLE – INTRODUCING THE MULTINATIONAL TIME USE STUDY SIMPLE FILE

Kimberly Fisher
University of Oxford

Jonathan Gershuny
University of Oxford

At time of writing, more than 170 publications use the Multinational Time Use Study (MTUS) archive, assembled by the Centre for Time Use Research at the University of Oxford. The MTUS includes more than three-quarters of a million diary-days from 68 surveys collected in 22 countries and spanning 60 years. In July 2013, CTUR released a new MTUS Simple File, representing a service not previously available to users as well as a return to principals that shaped the original MTUS file.

The MTUS arose during Jonathan Gershuny's early career research into "post-industrial society". Gershuny (1978, 1983) used expenditure diary surveys to test Daniel Bell's 1976 thesis that as societal wealth increases, economies shift from the production of goods to the production of services. Gershuny noticed that households decreased spending on services while increasing their service use by producing many final services for themselves outside the economic activity typically measured at the time (1983). Gershuny characterised post-industrial society as the growth in "knowledge work" service jobs alongside increasingly sophisticated manufacturing technology that permitted the expanding service employment while household spending on service consumption declined.

Gershuny's research built on ideas previously suggested by Robert Giffen (summarised in Gershuny 2000), and Hildegarde Kneeland (1929), who based her work on time diaries collected from samples of women in the United States from the 1920s through 1930s. Gershuny em-

barked on a search for available time diaries to explore the behaviour component of household self-servicing. Consistent with many researchers at the time, Gershuny compiled summary files containing total minutes per day devoted to various activities by individuals in national-level population surveys to assess alongside aggregate expenditure data.

Gershuny found the 1965 Multinational Comparative Time-Budget Research surveys (Szalai 1972), which restricted samples to working-aged people (18-65, though much of the analysis concentrated on the more limited age range of 20-59). Through extensive legwork, as well as personal interaction with John Robinson, staff of the BBC Audience Research Department, Peter Willmott and others, (Fisher and Gershuny 2013), Gershuny amassed a collection of British time diary surveys to harmonise with the Szalai surveys. Thus the basis for the MTUS was born.

The European Foundation for the Improvement of Living and Working Conditions funded the first release of the MTUS, covering seven countries (Gershuny 2000). The first MTUS file only included diaries from people aged 20-59, summary time in 41 activities, and 10 survey and demographic variables.

The initial MTUS harmonisation progressed rapidly. The Szalai project already offered harmonised classifications for activities, diary context, and background variables. Gershuny only needed to condense the Szalai activity codes for his post-industrial research, and adopt a limited range of background variables consistent with the variables available from the household expenditure surveys. The age of youngest diarists varies markedly from survey to survey, and quality of data collection also varies considerably among older diarists in some surveys. As the early MTUS only included diaries from people aged 20-59, the early MTUS sidestepped the problem of how to deal with differential youngest ages and variations in response rates and quality of diaries from the oldest diarists. The initial harmonisation process also included no effort to improve data or address inconsistencies. Gershuny and his main assistant, Sally Jones, created limited documentation. Complex research questions inspired the creation of an MTUS file reflecting the simple needs from the time diary data.

As more users have worked with the MTUS, and as the general sophistication of time use research has progressed over the decades, user interest in greater detail has expanded at a pace beyond the capacity to incorporate new files and new features in the MTUS. At the time of writing, CTUR included a tiny staff of one archivist, four researchers, one further researcher with substantial teaching commitments, and two PhD students. As this team undertakes a range of projects, of which MTUS is only one, and as many members do not contribute to the production of MTUS materials, CTUR has restricted capacity to expand these resources. In spite of these limitations, CTUR nevertheless has created a harmonised episode-level file (now covering 35 surveys from 12 countries), and an expanded range of household and diary variables. CTUR also has expanded the documentation for the MTUS. The main user guide now covers more than 150 pages of main text, with survey level supplementary documentation for each individual survey, separate documentation for child diaries and supplementary files, packages

to add variable and value labels in two languages, and metadata summaries of all surveys included in the MTUS - as well as of over 400 other time use surveys (Fisher et. al. 2013).

Additionally, the MTUS adds value through data cleaning and enhancement which makes use of the narrative properties of diary data (Fisher and Gershuny 2013). As some information spans column categories in diary instruments (for instance, the word "train" simultaneously conveys an activity, a mode of transport, and a location; or an activity description "shopping on-line" both reflects the activity of shopping and the context that this activity involves use of the internet), MTUS codes all detail provided by diarists or collected by the survey into all relevant domains. CTUR takes advantage of contemporary computing power not available to survey designers in the 1950s through the 1980s to undertake a variety of data cleaning tasks skipped as too time consuming and expensive in the past. In some cases, the MTUS team has recovered corrupted files and worked from some stored paper materials to recover information not stored in the electronic files to which most researchers have had access. As a result, the best available version for many of the surveys – particularly some of older surveys, included in the MTUS is the MTUS version.

Future developments will improve the delivery of the data and documentation. The US National Institutes of Health fund collaboration between CTUR, the Maryland Population Research Center and the Minnesota Population Center, to add all years of, first, USA-based time use surveys, then surveys from a limited range of other countries included in the MTUS to the Time Use Survey-X distribution mechanism (Hofferth, Flood and Fisher 2012). The ATUS-X system, developed by the two population centres, presently enables registered users to select only required sets of variables for sub-samples of surveys. The ATUS-X system additionally assists users to create customised time use variables that map together elements of the diary, as well as to map in additional variables from earlier waves of the longitudinal Current Population Survey (from which the American Time Use Survey (ATUS) sample is drawn). The USA component of the MTUS should be available to the time use community through the TUS-X access system by late 2014 or early 2015. The collaboration will add some further countries from the main MTUS files into the TUS-X distribution mechanism over subsequent years, and enhance the accessibility of information in the MTUS documentation (Hofferth, Flood and Fisher 2012).

In parallel, CTUR has raised grants to overhaul the MTUS website. The site already links users to publications using MTUS data (available in a searchable database). Current development work will transform the current pdf documentation files into searchable databases of survey, harmonisation process, and variable-level metadata. CTUR also will be amending the MTUS file distribution mechanisms to facilitating customised downloads of documentation.

An unfortunate by-product of the size of the CTUR team and the range of services supplied through the MTUS, however, is that the addition of new surveys and the upgrade of surveys only harmonised to the standards of a previous version, is slow. The process of converting original files into the full range of current MTUS outputs presently takes a minimum of five weeks.

Developing the harmonised episode file, new data distribution systems, data enhancement, and improved documentation facilities continue to require considerable effort and time resources.

As more surveys become available, more users want access to a wider range of recent data. At the same time, more users also hope to make use of the cross-time and historical change possibilities offered by the MTUS and eagerly await the upgrade of the older data. The lack of capacity to produce the range of desired outputs on our present resources has created a degree of frustration within CTUR.

This situation gave rise to the solution of the new MTUS Simple File. Until the release of this file, MTUS offered a wide range of files. The most used and best documented of these files offer the current best practice aggregate time use, demographic and episode data for some surveys. Older variations of the MTUS including only those surveys not yet upgraded to current best practice add to the number of total surveys available. Nevertheless, the standard of harmonisation in the older files is not as high quality as in the newer files. Some MTUS variables have changed, and the range of information has expanded. This had meant that matching the older and the newer files was not straight-forward, and required post-mapping effort. Less experienced users encountered some difficulties undertaking this task.

CTUR now documents the mapping of the older and newer files together (Appendix 3 of the MTUS User Guide), and highlights the shortcomings entailed. Algorithms now combine the older and all current best practice files into the single Simple File, where each row case represents one 24-hour time diary, including:

- the seven essential MTUS survey to diary level identifiers;
- three diary date variables;
- a limited range of twelve household- and person-level variables;
- summary time in twenty five activity categories (with four additional variables separating eating out or eating at school or work from other eating time; playing computer games from other computer and internet use; and child and adult care travel from other travel);
- for diarists in couples, the total minutes in the diary day that the diarist reported that her or his partner was present.

Thus, users now can directly download a single file with a maximum number of countries immediately ready for use.

The twenty-five activities reflect recent amendments to the MTUS (covering new activities, like using computers and the internet, as well as improvements to older time use categories, including splitting out interactive child care (such as playing with, reading to or teaching children) from the more routine physical and supervisory forms of care. The limited activity range nevertheless returns to an earlier principal of emphasising the activities with the highest number of research applications.

This Simple File, like the original MTUS, offers a limited range of demographic variables. In so doing, the MTUS Simple File provides an easily-used file, suitable for rapid production of figures for experienced time use researchers, as well as a less daunting file for use by people who have recently begun working with time use data. Offering a readily comprehensible file to novice users particularly matters with the current dearth of time use research text books (of which Michelson 2005 and Durán 2007 are the most recent) and the limited number of time use training courses.

The one complexity in the Simple File not present in the earliest MTUS versions is that the Simple File covers all diaries from all age groups. As a consequence, users of the Simple File will need to consider potential diary quality issues in relation to the diaries from the oldest respondents (as suggested in the survey documentation). Users also will have to account for the differential age of the youngest diarists across the surveys. The MTUS documentation on using the child dairies (http://www.timeuse.org/sites/ctur/files/1796/youth-supplement.pdf) will assist users in consideration of how to plan analysis of the daily activities of the younger contributors to the MTUS archives.

As well as being simple to use, the Simple File also is simple to create. CTUR chose not to include complex variables that can be time consuming to produce (such as the identifiers mapping diaries between household members) – thereby significantly reducing the staff training and total time required to add new surveys (or to upgrade older surveys). In future, CTUR will follow two data inclusion strategies:

- a precursor to full inclusion conversion that enables some surveys to be released in the simple format more rapidly, but which also facilitates full episode level conversion later (this process involves making only those episode and full aggregate file variables required for the Simple File – a single algorithm produces the Simple File format from the full MTUS files. This conversion procedure produces just enough of the full files to permit this algorithm to create the Simple File version of a survey); and
- a Simple File only-conversion to include some form of datasets we are not likely to otherwise have resources to fully convert.

In consequence, the Simple File maximises the range of available countries and time periods for researchers exploring changes in daily behaviours.

The time use research field has grown considerably in recent years, as evidenced by the rising volume of new time use publications, numbers of people joining the International Association for Time Use Research e-mail list, and the expanding range of countries conducting national time use surveys. The MTUS archive offers a basis against which researchers can monitor changes in behaviour trends over time, as well as assess the potential impacts of changes in the way time use surveys are collected over time. The MTUS Simple File both provides a straightforward entry point to time use research for people new to the field, while also facilitating ready access to production of basic statistics for experienced users. This new file structure also

will facilitate the more rapid release of more surveys in the MTUS format. If you are not already an MTUS user and wish to explore the archive, access is free for all academic and policy researchers, and can be arranged following the registration process on the MTUS website (http://www.timeuse.org/mtus/register), With the release of the Simple File, the MTUS project has come full circle in finding a way to make the most of early efforts and principals while also facilitating the future expansion of MTUS resources.

REFERENCES

Bell, D. (1976), The cultural contradictions of capitalism, New York, Basic Books.

Durán, M. A. (2007), El valor del tiempo – Cuántas horas te faltan al día?, Madrid, Espasa.

Fisher, K., Tucker, J., Altintas, E., Bennett, M., Jun, J. and members of the time use team (2013), Technical details of time use studies, release 13 July 2013, Oxford, Centre for Time Use Research, University of Oxford.

Fisher, K. and J. Gershuny, with contributions from E. Altintas and A. H. Gauthier. (2013), Multinational Time Use Study – Users' guide and documentation, Oxford, Centre for Time Use Research.

Gershuny, J. (1978), After industrial society? – Emerging self-service economy, London, Macmillan.

Gershuny, J. (2000), Changing times – Work and leisure in postindustrial society, Oxford, Oxford University Press.

Gershuny, J. (1983), Social innovation and the division of labour, Oxford, Oxford University Press.

Hofferth, S. L., Flood, S. and K. Fisher (2012), Time pieces – The time use data access system – the next phase of the American Time Use Survey Data Extract Builder (ATUS-X), in: *Electronic International Journal of Time Use Research*, Vol 9, No. 1, 151-155.

Kneeland, H. (1929), Woman's economic contribution in the home, in: *Annals of the American Academy for Political and Social Science*, Vol. 143, No. 1, 33-40.

Michelson, W. (2005), Time use – Expanding explanation in the social sciences, Boulder, Colorado, Paradigm Press.

Szalai, A. (1972), The use of time – Daily activities of urban and suburban populations in twelve countries, Paris, Mouton.

THE 2014-2015 UNITED KINGDOM TIME USE SURVEY

Kimberly Fisher
University of Oxford

Jonathan Gershuny
University of Oxford

The April 2013 meeting of the Eurostat Working Party on Time Use Surveys in Luxembourg offered depressing news for researchers interested in daily activities in the United Kingdom. At that time, the UK appeared on the list of countries with no plans to conduct a second round Harmonised European Time Use Survey (or any other form of official national population sample time use survey). While a range of factors dampened enthusiasm for further time use surveys in official circles, the Centre for Time Use Research at the University of Oxford has pursued a range of avenues to ensure the collection of contemporary time diary surveys in the UK. By the start of 2013, CTUR had formed collaborations with partners and raised funding for time use components in two longitudinal surveys. Together with the Centre for Longitudinal

Studies at the University of London Institute of Education (which, among other projects, manages the British Millennium Cohort Survey, MCS), CTUR raised resources to add a time diary to the battery of instruments the MCS participants will complete during the age 14 fieldwork in 2014. CTUR also is collaborating with the Institute for Social and Economic Research at the University of Essex to design a time diary which will be part of the innovation panel of the Understanding Society UK Household Longitudinal Study, which also will go into the field in 2014.

CTUR Director Jonathan Gershuny recently secured a European Research Council (ERC) Advanced Grant to fund a five-year time use research programme at CTUR, and now, thanks to a £3.7 million grant from the UK Economic and Social Research Council, CTUR has the funding to conduct a new national sample time use survey in the United Kingdom following the 2008 second phase Harmonised European Time Use Surveys (HETUS) guidelines issued by Eurostat. Tenders for conducting fieldwork will be received by the end of December 2013. The new ESRC funding also will support continued improvements to the Multinational Time Use Study (MTUS), as well as supplementing diaries samples with accelerometers, "SenseCam" technology (which provides a continuous video record of a diarist's experiences throughout the diary day), and potentially other additional devices. CTUR soon will be advertising for several new posts to work on time diary data collection and analysis.

CTUR staff currently are designing of the new UK HETUS instruments, with aims to optimise compatibility with the current second round HETUS project, with the 2000-01 first round UK HETUS, as well as with the longer sequence of UK data in the Multinational Time Use Study. A subsample of the new UK HETUS households will complete an affect field in their time diary alongside the activity and activity-context reporting. The ESRC grant requires this new survey to enter the field in April 2014, and for data collection to continue through March 2015. This survey will collect two diaries from all household members aged 8 and older in sampled private households, one diary on a week day or work/school day, and one diary on a weekend or non-work/school day. Diarists will record their activities in their own words.

As the new UK HETUS fieldwork will overlap the collection of time diaries in the two longitudinal surveys (the Millennium Cohort Survey and Understanding Society), those who analyse this data will have a chance to compliment diary analysis with longitudinal evidence. The UK HETUS survey design process already includes preparation for distribution to researchers through the MTUS, the TUS-X extract system (in which CTUR is collaborating with the Maryland Population Research Centre and the Minnesota Population Centre to provide customised variable construction and file download facilities), and through the UK Data Archive. 2014 and beyond will offer many new research opportunities for those with an interest in daily behaviour in the UK.

electronic ***International Journal of Time Use Research***
2013, Vol. 10, No. 1, 91-111 dx.doi.org/10.13085/eIJTUR.10.1.91-111

KEEPING AN EYE ON TIME USE IN THE NETHERLANDS – TOWARDS NEW TRENDS

Mariëlle Cloïn
Research Sector Care, Emancipation and Time Use
The Netherlands Institute for Social Research

The Dutch Time Use Survey is unique in its long time span: between 1975 and 2005, data were collected every five years. During that period, the method of data collection has remained largely unchanged. Owing to this long tradition, changes in time use could be followed over decades. Nevertheless, given the specific method used, the results were not readily comparable with the results of research in other countries.

That situation has changed. The 2005 round used the method that had been usual in the Netherlands up to that point; in 2006, the survey was carried out following the Harmonised European Time Use Surveys guidelines (Hetus) (European Communities, 2009). After comparing the two surveys (see Kamphuis et al., 2009, in Dutch) the decision was made to follow the Hetus guidelines in the future. The results of this 2006 Dutch Time Use Survey were also published in international comparative study (Cloïn, 2012). In March 2011, new fieldwork (in collaboration with Statistics Netherlands (cbs)) began, in accordance with the Hetus guidelines. For the TBO 2011, a representative sample of the Dutch population of just under 2000 respondents (10 years and older) filled out a diary for seven consecutive days, recording their activities in their own words. Afterwards, the activities were coded using the HETUS harmonized coding frame.

Based on the Dutch Time Use Survey (Tijdsbestedingsonderzoek, TBO 2011), a new report is published on November 26th 2013 by the Netherlands Institute for Social Research | SCP[1], focusing on time use in the Netherlands. The emphasis is on the results for the period 2006-2011. Since different research methods were followed up to and including 2005, the years 2005 and 2006 represent a double measurement. The data for the years 1975-2005 are included only in order to make it possible to study whether trends in time use span a longer period, whether earlier trends come to a halt or begin moving in the opposite direction. The findings relate to the Dutch population aged 12 years and older.

Time use in the Netherlands 2006-2011 – less time for obligations and no further decline in free time

On balance, the Dutch spent less time on obligations in 2011 than in 2006. In the population aged 12 years and older, there is a reduction from 42.8 to 41.2 hours per week (table 1). For the busier 20-64 year-olds, the reduction is from 46.7 to 45.9 hours per week. The long series of time use surveys (starting in 1975) has never before shown a significant reduction in the amount of time spent on obligations. Although the growth in this category of time use slackened somewhat after the turn of the millennium, the reduction between 2006 and 2011 marks the first actual decline. Also another 'constant' development in time use has come to a halt: free

[1] http://www.scp.nl/Publicaties/Alle_publicaties/Publicaties_2013/Met_het_oog_op_de_tijd with a summary in English

time is no longer decreasing. Especially around the turn of the millennium there was a reduction in the weekly amount of free time.

Table 1
Time use, Dutch population ≤ 12 years –
1975-2005 and 2006-2011 (in hours per week)

	1975	1980	1985	1990	1995	2000	2005	2006	2011
obligatory time	40,7	40,8	40,7	42,0	42,6	43,9	44,3	**42,8**	**41,2**
personal time	76,3	76,8	75,3	75,5	75,0	76,6	76,2	**76,9**	**77,7**
free time	47,9	47,0	49,0	47,2	47,3	44,8	44,7	46,9	47,8

Bold: difference between 2006 and 2011 is significant ($p < 0.05$)
a. The total doesn't add up to 168 hours, because of some unspecified time use activities and general categories such as filling in the diary (1.4 hours a week in 2006 and 1.3 hours per week in 2011 in total) Source: SCP (TBO'75-'05, TBO 2006) SCP & CBS (TBO 2011), own calculations.

Less cleaning and cooking

Where are people saving time and which people are they? The reduction in obligatory tasks derives not from a reduction in paid work but in the time spent on household chorus. Dutch people spent an average of 2.4 hours per week less on the household in 2011 (and especially the more routine domestic tasks such as cooking, cleaning, laundry, etc.) than in 2006 (17.9 hours in 2011 versus 20.3 hours in 2006). The reduction is actually slightly greater among 20-64 year-olds (not in table). Both men and women (in proportion to their input) spent less time on the household. The biggest reduction in time spent on these tasks is in households with young children aged up to four years and households with children of secondary school age (12-17 years). In addition, people with part-time jobs (1-34 hours per week paid work) and people who are not in paid work, in particular, have reduced the amount of time devoted to household tasks.

Table 2
Time spent on study, paid work, household tasks and childcare,
Dutch population ≥ 12 years, 1975-2005 and 2006-2011 (in hours per week)

	1975	1980	1985	1990	1995	2000	2005	2006	2011
study	6,7	7,3	7,2	6,9	6,4	5,5	5,7	3,1	3,7
paid work	14,8	14,0	14,1	16,6	17,3	19,4	19,7	19,3	19,6
Household & child care	19,1	19,5	19,4	18,5	18,9	19,0	18,9	**20,3**	**17,9**

Bold: difference between 2006 and 2011 is significant ($p < 0.05$)
a. The total doesn't add up to 168 hours, because of some unspecified time use activities and general categories such as filling in the diary (1.4 hours a week in 2006 and 1.3 hours per week in 2011 in total) Source: SCP (TBO'75-'05, TBO 2006) SCP & CBS (TBO 2011), own calculations.

There was no net change in the amount of time devoted to paid work. That, too, is a break with the past: for a long time, the time spent on paid work increased steadily. Now, this is the case only for some groups in society: parents with young children (up to four years old) and people who already work full-time (35 hours per week or more). Since parents with young children have also made a bigger than average reduction in the amount of time spent on household tasks, the total amount of time devoted to obligatory tasks remains unchanged. The reduction in the amount of time spent on the household by full-time workers is not enough to compensate for the increase in time spent on paid work, and as a result they have two hours per week' more obligations in 2011 than in 2006.

Personal and free time

The Dutch spend an average of 77.7 hours per week on personal time in 2011 (see table 1), making this the largest time use category. A substantial proportion of this time is taken up with sleeping, on which the average Dutch person aged 12 years and older spends 59.5 hours per week. The time spent on eating and drinking has remained unchanged in recent years.

In an average week, Dutch citizens aged 12 years and older have 47.8 hours' free time (table 1). Men have more free time than women and parents of (young) children have less free time than average. Full-time workers have almost two hours per week less free time per week in 2011 than in 2006, which is in line with the observation that their time spent on obligations has increased by two hours. By contrast, people who do not work or who work part-time have acquired more free time.

The amount of time devoted to media and ICT increased from 19.6 hours to 20.9 hours per week between 2006 and 2011. In line with expectations, the use of the Internet and/or computers has increased, from 2.8 hours to four hours per week. But watching television also increased from 12.7 hours to 14 hours per week: the ability to watch television on mobile devices and also to watch missed programmes via 'on demand ‘TV may have contributed to this increase. Reading newspapers, magazines or books showed a further decline over the period (from 3.9 hours to 2.5 hours per week). The time spent on social contacts (including online, e.g. via social media) fell from 8.6 hours per week in 2006 to 7.2 hours in 2011. This continues the trend seen in the period 1975-2005. The time devoted to online social contacts is the only form of social contact to show an increase between 2006 and 2011, though it is a modest increase. There has been a further reduction in the use of the telephone in recent years, as well as in visiting others. People do spend slightly more time going to parties and dinners at other people’s homes, but this is classified as recreational time: Dutch people had an average of 13.5 hours available in 2011 for activities that are classified as recreational activities and relaxation. The total time spent on these activities did not change between 2006 and 2011, though some shifts did take place within this category: people devoted slightly more time to going out and to sport, and slightly less to other pastimes (such as games, hobbies, gardening, looking after animals, amateur arts). Finally, the Dutch spent an average of two hours per week on various forms of social participation in 2011, a figure that has been stable for many years. This includes activities such

as volunteering, providing informal help to persons outside one's own household and practising a religion.

People with many obligatory tasks, such as people who work full-time, have found their lives becoming busier, contrary to the general trend in society. People who spend a lot of time on paid work also set aside less time for volunteering and providing informal support. Older people are expected to continue working for longer in the Netherlands. Sooner or later, however, they become involved in providing informal care for their partner or for others in their network. In the light of the government drive in the Netherlands for 'engaged citizens' who make an active contribution to the care and well-being of others, these developments beg the question of whether people are able to meet all the expectations being placed on them by the government. These and other questions will be addressed in SCP publications based on time use research in the coming years.

REFERENCES

Cloïn, M. (2012), A day with the Dutch – Time use in the Netherlands and fifteen other European countries, The Hague, The Netherlands Institute for Social Research (SCP), http://www.scp.nl/english/Publications /Publications_by_year/Publications_2012/A_day_with_the_Dutch.

European Communities (2009), Harmonised European time use surveys – 2008 guidelines, Luxembourg, Office for Official Publications of the European Communities.

Kamphuis, C., van den Dool, R., van den Broek, A., Stoop, I., Adelaar P. and J. de Haan (2009), tbo/eu en tbo/nl. Een vergelijking van twee methoden van tijdsbestedingsonderzoek, The Hague, Netherlands Institute for Social Research (SCP).

LIFE-STYLE CONCLUSIONS FROM SYNTHETIC WEEKLY TIME DIARIES AND STYLIZED QUESTIONS

Jonathan Gershuny
University of Oxford

John P. Robinson
University of Maryland

This joint project is picking up the long term discussion about the correlation between stylized and weekly diary time use data and offers some US results.

Generating respondent lifestyle profiles by correlating (or using data reduction programs like factor analysis) of respondent estimates of the time they spend on different activities is a conventional, simple and inexpensive approach to examining how different people connect their daily activities in a patterned and meaningful way. However, it is subject to potential exaggeration and distortion from the familiar response set of social desirability, particularly in Western societies in which "keeping busy" can be a "badge of honor". In other words, in giving overall activity time estimates without the normal time constraint of 168 hours a week makes it possi-

ble or likely that respondents, particularly those who consider themselves as "omnivores", will give estimates that total more than 168 hours a week. This would also probably lead to higher correlations between activities that would exaggerate how inter-connected people's lives actually are.

The time diary provides a way out of this difficulty, since all activities must sum to exactly 24 hours per day, or when aggregated across days of the week, to 168 hours per week. The problem however is that almost all diary studies collect data for single days, which is subject to distortion because single days vary so much from each other – particularly between weekdays and weekend days. Having a week's worth of diary data, then, helps to smooth out these unwanted and irregular sources of activity variation to provide a more "normal" or balanced accounting of time and activity.

However, collecting weekly time diaries is an expensive and cumbersome process. It also runs the risk of distortion due to survey burden and hence low response rates, particularly in today's presumably overworked and over-scheduled pattern of living. However, in the case of the US, it is possible to exploit an earlier high-quality diary study to generate a "synthetic" weekly diary, from a study done at the University of Michigan in 1975-76 (Juster and Stafford 1985). In its year-long panel design, respondents were contacted four times across the year, each time completing a full 24-hour diary. Two of the diaries were for a weekday, one for a Saturday and one for a Sunday. The synthetic week then is generated by multiplying the weekday diary by 5/2 (to approximate the five weekdays) and adding in the Saturday and Sunday diary figures.

As with any panel study, there was a problem of sample attrition. Of the original 1519 respondents, only about 2/3 remained in the sample for at least three of the four waves. Nonetheless, those who dropped out of the sample did not appear to be notably different in their activities from those who stayed in. These synthetic week data were then further adjusted by the archive analysis staff at the Time Research Centre at the University of Oxford to produce a synthetic week file, one that better represents all activity across a 168-hour week—though still somewhat underestimating the weekly spread of activities as a result of the missing three weekdays.

We employ this US 1975 synthetic week diary file to generate cross-activity correlations to compare to the parallel correlations generated from two separate high-quality surveys in which respondents gave estimates from "stylized" questions, such as "In the last year did you go out to the movies?" or "About how many hours a day do you personally watch television?".

The two estimate surveys were unfortunately conducted some years after this 1975 Michigan study, (there is however limited 1975 evidence producing similar results to those presented here). These were:

1. The 1982 Survey of Public Participation in the Arts (SPPA), conducted by the US Census Bureau, which asked more than 100 questions about attendance at various arts events (e.g. opera, plays) in the previous year. In order to put this arts exposure into further perspective, other SPPA questions asked about participation in other free-time activities, like TV, movie

going or sports activities. More than 5000 respondents answered these leisure activity questions, with a response rate of over 80%.

2. The 1993 General Social Survey (GSS), conducted by the National Opinion Research Center (NORC) at the University of Chicago. The GSS has been conducted since 1972 and is considered to be the premier monitor of American social trends. The response rate for this national probability survey was about 75%., with a sample of of 1596 respondents aged 18+.

While there are important differences in question wording in the two surveys, both did use the same activity time frame of the previous year. Further problems arise in aligning these questions with the coding categories in the 1975 diary study. However, the main focus of this analysis is not in estimating percentages of participation, but in terms of the correlations across activities and their inferences about the larger pattern of life styles that emerge. In other words, do people who report more work or TV in their estimates (or diaries) spend more time in religious activities, attending movies, or socializing with others?

Correlations across Activities: The correlations for the working-age population of 18-64 focus on the two activities of work and TV, first because they represent such large and meaningful amounts of time, and secondly because they are measured in common and unambiguous ways in the GSS and SPPA surveys. Both surveys asked respondents the identical question on TV viewing hours per day (with virtually the same average of 3 daily hours).

Work Hours: The first line in Table 1 for work hours illustrates the expected pattern of lower activity correlation in the 168-hour constrained time diary.

In the case of the negative diary correlations of -.10 among those 18-64 (and -12 for all adults) indicate the time constraint that if one works more hours, there is less time for religious activity. In contrast, the low figures of -.01 and .03 from the GSS estimate question suggest that working longer hours is not related to (or does not constrain) religious practice. Again, that may be due to those working longer hours being unable to realize that their working longer hours means they will less time for other activities, religion being just one of many non-work activities they may have to forego.

The same holds true for the next activity in Table 1, namely volunteering in the SPPA questionnaire, which shows a low positive correlation (.03) implying those employed working longer hours actually do a little more volunteer work than those working less. Again that conclusion is refuted in the weekly diaries, with their negative correlation of -.13.

The estimate correlations for event attendance activities are notably larger (and statistically significant), ranging between +.09 and +.22, suggesting that those working (and working longer hours) are more likely to attend movies, sports events, concerts and art museums. Again, in contrast, the diary correlations are not only lower but insignificant, indicating working longer hours do not affect attending such events.

Table 1
Activity correlations with work hours – Age 18-64

	Work hours		
Activity	GSS 1993	SPPA 1982	Diary 1975
Organizations			
Religion	-.01	NA	-.10
Volunteer	NA		
Other organizations	NA	.03	-.14
Attend			
Movies	.11	.13	-.01
Sports event	.17	.14	.05
Concerts/theatre	.13	.05	-.04
Art museums	.09	.07	-.02
Fitness			
Play sports	.19	.25	-.09
Exercise	.19	.08	-.09
Hunt/fish	.08	NA	-.02
Camping	.10	NA	NA
Outdoor	NA	NA	.01

Note: NA = Question not asked,
Source: General Social Survey (GSS) 1993, Survey of Public Participation in the Arts (SPPA) 1982, Weekly diary study done at the University of Michigan in 1975-76, own calculations.

TV Hours: The parallel analyses for TV, the main use of free time, are arrayed in Table 2. Again, mainly negative correlations are expected given the "zero-sum" nature of time, in that an increase in one use of time means that it must be offset by a decrease in another use of time. The conclusions here are largely in line with those in Table 2, but the magnitude of the numbers varies more across activities.

In the first line of Table 2 for religious activities, one does see the type of correspondence largely absent in Table 2. Both the GSS estimate and the diary show a modest significant negative correlation of around -.10, suggesting heavier viewers go to church less often. That is suggested as well in the diary for volunteer and other organizational activity, but the SPPA estimate question shows a very slight positive correlation. The next set of attendance estimate questions does show the expected pattern of negative correlations for all four GSS events, but only for movies and sports events in the SPPA estimate questions, and not for the SPPA questions on attending concerts or art museums.

This research has compared leisure activity intercorrelations in two US national studies using time estimate questions and then compared both to the patterns of activity correlation found in weekly time diaries. As expected, it finds considerable convergence in the two studies using respondent estimates, as correlated with reported work hours and hours watching TV.

Table 2
Activity correlations with TV hours – Age 18-64

	TV hours		
Activity	GSS 1993	SPPA 1982	Diary 1975
Organizations			
Religion	-.10(-.09)	NA	-.12
Volunteer	NA	.04	-.07
Other organizations	NA	NA	-.14
Attend			
Movies	-.13	-.24	-.01
Sports event	-.17	-.21	.05
Concerts/theatre	-.19	-.03	-.04
Art museums	-.16	-.05	-.02
Fitness			
Play sports	-.20	-.25	.04
Exercise	NA	-.14	.04
Hunt/fish	-.06	NA	-.02
Camping	-.18	NA	NA
Outdoor	NA	-.14	-.05

Note: NA = Question not asked,
Source: General Social Survey (GSS) 1993, Survey of Public Participation in the Arts (SPPA) 1982, Weekly diary study done at the University of Michigan in 1975-76, own calculations.

Work hours tended to correlate more positively with most leisure activities, while TV hours tended to correlate negatively with them. In contrast, the weekly time diary figures for these activities for both work hours and TV hours tended to correlate negatively. Moreover, the inter-correlations of specific leisure activities, like movies and sports, were also stronger in the respondent estimates.

electronic ***International Journal of Time Use Research***
2013, Vol. 10, No. 1, 91-111 dx.doi.org/10.13085/eIJTUR.10.1.91-111

Book notes

by Kimberly Fisher

Barberà, E. and P. Reimann
Assessment and evaluation of time factors in online teaching and learning (2013)

Publisher: IGI Global
ISBN: 978-14-666-4651-3
EISBN: 978-14-666-4652-0
DOI: 10.4018/978-1-4666-4651-3
Website: http://www.igi-global.com/book/assessment-evaluation-time-factors-online/77392
Languages Available: English

This e-book reflects practical advice compiled from research into the on-line academic courses provided by the E-Learn Center at the Universitat Oberta de Catalunya, Spain, as well as a multi-disciplinary investigation of the processes that operate in e-learning environments. The authors develop measures to reflect successful learning progress and programme implementation. Time features throughout the book. Arrangement of daily activities contribute to the success or failure of e-learning, and a range of time-based considerations underscore the development of any successful e-learning programme.

Bureau of Labor Statistics
Unpaid eldercare in the United States - 2011-2012 – Data from the American Time Use Survey (2013)

Publisher: United States Department of Labor, Bureau of Labor Statistics
Website: http://www.bls.gov/news.release/pdf/elcare.pdf
Languages Available: English

From 2011, the American Time Use Survey introduced a secondary eldercare module, repeated subsequently, asking respondents if they provide care to a person aged 65 or older who needs help with daily activities. If respondents report providing such care, as the ATUS does not collect general secondary activity information, the interviewer then asks if they did any secondary activity elder care during the diary day. If the respondents answer affirmatively, the interviewers goes through the diary episode by episode to ask when this secondary care took place. This report is the first official publication using this elder care information. The BLS reports that roughly 16% of the US population provides elder care at any given point, and around one quarter of carers are looking after an older person on any given day. Most people providing elder care are women aged 45-65. This report expands on the profile of carers and offers insights into the implications for other activities when people provide elder care.

Fellner, W. J.
Von der Güter- zur Aktivitätenökonomie – Zeitnutzung und endogene Präferenzen in einem Konsummodell , From goods based to activity based economics – Time use and endogenous preferences in a consumption modelnous preferences in a model of consumption (2013)

Publisher: Springer Gabler
ISBN: 978-3-658-04190-8
EISBN: 978-3-658-04191-5
DOI: 10.1007/978-3-658-04191-5
Website:
http://www.springer.com/springer+gabler/bwl/book/978-3-658-04190-8
Languages Available: German

This hard copy and e-book, based on Dr. Fellner's doctoral research, expands economic theories of the dynamics of consumption to add time constraints alongside financial resources and need (or perceived lack) of goods. People who have limited capacity to add consumption time in their daily schedules are less responsive to changes in prices than those with more flexible time schedules (as well as the financial income and the desire to change their consumption patterns). This book demonstrates the central role that time plays in individual as well and household engagement with the exchange of goods and services.

Henckel, D., Thomaier, S., Könecke, B., Zedda, R. and S. Stabilini (eds.)
Space–Time design of the public city (2013)

Contributing Authors: Boulin, J.-Y., Colleoni, M., Eldridge, A., Henckel, D., Rodríguez Gutiérrez, F., Gwiazdzinski, L., Könecke, B., Kuoppa, J., Mareggi, M., Mayr, A., Miciukiewicz, K., Mückenberger, U., Pottharst, M., Radicchi, A., Radoccia, R., Roberts, M., Stabilini, S., Thomaier, S., van Schaick, J., Vecchi, G., Vigar, G., Vilà, G., Zambianchi, M., Zanettichini, L. and R. Zedda
Publisher: Springer
ISBN: 978-94-007-6425-5
Website:
http://www.springer.com/earth+sciences+and+geography/geography/book/978-94-007-6424-8?wt_mc=Alerts.NBA.Aug-13_WEST_13422520
Languages Available: English

This book reflects contributions from a range of European academics working in the fields of architecture and design as well as social sciences who explore the relation between urban environments and daily behaviour in France, Germany and Italy. The twenty chapters make use a range of time use data to varying degrees to explore issues from use of public places, behaviours of teenagers, effects of working on Sundays on daily activities, night-time activities, and how public policies influence daily activities in cities.

Inbakaran C. and M.-L. van der Klooster (eds.)
2012 Time use – Australia, UK, USA, Canada, South America, and Europe (2013)

Contributing Authors: Antonopoulos, R., Fisher, K., Gershuny, J., Gracia, P., Harvey, A. S., Hoemke, M., Won Lee, C., Liddle, J., Masterson, T., Michelson, W., Millward, H., (Fredberg) Molén, M., Patulny, R., Robinson, J. P., Shahbazian, R., Sepahvand, M., Spinney, J., Suen, Y.-T., Sullivan, O., and A. Zacharias
Publisher: Deakin University
ISBN: 978-1906040895
Languages Available: English

This publication offers summaries of thirteen papers in production following the 2012 IATUR conference in Matsue, Japan. This volume covers some new fashions in the field: time use in recession, co-presence, for whom activities take place, domestic activities in households that outsource some domestic tasks to paid help, and comparative analysis of time and income poverty. The volume also covers longer standing research interests - cross-national comparisons, parenting, health, transport, national time use reports, and new areas in the field, including the environmental impact of behaviours, and time use among same sex couples.

Nabli F. and L. Ricroch
Plus souvent seul devant son Ėcran – Spending more time alone in front of tv screens) (2013)

Publisher: INSEE
ISSN: 0997 – 3192
Website:
http://www.insee.fr/fr/ffc/ipweb/ip1437/ip1437.pdf
Languages Available: French

This report uses the historical range of national French time use surveys to chart screen time in that country. INSEE reports that time in front of screens during non-working hours has increased significantly over recent decades. Television viewing grew steadily until the turn of the millennium, and since has plateaued. More recently, computer screen time has risen steadily, though the television remains the main source of screen time. This report explores the demographic, occupation, and employment status differences in screen time, noting an overall trend of people spending more time alone in front of screens.

OECD
Guidelines on measuring subjective well-being (2013)

Publisher: OECD
ISBN: 978-92-641-9164-8
DOI: 10.1787/9789264191655-en
Website: http://www.oecd-ilibrary.org/economics/oecd-guidelines-on-measuring-subjective-well-being_9789264191655-en
Languages Available: English

This report offers the first set of international guidelines for the collection and reporting of official statistics on subjective well-being to measure changes in quality of life within countries and to compare well-being between countries. These guidelines reflects the culmination of the research and consultation exercises conducted by the OECD as a part of its Better Life Initiative started in 2011. The guidelines cover a range of measures across 11 domains, and a range of question styles. One of the areas highlighted in the report is the need to measure work-life balance and daily time use patterns associated with different on the day affect responses.

Schulte, B.
Overwhelmed – Work, love, and play – When no one has the time (2014)

Publisher: Sarah Crichton Books
ISBN: 978-03-742-2844-6
Languages Available: English

Washington Post journalist and mother Brigid Schulte explores the work-life balance struggle which working parents in the USA face trying to schedule tasks they aspire to achieve as that country emerges from recession. As part of her research, Schulte attended the 2010 IATUR conference in Paris, where she interviewed a number of prominent authors in the field. She offers a hilarious account of dining out with John Robinson and Jonathan Gershuny. Though aimed at a general rather than an academic audience, this book does address the current debates in the time use literature relating to gendered divisions of work and the structure of contemporary work and family life patterns in the USA and European countries.

Sonck, N., and H. Fernee
Using smartphones in survey research – A multifunctional tool (2013)

Publisher: Sociaal en Cultureel Planbureau
ISBN: 978-90-377-0669-7
Website: http://www.scp.nl/english/Publications/Publications_by_year/Publications_2013/Using_smartphones_in_survey_research_a_multifunctional_tool
Languages Available: English

As the smartphone and related mobile devices spread to near ubiquitous presence in many countries, this book reports on experiments conducted by the Netherlands Institute for Social Research (SCP) and the research institute CentERdata at Tilburg University to harness the opportunities these devices present to time use researchers through diaries collected via apps. SCP builds on a long history of time diary re-

search which spans back to 1975. This book explores practical data collection challenges as well as new opportunities from GPS tracking and time-logged data which mobile devices collect, and explores data collection issues that will shape the future of time use research.

Varjonen, J. and K. Aalto
Kotitalouksien Palkaton Tuotanto ja Sen Muutokset 2001–2009, Tuöselosteita ja Esitelmiä 145 – Households' unpaid production and change 2001–2009 (2013)

Publisher: Kuluttajatutkimuskeskus / National Consumer Research Centre of Finland
ISBN: 978-951-698-264-2
Website:
http://www.kuluttajatutkimuskeskus.fi/files/5747/2013_145_tyoseloste_kotitalouksien_palkaton_tuotanto.pdf
Languages Available: Finnish

This report draws on household production data compiled by the collaboration of Statistics Finland and the National Consumer Research Centre in 2001, 2006, and 2009 to monitor changes in the production of goods and services by households in Finland. Only a fraction, roughly €12.5 billion of €82.6 billion worth of the goods and services which Finns produce annually, are measured by the System of National Accounts (SNA) and GDP. This report compiled a Satellite Account of Household Production, and then explores how households with different demographic characteristics adapted production patterns to the recent economic downturn.

Wolf, A.
The XX Factor – How working women are creating a new society (2013)

Publisher: Profile Books
ISBN: 978-18-466-8403-6
Languages Available: English

This book explores changes in the daily activity patterns of working women in European and North American countries. Wolf argues that in previous decades, most working women faced similar challenges from discrimination in the workplace and expectations of their activities in the home which resulted in a number of common experiences for all working women. She uses time use as well as a range of other data sources to argue that working women's lives have developed along diverging trajectories. On the one hand, some more skilled labour and educated professional women have moved closer to male-style day schedules both by shaping workplaces to accommodate their needs and by using paid outsourcing or delayed family life to meet working demands, which in turn has lead to the increasing social status of these women. On the other hand, women working in sectors where part-time employment predominates to facilitate mixing unpaid family activities with paid employment face continued truncated career and social capital development prospects. Wolf explores the potential social consequences of a growing gulf between these groups of working women.

electronic ***International Journal of Time Use Research***
2013, Vol. 10, No. 1, 91-111 dx.doi.org/10.13085/eIJTUR.10.1.91-111

New guidelines for harmonizing time use surveys

A major contribution to the time use survey literature, guiding future data collection, official reporting of survey results, and policy analysis using time use data, will be released soon by the United Nations Economic Commission for Europe (Geneva, Switzerland). The time use community should look out for these new Guidelines for Harmonizing Time-Use Surveys, which will be released in the coming weeks following the publication of this volume.