

# Hafele & Keating Tests; Did They Prove Anything?

A. G. Kelly PhD\*

\* HDS Energy Ltd, Celbridge, Co. Kildare, Ireland.

**Abstract.** The original test results were not published by Hafele & Keating, in their famous 1972 paper; they published figures that were radically different from the actual test results which are here published for the first time. An analysis of the real data shows that no credence can be given to the conclusions of Hafele & Keating.

**Key Words:** cesium clocks, relativity, accuracy, drift-rate,

## 1. Introduction

Hafele and Keating (1972) [1] (hereafter referred to as H & K) carried out experiments that purported to confirm the Theory of Special Relativity. The evidence provided was derived from the differences in time recorded by cesium clocks transported in aeroplanes, Eastward and Westward, around the Earth.

Recent university and other texts on Physics quote the tests e.g. Arfken *et al* (1989) [2], Beiser (1991) [3], Blatt (1992) [4], Cutnell and Johnson (1995) [5], Davies (1995) [6], Giancoli (1998) [7], Halliday *et al* (1997) [8] Ohanian (1989) [9] and Rindler (1991) [10]. A leader in Nature in 1972 [11] said that "*the agreement between theory and experiment was most satisfactory*". In the Science Citation Index their work has accumulated over 1000 references.

H & K avoided giving the actual test results in their paper; they gave figures that were radically altered from those results. These altered results gave the impression that they were consistent with the theory. The original test results are reproduced for the first time in this paper; these do not confirm the theory. The corrections made by H & K to the raw data, are shown to be totally unjustified.

It is also shown that the clocks used were not of sufficient stability to prove anything. The magnitude of the random alterations in performance, during the air transportation, were such as to make any result useless.

## 2. Test Details

H & K transported atomic clocks around the Earth in aeroplanes; they were sent firstly Eastward and then Westward. To minimise the effect of the variations in the Earth's magnetic field, the clocks were triple shielded. Four clocks were employed and the average of their times was used, to lessen the effect of changes in individual drift patterns relative to the standard clock-station at Washington, D. C.. The clocks used had serial numbers 120, 361, 408 and 447.

Figure 1 is a smoothed sketch of the graphical test results, which were in the original H & K 1972 paper. To appreciate the meandering nature of the results requires consulting that paper. The total test period was 26.5 days; the Eastward test began after ten days; the Westward test began nine days after the start of the Eastward test. The times of circumnavigation were 65.4 hours Eastward and 80.3 hours Westward. The vertical axis in

Figure 1 shows the difference in time, as read at hourly intervals, between each of the four clocks and that recorded by the standard station. The difference, between the average of the times recorded by the four clocks and that of the standard station, was used as the final result, and is shown as the heavy dotted line. As will be seen, H & K did not quote this average, but a radically altered version.

The forecast was that the clocks would lose (-40ns) on the Eastward journey and gain (+275ns) on the Westward journey, relative to the standard station.

### 3. Clock Stability

The times recorded by individual portable cesium clocks diverge or converge when measured against each other, or against the more accurate average of several larger clocks at a standard ground clock-station; this in turn has drift relative to other standard clock-stations. H & K referred to the fact that cesium clocks exhibited small unpredictable changes in rate. These rate changes were typically separated by at least 2 or 3 days for "good" clocks; some clocks had no rate change in the laboratory for several months.

Winkler *et al.* (1970) [12] reported that the time scale at Washington, D C (used by H & K as their standard) was obtained by averaging about 16 selected large cesium beam clocks. Clocks were replaced if their performance deteriorated. In a sample of 45 such clocks used at several stations, one failure per six clocks was experienced over two years; improvements were implemented as type faults were identified. During January 1970, three clocks had changed by +16ns, +18ns and -68ns per day. Two others were removed due to poor timekeeping, while nine had shown no change in the month.

One of the aims of the tests was the determination of the behaviour of portable clocks on scheduled flights around the globe. The Eastward and Westward flights had 13 and 15 landings/takeoffs respectively; there were several changes of aircraft.

For example, the Westward test brought the clocks from Washington to Dulles airport, then via Los Angeles, Honolulu, Guam, Okinawa, Hong Kong, Bangkok, Bombay, Tel Aviv, Athens, Rome, Shannon (an unscheduled fuel stop) Boston and Dulles, then by road back to the starting point.

Prior to 1968, a procedure had been adopted whereby the more stable clocks were given more weighting in the calculation of the standard time; thereafter, the simple average of the clocks was adopted. Winkler *et al.* gave the standard deviation of the mean of the assembly as 2ns to 4ns when tested every 3 hours over several 5 day periods. In that station, the clocks were housed in six vaults, free from vibration, with control of temperature and humidity, elaborate power supplies, vacuum systems and signal sources and a fixed magnetic field. Beehler *et al.* (1965) [13] record that the accuracy of smaller portable clocks is worse, by a factor of two, than large stationary clocks; they include variations in the magnetic field among the influences that contribute to inaccuracy of cesium clocks.

H & K claimed that they chose the four clocks because they showed a steady drift rate for at least 24 hrs before the tests. It was hoped that they would continue as a steady rate during the tests.

### 4. Changes in Drift-rate (nsec/hr)

The individual portable clocks used by H & K should have displayed a steady drift-rate, relative to the ground

clock-station. Three of the four clocks were so poor in this regard as to render them useless.

The original 1971 test report, prepared by Hafele, has been obtained by this author direct from the United States Naval Observatory (USNO), [14]. It is to be wondered why H & K avoided giving the actual test results in their 1972 paper. The drift-rates given in Table 1 are those from this report written a month after the tests were completed, and four months before the H & K papers were submitted for publication. From these drift-rates it is possible to analyse in detail the performance of the four clocks.

The drift-rates before and after a test can be compared to determine the change during a test. The Hafele 1971 report said *"Most people (myself included) would be reluctant to agree that the time gained by any one of these clocks is indicative of anything"* and *"the difference between theory and measurement is disturbing"*. Also, he said that, for a useful test, the drift rate of any clock should be constant over the whole period of the test. These reservations are not mentioned in the H & K 1972 paper. The relativistic effect should result in an incremental step change in the times recorded by each clock, before and after a flight test, but should not affect the drift rate.

**Table 1. Drift-rates of the clocks (ns per hour)**

Clock No	120	361	408	447
Before the Eastward test	-4.50	+2.66	-1.78	-7.16
After the Eastward Test	-8.89	+4.38	+3.22	-8.41
Before the Westward test	-8.88	+6.89	+4.84	-7.17
After the Westward test	-4.56	+3.97	+2.16	-9.42

Three of the clocks did not keep the same drift rate when on the ground, between the Eastward and Westward tests. There can be little confidence that the same clocks would perform far better while they were transported in passenger seats upon commercial planes.

Clock 120 was a disaster; it had a change (Table 1) from losing 4.50 ns per hour to losing 8.89 ns per hour on the Eastward trip; on the Westward trip it altered from losing 8.88 to losing 4.56 ns per hour.

An examination of Table 1 shows that, with the single exception of clock 447, the drift rates were so far from being steady as to render the results totally useless. The changes in drift-rates that occurred during a circumnavigation are derived from Table 1 and shown in Table 2. These alterations in drift-rate should be close to zero, to give confidence in any conclusions. An examination of the drift rates before and after the Eastward and Westward (Table 1) tests shows that huge differences emerged during the tests.

**Table 2. Alteration in drift-rates during tests**

Clock No	120	361	408	447
Eastward Test	-4.39	+1.72	+5.00	-1.25

Westward Test +4.31 -2.93 -2.68 -2.25

The wild swings in drift rate should have resulted in the whole test being declared a failure. The authors had recognised that a steady drift rate was a prerequisite of a meaningful test. The 1971 report says "*Particularly in the case of 361 after the eastbound flight, it is quite uncertain what the rate is after the flight*", and "*Portable cesium clocks cannot be expected to perform as well under travelling conditions as they do in the laboratory. Our results show that changes as large as 120 nsec/day may occur during trips with clocks that have shown considerably better performance in the laboratory*".

An alteration in drift-rate during a flight could be the result of gradual alteration, or of a series of sudden changes.

The Eastward circumnavigation time of 65.4 hours would accumulate the forecast theoretical alteration of -40ns at a change in drift-rate of -0.6ns per hour. The figures in Table 2 should, as stated above, all read zero, but any change that was not smaller than 0.6ns per hour by at least, say, a factor of at least 5 would obfuscate any result. This means that no clock was of any value in reaching any conclusion on the Eastward test; the accuracy would need to be better by two orders of magnitude. On the Westward test, the forecast change of +275ns would be accumulated in the trip time of 80.3 hours at a change in drift-rate of +3.4ns per hour. The change in drift rate would have to be less than 0.7 ns per hour to be of any use; again, no result can be used with any confidence.

Figure 2 depicts a sketch of H & K's enlarged view of the average of the four clocks (as given in the H & K paper) for the period immediately before and after the Westward test. A similar graph was given for the Eastward. As required by the theory, the trend lines that they drew for the periods before and after the tests were parallel, but with an incremental step change downwards and upwards for the Eastward and Westward tests respectively. The change due to the forecast relativistic effect would, of course, occur gradually during a flight, but can only be determined as the resulting step change that emerges following that test, because there was no way to compare with the ground station during a flight.

The starting time for the flights was determined by the departure of a commercial aeroplane. Had the Westward test begun some 12 hours earlier (see Figure 2), the trend before that test would be very different. Using the average of the four clocks, for the complete period shown by H & K on Figure 2, a time shift of approximately zero could have been reasonably deduced for the Eastward test; indeed the 1971 report described the Eastward result as "*consistently negative or near zero*".

The rate changes are random and could have occurred in either a + or - direction. Clock 120 altered in drift-rate by +4.39ns/h on the Eastward test and by -4.31ns/h on the Westward test; we should not say that this clock had an average drift-rate change of 0.04ns/h; indeed this was the clock with the most erratic performance. This is like saying that a watch, which gained ten hours in the first week and lost ten in the second, is a perfect timekeeper! From Figure 1, Clock 447 can be interpreted as having a small alteration in drift from 100 hours into the test period to the end of the Westward test. Had this clock, with the most steady performance, been chosen, the overall result would have been zero.

The trend shown in Figure 2 was derived from the *average* of the four clocks. **The results from the individual clocks was not disclosed; they are published here for the first time in Columns 2 and 5 of**

**Table 3.** Taking the mathematical average of Columns 2 or 5 is meaningless; on the Eastward trip, clock 408 gained 166ns, while the theory forecast a loss of 40ns; on the Westward trip clock 361 lost 44ns, while the theory forecast a gain of 275ns!

## 5. Test Results

### Table 3.

#### Original Test Results and H & K alterations (ns)

Eastward Westward

Clock **Test** First Second **Test** First Second

No **Results** Change Change **Results** Change Change

120 **-196** -52 -57 **+413** +240 +277

361 **-54** -110 -74 **-44** +74 +284

408 **+166** +3 -55 **+101** +209 +266

447 **-97** -56 -51 **+26** +116 +266

Average --- -54 -59 --- +160 +273

Notes : (1) The -59ns and +273ns averages derived by H & K are to be compared with the -40ns and +275ns predicted by the theory. (2) Standard Deviations were shown for the mean of the second method as 10 and 7 respectively; the 7 should have read 9.

The data shown as the "Test Results" in Table 3 are the ns alterations evinced by each clock, from what should be the result, at the drift-rate that was present at the start of the flight. For example, the drift rate of clock 361 before the Eastward flight was +2.66ns/hr, compared to the standard station at Washington.. The flight was 65.42 hours and, at that rate, this clock should have gone ahead of Washington by an accumulated +174ns; it had only gone +120ns, thus proving that it had lost 54ns on the flight.

As seen in Table 3, there are two instances when a clock altered in the *opposite* direction to the theory; viz. 408 on the Eastward test (shows a gain, marked G in Figure 1 above) and 361 on the Westward test (shows a loss, marked L). The H & K so-called 'result' of the Eastward test (-59ns) was less, by a factor of 2.8, than, and opposite in sign to, the time shift of one clock (408).

The first attempt by H & K to bring the results closer to the theoretical forecasts was to take the average of the drift rates before and after a flight, and assume that this average was the drift rate that applied throughout the flight. This is equivalent to assuming that one single sudden change in drift-rate occurred mid-way. Such an assumption would have some credence had the alteration in drift-rate been very small e.g. a change from +3.34 to +3.35ns/h, which would not significantly affect the end result. The actual drift-rates (Table 1) doubled (one clock from -4.5 to -8.9ns/h) or halved (+4.8 to +2.2ns/h; -8.9 to -4.6ns/h) or reversed (-1.8 to +3.2ns/h). The alterations made to the test results, using this first method, are seen in Columns 3 and 6 of

Table 3. Having made those changes, and from that produced Figure 2, H & K correctly dismissed that approach on the basis that it "*depended on the unlikely chance that only one rate change occurred during each trip and that this change occurred the midpoint of the trip*"; they added that there was no obvious method of estimating the experimental error of such an assumption. They had, as will be discussed later, actually identified seven alterations in drift-rate on one clock, four on another and two on a third.

Having dismissed this method, they still published graphs (corresponding to Figure 2 above), based upon that method, which they described as producing "*convincing qualitative results*". It was published because it looked convincing and not because it gave a legitimate picture of the test results. To the unsuspecting reader, these graphs looked like proof of the success of the tests.

H & K next used another method of altering the test data. It was not possible, during a flight test, to check the behaviour of clocks relative to the standard station; when a test ended, the measurement of drift, relative to the standard clock-station, resumed. However, comparisons that were made between the four clocks during flights were used by H & K, to decide whether one clock had undergone what was deemed to have been a sudden change in pattern; in such a case it was assumed that the behaviour of the other three was correct. The rationale was that "*the chance that two or more clocks will change rate by the same amount in the same direction at the same time is extremely remote*". Corrections were made for fourteen changes: clock 120 three changes Eastward and one Westward; clock 361 three Eastward and four Westward; clock 408 two Eastward; clock 447 one Eastward. These corrections were made after the 1971 report was produced. It might have been justifiable to ignore a single isolated sudden change on one clock during the complete 26.5 day period, but to have made corrections for fourteen such alterations in six days of flights and by amounts that exceed the forecast results by up to 5.5 times is breathtaking.

The USNO standard station had some years previously adopted a practice of replacing at intervals whichever clock was giving the worst performance. On a similar basis, the results of Clock 120 should have been disregarded. That erratic clock had contributed ***all of the alteration in time on the Eastward test and 83% on the Westward test***, as given in the 1971 report. Discounting this one totally unreliable clock, the results would have been within 5ns and 28ns of zero on the Eastward and Westward tests respectively. This is a result that could not be interpreted as proving any difference whatever between the two directions of flight.

The actual test results (Cols. 2 & 5 of Table 3) are very different from the altered figures produced by the second method (Cols. 4 & 7). The figures in Cols. 4 and 7 were the only ones published in 1972; these give the very misleading impression that the results were compatible, as if they were all of the forecast sign and within a narrow band.

Examples of how unreasonable were the corrections from the actual test results to the amended version are:

clock 408 (Eastward) 'corrected' from +166ns to -55ns;

clock 361 (Westward) 'corrected' from -44ns to +284ns.

Clock 447 was amended from +26ns to +266ns on the Westward test; this was by a factor of 10. Yet, the H & K paper said that *no significant changes in rate were found for clocks 408 and 447 during the westward trip*". This barefaced manipulation of the data was outrageous. Clock 447 was the single clock that had a pretty

steady drift-rate throughout the tests. The 1971 report stated that *"rate changes that are noticeably larger than those typical in the laboratory occurred for each clock during at least one of the trips, except for clock 447"*. Why then did they not use the results of this one stable clock and abandon the other three?

On the Eastward test, corrections of +3.5 and -5.5 times the forecast theoretical -40ns result were applied to two of the clocks; on the Westward test, where the forecast was +273ns, corrections of 0.5 to 1.2 times that amount were applied to three of the clocks.

H & K stated that the number of measured values was too small for a proper statistical analysis; nonetheless, they gave a standard deviation of the four results on either test, as quoted under Table 3. This gives the misleading impression that the results are grouped reliably closely.

Was the aim of these tests to fabricate a confirmation of the theory, or to give objective reliable experimental results?

## 6. Discussion

Bodily and Hyatt (1967) [15] stated that 2.5ns/hr as less than changes in drift-rate and random errors in mobile clocks. The average change for the four clocks in these tests (Table 2) was 3.07ns/h. Alterations less than 200ns Eastward or 250ns Westward are therefore of no significance whatever.

The 1971 report showed a graphical proof that any result below about 125ns could not be used on the Eastward test. Yet, H & K used the Eastward results, which were all below this threshold. That report stated that it was *"amusing that values for the Eastward trip were in excellent agreement with the theory, despite expectations that they would not be able to detect any definite effect"*. But, this is not true; the Eastward results varied from +166 to -196ns; to average such wildly divergent results is meaningless.

H & K recorded that previous tests, reported in 1970, had showed results that were normally distributed zero-centred and with a spread of about 60ns per day of travel. The 1971 report advocated the future use of better clocks and a circumnavigation with less ground time, which would probably reduce the standard deviation of the results by a factor of ten.

H & K concluded that there seemed to be little basis for further arguments about whether clocks would indicate the same time after a round trip. The clocks certainly altered during the circumnavigations, but the alterations that occurred were random and have no significance.

Earlier attempts to deduce the changes in drift rates from the graphs in the 1972 H & K paper were made by this author [16], and later found to have been done by Essen in 1977 [17]. Both concluded that the alterations in drift-rates of the clocks made the results useless. These attempts could reasonably have been discounted, on the basis that the original raw test data was not available to these authors. That excuse is now no longer valid.

## 7. Conclusions

The H & K tests prove nothing. The accuracy of the clocks would need to be two orders of magnitude better to give confidence in the results. The actual test results, which were not published, were changed by H & K

give the impression that they confirm the theory. Only one clock (447) had a fairly steady performance over the whole test period; taking its results gives no difference for the Eastward and the Westward tests.

## References

- [1] J.C. Hafele and R.E. Keating, *Science* 177, 166-168 and 168-170 (1972)
- [2] G.B. Arfken, D.F. Griffing, D.C. Kelly & J. Priest, *University Physics* 2 ed. (San Diego: Harcourt Brace Jovanov) p 842, 1989
- [3] A. Beiser, *Physics* 5 ed (Reading: Addison-Wesley) p 747 (1991)
- [4] F.J. Blatt, *Modern Physics* (New York: McGraw Hill) p 25 (1992)
- [5] J.D. Cutnell and K.W. Johnson, *Physics* 3 ed. (N. Y.Wiley) p 909 (1995)
- [6] P.C.W. Davies *About Time* (London: Viking) p 57 (1995)
- [7] D.C. Giancoli, *Physics* 5 ed. (London: Prentice-Hall) p 755 (1998)
- [8] D. Halliday, R. Resenick and J. Merrill, *Fundamentals of Physics* (New York:Wiley) p 960 (1997)
- [9] H.C. Ohanian, *Physics* 2nd ed. expanded.(W. W. Norton, N.Y., 1989)
- [10] W. Rindler *An Introduction to Special Relativity* 2 ed. (Oxford: Clarendon Press) p 29 (1991)
- [11] Leader, *Nature* 238, 2425 (1972)
- [12] G.M.R. Winkler, R.G. Hall and D.B. Percival, *Meteorologia* 6 No 4, 126-134 (1970)
- [13] R.E. Beehler, R.C. Mockler and J.M. Richardson, *Meteorologia* 1 No 3, 114-131 (1965)
- [14] Proc. 3rd Dept. Def. PTTI Meeting 261-288 (1971)
- [15] L.N. Bodily and R.C. Hyatt, *Hewlett Packard J.* 19, No 4 12-20 (1967)
- [16] A.G. Kelly, *Monograph* No 3, *Inst. Engrs. Irel.* (1996)\*
- [17] L. Essen, *Creation Res. Society Quarterly* Vol 14, 46, (1977)

\* Post free, Institution of Engineers of Ireland,, 22 Clyde Rd, Dublin 4, Ireland.

## List of Figure Captions

### Figure 1, Sketch of Results Given by H & K

### Figure 2. Enlarged View

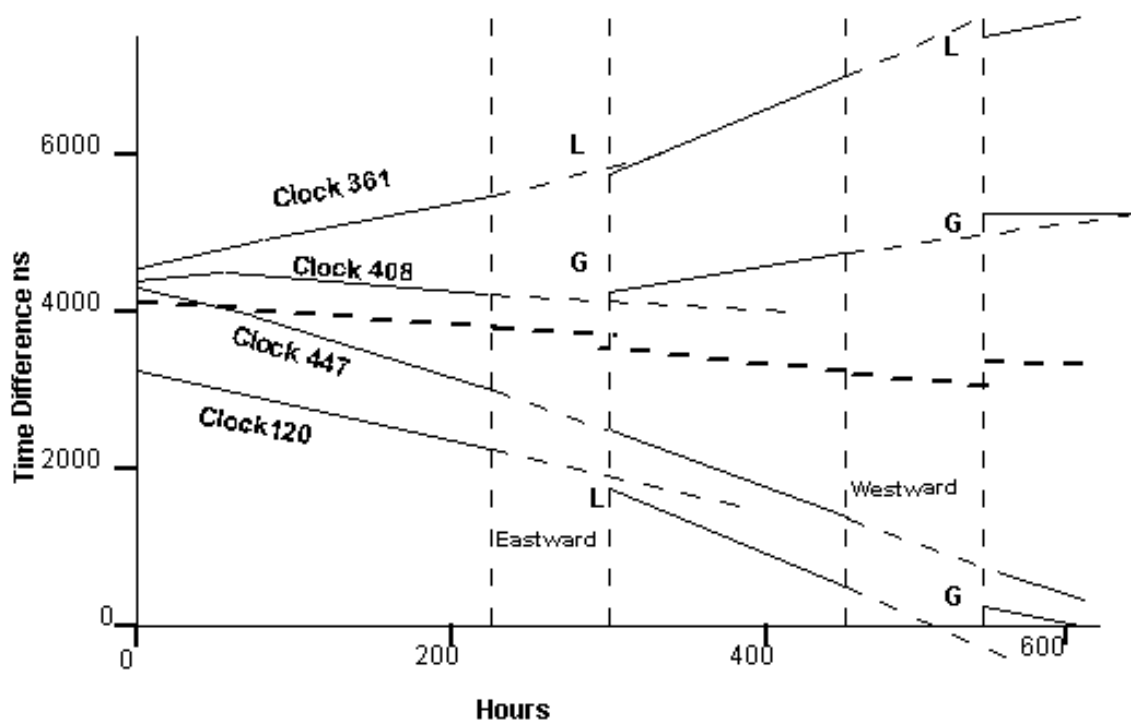
### List of Tables



**Table 1. Drift-rate of Clocks (ns per hour)**

**Table 2. Alteration in Drift-Rates During Tests**

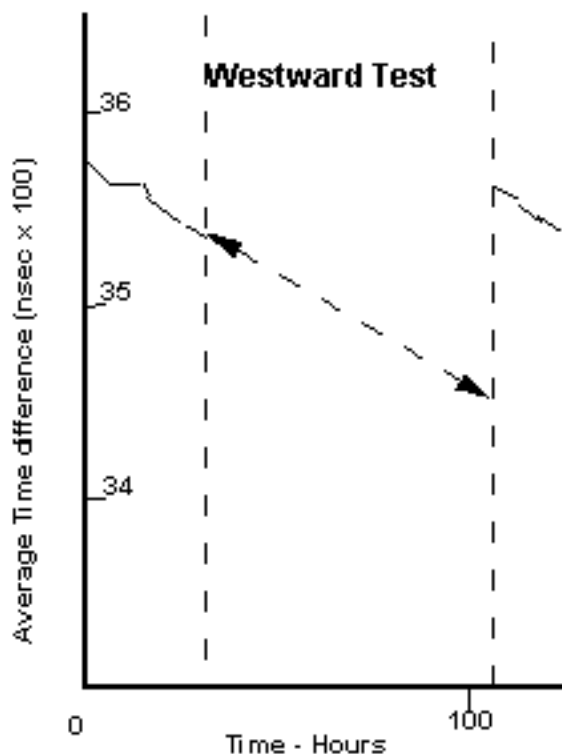
**Table 3. Original Test Results and H & K alterations (ns)**



L= Loss; G = Gain.

**Figure 1**

Sketch of Results Given by H & K



**Figure 2**

**Enlarged View**

A. G. Kelly.

HDS Energy Ltd.,

Celbridge,

Co. Kildare,

IRELAND.

**French Translation of Summary**

Résumé: Les résultats de l'essai original n'ont pas été publiés par Hafele et Keating, dans leur célèbre article de 1972. Ces résultats sont maintenant disponibles et sont publiés dans cet article. On y montre que des changements radicaux apportés par les auteurs aux données d'essai n'étaient pas justifiés. Une analyse des données confirme que les conclusions d'Hafele et Keating sont fausses. On ne peut donner aucune croyance aux déductions fondées sur ces essais.